IEOR 222

Financial Engineering

Lecture notes

Alberto Gennaro

https://github.com/GennaroAlberto/ieor222-financial-engineering

Built October 31, 2026

Contents

About these notes **7**

I Probability and discrete-time markets **9**

1 Probability as information **10**

1.1 A finite model of an economy . . . 10

1.2 Information . . . 11

1.2.1 σ-algebras as information . . . 11

1.2.2 Partitions . . . 12

1.2.3 Filtrations and adapted processes . . . 13

1.3 Independence . . . 13

1.4 Conditional expectation . . . 14

1.4.1 Conditioning on a partition . . . 14

1.4.2 The general definition . . . 15

1.4.3 Properties . . . 15

1.4.4 The independence lemma . . . 17

1.4.5 The best predictor . . . 18

1.5 Changing the probability measure . . . 19

1.5.1 The Gaussian shift . . . 21

Further reading . . . 21

2 Martingales and random walks **22**

2.1 Martingales . . . 22

2.2 Trading strategies and martingale transforms . . . 25

2.3 Stopping times and optional stopping . . . 27

2.4 Gambler's ruin . . . 29
Further reading . . . 32

3 Arbitrage and static replication 33
3.1 Assets, positions and payoffs . . . 33
3.2 Arbitrage . . . 35
3.3 Price relations from static arbitrage . . . 36
3.4 Static replication . . . 40
Further reading . . . 44

4 The binomial model 45
4.1 The model . . . 45
4.2 One period . . . 46
4.2.1 The risk-neutral measure . . . 47
4.3 N periods . . . 48
4.4 The fundamental theorem of asset pricing . . . 50
4.5 American options . . . 51
4.6 The trinomial model: a first incomplete market . . . 54
Further reading . . . 56

II Brownian motion and the Black–Scholes formula 57

5 Brownian motion 58
5.1 The symmetric random walk . . . 58
5.2 Scaling the walk . . . 59
5.3 Brownian motion . . . 61
5.4 Filtrations and martingales . . . 62
5.5 Quadratic variation . . . 63
Further reading . . . 66

6 From the binomial model to Black–Scholes 67
6.1 Geometric Brownian motion . . . 67
6.2 The binomial model in the limit . . . 68
6.3 Risk-neutral pricing in continuous time . . . 71

6.4 The Black–Scholes formula . . . 73
6.5 The Greeks . . . 74
6.6 At the money . . . 76
Further reading . . . 78

III Stochastic calculus **79**

7 The Itô integral **80**
7.1 Gains from trading . . . 80
7.2 Simple integrands . . . 81
7.3 Square-integrable integrands . . . 83
7.4 Deterministic integrands . . . 86
7.5 The integral of W against itself . . . 87
7.6 Quadratic variation and covariation . . . 88
7.7 Stopping and localization . . . 90
Further reading . . . 91

8 Itô's formula **92**
8.1 Variations, and what survives a Taylor expansion . . . 92
8.2 Itô's formula for Brownian motion . . . 93
8.3 Itô processes . . . 97
8.4 Several Brownian motions . . . 100
Further reading . . . 103

9 Stochastic differential equations **104**
9.1 Stochastic differential equations . . . 104
9.2 Existence and uniqueness . . . 105
9.3 Linear equations . . . 108
9.4 The Ornstein–Uhlenbeck process and the Vasicek model . . . 109
9.5 The Markov property and the generator . . . 110
9.6 Local volatility and the CEV model . . . 112
9.7 The Euler–Maruyama scheme . . . 113
Further reading . . . 115

IV Pricing and hedging **116**

10 The Black–Scholes equation by hedging **117**

10.1 Self-financing strategies . . . 118
10.2 The Black–Scholes equation . . . 119
10.3 Verification and the Black–Scholes formula . . . 122
10.4 Hedging in discrete time . . . 124
Further reading . . . 127

11 Change of measure and risk-neutral pricing **129**

11.1 Densities on a filtration . . . 129
11.2 Girsanov's theorem . . . 130
11.3 Risk-neutral pricing . . . 133
11.4 Martingale representation and completeness . . . 135
11.5 Change of numeraire . . . 138
11.6 Dividends and exchange rates . . . 139
Further reading . . . 140

12 The Feynman–Kac formula **141**

12.1 From the Markov property to an equation . . . 142
12.2 The Feynman–Kac formula . . . 142
12.3 Pricing equations for local volatility models . . . 144
12.4 The heat equation and the Black–Scholes equation . . . 146
12.5 Transition densities and Kolmogorov's equations . . . 148
12.6 Digital and pay-later options . . . 151
Further reading . . . 152

13 Exotic options **153**

13.1 Path-dependent claims . . . 153
13.2 The reflection principle . . . 154
13.3 Barrier options . . . 156
13.4 Lookback options . . . 159
13.5 Asian options . . . 160
13.6 The perpetual American put . . . 163

Further reading . . . 164

14 Short-rate models **165**

14.1 Bonds, yields and forward rates . . . 165
14.2 Short-rate models . . . 166
14.3 Affine term structures . . . 167
14.4 The Vasicek model . . . 170
14.5 The Cox–Ingersoll–Ross model . . . 172
14.6 The forward measure and bond options . . . 174
14.7 Beyond short-rate models . . . 176
Further reading . . . 176

15 Implied volatility and the smile **177**

15.1 The limits of the Black–Scholes model . . . 177
15.2 Implied volatility . . . 178
15.3 Static arbitrage in implied volatility . . . 180
15.4 Local volatility and Dupire's formula . . . 182
15.5 Stochastic volatility and the Heston model . . . 183
15.6 One model for all prices . . . 185
15.7 Variance swaps and the VIX . . . 187
Further reading . . . 188

16 Numerical methods **189**

16.1 Monte Carlo . . . 189
16.1.1 Discretized dynamics . . . 190
16.2 Finite differences . . . 191
16.3 The Black–Scholes equation on a grid . . . 196
16.4 The American put on a grid . . . 197
16.5 Physics-informed neural networks . . . 199
Further reading . . . 200

V Extras **201**

17 Arbitrage-free option surfaces **202**

17.1 Why an arbitrage-free surface . . . 202
17.2 Static arbitrage on a grid . . . 204
17.3 SANOS . . . 206
17.3.1 Calibration by linear programming . . . 207
17.4 Discrete local volatility . . . 209
17.5 DYSANOS: generating arbitrage-free surfaces . . . 210
Further reading . . . 213

18 Hawkes processes and limit order books **214**
18.1 Point processes and intensities . . . 214
18.2 The Hawkes process . . . 216
18.3 Simulation by thinning . . . 219
18.4 Likelihood and goodness of fit . . . 220
18.5 Order flow in a limit order book . . . 221
18.6 Nearly unstable Hawkes processes and rough volatility . . . 223
Further reading . . . 226

19 Rough volatility **227**
19.1 Fractional Brownian motion . . . 227
19.2 Volatility is rough . . . 230
19.3 Rough Bergomi and the at-the-money skew . . . 232
19.4 Simulation . . . 234
19.5 The rough Heston model . . . 236
Further reading . . . 237

20 Signatures and signature-based models **239**
20.1 The signature of a path . . . 240
20.2 Signatures of Brownian paths . . . 244
20.3 Signature-based models . . . 247
20.4 Calibration . . . 249
Further reading . . . 251

About these notes

These are lecture notes for IEOR 222, *Financial Engineering*, at UC Berkeley: the mathematics of pricing and hedging derivatives, from discrete-time models to stochastic calculus, risk-neutral pricing and the numerical methods used in practice.

They began as the notes I took when I followed IEOR 222 as a student at UC Berkeley. I later taught the course's discussion sections as graduate student instructor (GSI), and the worksheets, exercises and explanations I prepared for those sections, together with my student notes, became the material of this book. The general flow of the course follows the lectures of Professor Thibaut Mastrolia: I am grateful to him for it, and for his great teaching, which shaped how I understand this subject. The text, the proofs, the demos and the wording of the exercises were written for this repository; any mistakes are mine.

How to use them. Each chapter has three parts. The notes, in this document, develop the theory with complete proofs wherever the argument teaches something, and with precise references where it does not. The exercises are collected in a separate document; solutions are not published, so that the course is not spoiled for the next class. The interactive demos, marked ▶, are Python notebooks that open in Google Colab in one click: they let you move the parameters of each result and see it happen.

Extras. Part V collects topics I would teach if the semester were longer: arbitrage-free option surfaces built from market quotes, Hawkes processes and limit order books, rough volatility, and signatures. They assume the first four parts.

Prerequisites. Undergraduate probability (independence, the central limit theorem, Gaussian vectors, characteristic functions), multivariable calculus, and enough Python to read short scripts. A few results from measure theory are used without proof and stated where they first appear: the monotone and dominated convergence theorems, the Borel–Cantelli lemma, Dynkin's π–λ theorem, and the Tonelli and Fubini theorems for exchanging integrals. Beyond these, no measure theory is assumed: the measure-theoretic ideas the course needs are introduced in Chapter 1.

Corrections. Please report errors, typos and unclear passages as issues on the GitHub repository.

Notation

$(\Omega, \mathcal{F}, \mathbb{P})$	probability space; $\mathbb{P}$ is the real-world (historical) measure
$\mathbb{Q}$, $\mathbb{E}^{\mathbb{Q}}$	a risk-neutral measure and its expectation
$(\mathcal{F}_t)_{t\geq 0}$	filtration: the information available at time t
W	a Brownian motion
$\mathcal{N}(m, v)$, $N(\cdot)$	normal law with mean m and variance v; standard normal CDF
$\mathbb{N}$	$\{0, 1, 2, \dots\}$
x^+, $x \wedge y$	$\max(x, 0)$; $\min(x, y)$
$\mathbf{1}_A$	indicator of the event A
$[X, X]_t$	quadratic variation of X on $[0, t]$
r, σ, K, T	interest rate, volatility, strike, maturity

Part I

Probability and discrete-time markets

Chapter 1

Probability as information

A derivative pays an amount that depends on how prices move until its maturity. To price and hedge it we must say what can happen, how likely each scenario is, and what is known at each date. A probability space takes care of the first two. This chapter is mostly about the third: information is a σ-algebra, the best estimate of a random quantity given some information is a conditional expectation, and a trading strategy may only use the information available when it is chosen. The last section introduces an idea that runs through the whole course: prices are expectations under a measure other than the real-world one, obtained by reweighting the scenarios.

We work mostly on finite probability spaces, where every statement reduces to a finite computation, and we say at each step what changes on a general space.

1.1 A finite model of an economy

Toss a coin n times. Each toss should land heads (H) with probability $p \in (0,1)$ and tails (T) with probability $q = 1-p$, independently of the others (Theorem 1.11 checks that the model below does this). An outcome is a word $\omega = \omega_1 \cdots \omega_n$ in the letters H and T, and the set of outcomes is $\Omega_n = \{H,T\}^n$. Write $H_k(\omega)$ for the number of heads among the first k tosses, and set

$$\mathbb{P}(\omega) = p^{H_n(\omega)} q^{\,n-H_n(\omega)}, \qquad \mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega) \quad (A \subseteq \Omega_n), \tag{1.1}$$

where $\mathbb{P}(\omega)$ is short for $\mathbb{P}(\{\omega\})$. We call $\mathbb{P}$ the *p-coin*. To turn the coin into a market, fix $S_0 > 0$ and $0 < d < u$, and let a stock price be multiplied by u after each head and by d after each tail:

$$S_k(\omega) = S_0\, u^{H_k(\omega)} d^{\,k-H_k(\omega)}, \qquad k = 0,1,\dots,n. \tag{1.2}$$

This is the binomial model of Chapter 4. In our smallest example, $n=2$, $S_0 = 4$, $u = \frac{3}{2}$ and $d = \frac{1}{2}$: then S_1 is 6 or 2 according to the first toss, and $S_2(HH) = 9$, $S_2(HT) = S_2(TH) = 3$, $S_2(TT) = 1$.

Definition 1.1 (σ-algebra, probability space)**.** Let Ω be a nonempty set. A collection $\mathcal{F}$ of subsets of Ω is a *σ-algebra* if (i) $\Omega \in \mathcal{F}$; (ii) $A \in \mathcal{F}$ implies $A^c = \Omega \setminus A \in \mathcal{F}$; (iii) $A_1, A_2, \dots \in \mathcal{F}$ implies $\bigcup_{j \ge 1} A_j \in \mathcal{F}$. The sets in $\mathcal{F}$ are *events*, and $(\Omega, \mathcal{F})$ is a *measurable space*. A *probability*

measure on $(\Omega, \mathcal{F})$ is a map $\mathbb{P} : \mathcal{F} \to [0,1]$ with $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\bigcup_j A_j) = \sum_j \mathbb{P}(A_j)$ for every sequence of pairwise disjoint events (*countable additivity*). The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space*.

On a finite Ω we take $\mathcal{F} = 2^\Omega$, and a probability measure amounts to weights $\mathbb{P}(\omega) \geq 0$ summing to one, as in (1.1). On $[0,1)$, no countably additive probability defined on all subsets is invariant under translations mod 1 (Vitali; see Williams (1991), Chapter 1 and Appendix A1), which is why $\mathcal{F}$ may be smaller than the power set. Countable unions are needed so that limiting events, such as $\bigcup_k \{S_k > b\}$ with infinitely many dates, are events (Exercise 1.1 shows that finite unions are not enough). In finance, $(\Omega, \mathcal{F}, \mathbb{P})$ models the economy: an outcome is a complete scenario up to the horizon, and $\mathbb{P}$, the *real-world* or *historical* probability, gives the frequencies of the scenarios. Other measures will appear; the letter $\mathbb{P}$ is reserved for this one.

1.2 Information

1.2.1 σ-algebras as information

Suppose that an outcome ω has been drawn but we are told only the first toss. Some events are then *resolved*: we know for sure whether they occurred. In the two-toss model, $A_H = \{HH, HT\}$ and $A_T = \{TH, TT\}$ are resolved after the first toss, while $\{HH\}$ is not. The resolved events form a σ-algebra: if we know whether A occurred, we know whether A^c did, and if we know it for each of $A_1, A_2, \ldots$, we know whether one of them did. We therefore model *information* by a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, the events that the information decides. In the two-toss model, the information at dates 0, 1, 2 is

$$\mathcal{F}_0 = \{\emptyset, \Omega_2\}, \qquad \mathcal{F}_1 = \{\emptyset, A_H, A_T, \Omega_2\}, \qquad \mathcal{F}_2 = 2^{\Omega_2} \quad (16 \text{ events}),$$

and more information means more events: $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2$.

The *σ-algebra generated* by a collection $\mathcal{C}$ of subsets, $\sigma(\mathcal{C})$, is the intersection of all σ-algebras containing $\mathcal{C}$, hence the smallest one: the information obtained by learning, for each set in $\mathcal{C}$, whether it occurred. The *Borel σ-algebra* $\mathcal{B}(\mathbb{R}^k)$ is generated by the rectangles of intervals, and $f : \mathbb{R}^k \to \mathbb{R}$ is *Borel* if $f^{-1}(B) \in \mathcal{B}(\mathbb{R}^k)$ for $B \in \mathcal{B}(\mathbb{R})$. Continuous functions, indicators of Borel sets, and sums, products and limits of Borel functions are Borel (Durrett (2019), Section 1.3).

Definition 1.2 (Random variables, $\sigma(X)$, measurability)**.** Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra.

(i) A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that $\{X \in B\} := \{\omega : X(\omega) \in B\} \in \mathcal{F}$ for every Borel set $B \subseteq \mathbb{R}$. A *random vector* $X : \Omega \to \mathbb{R}^k$ satisfies this for every $B \in \mathcal{B}(\mathbb{R}^k)$.

(ii) The *σ-algebra generated by X* is $\sigma(X) = \{\{X \in B\} : B \in \mathcal{B}(\mathbb{R})\}$ (with $\mathcal{B}(\mathbb{R}^k)$ for a random vector), and $\sigma(X_1, \ldots, X_k) = \sigma((X_1, \ldots, X_k))$.

(iii) X is *$\mathcal{G}$-measurable* if $\sigma(X) \subseteq \mathcal{G}$, that is, $\{X \in B\} \in \mathcal{G}$ for every Borel set B.

Preimages commute with complements and unions, so $\sigma(X)$ is a σ-algebra (with intervals alone in place of Borel sets, it would not be closed under unions). It is the information revealed by observing X, and "X is $\mathcal{G}$-measurable" reads: *X is known once $\mathcal{G}$ is known.* It suffices that

$\{X \le b\} \in \mathcal{G}$ for all b, because the sets B with $\{X \in B\} \in \mathcal{G}$ form a σ-algebra. Likewise, with rectangles, a random vector is $\mathcal{G}$-measurable if and only if each coordinate is, and $\sigma(X_1, \dots, X_k)$ is the smallest σ-algebra containing every $\sigma(X_i)$.

1.2.2 Partitions

A *partition* of Ω is a finite collection $\mathcal{P} = \{A_1, \dots, A_m\}$ of nonempty, pairwise disjoint events with union Ω; its elements are *atoms*. The unions of atoms, including $\emptyset$, form a σ-algebra with 2^m elements, which is $\sigma(\mathcal{P})$. Knowing $\sigma(\mathcal{P})$ means knowing which atom contains ω, and nothing more.

Lemma 1.3 (Information on a finite space)**.** *(a) If a σ-algebra $\mathcal{G}$ has finitely many elements (for instance, if Ω is finite), there is exactly one partition $\mathcal{P}$ with $\mathcal{G} = \sigma(\mathcal{P})$; its atoms are the minimal nonempty sets of $\mathcal{G}$.*

(b) A random variable is $\sigma(\mathcal{P})$-measurable if and only if it is constant on each atom of $\mathcal{P}$.

(c) If X takes finitely many values, $\sigma(X)$ is generated by the partition into the level sets $\{X = x\}$.

Proof. (a) Let $A(\omega)$ be the intersection of the finitely many sets of $\mathcal{G}$ containing ω, the smallest such set. If $\omega' \in A(\omega)$, then $A(\omega') \subseteq A(\omega)$; and $\omega \in A(\omega')$, since otherwise $A(\omega) \setminus A(\omega')$ would be a smaller set of $\mathcal{G}$ containing ω. So $A(\omega') = A(\omega)$. Hence two such sets are equal or disjoint (a common point ω'' gives $A(\omega) = A(\omega'') = A(\omega')$): they form a partition, and each $A \in \mathcal{G}$ is the union of the $A(\omega)$ with $\omega \in A$. In any partition generating $\mathcal{G}$ the atoms are the minimal nonempty sets of $\mathcal{G}$, which gives uniqueness. (b) If $X = c_i$ on A_i, then $\{X \in B\} = \bigcup_{i:\, c_i \in B} A_i$. Conversely, if $X(\omega) \neq X(\omega')$ with ω, ω' in an atom A, then $\{X = X(\omega)\}$ is a union of atoms containing ω, hence containing $A \ni \omega'$, a contradiction. (c) The level sets lie in $\sigma(X)$, and by (b) X is measurable with respect to the σ-algebra they generate. □

On a finite space, information is a partition, and more information is a finer one: $\mathcal{G} \subseteq \mathcal{H}$ if and only if each atom of $\mathcal{G}$ is a union of atoms of $\mathcal{H}$.

Example 1.4 (What a price reveals)**.** In the two-toss model, the atoms of $\mathcal{F}_1$ are A_H and A_T, and those of $\mathcal{F}_2$ are the singletons. Since S_2 takes the values 9, 3, 1, the atoms of $\sigma(S_2)$ are $\{HH\}$, $\{HT, TH\}$, $\{TT\}$: observing S_2 reveals how many heads occurred, but not in which order. So S_2 is $\mathcal{F}_2$- but not $\mathcal{F}_1$-measurable; S_1 is not $\sigma(S_2)$-measurable, since it differs on HT and TH; and neither of $\mathcal{F}_1$ and $\sigma(S_2)$ contains the other.

Proposition 1.5 (Functions of known quantities are known)**.** *If $X_1, \dots, X_k$ are $\mathcal{G}$-measurable and $f : \mathbb{R}^k \to \mathbb{R}$ is Borel, then $f(X_1, \dots, X_k)$ is $\mathcal{G}$-measurable.*

Proof. The vector $X = (X_1, \dots, X_k)$ is $\mathcal{G}$-measurable, and $\{f(X) \in B\} = \{X \in f^{-1}(B)\}$ with $f^{-1}(B)$ Borel. □

Conversely, by the *Doob–Dynkin lemma*, a $\sigma(Y)$-measurable X is $f(Y)$ for a Borel f. On a finite space, X is constant on each level set $\{Y = y\}$ (Theorem 1.3): let $f(y)$ be that constant, and 0 off the range of Y. The general case is Exercise 1.4.

Remark 1.6 (What changes on a general space). On an infinite Ω, a σ-algebra need not come from a partition, and atoms may carry no probability: if X has a density, each $\{X = x\}$ is null, though observing X is very informative. Information is then described by events, and Section 1.4 defines conditional expectations through events.

1.2.3 Filtrations and adapted processes

Definition 1.7 (Filtration)**.** Let $\mathbb{T} \subseteq [0, \infty)$ be a set of dates, such as $\{0, \dots, n\}$, $\mathbb{N}$ or $[0, T]$. A *filtration* on $(\Omega, \mathcal{F})$ is a family $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ of σ-algebras with $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ whenever $s \le t$.

Example 1.8 (The coin-toss filtration)**.** On Ω_n, the information carried by the first k tosses is the σ-algebra $\mathcal{F}_k$ generated by the partition into the 2^k sets $A_w = \{\omega : \omega_1 \cdots \omega_k = w\}$, $w \in \{H, T\}^k$. Then $\mathcal{F}_0 = \{\emptyset, \Omega_n\}$, $\mathcal{F}_n = 2^{\Omega_n}$, and $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ because each atom splits in two, $A_w = A_{wH} \cup A_{wT}$: the atoms of $\mathcal{F}_k$ are the nodes at depth k of the binary tree of the tosses. The information carried by the later tosses is the σ-algebra $\mathcal{G}_k$ generated by the sets $\{\omega : \omega_{k+1} \cdots \omega_n = w'\}$, $w' \in \{H, T\}^{n-k}$.

Definition 1.9 (Stochastic process; adapted and predictable)**.** A *stochastic process* is a family $X = (X_t)_{t \in \mathbb{T}}$ of random variables; for fixed ω, $t \mapsto X_t(\omega)$ is a *path*. The process is *adapted* to $(\mathcal{F}_t)_{t \in \mathbb{T}}$ if X_t is $\mathcal{F}_t$-measurable for every t. In discrete time, $(X_k)_{k \ge 1}$ is *predictable* if X_k is $\mathcal{F}_{k-1}$-measurable for every $k \ge 1$.

The *natural filtration* $\mathcal{F}^X_t = \sigma(X_s : s \le t)$, the smallest σ-algebra containing every $\sigma(X_s)$ with $s \le t$, is the smallest filtration to which X is adapted. In the coin-toss model the price is adapted but not predictable, and since $u \ne d$, S_k / S_{k-1} reveals the k-th toss, so $\mathcal{F}^S_k = \mathcal{F}_k$. Adaptedness is the form of a constraint every trader faces: the number Δ_k of shares held from date k to $k+1$ must be chosen with what is known at date k. Without it there would be nothing to study: knowing ω, one would buy whatever is about to go up.

1.3 Independence

Measurability means that $\mathcal{G}$ determines X. Independence is the opposite extreme: $\mathcal{G}$ says nothing about X.

Definition 1.10 (Independence)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

(i) σ-algebras $\mathcal{G}_1, \dots, \mathcal{G}_m \subseteq \mathcal{F}$ are *independent* if $\mathbb{P}(A_1 \cap \cdots \cap A_m) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_m)$ for all $A_i \in \mathcal{G}_i$. An infinite family is independent if every finite subfamily is.

(ii) Random variables or vectors $X_1, \dots, X_m$ are *independent* if $\sigma(X_1), \dots, \sigma(X_m)$ are, that is, $\mathbb{P}(X_1 \in B_1, \dots, X_m \in B_m) = \prod_i \mathbb{P}(X_i \in B_i)$ for all Borel B_i. A random variable X is *independent of* $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent. Events are independent if their indicators are.

Functions of independent variables are independent, since $\sigma(f(X)) \subseteq \sigma(X)$. Independence is a property of the measure: the same variables can be independent under one measure and not under another (Exercise 1.2). For variables with finitely many values it suffices to check $\mathbb{P}(X_1 = x_1, \dots, X_m = x_m) = \prod_i \mathbb{P}(X_i = x_i)$ and sum over $x_i \in B_i$; in general, half-lines $B_i = (-\infty, b_i]$ suffice, by Dynkin's π–λ theorem (Durrett (2019), Section 2.1).

Lemma 1.11 (Past and future tosses)**.** *Under the p-coin the tosses are independent with $\mathbb{P}(\omega_j = H) = p$, and for every k the σ-algebras $\mathcal{F}_k$ and $\mathcal{G}_k$ are independent.*

Proof. For a word v of length ℓ with h heads, let $\pi(v) = p^h q^{\ell-h}$. For words w, w' of lengths k, $n-k$, the event $A_w \cap \{\omega_{k+1}\cdots\omega_n = w'\}$ is the outcome ww', of probability $\pi(w)\pi(w')$. Summing over w', or over w, gives $\mathbb{P}(A_w) = \pi(w)$ and $\mathbb{P}(\omega_{k+1}\cdots\omega_n = w') = \pi(w')$, since $(p+q)^\ell = 1$. So the product rule holds for atoms, and by summation for all $A \in \mathcal{F}_k$, $B \in \mathcal{G}_k$. Summing $\mathbb{P}(A_w) = \pi(w)$ over the words of length j ending in H gives $\mathbb{P}(\omega_j = H) = p$, so $\mathbb{P}(\omega_1 = a_1, \ldots, \omega_n = a_n) = \pi(a_1\cdots a_n) = \prod_j \mathbb{P}(\omega_j = a_j)$: the tosses are independent. □

In particular $H_n - H_k$, the number of heads among the tosses after date k, is $\mathcal{G}_k$-measurable, independent of $\mathcal{F}_k$, and binomial with parameters $n-k$, p.

1.4 Conditional expectation

Between the two extremes, information tells us something about X without determining it. The conditional expectation is the best estimate of X that the information allows.

1.4.1 Conditioning on a partition

For an event A with $\mathbb{P}(A) > 0$, the elementary conditional expectation $\mathbb{E}[X \mid A] = \mathbb{E}[X\mathbf{1}_A]/\mathbb{P}(A)$ is the average of X over A.

Definition 1.12 (Conditional expectation given a partition)**.** Let $\mathcal{P} = \{A_1, \ldots, A_m\}$ be a partition of Ω into events, $\mathcal{G} = \sigma(\mathcal{P})$, and X integrable. The *conditional expectation of X given $\mathcal{G}$* is

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{i:\,\mathbb{P}(A_i)>0} \frac{\mathbb{E}[X\mathbf{1}_{A_i}]}{\mathbb{P}(A_i)}\,\mathbf{1}_{A_i}. \tag{1.3}$$

If Y takes finitely many values, $\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)] = \varphi(Y)$, where $\varphi(y) = \mathbb{E}[X \mid Y = y]$ if $\mathbb{P}(Y = y) > 0$ and $\varphi(y) = 0$ otherwise.

It is a random variable, not a number: once we learn which atom contains ω, our estimate of X is its average over that atom. On null atoms any value would do.

Example 1.13 (Two tosses)**.** Under the p-coin, $\mathbb{E}[S_2 \mid \mathcal{F}_1] = (9p^2 + 3pq)/p = 9p + 3q$ on A_H and $(3pq + q^2)/q = 3p + q$ on A_T. As S_1 is 6 on A_H and 2 on A_T,

$$\mathbb{E}[S_2 \mid \mathcal{F}_1] = (pu + qd)\,S_1 = (\tfrac{1}{2} + p)S_1, \qquad \mathbb{E}[S_2] = (\tfrac{1}{2} + p)^2 S_0.$$

For $p = \frac{1}{2}$, the forecast of tomorrow's price is today's price. Conditioning can also look backwards: on $\{S_2 = 3\} = \{HT, TH\}$, $\mathbb{E}[S_1 \mid S_2] = (6pq + 2pq)/(2pq) = 4$ whatever p.

The variable (1.3) (i) is $\mathcal{G}$-measurable, being constant on atoms, and (ii) satisfies $\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}]\mathbf{1}_A\big] = \mathbb{E}[X\mathbf{1}_A]$ for every $A \in \mathcal{G}$: as A is a disjoint union of atoms, take $A = A_i$, where both sides equal $\mathbb{E}[X\mathbf{1}_{A_i}]$. Conversely, an integrable Y' with (i) and (ii) equals a constant c_i on each A_i (Theorem 1.3), and (ii) with $A = A_i$ gives $c_i\mathbb{P}(A_i) = \mathbb{E}[X\mathbf{1}_{A_i}]$: Y' agrees with (1.3) on every atom of positive probability, that is, almost surely.

1.4.2 The general definition

Properties (i) and (ii) make sense for any σ-algebra, with no atoms in sight, and they determine the conditional expectation up to a null set. They become the definition.

Definition 1.14 (Conditional expectation)**.** Let X be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. A *conditional expectation of X given $\mathcal{G}$* is an integrable random variable Y such that

(i) Y is $\mathcal{G}$-measurable;

(ii) $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$ for every $A \in \mathcal{G}$ *(partial averaging)*.

We write $Y = \mathbb{E}[X \mid \mathcal{G}]$, $\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \sigma(Y)]$ for a random variable or vector Y, and $\mathbb{P}(A \mid \mathcal{G}) = \mathbb{E}[\mathbf{1}_A \mid \mathcal{G}]$.

The estimate uses only the information $\mathcal{G}$, and on every event that $\mathcal{G}$ decides it has the same average as X; with $A = \Omega$, $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$.

Theorem 1.15 (Existence and uniqueness)**.** *Let X be integrable and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. A conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ exists; any two are equal almost surely; and $\mathbb{E}|\mathbb{E}[X \mid \mathcal{G}]| \leq \mathbb{E}|X|$.*

Proof. Uniqueness. Let Y, Y' be two conditional expectations and $A = \{Y > Y'\}$, which is in $\mathcal{G}$ by Theorem 1.5. By (ii), $W = (Y - Y')\mathbf{1}_A \geq 0$ has $\mathbb{E}[W] = \mathbb{E}[X\mathbf{1}_A] - \mathbb{E}[X\mathbf{1}_A] = 0$. Then $W = 0$ almost surely: by Markov's inequality $\mathbb{P}(W \geq 1/j) \leq j\,\mathbb{E}[W] = 0$ for all $j \geq 1$, and $\{W > 0\} = \bigcup_j \{W \geq 1/j\}$. As $W > 0$ on A, $\mathbb{P}(A) = 0$; exchanging Y and Y' gives $\mathbb{P}(Y < Y') = 0$. *The bound.* With $A = \{Y > 0\} \in \mathcal{G}$, $\mathbb{E}|Y| = \mathbb{E}[Y\mathbf{1}_A] - \mathbb{E}[Y\mathbf{1}_{A^c}] = \mathbb{E}[X\mathbf{1}_A] - \mathbb{E}[X\mathbf{1}_{A^c}] \leq \mathbb{E}|X|$. *Existence.* For partitions, and so on finite spaces, (1.3) works. In general, for $X \geq 0$, $\nu(A) = \mathbb{E}[X\mathbf{1}_A]$ is a finite measure on $(\Omega, \mathcal{G})$ vanishing on $\mathbb{P}$-null sets, and its Radon–Nikodym density with respect to $\mathbb{P}$ on $\mathcal{G}$ satisfies (i) and (ii); for general X, use $X = X^+ - X^-$. See Durrett (2019), Section 4.1, or Williams (1991), Sections 6.11 and 9.5, which avoids Radon–Nikodym by starting from the existence of orthogonal projections in L^2; compare Theorem 1.21. □

Every identity or inequality between conditional expectations holds almost surely; we will not repeat it. If every outcome of a finite space has positive probability, "almost surely" means "everywhere".

1.4.3 Properties

Proposition 1.16 (Properties of conditional expectation)**.** *Let X and Y be integrable, $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ σ-algebras, and $a, b \in \mathbb{R}$.*

(i) Linearity: $\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$.

(ii) Positivity: *if $X \geq 0$, then $\mathbb{E}[X \mid \mathcal{G}] \geq 0$. Hence $X \leq Y$ implies $\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}]$, and $|\mathbb{E}[X \mid \mathcal{G}]| \leq \mathbb{E}[|X| \,|\, \mathcal{G}]$.*

(iii) Taking out what is known: *if Z is $\mathcal{G}$-measurable and ZX is integrable (for instance, if Z is bounded), then $\mathbb{E}[ZX \mid \mathcal{G}] = Z\,\mathbb{E}[X \mid \mathcal{G}]$. In particular, $\mathbb{E}[Z \mid \mathcal{G}] = Z$ for every integrable $\mathcal{G}$-measurable Z.*

(iv) Tower property: *if* $\mathcal{H} \subseteq \mathcal{G}$, *then* $\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \,\big|\, \mathcal{H}\big] = \mathbb{E}[X \mid \mathcal{H}]$.

(v) Independence: *if* X *is independent of* $\mathcal{G}$, *then* $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$.

(vi) Conditional Jensen: *if* $\phi : \mathbb{R} \to \mathbb{R}$ *is convex and* $\phi(X)$ *is integrable, then* $\phi(\mathbb{E}[X \mid \mathcal{G}]) \le \mathbb{E}[\phi(X) \mid \mathcal{G}]$.

Proof. Each time we check (i) and (ii) of Theorem 1.14 for a candidate and conclude by uniqueness. (i) $a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$ is $\mathcal{G}$-measurable, and its integral over $A \in \mathcal{G}$ is $a\mathbb{E}[X\mathbf{1}_A] + b\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[(aX + bY)\mathbf{1}_A]$.

(ii) Let $W = \mathbb{E}[X \mid \mathcal{G}]$ and $A = \{W < 0\} \in \mathcal{G}$. Then $W\mathbf{1}_A \le 0$ while $\mathbb{E}[W\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A] \ge 0$, so $W\mathbf{1}_A = 0$ almost surely and $\mathbb{P}(A) = 0$. The other claims follow by applying this to $Y - X$ and to $|X| \pm X$. (iii) Let $W = \mathbb{E}[X \mid \mathcal{G}]$. If $Z = \mathbf{1}_B$ with $B \in \mathcal{G}$, then for $A \in \mathcal{G}$, $\mathbb{E}[\mathbf{1}_B W\mathbf{1}_A] = \mathbb{E}[W\mathbf{1}_{A\cap B}] = \mathbb{E}[X\mathbf{1}_{A\cap B}] = \mathbb{E}[\mathbf{1}_B X\mathbf{1}_A]$. By linearity the claim holds for every $\mathcal{G}$-measurable Z with finitely many values, which on a finite space is every Z. In general, let first $X, Z \ge 0$, and let $Z_j = \min(j, 2^{-j}\lfloor 2^j Z\rfloor)$; these take finitely many values, are $\mathcal{G}$-measurable, and increase to Z. For $A \in \mathcal{G}$, $\mathbb{E}[Z_j W\mathbf{1}_A] = \mathbb{E}[Z_j X\mathbf{1}_A]$, and $W \ge 0$ by (ii), so monotone convergence gives

$$\mathbb{E}[ZW\mathbf{1}_A] = \mathbb{E}[ZX\mathbf{1}_A], \tag{1.4}$$

an identity in $[0, \infty]$ that holds even if ZX is not integrable. Here it is, so $A = \Omega$ shows ZW integrable, and $ZW = \mathbb{E}[ZX \mid \mathcal{G}]$. For general signs, each of the four products $Z^{\pm}X^{\pm}$ is integrable, being at most $|ZX|$; the nonnegative case and (i) give $\mathbb{E}[ZX \mid \mathcal{G}] = (Z^+ - Z^-)(\mathbb{E}[X^+ \mid \mathcal{G}] - \mathbb{E}[X^- \mid \mathcal{G}]) = ZW$.

(iv) $\mathbb{E}[X \mid \mathcal{H}]$ is $\mathcal{H}$-measurable, and for $A \in \mathcal{H} \subseteq \mathcal{G}$, $\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{H}]\mathbf{1}_A\big] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}]\mathbf{1}_A\big]$, which is the defining property of $\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \,\big|\, \mathcal{H}\big]$.

(v) We need $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X]\mathbb{P}(A)$ for $A \in \mathcal{G}$. If X takes finitely many values, this is the product rule for $\{X = x\}$ and A, summed over x. For $X \ge 0$, apply it to the functions $X_j = \min(j, 2^{-j}\lfloor 2^j X\rfloor)$ of X, which increase to X, and use monotone convergence. In general, use $X = X^+ - X^-$.

(vi) A convex ϕ on $\mathbb{R}$ is continuous and lies above its tangent lines: with a_c the right derivative at c, $\phi(x) \ge \phi(c) + a_c(x - c)$ for all x, and $c \mapsto a_c$ is nondecreasing, hence bounded on bounded sets. By (i) and (ii), for each c, $\mathbb{E}[\phi(X) \mid \mathcal{G}] \ge \phi(c) + a_c\big(\mathbb{E}[X \mid \mathcal{G}] - c\big)$ outside a null set. Outside the union of the countably many null sets for rational c, this holds for all rational c at once; letting $c \to \mathbb{E}[X \mid \mathcal{G}]$ along the rationals, the right side tends to $\phi(\mathbb{E}[X \mid \mathcal{G}])$. □

On a finite space, $\mathbb{E}[\,\cdot \mid \mathcal{G}]$ is on each atom A_i of positive probability the expectation under $\mathbb{P}(\,\cdot \mid A_i)$, so (i), (ii) and (vi) are properties of ordinary expectations, atom by atom. If $\mathcal{G} \subseteq \mathcal{H}$, (iii) gives $\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \,\big|\, \mathcal{H}\big] = \mathbb{E}[X \mid \mathcal{G}]$, so with (iv) the smaller σ-algebra wins in either order; when neither contains the other, the order matters (Exercise 1.3).

Remark 1.17 (Nonnegative variables). For $X \ge 0$, not necessarily integrable, $\mathbb{E}[X \mid \mathcal{G}]$ is defined with values in $[0, \infty]$ as the limit of $\mathbb{E}[\min(X, j) \mid \mathcal{G}]$, which is nondecreasing in j by positivity. By monotone convergence on both sides, it is $\mathcal{G}$-measurable and satisfies (ii) of Theorem 1.14 as an identity in $[0, \infty]$. Any other $\mathcal{G}$-measurable $Y' \ge 0$ with this property agrees with it almost surely: with Y the limit and $A = \{Y' < Y,\ Y' \le m\} \in \mathcal{G}$, $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[Y'\mathbf{1}_A] \le m$, so $(Y - Y')\mathbf{1}_A \ge 0$ has zero mean and $\mathbb{P}(A) = 0$; let $m \to \infty$ and exchange Y, Y'. When X is integrable this is the usual $\mathbb{E}[X \mid \mathcal{G}]$, and (1.4) says that $\mathbb{E}[ZX \mid \mathcal{G}] = Z\,\mathbb{E}[X \mid \mathcal{G}]$ for all $\mathcal{G}$-measurable $Z \ge 0$.

Example 1.18 (A path-dependent payoff)**.** Figure 1.1 shows the running maximum $X = \max_{j\le 4} S_j$ of the stock (1.2) with $S_0 = 4$, $u = \frac{3}{2}$, $d = \frac{1}{2}$ under the 0.6-coin, and $\mathbb{E}[X \mid \mathcal{F}_k]$ for $k \le 3$, computed with `code/src/ieor222/cointoss.py`. Each step function averages the next one over coarser atoms: the tower property. On A_{HT} and A_{TH} the price is the same, $S_2 = 3$, yet $\mathbb{E}[X \mid \mathcal{F}_2]$ is 6.27 on the first and 5.11 on the second: a path that went up first has recorded a higher maximum, and $\mathbb{E}[X \mid \mathcal{F}_2]$ is not a function of S_2 alone.

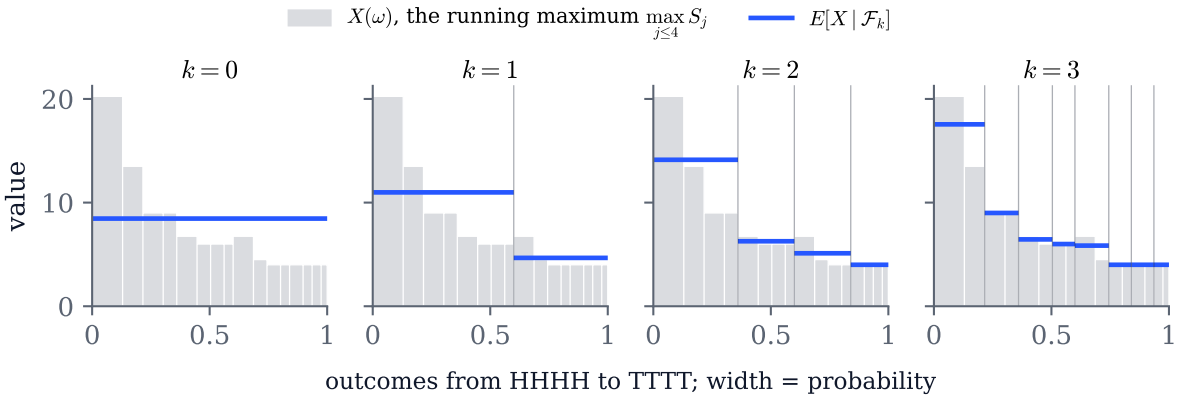

Figure 1.1: The running maximum X of Theorem 1.18 on the 16 outcomes of four tosses, each with width equal to its probability, and $\mathbb{E}[X \mid \mathcal{F}_k]$ for $k = 0, \dots, 3$. Vertical lines separate the atoms of $\mathcal{F}_k$; on each atom the blue level is the probability-weighted average of X, so the areas above and below it balance.

▶ Interactive demo: conditional expectations on the coin-toss tree, for any n, p and payoff.

1.4.4 The independence lemma

To condition on $\mathcal{G}$, freeze what $\mathcal{G}$ knows and average over what it does not.

Lemma 1.19 (Independence lemma)**.** *Let $\mathcal{G} \subseteq \mathcal{F}$, let X be a $\mathcal{G}$-measurable random vector in $\mathbb{R}^k$ and Y a random vector in $\mathbb{R}^l$ independent of $\mathcal{G}$, and let $f : \mathbb{R}^{k+l} \to \mathbb{R}$ be a Borel function such that $f \ge 0$ or $f(X, Y)$ is integrable. Then*

$$\mathbb{E}[f(X, Y) \,|\, \mathcal{G}] = g(X), \qquad \text{where } g(x) = \mathbb{E}[f(x, Y)].$$

If $f \ge 0$, g takes values in $[0, \infty]$ and the left side is understood as in Theorem 1.17. If $f(X, Y)$ is integrable, $f(x, Y)$ is integrable for $\mathbb{P}_X$-almost every x, where $\mathbb{P}_X$ is the law of X; g is defined at those x and set to 0 elsewhere.

Proof. Finitely many values (all that is needed on a finite space). Let X take the values x_i and Y the values y_j. Then $g(x) = \sum_j f(x, y_j)\mathbb{P}(Y = y_j)$ is Borel, so $g(X)$ is $\mathcal{G}$-measurable. For $A \in \mathcal{G}$, the event $A \cap \{X = x_i\} \in \mathcal{G}$ is independent of $\{Y = y_j\}$, so

$$\mathbb{E}[f(X, Y)\mathbf{1}_A] = \sum_{i,j} f(x_i, y_j)\mathbb{P}(A \cap \{X = x_i\})\mathbb{P}(Y = y_j) = \sum_i g(x_i)\mathbb{P}(A \cap \{X = x_i\}),$$

which is $\mathbb{E}[g(X)\mathbf{1}_A]$. *General case.* Let $\mathcal{D}$ be the class of Borel sets $D \subseteq \mathbb{R}^{k+l}$ such that $f = \mathbf{1}_D$ satisfies the lemma with $g_D(x) = \mathbb{P}((x, Y) \in D)$ Borel. For a rectangle $B \times C$, $g(x) = \mathbf{1}_B(x)\mathbb{P}(Y \in C)$ and, for $A \in \mathcal{G}$, $\mathbb{E}[\mathbf{1}_B(X)\mathbf{1}_C(Y)\mathbf{1}_A] = \mathbb{P}(A \cap \{X \in B\})\mathbb{P}(Y \in C) = \mathbb{E}[g(X)\mathbf{1}_A]$ by independence. So $\mathcal{D}$ contains the rectangles, a class closed under intersections, and $\mathbb{R}^{k+l}$; it is closed under proper differences (subtract the identities) and increasing unions (monotone convergence, in g and in the identity). By Dynkin's π–λ theorem, $\mathcal{D} = \mathcal{B}(\mathbb{R}^{k+l})$. By linearity the lemma holds for nonnegative Borel f with finitely many values. For Borel $f \geq 0$, take such f_j increasing to f: the g_j increase to g, which is Borel, and monotone convergence in $\mathbb{E}[f_j(X, Y)\mathbf{1}_A] = \mathbb{E}[g_j(X)\mathbf{1}_A]$ gives $\mathbb{E}[f(X, Y)\mathbf{1}_A] = \mathbb{E}[g(X)\mathbf{1}_A]$ in $[0, \infty]$ for all $A \in \mathcal{G}$, which is the claim. If $f(X, Y)$ is integrable, apply this to $f^{\pm}$: $\mathbb{E}[g^{\pm}(X)] = \mathbb{E}[f^{\pm}(X, Y)] < \infty$, so the Borel set N where g^+ or g^- is infinite has $\mathbb{P}(X \in N) = 0$. Off N, $f(x, Y)$ is integrable and $g = g^+ - g^-$; so $g(X) = g^+(X) - g^-(X)$ almost surely, and linearity gives the claim. □

The lemma is how option prices at intermediate dates are computed throughout the course.

Example 1.20 (Forecasting a function of the price)**.** Fix $k \leq n$ and a function h, and write $S_n = S_k R$ with $R = u^{H_n - H_k} d^{(n-k)-(H_n - H_k)}$. Here S_k is $\mathcal{F}_k$-measurable, and R is $\mathcal{G}_k$-measurable, hence independent of $\mathcal{F}_k$, with $H_n - H_k$ binomial (Theorem 1.11). The independence lemma with $f(x, y) = h(xy)$ gives

$$\mathbb{E}[h(S_n) \,|\, \mathcal{F}_k] = g(S_k), \qquad g(x) = \sum_{j=0}^{n-k} \binom{n-k}{j} p^j q^{n-k-j}\, h(x\, u^j d^{n-k-j}). \tag{1.5}$$

The forecast depends on the past only through the current price (the Markov property); with $h(x) = x$, $\mathbb{E}[S_n \mid \mathcal{F}_k] = (pu + qd)^{n-k} S_k$. With p replaced by the risk-neutral probability, (1.5) is the binomial pricing formula of Chapter 4.

1.4.5 The best predictor

Write L^2 for the square-integrable random variables, and $L^2(\mathcal{G})$ for those that are also $\mathcal{G}$-measurable.

Proposition 1.21 (Conditional expectation as orthogonal projection)**.** *Let $X \in L^2$ and $\mathcal{G} \subseteq \mathcal{F}$. Then $\mathbb{E}[X \mid \mathcal{G}] \in L^2(\mathcal{G})$,*

$$\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])\, Z] = 0 \qquad \text{for all } Z \in L^2(\mathcal{G}), \tag{1.6}$$

and for every $Z \in L^2(\mathcal{G})$,

$$\mathbb{E}[(X - Z)^2] = \mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])^2] + \mathbb{E}[(\mathbb{E}[X \mid \mathcal{G}] - Z)^2]. \tag{1.7}$$

Hence $\mathbb{E}[X \mid \mathcal{G}]$ minimizes $\mathbb{E}[(X - Z)^2]$ over $L^2(\mathcal{G})$, and any other minimizer equals it almost surely.

Proof. Let $Y = \mathbb{E}[X \mid \mathcal{G}]$. Conditional Jensen with $\phi(x) = x^2$ gives $Y^2 \leq \mathbb{E}[X^2 \mid \mathcal{G}]$, so $\mathbb{E}[Y^2] \leq \mathbb{E}[X^2]$. For $Z \in L^2(\mathcal{G})$, ZX is integrable since $|ZX| \leq \frac{1}{2}(Z^2 + X^2)$, and by the tower property and taking out what is known, $\mathbb{E}[ZX] = \mathbb{E}[\mathbb{E}[ZX \mid \mathcal{G}]] = \mathbb{E}[ZY]$: this is (1.6). Expanding $(X - Z)^2 = \big((X - Y) + (Y - Z)\big)^2$, the cross term vanishes by (1.6) applied to $Y - Z$, which gives (1.7). Its last term is zero only if $Z = Y$ almost surely. □

So $\mathbb{E}[X \mid \mathcal{G}]$ is the orthogonal projection of X onto $L^2(\mathcal{G})$: the best mean-square prediction of X that uses only the information $\mathcal{G}$. For trivial $\mathcal{G}$ the predictors are constants, and the best is $\mathbb{E}[X]$, with error $\mathrm{Var}(X)$. Taking $Z = \mathbb{E}[X]$ in (1.7) splits the variance into residual and explained parts, $\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X \mid \mathcal{G}])^2\big] + \mathrm{Var}\left(\mathbb{E}[X \mid \mathcal{G}]\right)$ (Exercise 1.6). For $\mathcal{G} = \sigma(Y)$, $L^2(\mathcal{G})$ consists of the square-integrable functions of Y (Doob–Dynkin): $\mathbb{E}[X \mid Y]$ is the best predictor among *all* functions of Y, the target of nonparametric regression. Linear regression projects onto the affine functions $a + bY$ only, and by (1.7) its mean squared error is larger by $\mathbb{E}\big[(\mathbb{E}[X \mid Y] - a - bY)^2\big]$; the two agree exactly when $\mathbb{E}[X \mid Y]$ is affine, as for Gaussian vectors (Exercise 1.5). Figure 1.2 shows a case where they do not.

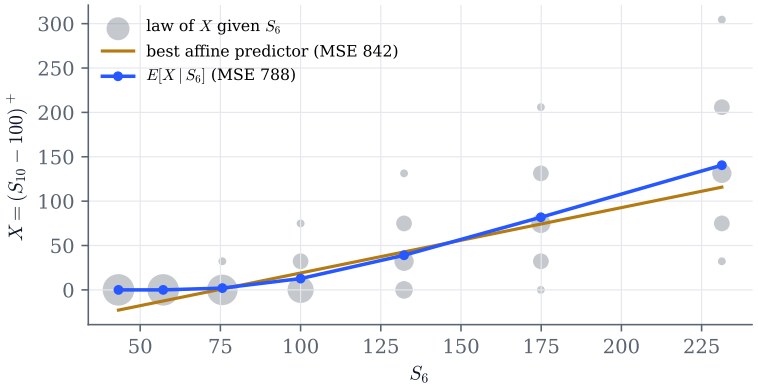

Figure 1.2: The call payoff $X = (S_{10} - 100)^+$ ($S_0 = 100$, $u = 1.15$, $d = 1/u$, $p = \frac{1}{2}$) against S_6. Grey: the conditional law of X given S_6, with area proportional to probability. Blue: $\mathbb{E}[X \mid S_6]$, from (1.5). Gold: the best affine predictor, with a larger mean squared error (MSE).

▶ Interactive demo: polynomial regressions of a payoff on S_k that reach $\mathbb{E}[X \mid S_k]$.

1.5 Changing the probability measure

In the binomial model, prices will be expectations under a coin whose probability of heads differs from the real-world p (Chapter 4); in continuous time, under a measure that changes the drift of the stock (Chapter 11). Outcomes, events and information stay the same; only the weights of the outcomes change.

Definition 1.22 (Equivalent measures; Radon–Nikodym derivative)**.** Probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ are *equivalent* if they have the same null sets: $\mathbb{P}(A) = 0$ if and only if $\mathbb{Q}(A) = 0$. Let Ω be finite, $\mathcal{F} = 2^\Omega$, and $\mathbb{P}(\omega) > 0$ for every ω. Then $\mathbb{Q}$ is equivalent to $\mathbb{P}$ if and only if $\mathbb{Q}(\omega) > 0$ for every ω, and the *Radon–Nikodym derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$ is the random variable

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(\omega) = Z(\omega) := \frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)}.$$

(If every $\mathbb{Q}(\omega) > 0$, the only null set of either measure is $\emptyset$; outcomes that $\mathbb{P}$ rules out can be removed from Ω.) Equivalent measures agree on what is possible, not on how likely it is. Each outcome is reweighted by Z: $\mathbb{Q}(\omega) = Z(\omega)\mathbb{P}(\omega)$, so $Z > 0$ and $\mathbb{E}[Z] = 1$. For example, the p-coin

$\mathbb{P}$ and the $\tilde{p}$-coin $\mathbb{Q}$ on Ω_n, with $p, \tilde{p} \in (0,1)$ and $\tilde{q} = 1 - \tilde{p}$, are equivalent, and

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \Big(\frac{\tilde{p}}{p}\Big)^{H_n} \Big(\frac{\tilde{q}}{q}\Big)^{n-H_n}. \tag{1.8}$$

Proposition 1.23 (Change of measure)**.** *Let $Z \geq 0$ be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[Z] = 1$, and let $\mathbb{Q}(A) = \mathbb{E}[Z\mathbf{1}_A]$ for $A \in \mathcal{F}$.*

(a) $\mathbb{Q}$ is a probability measure, every $\mathbb{P}$-null set is $\mathbb{Q}$-null, and

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}[ZX] \qquad \text{for every } X \geq 0 \text{ and every } X \text{ with } \mathbb{E}[Z|X|] < \infty. \tag{1.9}$$

(b) If $Z > 0$ almost surely, then $\mathbb{Q}$ is equivalent to $\mathbb{P}$, and $\mathbb{E}[X] = \mathbb{E}^{\mathbb{Q}}[X/Z]$ for $X \geq 0$ and for integrable X: $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q} = 1/Z$.

(c) On a finite space as in Theorem 1.22, every $\mathbb{Q}$ equivalent to $\mathbb{P}$ is of this form, with $Z = \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$.

Proof. On a finite space, $\mathbb{E}^{\mathbb{Q}}[X] = \sum_\omega X(\omega)Z(\omega)\mathbb{P}(\omega) = \mathbb{E}[ZX]$, which is (c) and (1.9). In general, $\mathbb{Q}(\Omega) = \mathbb{E}[Z] = 1$, and countable additivity follows from $Z\mathbf{1}_{\bigcup_j A_j} = \sum_j Z\mathbf{1}_{A_j}$ (disjoint A_j) by monotone convergence. If $\mathbb{P}(A) = 0$, then $Z\mathbf{1}_A = 0$ almost surely and $\mathbb{Q}(A) = 0$. Identity (1.9) holds for indicators, then for finitely many values by linearity, for $X \geq 0$ by monotone convergence, and in general by $X = X^+ - X^-$. (b) If $\mathbb{Q}(A) = 0$, then $Z\mathbf{1}_A = 0$ almost surely, so $\mathbb{P}(A) = 0$; and (1.9) for X/Z gives $\mathbb{E}^{\mathbb{Q}}[X/Z] = \mathbb{E}[X]$. □

On a general space, the converse is the Radon–Nikodym theorem: if $\mathbb{Q}$ is equivalent to $\mathbb{P}$, there is $Z > 0$, unique almost surely, with $\mathbb{E}[Z] = 1$ and $\mathbb{Q}(A) = \mathbb{E}[Z\mathbf{1}_A]$ for all $A \in \mathcal{F}$, written $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ (Shreve (2004b), Section 1.6). A change of measure leaves alone events, random variables, filtrations and measurability, and between equivalent measures "almost surely" means the same thing. What changes are expectations, laws, independence, and conditional expectations, which transform as follows.

Proposition 1.24 (Bayes' formula)**.** *Let Ω be finite, $\mathcal{F} = 2^\Omega$, $\mathbb{P}$ and $\mathbb{Q}$ equivalent with $\mathbb{P}(\omega) > 0$ for every ω, and $Z = \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$. For every σ-algebra $\mathcal{G}$ and every random variable X,*

$$\mathbb{E}^{\mathbb{Q}}[X \mid \mathcal{G}] = \frac{\mathbb{E}[ZX \mid \mathcal{G}]}{\mathbb{E}[Z \mid \mathcal{G}]}. \tag{1.10}$$

Proof. By Theorem 1.3, $\mathcal{G} = \sigma(\mathcal{P})$ for a partition $\mathcal{P}$, whose atoms have positive probability under both measures. On an atom A, by (1.3) under $\mathbb{Q}$ and (1.9),

$$\mathbb{E}^{\mathbb{Q}}[X \mid \mathcal{G}] = \frac{\mathbb{E}^{\mathbb{Q}}[X\mathbf{1}_A]}{\mathbb{Q}(A)} = \frac{\mathbb{E}[ZX\mathbf{1}_A]}{\mathbb{E}[Z\mathbf{1}_A]} = \frac{\mathbb{E}[ZX\mathbf{1}_A]/\mathbb{P}(A)}{\mathbb{E}[Z\mathbf{1}_A]/\mathbb{P}(A)} = \frac{\mathbb{E}[ZX \mid \mathcal{G}]}{\mathbb{E}[Z \mid \mathcal{G}]}.$$ □

On each atom, the $\mathbb{Q}$-average of X is the Z-weighted $\mathbb{P}$-average, normalized by the average weight. On a general space, with Z as in Theorem 1.23 and $\mathbb{E}^{\mathbb{Q}}|X| < \infty$, (1.10) holds $\mathbb{Q}$-almost surely (Exercise 1.8; Shreve (2004b), Lemma 5.2.2, is the special case $Z > 0$ in density-process form).

Remark 1.25 (The density process). For $A \in \mathcal{G}$, $\mathbb{Q}(A) = \mathbb{E}[\mathbb{E}[Z \mid \mathcal{G}]\mathbf{1}_A]$: on the information $\mathcal{G}$, the density of $\mathbb{Q}$ is $\mathbb{E}[Z \mid \mathcal{G}]$. Given a filtration, the *density process* $Z_k = \mathbb{E}[Z \mid \mathcal{F}_k]$ satisfies

$\mathbb{E}[Z_j \mid \mathcal{F}_k] = Z_k$ for $k \le j$ (a martingale, Chapter 2). For $\mathcal{F}_j$-measurable X and $k \le j$, $\mathbb{E}[ZX \mid \mathcal{F}_k] = \mathbb{E}\big[\mathbb{E}[ZX \mid \mathcal{F}_j] \,\big|\, \mathcal{F}_k\big] = \mathbb{E}[Z_j X \mid \mathcal{F}_k]$, so Bayes' formula reads $\mathbb{E}^{\mathbb{Q}}[X \mid \mathcal{F}_k] = \mathbb{E}[Z_j X \mid \mathcal{F}_k]/Z_k$: the form used for Girsanov's theorem. For the two coins, see Exercise 1.7.

1.5.1 The Gaussian shift

On a general space, the most important change of measure is the following.

Proposition 1.26 (Gaussian shift)**.** *Let $X \sim \mathcal{N}(0,1)$ under $\mathbb{P}$, $\theta \in \mathbb{R}$, and*

$$Z = \exp\Big(\theta X - \frac{\theta^2}{2}\Big), \qquad \mathbb{Q}(A) = \mathbb{E}[Z\mathbf{1}_A], \quad A \in \mathcal{F}.$$

Then $Z > 0$, $\mathbb{E}[Z] = 1$, $\mathbb{Q}$ is equivalent to $\mathbb{P}$, and $X \sim \mathcal{N}(\theta, 1)$ under $\mathbb{Q}$; equivalently, $X - \theta$ is standard normal under $\mathbb{Q}$. More precisely, $\mathbb{E}^{\mathbb{Q}}[h(X)] = \mathbb{E}[h(X+\theta)]$ for every Borel $h \ge 0$ and every Borel h with $\mathbb{E}|h(X+\theta)| < \infty$.

Proof. Let φ be the standard normal density. Completing the square gives $e^{\theta x - \theta^2/2}\varphi(x) = \varphi(x-\theta)$, so for Borel $h \ge 0$, as X has density φ under $\mathbb{P}$,

$$\mathbb{E}\big[Z\,h(X)\big] = \int_{\mathbb{R}} h(x)\, e^{\theta x - \theta^2/2}\varphi(x)\,\mathrm{d}x = \int_{\mathbb{R}} h(x)\varphi(x-\theta)\,\mathrm{d}x = \mathbb{E}\big[h(X+\theta)\big].$$

With $h = 1$, $\mathbb{E}[Z] = 1$, so Theorem 1.23 applies, and $Z > 0$ makes $\mathbb{Q}$ equivalent to $\mathbb{P}$. By (1.9) the left side is $\mathbb{E}^{\mathbb{Q}}[h(X)]$. With $h = \mathbf{1}_{(-\infty,b]}$, $\mathbb{Q}(X \le b) = \mathbb{P}(X + \theta \le b)$, the $\mathcal{N}(\theta,1)$ distribution function at b. For integrable $h(X+\theta)$, use $h = h^+ - h^-$. □

For $\theta > 0$, the weight Z exceeds one exactly where $X > \theta/2$ (for $\theta < 0$, where $X < \theta/2$), and it moves probability in exactly the way that shifts the Gaussian without deforming it. Averaging $Z\,h(X)$ over samples drawn under $\mathbb{P}$ estimates $\mathbb{E}^{\mathbb{Q}}[h(X)]$: this is *importance sampling* (Exercise 1.9). The cost: $\mathbb{E}[Z^2] = e^{\theta^2}$, so for large $|\theta|$ a few samples carry most of the weight.

Remark 1.27 (A first look at Girsanov's theorem). Let W be a Brownian motion with a filtration $(\mathcal{F}_t)$ (Theorems 5.6 and 5.11), $T > 0$, $\theta \in \mathbb{R}$, and $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = Z_T = \exp(\theta W_T - \frac{\theta^2}{2}T)$. For $0 < t \le T$, the density of $\mathbb{Q}$ on $\mathcal{F}_t$ is $\mathbb{E}[Z_T \mid \mathcal{F}_t] = \exp(\theta W_t - \frac{\theta^2}{2}t)$ (Theorem 5.12), which is the density of Theorem 1.26 for $X = W_t/\sqrt{t}$ and shift $\theta\sqrt{t}$; so $W_t \sim \mathcal{N}(\theta t, t)$ under $\mathbb{Q}$. Girsanov's theorem (Chapter 11) shows much more: $(W_t - \theta t)_{t \le T}$ is a $\mathbb{Q}$-Brownian motion. An exponential change of measure adds a drift; this is how risk-neutral pricing replaces the expected return of a stock by the interest rate.

▶ Interactive demo: reweighting a fair coin into a biased one, and $\mathcal{N}(0,1)$ samples into $\mathcal{N}(\theta,1)$ samples, by $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$.

Further reading

Shreve (2004a), Chapters 2–3, and Shreve (2004b), Chapters 1–2, cover this material on coin-toss and general spaces; Williams (1991) and Durrett (2019) give the measure theory in full; Föllmer and Schied (2016), Chapter 1, brings equivalent measures to arbitrage.

Chapter 2

Martingales and random walks

A martingale is the mathematical form of a fair game: given everything known today, the expected value of tomorrow's position is today's position. Two facts make martingales the central objects of this course. In a market without arbitrage, discounted prices are martingales under a risk-neutral probability (Chapter 4), so that pricing a derivative comes down to computing a conditional expectation. And the gains of a trading strategy in a martingale are again a martingale: in a fair game no strategy, however clever, has a positive expected gain, as long as it uses only past information and does not take unbounded risks. The second fact is the discrete-time ancestor of the Itô integral (Chapter 7).

This chapter develops both ideas in discrete time, together with stopping times, the optional stopping theorem, and the gambler's ruin problem: how likely a random walk is to reach a target before a ruin level, and how long that takes. The same computations evaluate stop-loss and take-profit rules, and they return in continuous time for barrier options.

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(\mathcal{F}_n)_{n\in\mathbb{N}}$ a filtration (Theorem 1.7). We use the properties of conditional expectation collected in Theorem 1.16 (linearity, positivity, taking out what is known, the tower property, independence) and the independence lemma, Theorem 1.19. Identities between conditional expectations hold almost surely; we do not repeat it each time.

2.1 Martingales

Definition 2.1 (Martingale)**.** A process $M = (M_n)_{n\in\mathbb{N}}$ is a *martingale* with respect to $(\mathcal{F}_n)$ and $\mathbb{P}$ if

(i) M is adapted (Theorem 1.9): M_n is $\mathcal{F}_n$-measurable for every n;

(ii) M is integrable: $\mathbb{E}|M_n| < \infty$ for every n;

(iii) $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n$ for every $n \in \mathbb{N}$.

It is a *submartingale* if (iii) is replaced by $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \geq M_n$, and a *supermartingale* if it is replaced by $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \leq M_n$.

Read M_n as the fortune of a player after n rounds of a game. Condition (iii) says that, whatever has happened so far, the expected fortune after the next round is the current one: the game is

fair. A supermartingale is an unfavourable game (it is *super* for the casino), a submartingale a favourable one. A process with a finite horizon, $(M_n)_{0\le n\le N}$, is a martingale if (i)–(ii) hold for $n \le N$ and (iii) for $n < N$; everything below applies to it unchanged.

Remark 2.2 (First consequences).

(a) *Several steps at once.* If M is a martingale, then $\mathbb{E}[M_m \mid \mathcal{F}_n] = M_n$ for all $m \ge n$. By induction on m, the tower property gives $\mathbb{E}[M_{m+1} \mid \mathcal{F}_n] = \mathbb{E}\big[\mathbb{E}[M_{m+1} \mid \mathcal{F}_m] \,\big|\, \mathcal{F}_n\big] = \mathbb{E}[M_m \mid \mathcal{F}_n] = M_n$. The same computation, with positivity, gives $\mathbb{E}[M_m \mid \mathcal{F}_n] \ge M_n$ for a submartingale and $\le$ for a supermartingale.

(b) *Means.* Taking expectations, $\mathbb{E}[M_n] = \mathbb{E}[M_0]$ for all n if M is a martingale; the mean is nondecreasing for a submartingale and nonincreasing for a supermartingale.

(c) *The filtration matters.* If M is a martingale with respect to $(\mathcal{F}_n)$ and is adapted to a smaller filtration $(\mathcal{G}_n)$, with $\mathcal{G}_n \subseteq \mathcal{F}_n$, then it is a martingale with respect to $(\mathcal{G}_n)$: $\mathbb{E}[M_{n+1} \mid \mathcal{G}_n] = \mathbb{E}\big[\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \,\big|\, \mathcal{G}_n\big] = \mathbb{E}[M_n \mid \mathcal{G}_n] = M_n$. In particular it is a martingale for its own history $\sigma(M_0, \dots, M_n)$. A larger filtration can destroy the property: a price that is a martingale for the public is not one for an insider who knows the next move.

(d) *The probability matters.* Changing $\mathbb{P}$ changes conditional expectations. Discounted stock prices are martingales under a risk-neutral probability $\mathbb{Q}$, and in general not under the real-world probability $\mathbb{P}$ (Theorem 4.8); under $\mathbb{Q}$, so is the discounted price of any derivative, a conditional expectation of its discounted payoff (Exercise 2.1).

Sums, products and exponentials

Let $\xi_1, \xi_2, \dots$ be random variables and $(\mathcal{F}_n)$ a filtration such that

$$\text{for every } n \ge 1, \quad \xi_n \text{ is } \mathcal{F}_n\text{-measurable and independent of } \mathcal{F}_{n-1}. \tag{2.1}$$

The ξ_n are then independent. Conversely, if $\xi_1, \xi_2, \dots$ are independent, (2.1) holds for their natural filtration $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$, since ξ_n is independent of the vector $(\xi_1, \dots, \xi_{n-1})$, hence of the σ-algebra it generates, by Dynkin's π–λ theorem (Durrett (2019), Section 2.1). Larger filtrations are allowed as long as they do not reveal future ξ's. Fix $x \in \mathbb{R}$ and let

$$S_0 = x, \qquad S_n = x + \xi_1 + \cdots + \xi_n, \quad n \ge 1.$$

Proposition 2.3 (Sums, products, exponentials)**.** *Assume* (2.1).

(i) If every ξ_k is integrable with mean μ_k, then $\big(S_n - \sum_{k=1}^n \mu_k\big)_n$ is a martingale. In particular S is a martingale if the ξ_k have mean 0.

(ii) If every ξ_k has mean 0 and finite variance σ_k^2, then $\big(S_n^2 - \sum_{k=1}^n \sigma_k^2\big)_n$ is a martingale.

(iii) If every ξ_k is integrable with mean 1, then $M_0 = 1$, $M_n = \xi_1 \xi_2 \cdots \xi_n$ defines a martingale.

(iv) If $\theta \in \mathbb{R}$ is such that $\varphi_k(\theta) = \mathbb{E}[e^{\theta \xi_k}] < \infty$ for every k, then $Z_n = e^{\theta S_n}/(\varphi_1(\theta) \cdots \varphi_n(\theta))$ defines a martingale. For $\xi_k \sim \mathcal{N}(0, \sigma^2)$ this is $Z_n = \exp\big(\theta S_n - \frac{1}{2}\theta^2 \sigma^2 n\big)$, for every $\theta \in \mathbb{R}$.

Proof. Each process is a function of $(\xi_1, \dots, \xi_n)$, hence adapted. In each case we write the value at time $n+1$ as $f(Y_n, \xi_{n+1})$, with Y_n an $\mathcal{F}_n$-measurable variable and ξ_{n+1} independent of $\mathcal{F}_n$, and apply the independence lemma: $\mathbb{E}[f(Y_n, \xi_{n+1}) \mid \mathcal{F}_n] = g(Y_n)$ with $g(y) = \mathbb{E}[f(y, \xi_{n+1})]$.

(i) S_n is integrable as a sum of integrable variables. With $f(s,y) = s+y$, $\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = S_n + \mu_{n+1}$; subtract $\sum_{k \le n+1} \mu_k$ from both sides.

(ii) S_n is square integrable as a sum of square-integrable variables. With $f(s,y) = (s+y)^2$, $g(s) = s^2 + 2s\,\mathbb{E}[\xi_{n+1}] + \mathbb{E}[\xi_{n+1}^2] = s^2 + \sigma_{n+1}^2$, so $\mathbb{E}[S_{n+1}^2 \mid \mathcal{F}_n] = S_n^2 + \sigma_{n+1}^2$.

(iii) Since $|\xi_1|, \dots, |\xi_n|$ are independent, $\mathbb{E}|M_n| = \prod_{k \le n} \mathbb{E}|\xi_k| < \infty$. With $f(m,y) = my$, $g(m) = m\,\mathbb{E}[\xi_{n+1}] = m$, so $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n$.

(iv) $Z_n = e^{\theta x} \prod_{k \le n} \left(e^{\theta \xi_k} / \varphi_k(\theta)\right)$ is a constant times a product of independent mean-one factors, so (iii) applies. For $\xi \sim \mathcal{N}(0, \sigma^2)$, $\mathbb{E}[e^{\theta \xi}] = e^{\theta^2 \sigma^2 / 2}$. □

The random walk

Let $X_1, X_2, \dots$ be independent with $\mathbb{P}(X_k = 1) = p$ and $\mathbb{P}(X_k = -1) = q := 1 - p$, where $p \in (0,1)$. Fix $x \in \mathbb{Z}$, let $S_0 = x$, $S_n = x + X_1 + \cdots + X_n$, and let $(\mathcal{F}_n)$ be the natural filtration of the X_k. This is the *random walk* started at x, *symmetric* if $p = \frac{1}{2}$ and *biased* otherwise: the fortune of a gambler who bets one dollar per round on a coin that lands heads with probability p, or a price that moves one tick up or down per period. Each step has mean $p - q$ and variance $1 - (p-q)^2 = 4pq$.

Corollary 2.4 (Martingales of the random walk)**.**

(a) $S_n - n(p-q)$ *is a martingale. The walk itself is a martingale if and only if* $p = \frac{1}{2}$*, a submartingale if and only if* $p \ge \frac{1}{2}$*, and a supermartingale if and only if* $p \le \frac{1}{2}$*.*

(b) $(q/p)^{S_n}$ *is a martingale.*

(c) $\big(S_n - n(p-q)\big)^2 - 4pq\,n$ *is a martingale; for* $p = \frac{1}{2}$*, so is* $S_n^2 - n$*.*

(d) For every $\theta \in \mathbb{R}$*,* $e^{\theta S_n} / \left(pe^{\theta} + qe^{-\theta}\right)^n$ *is a martingale.*

Proof. (a) By Theorem 2.3(i), $\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = S_n + (p-q)$, which is equal to, at least, or at most S_n according to the sign of $p-q$. (b) $(q/p)^{S_n} = (q/p)^x \prod_{k \le n} (q/p)^{X_k}$, and $\mathbb{E}\big[(q/p)^{X_k}\big] = p \cdot \frac{q}{p} + q \cdot \frac{p}{q} = 1$: apply Theorem 2.3(iii). (c) Let $Y_n = S_n - x - n(p-q)$, the sum of the centred steps $X_k - (p-q)$, which have variance $4pq$. By Theorem 2.3(i) and (ii), Y and $Y_n^2 - 4pqn$ are martingales, hence so is $(x + Y_n)^2 - 4pqn = Y_n^2 - 4pqn + 2xY_n + x^2$. (d) Theorem 2.3(iv) with $\varphi(\theta) = pe^{\theta} + qe^{-\theta}$. □

A biased walk is not a martingale: it drifts. Subtracting the drift $n(p-q)$ is one way to obtain a martingale; the exponential transformation in (b) is another, and it is the one that solves the gambler's ruin problem in Section 2.4.

Remark 2.5 (The martingale (b) as a density). Fix n. The variable $Z_n = (q/p)^{S_n - x}$ is positive with $\mathbb{E}[Z_n] = 1$ by (b), so $\mathbb{P}'(A) = \mathbb{E}[Z_n \mathbf{1}_A]$, $A \in \mathcal{F}_n$, defines a probability on $\mathcal{F}_n$ equivalent to $\mathbb{P}$ (Theorem 1.23). A path of n steps with u up-steps and d down-steps has probability $p^u q^d$ under $\mathbb{P}$, hence $(q/p)^{u-d} p^u q^d = q^u p^d$ under $\mathbb{P}'$: under $\mathbb{P}'$ the steps are still independent, but go up with probability q and down with probability p. Because Z is a martingale, these definitions

agree as n varies: Z is the density process of the change of measure that flips the bias of the coin (Theorem 1.25). Density processes are the basis of the change of measure in Chapter 11.

Remark 2.6 (A martingale whose paths die). Fix $0 < a < 1$, and let the ξ_k be independent and equal to $1 + a$ or $1 - a$ with probability $\frac{1}{2}$ each. An investor who, each period, puts all her wealth in a fair bet that returns $+a$ or $-a$ has wealth $M_n = \xi_1 \cdots \xi_n$ (starting from 1). By Theorem 2.3(iii), M is a martingale and $\mathbb{E}[M_n] = 1$ for all n. Yet $\log M_n = \sum_{k \le n} \log \xi_k$ is a sum of independent, identically distributed, bounded variables with mean $\frac{1}{2}\log(1 - a^2) < 0$, so by the strong law of large numbers (Durrett (2019), Theorem 2.4.1) $\frac{1}{n}\log M_n \to \frac{1}{2}\log(1 - a^2)$ almost surely: $M_n \to 0$ almost surely, at the exponential rate $(1 - a^2)^{n/2}$ (Figure 2.1). The mean is carried by rarer and rarer paths with many more gains than losses. For $a = 1$ the picture is extreme: $M_n = 2^n$ with probability 2^{-n} and $M_n = 0$ otherwise. The gap between the expected return 0 and the typical growth rate $\frac{1}{2}\log(1 - a^2) \approx -\frac{a^2}{2}$ is the *volatility drag* of a risky investment; Theorem 5.14 is its continuous-time version.

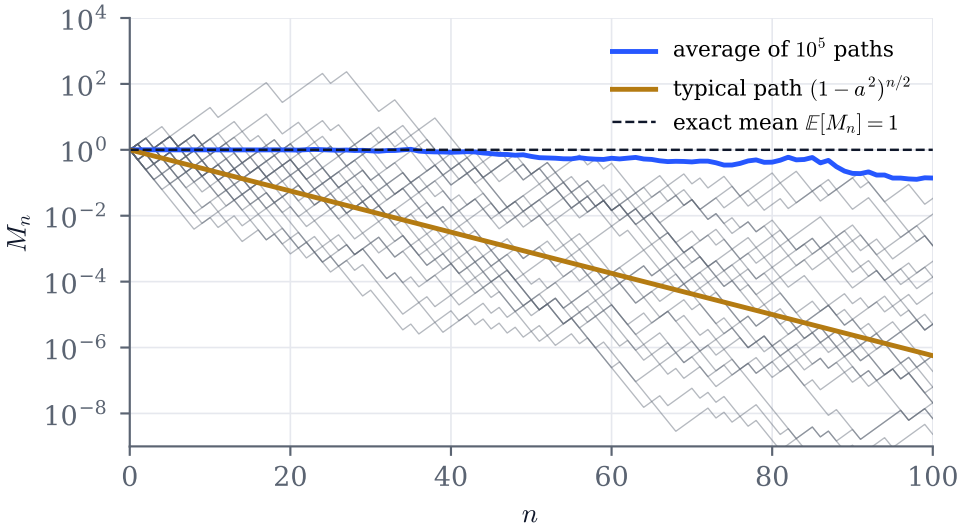

Figure 2.1: The martingale $M_n = \xi_1 \cdots \xi_n$ with $\xi_k = 1 \pm \frac{1}{2}$ (Theorem 2.6). Grey: 25 paths. Almost every path decays like the median $(1 - a^2)^{n/2}$, although $\mathbb{E}[M_n] = 1$. Even an average of 10^5 paths eventually falls below 1: the paths that carry the mean are too rare to be sampled.

▶ Interactive demo: sample means of S_n, $S_n - n(p - q)$ and $(q/p)^{S_n}$ for a walk with the bias p of your choice; and products of mean-one factors whose mean stays at one while their paths go to zero.

2.2 Trading strategies and martingale transforms

Consider a market with a riskless asset paying zero interest and a risky asset whose price at time n is X_n (Chapter 4 removes the zero-interest assumption by working with discounted prices). A trader decides at time $k - 1$ how many units H_k of the risky asset to hold over the period from $k - 1$ to k. She may use everything she knows at time $k - 1$, and nothing more.

Definition 2.7 (Predictable strategy, martingale transform)**.** A process $H = (H_n)_{n\ge 1}$ is *predictable* if H_n is $\mathcal{F}_{n-1}$-measurable for every $n \ge 1$. For an adapted process X and a predictable process H, the *martingale transform* of X by H is

$$(H \cdot X)_0 = 0, \qquad (H \cdot X)_n = \sum_{k=1}^{n} H_k\,(X_k - X_{k-1}), \quad n \ge 1.$$

$H_k(X_k - X_{k-1})$ is the profit or loss on the position held during the k-th period, so $(H \cdot X)_n$ is the trader's cumulative gain. If she starts with wealth V_0 and finances her purchases by borrowing and lending at zero interest, adding and withdrawing no money afterwards, her wealth at time n is $V_n = V_0 + (H \cdot X)_n$: the strategy is *self-financing*. For a gambler, $X_k - X_{k-1}$ is the result of round k per dollar staked and H_k is the stake.

Theorem 2.8 (You cannot beat a fair game)**.** *Let H be predictable and bounded: $|H_n| \le C_n$ for some constants C_n, $n \ge 1$.*

(i) If M is a martingale, then $H \cdot M$ is a martingale with $(H \cdot M)_0 = 0$. In particular $\mathbb{E}[(H \cdot M)_n] = 0$ for every n.

(ii) If M is a supermartingale (respectively a submartingale) and $H_n \ge 0$ for every n, then $H \cdot M$ is a supermartingale (respectively a submartingale), and $\mathbb{E}[(H \cdot M)_n] \le 0$ (respectively ≥ 0).

Proof. $(H \cdot M)_n$ is built from $H_1, \dots, H_n$ and $M_0, \dots, M_n$, all $\mathcal{F}_n$-measurable, and $|(H \cdot M)_n| \le \sum_{k\le n} C_k(|M_k| + |M_{k-1}|)$ is integrable. Since H_{n+1} is bounded and $\mathcal{F}_n$-measurable, taking out what is known gives

$$\mathbb{E}\big[(H \cdot M)_{n+1} - (H \cdot M)_n \,\big|\, \mathcal{F}_n\big] = \mathbb{E}\big[H_{n+1}(M_{n+1} - M_n) \,\big|\, \mathcal{F}_n\big] = H_{n+1}\big(\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] - M_n\big).$$

For a martingale the bracket vanishes. For a supermartingale it is ≤ 0, and so is its product with $H_{n+1} \ge 0$; for a submartingale both inequalities reverse. □

The theorem speaks about all strategies at once. Doubling the stake after a loss, betting more during a winning streak, following a technical indicator: no rule that chooses the stakes from the past changes the zero expected gain of a fair game, or makes an unfavourable game favourable. Predictability carries the weight. If H_k could depend on X_k, the trader could hold one unit exactly in the periods in which the price goes up, and collect $\sum_k (X_k - X_{k-1})^+$; for a martingale, this has positive mean unless the price never moves. The bound C_n may grow with n, so it costs nothing at a fixed horizon; Section 2.3 shows what goes wrong when the horizon itself is unbounded.

Remark 2.9 (A converse). Let X be adapted and integrable, and suppose that $\mathbb{E}[(H \cdot X)_n] = 0$ for every bounded predictable H and every n. Then X is a martingale. Indeed, fix $m \ge 1$ and $A \in \mathcal{F}_{m-1}$, and let $H_m = \mathbf{1}_A$ and $H_k = 0$ for $k \ne m$. Then $0 = \mathbb{E}[(H \cdot X)_m] = \mathbb{E}\big[\mathbf{1}_A(X_m - X_{m-1})\big]$, so $\mathbb{E}[\mathbf{1}_A X_m] = \mathbb{E}[\mathbf{1}_A X_{m-1}]$ for every $A \in \mathcal{F}_{m-1}$. As X_{m-1} is $\mathcal{F}_{m-1}$-measurable, this is the partial-averaging property that characterizes $\mathbb{E}[X_m \mid \mathcal{F}_{m-1}] = X_{m-1}$ (Theorem 1.14). A price process is thus a martingale exactly when no trading strategy has a nonzero expected gain. This equivalence is the core of the fundamental theorem of asset pricing (Theorem 4.9).

Remark 2.10 (The discrete ancestor of the Itô integral). When X is a symmetric random walk, or a Brownian motion observed at times $t_0 < t_1 < \cdots$, the transform $\sum_k H_k(X_{t_k} - X_{t_{k-1}})$

is a Riemann–Stieltjes sum whose integrand is evaluated at the left end of each interval, as predictability requires. The Itô integral $\int_0^t \Delta_u \, \mathrm{d}W_u$ of Chapter 7 is the limit of such sums as the grid is refined, and Theorem 2.8 survives the limit: Itô integrals of suitable integrands are martingales. The left end is not a matter of convention. For the symmetric walk started at 0, $\sum_{k\le n} S_{k-1}(S_k - S_{k-1}) = \frac{1}{2}(S_n^2 - n)$, the discrete form of $\int_0^t W \, \mathrm{d}W = \frac{1}{2}(W_t^2 - t)$ (Exercise 2.2); evaluating at the right end gives $\frac{1}{2}(S_n^2 + n)$ instead. The two differ by $[S,S]_n = n$, the quadratic variation of the walk (Theorem 5.1).

2.3 Stopping times and optional stopping

A gambler leaving the table, a trader closing a position at a stop-loss, the holder of an American option exercising it: each chooses a random time. The time is random, but the decision to stop at time n may only use the information available at time n.

Definition 2.11 (Stopping time)**.** A random variable τ with values in $\mathbb{N} \cup \{\infty\}$ is a *stopping time* if $\{\tau \le n\} \in \mathcal{F}_n$ for every $n \in \mathbb{N}$. Equivalently, $\{\tau = n\} \in \mathcal{F}_n$ for every n, since $\{\tau = n\} = \{\tau \le n\} \setminus \{\tau \le n-1\}$ and $\{\tau \le n\} = \bigcup_{k\le n}\{\tau = k\}$.

Constant times are stopping times. So are *first entrance times* $\tau_A = \inf\{n \ge 0 : X_n \in A\}$ of an adapted process X into a Borel set A, with $\inf \emptyset = \infty$, because $\{\tau_A \le n\} = \bigcup_{k=0}^n \{X_k \in A\} \in \mathcal{F}_n$. The exit time of a walk from an interval is of this type. By contrast, the last visit of the walk to 0 before time N, or the time at which it reaches its maximum over $\{0, \dots, N\}$, are not stopping times: deciding requires seeing the future (Exercise 2.4).

For a process X and a stopping time τ, the *stopped process* X^τ is $X_n^\tau = X_{n\wedge\tau}$: it follows X up to time τ and stays frozen at X_τ afterwards. On $\{\tau < \infty\}$ we write $X_\tau(\omega) = X_{\tau(\omega)}(\omega)$.

Proposition 2.12 (Stopped martingales)**.** *If M is a martingale (respectively a super- or submartingale) and τ is a stopping time, then M^τ is a martingale (respectively a super- or submartingale). In particular $\mathbb{E}[M_{n\wedge\tau}] = \mathbb{E}[M_0]$ for every n (respectively $\le$, $\ge$).*

Proof. Stopping is a trading strategy: hold one unit until τ and nothing afterwards. With $H_k = \mathbf{1}_{\{\tau \ge k\}}$, the sum telescopes:

$$M_{n\wedge\tau} = M_0 + \sum_{k=1}^{n\wedge\tau} (M_k - M_{k-1}) = M_0 + \sum_{k=1}^{n} \mathbf{1}_{\{\tau\ge k\}}(M_k - M_{k-1}) = M_0 + (H \cdot M)_n.$$

H is predictable, because $\{\tau \ge k\} = \{\tau \le k-1\}^c \in \mathcal{F}_{k-1}$, and it takes values in $\{0,1\}$. By Theorem 2.8, $H \cdot M$ is a martingale (respectively a super- or submartingale) started at 0, and adding the integrable $\mathcal{F}_0$-measurable variable M_0 preserves the property. □

So $\mathbb{E}[M_{n\wedge\tau}] = \mathbb{E}[M_0]$ at every fixed n. If $\tau < \infty$, then $M_{n\wedge\tau} \to M_\tau$ as $n \to \infty$, and it is tempting to conclude that $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$. That can fail.

Example 2.13 (The doubling strategy)**.** A gambler plays a fair game ($p = \frac{1}{2}$). She stakes 1 in the first round, doubles her stake after every loss, and stops after her first win. Formally, let S be the symmetric walk started at 0, $T = \inf\{n \ge 1 : X_n = 1\}$ the round of the first win, and $H_k = 2^{k-1}\mathbf{1}_{\{T\ge k\}}$, which is predictable since $\{T \ge k\} = \{X_1 = \cdots = X_{k-1} = -1\} \in \mathcal{F}_{k-1}$.

Her gains $G = H \cdot S$ satisfy $G_n = -(1 + 2 + \cdots + 2^{n-1}) = -(2^n - 1)$ for $n < T$, and $G_n = -(2^{T-1} - 1) + 2^{T-1} = 1$ for $n \geq T$. Since $\mathbb{P}(T > n) = 2^{-n}$, T is finite almost surely (indeed $\mathbb{E}[T] = 2$), so $G_T = 1$ almost surely and $\mathbb{E}[G_T] = 1 \neq 0 = \mathbb{E}[G_0]$. At every fixed n, however,

$$\mathbb{E}[G_n] = (1 - 2^{-n}) \cdot 1 + 2^{-n} \cdot (1 - 2^n) = 0,$$

as Theorem 2.8 demands: the rare run of n losses costs exactly what the frequent small wins earn. What fails is the passage $n \to \infty$. The losses $G_{n \wedge T}$ are unbounded below, and so are the stakes needed to carry the strategy through. With a finite bankroll $2^m - 1$, the gambler can survive only m losses in a row and must stop at $T \wedge m$, a bounded stopping time. She then wins 1 with probability $1 - 2^{-m}$ and loses $2^m - 1$ with probability 2^{-m}: the expected gain is 0. At a casino, where the game is unfavourable, the same strategy has a negative expected gain that grows with the bankroll (Exercise 2.3).

The next theorem gives conditions under which the passage to the limit is legitimate.

Theorem 2.14 (Optional stopping)**.** *Let M be a martingale and τ a stopping time. In each of the following cases, M_τ is integrable and $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$:*

(a) τ is bounded: $\tau \leq N$ for a constant N;

(b) $\tau < \infty$ almost surely, and the stopped process is bounded: $|M_{n \wedge \tau}| \leq K$ for all n, for a constant K;

(c) $\mathbb{E}[\tau] < \infty$, and the increments are bounded in conditional mean: there is a constant K with $\mathbb{E}[|M_k - M_{k-1}| \,|\, \mathcal{F}_{k-1}] \leq K$ for every $k \geq 1$ (for instance, $|M_k - M_{k-1}| \leq K$).

If M is a supermartingale (respectively a submartingale), the same holds with $\mathbb{E}[M_\tau] \leq \mathbb{E}[M_0]$ (respectively $\geq$).

Proof. In each case we let $n \to \infty$ in $\mathbb{E}[M_{n \wedge \tau}] = \mathbb{E}[M_0]$ (Theorem 2.12), or in the corresponding inequality.

(a) For $n \geq N$, $M_{n \wedge \tau} = M_\tau$, so there is nothing to prove.

(b) Since $\tau < \infty$ almost surely, $M_{n \wedge \tau} = M_\tau$ for all $n \geq \tau$, so $M_{n \wedge \tau} \to M_\tau$ almost surely, and $|M_\tau| \leq K$. Dominated convergence gives $\mathbb{E}[M_{n \wedge \tau}] \to \mathbb{E}[M_\tau]$.

(c) Let $Y = \sum_{k \geq 1} \mathbf{1}_{\{\tau \geq k\}} |M_k - M_{k-1}|$. By monotone convergence, then the tower property and taking out the bounded $\mathcal{F}_{k-1}$-measurable factor $\mathbf{1}_{\{\tau \geq k\}}$, then positivity,

$$\mathbb{E}[Y] = \sum_{k \geq 1} \mathbb{E}\Big[\mathbf{1}_{\{\tau \geq k\}} \, \mathbb{E}[|M_k - M_{k-1}| \,|\, \mathcal{F}_{k-1}]\Big] \leq K \sum_{k \geq 1} \mathbb{P}(\tau \geq k) = K \, \mathbb{E}[\tau] < \infty,$$

where the last equality is monotone convergence applied to $\tau = \sum_{k \geq 1} \mathbf{1}_{\{\tau \geq k\}}$. Writing $M_{n \wedge \tau} = M_0 + \sum_{k=1}^{n} \mathbf{1}_{\{\tau \geq k\}} (M_k - M_{k-1})$ as in the proof of Theorem 2.12, we get $|M_{n \wedge \tau}| \leq |M_0| + Y$ for every n, and also $|M_\tau| \leq |M_0| + Y$. As $\mathbb{E}[\tau] < \infty$, τ is finite almost surely, so $M_{n \wedge \tau} \to M_\tau$ almost surely, and dominated convergence concludes. □

The three conditions bound the stopped process by an integrable variable, in time (a), in size (b), or through finite expected duration and moderate stakes (c); each of the three suffices on its own. Read as a statement about games, the theorem says that a fair game remains fair under

any stopping rule that has bounded time, or bounded credit, or finite expected duration with bounded stakes, and that under such a rule an unfavourable game offers no way to quit ahead on average. The doubling strategy has $\mathbb{E}[T] < \infty$, but its increments 2^{k-1} are unbounded and (c) fails; so does (b). In continuous time, strategies of this kind can be run within a finite time interval, which is why continuous-time models restrict the admissible strategies, for instance to those whose wealth is bounded below (Harrison and Pliska (1981); Chapter 11).

The optional stopping theorem turns martingales into identities for random sums.

Proposition 2.15 (Wald's identities)**.** *Assume (2.1) with $\xi_1, \xi_2, \dots$ identically distributed and integrable, let $\mu = \mathbb{E}[\xi_1]$, $S_n = \xi_1 + \cdots + \xi_n$, and let τ be a stopping time with $\mathbb{E}[\tau] < \infty$. Then*

(i) S_τ is integrable and $\mathbb{E}[S_\tau] = \mu\,\mathbb{E}[\tau]$;

(ii) if moreover $\sigma^2 = \mathrm{Var}(\xi_1) < \infty$, then $\mathbb{E}\big[(S_\tau - \mu\tau)^2\big] = \sigma^2\,\mathbb{E}[\tau]$.

Proof. Let $\eta_k = \xi_k - \mu$ and $M_n = S_n - n\mu = \eta_1 + \cdots + \eta_n$, a martingale by Theorem 2.3(i).

(i) Since η_k is independent of $\mathcal{F}_{k-1}$, $\mathbb{E}\big[|M_k - M_{k-1}| \,\big|\, \mathcal{F}_{k-1}\big] = \mathbb{E}|\eta_k| \le 2\mathbb{E}|\xi_1|$. By Theorem 2.14(c), $\mathbb{E}[M_\tau] = \mathbb{E}[M_0] = 0$, that is, $\mathbb{E}[S_\tau - \mu\tau] = 0$. As $\mu\tau$ is integrable, so is S_τ, and $\mathbb{E}[S_\tau] = \mu\mathbb{E}[\tau]$.

(ii) For $m < n$, $M_{n\wedge\tau} - M_{m\wedge\tau} = \sum_{k=m+1}^{n} \mathbf{1}_{\{\tau \ge k\}}\eta_k$. Expand the square. For $j < k$, the variable $U = \mathbf{1}_{\{\tau \ge k\}}\eta_j$ is $\mathcal{F}_{k-1}$-measurable and integrable, and η_k is independent of $\mathcal{F}_{k-1}$, so $\mathbb{E}\big[\mathbf{1}_{\{\tau \ge j\}}\eta_j \mathbf{1}_{\{\tau \ge k\}}\eta_k\big] = \mathbb{E}[U\eta_k] = \mathbb{E}[U]\,\mathbb{E}[\eta_k] = 0$. On the diagonal, $\{\tau \ge k\} \in \mathcal{F}_{k-1}$ is independent of η_k, so $\mathbb{E}\big[\mathbf{1}_{\{\tau \ge k\}}\eta_k^2\big] = \sigma^2\mathbb{P}(\tau \ge k)$. Hence

$$\mathbb{E}\big[(M_{n\wedge\tau} - M_{m\wedge\tau})^2\big] = \sigma^2 \sum_{k=m+1}^{n} \mathbb{P}(\tau \ge k) = \sigma^2\,\mathbb{E}[n \wedge \tau - m \wedge \tau]. \tag{2.2}$$

Let $n \to \infty$: $M_{n\wedge\tau} \to M_\tau$ almost surely, and Fatou's lemma (a consequence of monotone convergence) gives $\mathbb{E}\big[(M_\tau - M_{m\wedge\tau})^2\big] \le \sigma^2\mathbb{E}[\tau - m \wedge \tau]$. The right side tends to 0 as $m \to \infty$ by dominated convergence ($0 \le \tau - m \wedge \tau \le \tau$). So $M_{m\wedge\tau} \to M_\tau$ in L^2, and therefore $\mathbb{E}[M_{m\wedge\tau}^2] \to \mathbb{E}[M_\tau^2]$. By (2.2) with the pair $(0, m)$ in place of (m, n), $\mathbb{E}[M_{m\wedge\tau}^2] = \sigma^2\mathbb{E}[m \wedge \tau]$, which tends to $\sigma^2\mathbb{E}[\tau]$ by monotone convergence. Hence $\mathbb{E}[(S_\tau - \mu\tau)^2] = \mathbb{E}[M_\tau^2] = \sigma^2\mathbb{E}[\tau]$. □

2.4 Gambler's ruin

Let S be the random walk of Theorem 2.4, with up-probability $p \in (0, 1)$ and $q = 1 - p$, started at x. Fix integers $a < b$ with $a \le x \le b$, and let

$$\tau = \inf\{n \ge 0 : S_n \notin (a, b)\} = \inf\{n \ge 0 : S_n \in \{a, b\}\},$$

the exit time from (a, b); the two infima agree because a walk with steps ± 1 cannot jump over a or b. A gambler with capital $x - a$ who plays until she is ruined or her fortune reaches b, and a trader with a stop-loss at a and a take-profit at b, both face the questions: how likely is the exit at b, and how long does it take?

Lemma 2.16 (The walk leaves every interval)**.** *There are constants C and $\varrho < 1$, depending on p, a and b, such that $\mathbb{P}(\tau > n) \le C\varrho^n$ for every n. In particular $\tau < \infty$ almost surely, and $\mathbb{E}[\tau^j] < \infty$ for every $j \ge 1$.*

Proof. If $x \in \{a, b\}$, then $\tau = 0$. Otherwise $b - a \geq 2$; let $L = b - a - 1 \geq 1$ and let A_i be the event that the L steps $X_{(i-1)L+1}, \ldots, X_{iL}$ all equal $+1$. If $\tau > (i-1)L$, the walk is at a level at least $a + 1$ at time $(i-1)L$, and on A_i it climbs, one step at a time, to a level at least $a + 1 + L = b$ by time iL; then $\tau \leq iL$. Hence $\{\tau > mL\} \subseteq A_1^c \cap \cdots \cap A_m^c$, and since these events involve disjoint blocks of independent steps, $\mathbb{P}(\tau > mL) \leq (1 - p^L)^m$. For any n, with $m = \lfloor n/L \rfloor \geq n/L - 1$, we get $\mathbb{P}(\tau > n) \leq (1 - p^L)^{n/L - 1}$: take $\varrho = (1 - p^L)^{1/L}$ and $C = (1 - p^L)^{-1}$. Then $\mathbb{P}(\tau = \infty) \leq \mathbb{P}(\tau > n) \to 0$, and, by monotone convergence, $\mathbb{E}[\tau^j] = \sum_{n \geq 0} \left((n+1)^j - n^j\right) \mathbb{P}(\tau > n) \leq C \sum_{n \geq 0} (n+1)^j \varrho^n < \infty$. □

Theorem 2.17 (Gambler's ruin)**.** *Let $a < b$ and $a \leq x \leq b$ be integers. The walk started at x leaves (a, b) almost surely, and*

(i) if $p = \frac{1}{2}$, then $\mathbb{P}(S_\tau = b) = \dfrac{x - a}{b - a}$;

(ii) if $p \neq \frac{1}{2}$, then, with $r = q/p$, $\mathbb{P}(S_\tau = b) = \dfrac{r^x - r^a}{r^b - r^a}$.

Proof. By Theorem 2.16, $\tau < \infty$ almost surely and $S_\tau \in \{a, b\}$; write $\pi = \mathbb{P}(S_\tau = b)$. Up to time τ the walk stays in $[a, b]$, so the stopped processes below are bounded and Theorem 2.14(b) applies.

(i) S is a martingale (Theorem 2.4(a)) and $|S_{n \wedge \tau}| \leq \max(|a|, |b|)$. So $x = \mathbb{E}[S_\tau] = b\pi + a(1 - \pi)$, and $\pi = (x - a)/(b - a)$.

(ii) r^{S_n} is a martingale (Theorem 2.4(b)), and $0 < r^{S_{n \wedge \tau}} \leq \max(r^a, r^b)$. So $r^x = \mathbb{E}[r^{S_\tau}] = r^b \pi + r^a (1 - \pi)$. Since $r \neq 1$ and $a \neq b$, $r^b \neq r^a$, and solving for π gives (ii). □

As $p \to \frac{1}{2}$, $r \to 1$ and the formula in (ii) tends to the one in (i) (l'Hôpital's rule in r). The difference between the two is dramatic far from $p = \frac{1}{2}$, and already noticeable near it. At roulette, a bet on red wins with probability $p = \frac{18}{37}$, so $r = \frac{19}{18}$. A player who starts with $x = 100$ dollars, with $a = 0$, and bets one dollar per spin until she is ruined or has doubled her money ($b = 200$) succeeds with probability $(r^{100} - 1)/(r^{200} - 1) \approx 0.0045$, against $\frac{1}{2}$ in a fair game. Betting the 100 dollars on a single spin succeeds with probability $\frac{18}{37} \approx 0.486$. In an unfavourable game, timid play is far worse than bold play, which is optimal in subfair red-and-black (Dubins and Savage, 1965).

Proposition 2.18 (Expected duration)**.** *In the setting of Theorem 2.17,*

(i) if $p = \frac{1}{2}$, then $\mathbb{E}[\tau] = (x - a)(b - x)$;

(ii) if $p \neq \frac{1}{2}$, then $\mathbb{E}[\tau] = \dfrac{(b - a)\,\mathbb{P}(S_\tau = b) - (x - a)}{p - q}$, with $\mathbb{P}(S_\tau = b)$ given by Theorem 2.17(ii).

Proof. $\mathbb{E}[\tau] < \infty$ by Theorem 2.16, so Wald's identities (Theorem 2.15) apply to $S_n - x = X_1 + \cdots + X_n$, with $\mu = p - q$ and $\sigma^2 = 4pq$. Write $\pi = \mathbb{P}(S_\tau = b)$.

(ii) By Theorem 2.15(i), $(p - q)\mathbb{E}[\tau] = \mathbb{E}[S_\tau - x] = (b - x)\pi + (a - x)(1 - \pi) = (b - a)\pi - (x - a)$. Divide by $p - q \neq 0$.

(i) Here $\mu = 0$ and $\sigma^2 = 1$, and Theorem 2.15(ii) gives, with $\pi = (x-a)/(b-a)$,

$$\mathbb{E}[\tau] = \mathbb{E}[(S_\tau - x)^2] = (b-x)^2\pi + (x-a)^2(1-\pi)$$
$$= \frac{(b-x)(x-a)\big((b-x)+(x-a)\big)}{b-a} = (x-a)(b-x). \qquad \square$$

From the midpoint of an interval of length $d = b - a$, the symmetric walk needs on average $d^2/4$ steps to leave it: distance grows like the square root of time, the diffusive scaling behind the step size $1/\sqrt{n}$ of Theorem 5.3. For Brownian motion, the exit time from (a, b) started at 0 has mean $-ab$ (Exercise 5.6), the continuous counterpart of (i). Other martingales of the walk give the variance of τ and its whole law (Exercise 2.6, Exercise 2.7); Exercise 2.8 checks the formulas by simulation.

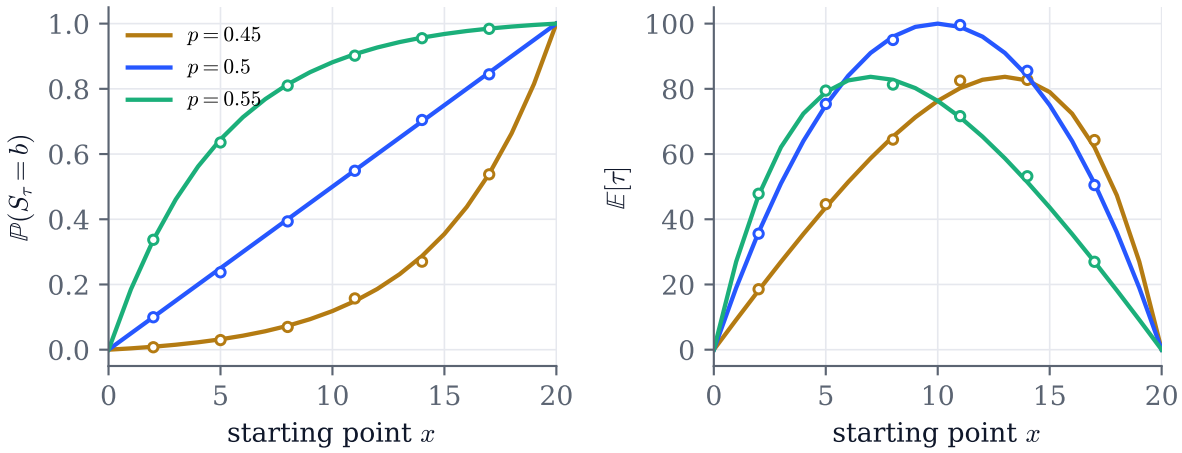

Figure 2.2: Gambler's ruin on $(a, b) = (0, 20)$. Left: the probability of exiting at b (Theorem 2.17). Right: the expected duration (Theorem 2.18). Lines: the formulas; circles: Monte Carlo estimates from 4000 walks per point. A bias of five percentage points either way moves the exit probability from the centre by about 0.4.

▶ Interactive demo: gambler's ruin with sliders for p, the starting point and the target: simulated exit probabilities and durations against the formulas.

Letting the target recede to infinity tells us whether the walk ever reaches a given level.

Corollary 2.19 (Hitting a level)**.** *Let $a < x$ be integers and $\sigma_a = \inf\{n \geq 0 : S_n = a\}$. Then*

$$\mathbb{P}(\sigma_a < \infty) = 1 \quad \text{if } p \leq \tfrac{1}{2}, \qquad \mathbb{P}(\sigma_a < \infty) = (q/p)^{x-a} \quad \text{if } p > \tfrac{1}{2}.$$

In particular, the symmetric walk reaches every level almost surely (by symmetry, levels above x as well), but the expected time to reach a level other than x is infinite.

Proof. For $b > x$, let E_b be the event that the walk hits a before b. To reach $b+1$ the walk must first pass through b, so $E_b \subseteq E_{b+1}$. Each E_b is contained in $\{\sigma_a < \infty\}$; conversely, if $\sigma_a < \infty$, the walk stays below some level up to time σ_a, and E_b occurs for every b above that level. So $\{\sigma_a < \infty\} = \bigcup_b E_b$, and by continuity of probability along increasing events,

$\mathbb{P}(\sigma_a < \infty) = \lim_{b\to\infty}(1 - \mathbb{P}(S_\tau = b))$ with τ the exit time from (a, b). By Theorem 2.17 this limit is $\lim_b \frac{b-x}{b-a} = 1$ if $p = \frac{1}{2}$, and $\lim_b \frac{r^b - r^x}{r^b - r^a}$ if $p \neq \frac{1}{2}$, which equals 1 if $r > 1$ ($p < \frac{1}{2}$) and $r^x/r^a = (q/p)^{x-a}$ if $r < 1$ ($p > \frac{1}{2}$).

For $p = \frac{1}{2}$, suppose that a level $c \neq x$ had $\mathbb{E}[\sigma_c] < \infty$. The walk is a martingale with increments bounded by 1, so Theorem 2.14(c) would give $\mathbb{E}[S_{\sigma_c}] = x$, while $S_{\sigma_c} = c$. □

The last statement is the doubling paradox in another guise. For the symmetric walk started at 0, the rule "bet one dollar per round until you are one dollar ahead" succeeds almost surely, with bounded stakes, yet $\mathbb{E}[S_{n\wedge\sigma_1}] = 0$ for every n. The price is time and credit: $\mathbb{E}[\sigma_1] = \infty$, and the stopped walk is unbounded below. A gambler whose credit stops at $a < 0$ succeeds with probability $\frac{-a}{1-a}$ and otherwise loses $-a$, for an expected gain of exactly zero. In a favourable game, by contrast, every level above the start is reached in finite expected time (Exercise 2.5).

Further reading

Williams (1991), Chapters 10–12, is the classic short introduction to discrete-time martingales; optional stopping and the hitting times of the simple random walk are in Sections 10.9–10.12. Durrett (2019), Chapter 4, covers the same ground in more depth, with random-walk applications in Section 4.8. Shreve (2004a), Chapters 2 and 5, treats martingales and the first-passage times of the random walk with the coin-toss notation used in finance. Grimmett and Stirzaker (2020), Chapter 12, has many more examples, and Föllmer and Schied (2016), Chapter 5, builds discrete-time arbitrage theory on martingale transforms.

Chapter 3

Arbitrage and static replication

How much can be said about the price of an option without any model for the price of the underlying asset? More than one might expect. If no portfolio can produce something out of nothing, the prices of related contracts must satisfy a whole network of relations: the forward price is fixed, calls and puts are tied by parity, and the price of a call, as a function of the strike, is nonincreasing, convex, and squeezed between two lines. Every relation is proved in the same way: compare two payoffs state by state, then compare their prices. And every relation comes with a trade that makes money when the relation fails.

The chapter ends with static replication. A smooth payoff $f(S_T)$ is a combination of a bond, a forward and a continuum of puts and calls (the Carr–Madan formula), so its price is determined by option prices across strikes. Read the other way, these prices determine the distribution that the market implicitly assigns to S_T. What no-arbitrage arguments cannot do is single out one price for a call: that takes a model, and Chapter 4 builds the first one.

3.1 Assets, positions and payoffs

There are two dates: today, 0, and a fixed horizon $T > 0$. What happens at T is uncertain, and the uncertainty is described by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is the real-world measure. In this chapter the probabilities themselves play almost no role: only which events have probability zero will matter. The following assets are traded at time 0, at prices known at time 0.

- *The zero-coupon bond* pays 1 at time T in every state and costs $B > 0$ at time 0. Buying bonds is lending; selling bonds is borrowing.
- *The stock* costs $S_0 > 0$ at time 0 and is worth S_T at time T, a random variable with $S_T \geq 0$ (limited liability: a share cannot be worth less than nothing). It pays no dividend between 0 and T.
- *European options.* A *call* with strike $K \geq 0$ and maturity T gives its holder the right, but not the obligation, to buy one share at time T for the price K. The holder exercises it exactly when $S_T > K$, so the call pays $(S_T - K)^+$. A *put* gives the right to sell one share for K and pays $(K - S_T)^+$. Their prices at time 0 are $C(K)$ and $P(K)$; the maturity is

the same for all options and is left out of the notation.

- *Forward contracts.* A forward with delivery price K is the obligation to buy one share at time T for K. It pays $S_T - K$, which can be negative.

Interest convention. Interest is compounded once per unit of time, as in the binomial model of Chapter 4, where the unit of time is one period: a dollar deposited at time 0 at the rate $r > -1$ grows to $(1+r)^T$ at time T, so

$$B = (1+r)^{-T}. \tag{3.1}$$

With continuous compounding, used from Chapter 6 on, $B = e^{-rT}$ instead. Every result of this chapter is stated in terms of B, so it holds verbatim under either convention. Usually $r \geq 0$, that is $B \leq 1$, but we never use this unless we say so.

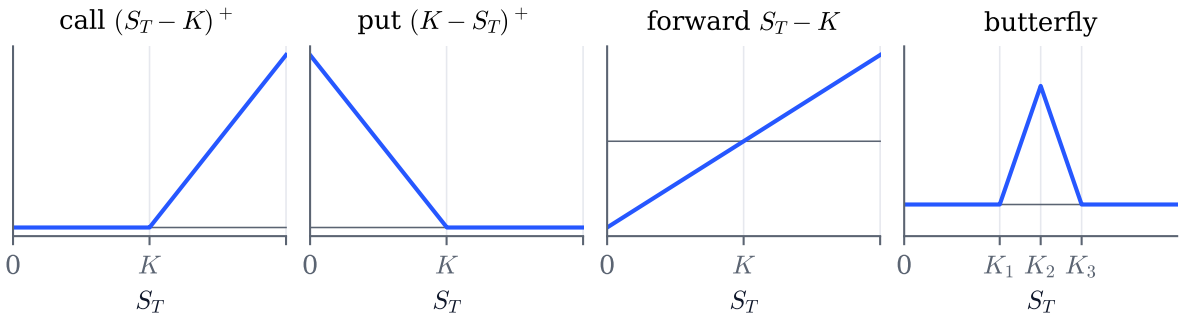

Figure 3.1: Payoffs at time T, as functions of S_T, of a long call, a long put and a long forward with strike K, and of a butterfly: long one call at K_1 and one at K_3, short two at $K_2 = \frac{1}{2}(K_1 + K_3)$. The payoff of a short position is the negative.

Frictionless markets and static portfolios. We work in an idealized market: (i) every asset can be bought or sold in any real quantity; (ii) assets are bought and sold at the same price (no bid–ask spread) and without transaction costs; (iii) every contract is honoured. A *static portfolio* holds fixed quantities $\theta_1, \dots, \theta_n$ of traded assets from time 0 to time T. If asset i costs π_i and pays X_i, assumptions (i) and (ii) make the cost and the payoff of the portfolio linear:

$$v = \sum_{i=1}^{n} \theta_i \pi_i, \qquad V = \sum_{i=1}^{n} \theta_i X_i.$$

A holding $\theta_i > 0$ is a *long* position, and $\theta_i < 0$ a *short* one. Being short an option means having written it: the writer of a call receives $C(K)$ today and pays $(S_T - K)^+$ at T. Being short the stock means having borrowed a share and sold it for S_0; at time T the short seller buys a share back for S_T and returns it to the lender, so the payoff is $-S_T$.

3.2 Arbitrage

Definition 3.1 (Arbitrage)**.** A static portfolio with cost v and payoff V is an *arbitrage* if $V \geq 0$ almost surely and either

(a) $v < 0$, or

(b) $v = 0$ and $\mathbb{P}(V > 0) > 0$.

The market is *free of arbitrage* if no portfolio of traded assets is an arbitrage.

An arbitrage of type (a) brings money today against a position that is never a liability; one of type (b) is a free lottery ticket. Since the bond is traded, an arbitrage of type (a) yields one of type (b): adding $-v/B$ bonds gives a portfolio that costs nothing and pays $V - v/B \geq -v/B > 0$ almost surely. (The converse fails in general: a lottery ticket need not bring money today.) So the market is free of arbitrage exactly when every portfolio that costs nothing and never loses has a payoff equal to zero almost surely: there is no free lunch.

Remark 3.2 (What the probability does). $\mathbb{P}$ enters Theorem 3.1 only through its null sets. Replacing $\mathbb{P}$ by an equivalent measure (Theorem 1.22) changes neither the arbitrages nor anything deduced from their absence. In particular, the probabilities that investors assign to future states, and their attitudes towards risk, play no role. The relations of Section 3.3 follow from inequalities between payoffs that hold in every state, so they hold whatever $\mathbb{P}$ is: they are *model-free*.

Proposition 3.3 (No dominance and the law of one price)**.** *Assume that the market is free of arbitrage, and consider two portfolios with costs $v, \tilde{v}$ and payoffs $V, \tilde{V}$.*

(i) (No dominance) If $V \geq \tilde{V}$ almost surely, then $v \geq \tilde{v}$. If moreover $\mathbb{P}(V > \tilde{V}) > 0$, then $v > \tilde{v}$.

(ii) (Law of one price) If $V = \tilde{V}$ almost surely, then $v = \tilde{v}$.

In particular, a portfolio whose payoff is nonnegative has a nonnegative cost; and if a portfolio replicates *a traded asset, that is, has the same payoff almost surely, its cost is the price of the asset.*

Proof. (i) Hold the first portfolio and short the second. This costs $v - \tilde{v}$ and pays $V - \tilde{V} \geq 0$ almost surely. If $v < \tilde{v}$, it is an arbitrage of type (a); if $v = \tilde{v}$ and $\mathbb{P}(V > \tilde{V}) > 0$, it is an arbitrage of type (b). Both are excluded. (ii) Apply (i) twice, exchanging the roles of the two portfolios. The last two claims are (i) with the empty portfolio as the second portfolio, and (ii) applied to the replicating portfolio and to the portfolio that holds one unit of the asset. □

Remark 3.4 (From a payoff inequality to a price relation, and back). Every result of Section 3.3 follows the same recipe. To prove a price relation, find two portfolios whose payoffs are ordered in every state, and apply Theorem 3.3. To exploit a violation of the relation, run the proof backwards: if $V \geq \tilde{V}$ in every state but the quotes have $v < \tilde{v}$, buy the cheap portfolio and sell the expensive one. This collects $\tilde{v} - v > 0$ today, and the position that remains, $V - \tilde{V}$, is never a liability. The art is in finding the two portfolios.

A second way to rule out arbitrage is to exhibit prices that are expectations.

Definition 3.5 (Pricing measure)**.** A probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$, equivalent to $\mathbb{P}$, is a *pricing measure* if every traded payoff X is $\mathbb{Q}$-integrable and has price $B\,\mathbb{E}^{\mathbb{Q}}[X]$ at time 0.

For the stock, $B\,\mathbb{E}^{\mathbb{Q}}[S_T] = S_0$: under $\mathbb{Q}$ the stock earns on average the same return as the bond. This is why a pricing measure is also called a *risk-neutral measure*. When Ω is finite and every state has positive probability, $q(\omega) = B\,\mathbb{Q}(\{\omega\})$ is the price that $\mathbb{Q}$ assigns to a claim paying 1 in the state ω and nothing otherwise (it is a market price if that claim is traded or replicable); these numbers are called *state prices*.

Lemma 3.6. *If a pricing measure exists, the market is free of arbitrage. It remains so after adding a new asset with a $\mathbb{Q}$-integrable payoff X and price $B\,\mathbb{E}^{\mathbb{Q}}[X]$.*

Proof. By linearity, a portfolio with payoff V costs $v = B\,\mathbb{E}^{\mathbb{Q}}[V]$. If $V \geq 0$ $\mathbb{P}$-almost surely, then also $\mathbb{Q}$-almost surely, because the two measures have the same null sets; hence $v \geq 0$, which excludes type (a). If moreover $v = 0$, then $\mathbb{E}^{\mathbb{Q}}[V] = 0$ for a nonnegative V, so $V = 0$ $\mathbb{Q}$-almost surely, hence $\mathbb{P}$-almost surely, which excludes type (b). For the second claim, note that $\mathbb{Q}$ is still a pricing measure for the enlarged market. □

The converse is the *fundamental theorem of asset pricing*: in a one-period market with finitely many traded assets, absence of arbitrage implies that a pricing measure exists. It is proved for the binomial model in Chapter 4 (Theorem 4.9) and in general in Föllmer and Schied (2016), Section 1.2. Harrison and Pliska (1981) proved the multi-period version on a finite probability space and developed its continuous-time counterpart, and Dalang et al. (1990) extended the discrete-time theorem to general probability spaces. We do not need it here. The relations of Section 3.3 use the absence of arbitrage directly, and pricing measures appear in Section 3.4 only as an assumption.

3.3 Price relations from static arbitrage

In this section the market of Section 3.1 is free of arbitrage. The relations were derived systematically by Merton (1973) from dominance arguments alone, without any model for the stock; put–call parity goes back at least to Stoll (1969).

Proposition 3.7 (Forward price)**.** *A forward with delivery price K is worth $S_0 - BK$ at time 0. In particular, the delivery price that makes it free to enter, the* forward price*, is*

$$F = \frac{S_0}{B} = S_0(1+r)^T.$$

Proof. One share together with $-K$ bonds (a loan of BK) pays $S_T - K$ at time T, the payoff of the forward. By the law of one price the forward is worth $S_0 - BK$, which vanishes exactly when $K = S_0/B$. □

The trade behind the proof is the *cash-and-carry*. Suppose that a dealer agrees today, at no cost, to buy the share at time T for $K > F$. Borrow S_0 (sell $S_0/B = F$ bonds), buy the share, and sell it forward to the dealer. Nothing is paid today; at time T the share is delivered for K and the loan is repaid with F, a sure profit of $K - F$. If $K < F$, do the opposite: short the share, lend the proceeds, and buy forward.

Remark 3.8 (Forwards and futures). A *futures* contract has the payoff profile of a forward, but gains and losses are paid every day as the futures price moves (marking to market). Futures

and forward prices coincide when interest rates are deterministic and can differ when they are random; see Hull (2021), Chapter 5. Only forwards appear in these notes.

Proposition 3.9 (Put–call parity)**.** *For every strike $K \geq 0$,*

$$C(K) - P(K) = S_0 - BK.$$

Proof. For every real x, $(x-K)^+ - (K-x)^+ = x - K$. So a long call and a short put with strike K pay $S_T - K$, the payoff of a forward, and by the law of one price they cost what the forward costs, $S_0 - BK$ by Theorem 3.7. □

At the forward price $K = F$ the right-hand side vanishes: a call and a put struck at the forward have the same price. If the quotes have $C(K) - P(K) > S_0 - BK$, the call is expensive relative to the rest: sell the call, buy the put, buy one share and borrow BK. This collects $C(K) - P(K) - S_0 + BK > 0$ today, and the position pays $-(S_T - K)^+ + (K - S_T)^+ + S_T - K = 0$ in every state. The reverse inequality is traded in reverse.

Proposition 3.10 (Bounds)**.** *For every strike $K \geq 0$,*

$$(S_0 - BK)^+ \leq C(K) \leq S_0, \qquad (BK - S_0)^+ \leq P(K) \leq BK.$$

Proof. For every $x \geq 0$,

$$0 \leq (x-K)^+, \quad x - K \leq (x-K)^+ \leq x, \qquad 0 \leq (K-x)^+, \quad K - x \leq (K-x)^+ \leq K,$$

where both upper bounds use $x \geq 0$ (and the first one also $K \geq 0$). At $x = S_T$, each inequality compares an option with nothing, with a long or short forward, with one share, or with K bonds, and Theorem 3.3 turns it into the corresponding inequality between prices. For example, $(S_T - K)^+ \geq S_T - K$ and Theorem 3.7 give $C(K) \geq S_0 - BK$. □

Each bound has its trade. If $C(K) < S_0 - BK$, buy the call, short one share and lend BK: this collects $S_0 - BK - C(K) > 0$, and the position pays $(S_T - K)^+ - S_T + K = (K - S_T)^+ \geq 0$. If $C(K) > S_0$, sell the call and buy one share: this collects $C(K) - S_0$, and the position pays $S_T - (S_T - K)^+ = \min(S_T, K) \geq 0$.

Remark 3.11 (Strict bounds and early exercise). The second half of Theorem 3.3(i) makes the bounds strict when the payoffs differ with positive probability: $C(K) > 0$ if $\mathbb{P}(S_T > K) > 0$, and $C(K) > S_0 - BK$ if $\mathbb{P}(S_T < K) > 0$. When $B \leq 1$, the lower bound gives $C(K) \geq S_0 - BK \geq S_0 - K$: a European call on a stock without dividends is worth at least what exercising it immediately would bring. Applied at every date before maturity, provided the bond price for the remaining maturity is at most 1 at every date (interest rates stay nonnegative), this observation of Merton (1973) shows that nothing is gained by exercising an American call on such a stock before maturity; American options are the subject of Chapter 4. Nothing similar holds for puts, whose lower bound $BK - S_0$ lies below the exercise value $K - S_0$ when $B < 1$, nor for calls on a stock that pays dividends (Exercise 3.2).

The last group of relations concerns the shape of the price as a function of the strike.

Proposition 3.12 (Prices as functions of the strike)**.** *Let $0 \leq K_1 < K_2$ and $\lambda \in [0,1]$. Then*

(i) (monotonicity) $C(K_1) \geq C(K_2)$ and $P(K_1) \leq P(K_2)$;

(ii) (slope bounds) $C(K_1) - C(K_2) \le B(K_2 - K_1)$ *and* $P(K_2) - P(K_1) \le B(K_2 - K_1)$*;*

(iii) (convexity) $C(\lambda K_1 + (1-\lambda)K_2) \le \lambda C(K_1) + (1-\lambda)C(K_2)$*, and the same holds for* P*. Moreover* $C(0) = S_0$ *and* $P(0) = 0$*. In words: on* $[0, \infty)$*, the map* $K \mapsto C(K)$ *is convex and nonincreasing, with difference quotients in* $[-B, 0]$*, and* $K \mapsto P(K)$ *is convex and nondecreasing, with difference quotients in* $[0, B]$*.*

Proof. Fix $x \ge 0$ and look at the call payoff $K \mapsto (x-K)^+$ as a function of the strike. It is nonincreasing; it is convex, as the maximum of the two affine functions 0 and $x - K$; and

$$(x-K_1)^+ - (x-K_2)^+ = \min\left((x-K_1)^+,\, K_2 - K_1\right) \in [0, K_2 - K_1]. \tag{3.2}$$

At $x = S_T$ these properties say, in every state: (i) $(S_T - K_1)^+ \ge (S_T - K_2)^+$; (ii) the *call spread*, long a call at K_1 and short a call at K_2, pays at most $K_2 - K_1$, the payoff of $K_2 - K_1$ bonds; (iii) with $K_\lambda = \lambda K_1 + (1-\lambda)K_2$,

$$\lambda(S_T - K_1)^+ + (1-\lambda)(S_T - K_2)^+ - (S_T - K_\lambda)^+ \ge 0. \tag{3.3}$$

Theorem 3.3 turns each of them into the claimed relation for calls. A call with strike 0 pays S_T, so $C(0) = S_0$ by the law of one price; a put with strike 0 pays nothing, so $P(0) = 0$. The put relations follow in the same way from $K \mapsto (K-x)^+$, or from parity: $P(K) = C(K) - S_0 + BK$ differs from $C(K)$ by an affine function of K with slope B. This preserves convexity, turns (ii) for calls into (i) for puts, and turns (i) for calls into (ii) for puts. □

Run backwards, as in Theorem 3.4, the proof gives the trades. If (i) fails, buy the call at K_1 and sell the call at K_2: a call spread that costs less than nothing. If (ii) fails, sell that call spread and lend $B(K_2 - K_1)$. If (iii) fails, buy the portfolio in (3.3).

The portfolio in (3.3) is a *butterfly*: long λ calls at K_1 and $1-\lambda$ calls at K_2, short one call at the strike K_λ in between. Its payoff is a tent, zero outside (K_1, K_2) and largest at K_λ (Figure 3.1). Convexity of C says that a butterfly never has a negative price.

Example 3.13 (A butterfly arbitrage)**.** Let $S_0 = 100$ and $B = 0.97$, and suppose that calls with strikes 95, 100 and 105 are quoted at 8.00, 5.50 and 2.60. The bounds, monotonicity and the slope bounds all hold, but $\frac{1}{2}(8.00 + 2.60) = 5.30 < 5.50$: the middle call is too expensive. Sell two calls with strike 100 and buy one call with strike 95 and one with strike 105. This collects $11.00 - 10.60 = 0.40$ today. At time T the position pays $(S_T - 95)^+ - 2(S_T - 100)^+ + (S_T - 105)^+$, which is zero outside $(95, 105)$, rises to 5 at $S_T = 100$, and is never negative.

Checking a quoted curve. In practice calls are quoted at finitely many strikes $0 < K_1 < \cdots < K_n$. Put $K_0 = 0$ and $c_0 = S_0$ (the stock is the call with strike 0), let c_i be the quote at K_i, and let

$$s_i = \frac{c_i - c_{i-1}}{K_i - K_{i-1}}, \qquad i = 1, \dots, n,$$

be the slopes of the piecewise-linear interpolation of the quotes. Theorems 3.10 and 3.12 require

$$-B \le s_1 \le s_2 \le \cdots \le s_n \le 0 \qquad \text{and} \qquad c_n \ge 0. \tag{3.4}$$

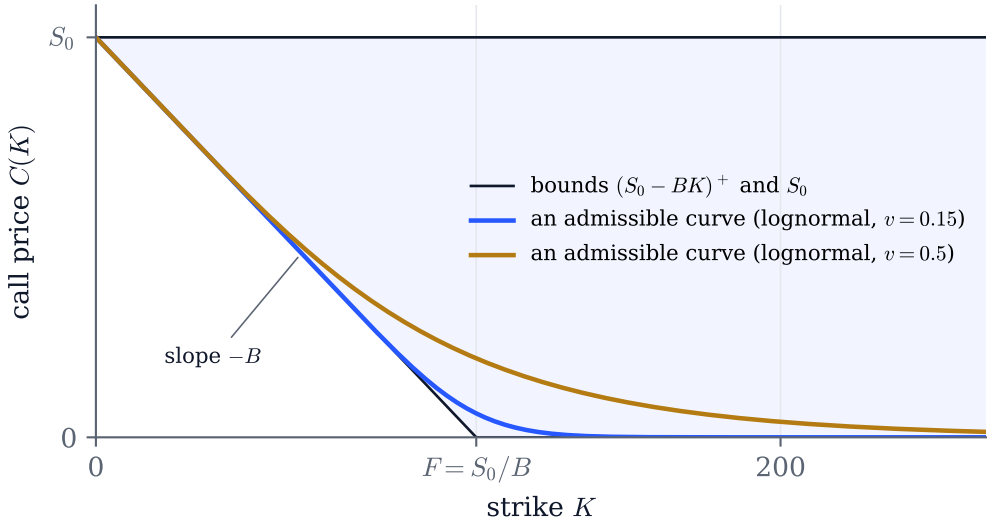

Figure 3.2: No-arbitrage constraints on the call price as a function of the strike. Every admissible curve starts at $C(0) = S_0$, stays in the shaded region between $(S_0 - BK)^+$ and S_0, and is convex and nonincreasing with slopes in $[-B, 0]$ (Theorems 3.10 and 3.12). Two such curves are shown, from the lognormal model of Chapter 6 with total volatility $v = 0.15$ and $v = 0.5$; here $S_0 = 100$ and $B = 0.9$.

Conversely, these finitely many conditions imply every inequality of Theorems 3.10 and 3.12 between quoted strikes: a piecewise-linear function with nondecreasing slopes is convex, and, for instance, $c_i = S_0 + \sum_{j \le i} s_j(K_j - K_{j-1}) \ge S_0 - BK_i$. Whether quotes satisfying (3.4) allow an arbitrage depends on $\mathbb{P}$: with $S_0 = 100$, $B = 1$ and $C(100) = 5$ they hold, yet if $\mathbb{P}(S_T \le 100) = 1$, selling the call is an arbitrage. The right question is whether there is always a model in which such quotes are arbitrage-free. Yes, except in one boundary case.

Proposition 3.14 (The relations are complete)**.** *Let $c_0 = S_0, c_1, \dots, c_n$ satisfy (3.4), with $c_n = 0$ if $s_n = 0$. Then there are finitely many points $x_j \ge 0$ and weights $q_j > 0$ such that*

$$\sum_j q_j = B, \qquad \sum_j q_j x_j = S_0, \qquad \sum_j q_j (x_j - K_i)^+ = c_i \quad (i = 1, \dots, n).$$

Consequently, in the model where S_T takes the values x_j with probabilities q_j/B, $\mathbb{Q} = \mathbb{P}$ is a pricing measure for the bond, the stock and the calls with prices c_i, and this market is free of arbitrage.

Proof. Put weight $B + s_1$ at 0, weight $s_{i+1} - s_i$ at K_i for $1 \le i \le n-1$, and, if $s_n < 0$, weight $-s_n$ at $x_* = K_n + c_n/(-s_n)$. By (3.4) all weights are nonnegative; discard those that vanish. The total weight telescopes to B. Let $G(K) = \sum_j q_j (x_j - K)^+$. On each interval $[K_{i-1}, K_i]$, $1 \le i \le n$, no point x_j lies strictly inside, so G is affine there, with slope

$$-\sum_{x_j \ge K_i} q_j = -\Big(\sum_{l=i}^{n-1} (s_{l+1} - s_l) - s_n\Big) = s_i.$$

Moreover $G(K_n) = -s_n(x_* - K_n) = c_n$ if $s_n < 0$, and $G(K_n) = 0 = c_n$ if $s_n = 0$. Two functions

that are affine with the same slope on each $[K_{i-1}, K_i]$ and agree at K_n agree at every K_i. Hence $G(K_i) = c_i$ for $i \geq 1$, and $G(0) = \sum_j q_j x_j = c_0 = S_0$. The last claim is Theorem 3.6. □

If K_{i-1}, K_i, K_{i+1} are equally spaced (for $i = 1$ this means $K_1 = K_2 - K_1$, since $K_0 = 0$), the weight $s_{i+1} - s_i$ at K_i is $(c_{i+1} - 2c_i + c_{i-1})/(K_{i+1} - K_i)$: the price of a butterfly, divided by the spacing. In the boundary case $s_n = 0 < c_n$ the last call spread costs nothing and pays off when $S_T > K_{n-1}$; it is an arbitrage in every model in which $\mathbb{P}(S_T > K_{n-1}) > 0$, and in the other models selling the last call is one, but there need not be a single portfolio that is an arbitrage in every model (for an instance, see Exercise 3.8). Apart from this case, every finite set of quotes satisfying (3.4), that is, compatible with Figure 3.2, is consistent with some model: no-arbitrage alone gives relations between prices, not prices. (For a continuum of strikes one must add $C(K) \to 0$ as $K \to \infty$, which holds under any pricing measure by dominated convergence.)

▶ Interactive demo: perturb a quoted call-price curve and watch which relations break; for each violation the checker returns the arbitrage portfolio, its profit and its payoff.

3.4 Static replication

The relations of Section 3.3 came from payoffs built out of a few calls. Conversely, many payoffs can be built out of calls. We start with piecewise-linear ones.

Lemma 3.15 (Piecewise-linear payoffs)**.** *Let $g : [0, \infty) \to \mathbb{R}$ be continuous and piecewise linear, with kinks only at $0 < K_1 < \cdots < K_m$, and let $w_i = g'(K_i+) - g'(K_i-)$ be the change of slope at K_i. Then*

$$g(x) = g(0) + g'(0+)\, x + \sum_{i=1}^{m} w_i (x - K_i)^+ \qquad \text{for all } x \geq 0.$$

Hence $g(S_T)$ is replicated by $g(0)$ bonds, $g'(0+)$ shares and w_i calls with strike K_i, and its price is $Bg(0) + g'(0+)S_0 + \sum_i w_i C(K_i)$.

Proof. Both sides are continuous and piecewise linear with kinks only at the K_i. They agree at 0 and have the same slope on $[0, K_1]$, and their slopes jump by the same amount w_i at each K_i, because $(x - K_i)^+$ is linear away from K_i and its slope jumps by 1 there. So they have the same slope on every interval and agree everywhere. The price follows from the law of one price. □

The forward, the call spread and the butterfly are examples; Exercise 3.1 has more. A smooth payoff is a limit of piecewise-linear ones, in which the changes of slope w_i become $f''(K)\,\mathrm{d}K$. It is convenient to expand around a point $\kappa > 0$ rather than around 0, with puts below κ and calls above.

Theorem 3.16 (Carr–Madan formula)**.** *Let $f : (0, \infty) \to \mathbb{R}$ be twice continuously differentiable and $\kappa > 0$. For every $x > 0$,*

$$f(x) = f(\kappa) + f'(\kappa)(x - \kappa) + \int_0^\kappa f''(K)\,(K - x)^+ \,\mathrm{d}K + \int_\kappa^\infty f''(K)\,(x - K)^+ \,\mathrm{d}K. \tag{3.5}$$

For each x, at most one of the two integrals is nonzero, and it is the integral of a continuous function over the bounded interval between x and κ. If f is twice continuously differentiable

on $[0,\infty)$, (3.5) *also holds at* $x=0$.

Proof. We use Taylor's formula with integral remainder: for $x,\kappa>0$,

$$f(x)=f(\kappa)+f'(\kappa)(x-\kappa)+\int_\kappa^x f''(K)\,(x-K)\,\mathrm{d}K, \tag{3.6}$$

which holds because an integration by parts gives

$$\int_\kappa^x f''(K)(x-K)\,\mathrm{d}K=\Big[f'(K)(x-K)\Big]_{K=\kappa}^{K=x}+\int_\kappa^x f'(K)\,\mathrm{d}K=-f'(\kappa)(x-\kappa)+f(x)-f(\kappa).$$

If $x\ge\kappa$, then $x-K=(x-K)^+$ for $K\in[\kappa,x]$ and $(x-K)^+=0$ for $K>x$, so the remainder in (3.6) is the call integral of (3.5); the put integral vanishes because $(K-x)^+=0$ for $K\le\kappa\le x$. If $x<\kappa$, the remainder is $\int_x^\kappa f''(K)(K-x)\,\mathrm{d}K$, which is the put integral because $(K-x)^+=0$ for $K\le x$; the call integral vanishes because $(x-K)^+=0$ for $K\ge\kappa>x$. If f is twice continuously differentiable on $[0,\infty)$, the same computation applies at $x=0$. □

Read as a portfolio, (3.5) replicates $f(S_T)$ statically with $f(\kappa)$ bonds, $f'(\kappa)$ forwards with delivery price κ, and, for every strike K, $f''(K)\,\mathrm{d}K$ puts if $K<\kappa$ and calls if $K>\kappa$. All the options are out of the money with respect to κ. The usual choice is the forward price, $\kappa=F$, for which the forwards cost nothing (Theorem 3.7). The proof uses only Taylor's formula, which holds more generally whenever f' is locally absolutely continuous on $(0,\infty)$, with f'' its derivative defined almost everywhere. The piecewise-linear payoffs of Theorem 3.15 are the discrete counterpart: a kink of size w_i at K_i plays the role of $f''(K)\,\mathrm{d}K$.

Finitely many strikes. Markets quote finitely many strikes. Let $k_0<k_1<\cdots<k_m$ be a grid of positive strikes with spacing at most h, and let f_h be the continuous piecewise-linear function equal to f at the grid points and extended linearly beyond k_0 and k_m. By Theorem 3.15, $f_h(S_T)$ is replicated exactly by bonds, shares and calls at the grid strikes; by parity, the calls with strikes below κ can be exchanged for puts, as in (3.5). The weight at k_i is the change of slope of f_h, which for equal spacing is $\big(f(k_{i+1})-2f(k_i)+f(k_{i-1})\big)/h\approx f''(k_i)\,h$: a Riemann sum for the strip. The replication error vanishes at the grid points and, between two of them, is bounded by

$$|f(x)-f_h(x)|\le\frac{h^2}{8}\max_{[k_{i-1},k_i]}|f''|,\qquad x\in[k_{i-1},k_i]. \tag{3.7}$$

To see this, take $x\in(a,b)=(k_{i-1},k_i)$ and $E=f(x)-f_h(x)$. The function $\varphi(t)=f(t)-f_h(t)-E\frac{(t-a)(t-b)}{(x-a)(x-b)}$ vanishes at a, x and b, so by Rolle's theorem, applied twice, $\varphi''(\xi)=0$ for some $\xi\in(a,b)$. That is, $E=\frac12 f''(\xi)(x-a)(x-b)$, and $|(x-a)(x-b)|\le(b-a)^2/4$. Outside $[k_0,k_m]$ the error is that of a linear extrapolation and need not be small: the strikes must cover the region where S_T is likely to end up (Figure 3.3).

▶ Interactive demo: replicate a smooth payoff with a finite strip of puts and calls, and move the strike spacing and range to see the replication error.

To turn (3.5) into a price, we need to price a portfolio holding a continuum of options. A pricing measure makes this precise.

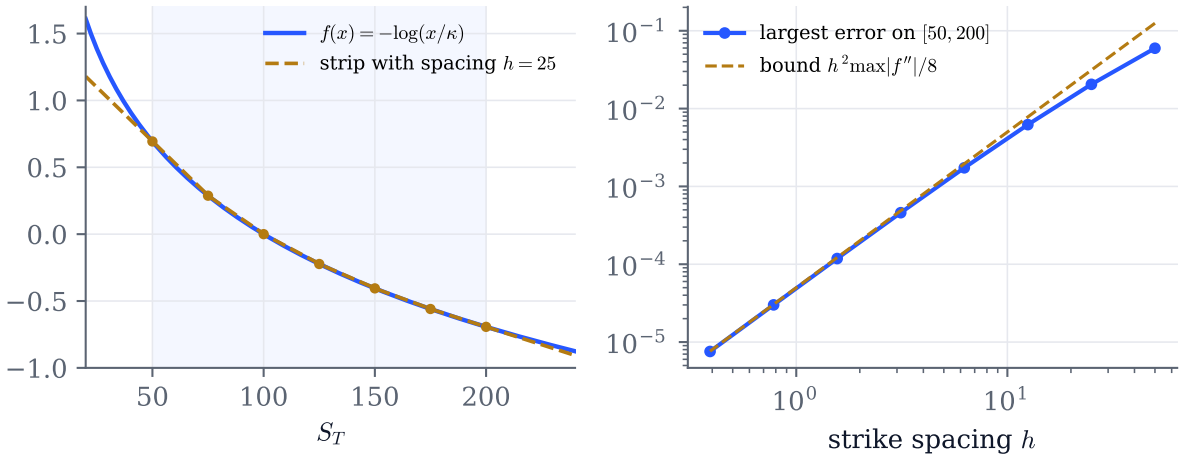

Figure 3.3: Static replication of $f(x) = -\log(x/\kappa)$, $\kappa = 100$, with strikes on the grid $50, 75, \ldots, 200$ (left). The replicated payoff matches f at the strikes, is close to it in between, and is a linear extrapolation outside the grid. Right: the largest error on $[50, 200]$ as the spacing h shrinks, against the bound (3.7).

Corollary 3.17 (Pricing with a strip of options)**.** *Let $\mathbb{Q}$ be a pricing measure for a market in which calls and puts with every strike $K > 0$ are traded, and assume $\mathbb{Q}(S_T > 0) = 1$. Let f and κ be as in Theorem 3.16, and assume that*

$$\int_0^\kappa |f''(K)|\, P(K)\, \mathrm{d}K + \int_\kappa^\infty |f''(K)|\, C(K)\, \mathrm{d}K < \infty. \tag{3.8}$$

Then $f(S_T)$ is $\mathbb{Q}$-integrable, and its price $B\,\mathbb{E}^{\mathbb{Q}}[f(S_T)]$ (Theorem 3.6) is

$$Bf(\kappa) + f'(\kappa)(S_0 - B\kappa) + \int_0^\kappa f''(K)\, P(K)\, \mathrm{d}K + \int_\kappa^\infty f''(K)\, C(K)\, \mathrm{d}K. \tag{3.9}$$

Proof. We use two results on exchanging an expectation with an integral in K, for a function $g(K, \omega)$ that is jointly measurable (here it is continuous in K): *Tonelli's theorem*, by which $\mathbb{E}^{\mathbb{Q}}\big[\int g\, \mathrm{d}K\big] = \int \mathbb{E}^{\mathbb{Q}}[g]\, \mathrm{d}K$ whenever $g \geq 0$, and *Fubini's theorem*, by which the same identity holds whenever $\int \mathbb{E}^{\mathbb{Q}}[|g|]\, \mathrm{d}K < \infty$; see Durrett (2019), Section 1.7. Write $X = S_T$, so that $P(K) = B\,\mathbb{E}^{\mathbb{Q}}[(K - X)^+]$ and $C(K) = B\,\mathbb{E}^{\mathbb{Q}}[(X - K)^+]$. By Tonelli's theorem, $\mathbb{E}^{\mathbb{Q}}\big[\int_0^\kappa |f''(K)|(K - X)^+\, \mathrm{d}K\big] = B^{-1} \int_0^\kappa |f''(K)| P(K)\, \mathrm{d}K < \infty$, and likewise for the call integral. So both integrals in (3.5), evaluated at $x = X$, are $\mathbb{Q}$-integrable, and so is $f(X)$ by (3.5), since X is. Take $\mathbb{E}^{\mathbb{Q}}$ in (3.5), exchange $\mathbb{E}^{\mathbb{Q}}$ with the integrals in K by Fubini's theorem, multiply by B and use $B\,\mathbb{E}^{\mathbb{Q}}[X] = S_0$. □

The price (3.9) involves only market prices. A pricing measure is needed to give a meaning to a portfolio of infinitely many options, but which one it is does not matter: all pricing measures that reproduce the same option prices give $f(S_T)$ the same price. Read in the other direction, option prices across strikes pin down the law of S_T under $\mathbb{Q}$.

Proposition 3.18 (Breeden–Litzenberger)**.** *Let $\mathbb{Q}$ be a pricing measure for a market in which calls with every strike $K > 0$ are traded.*

(i) For every $K > 0$, the call price has one-sided derivatives

$$\partial^+ C(K) = -B\,\mathbb{Q}(S_T > K), \qquad \partial^- C(K) = -B\,\mathbb{Q}(S_T \geq K).$$

(ii) If the law of S_T under $\mathbb{Q}$ has a density q_T that is continuous on an open interval $I \subseteq (0,\infty)$, then C is twice continuously differentiable on I and

$$C''(K) = B\,q_T(K), \qquad K \in I.$$

(iii) Conversely, if C is differentiable on an open interval $I \subseteq (0,\infty)$ and C' is locally absolutely continuous on I, then on I the law of S_T under $\mathbb{Q}$ has the density C''/B: $\mathbb{Q}(a < S_T \leq b) = B^{-1}\int_a^b C''(K)\,\mathrm{d}K$ for $a < b$ in I.

Proof. (i) For $h > 0$, (3.2) gives

$$\frac{C(K+h) - C(K)}{h} = -B\,\mathbb{E}^{\mathbb{Q}}\Big[\frac{\min\big((S_T - K)^+, h\big)}{h}\Big].$$

The variable inside the expectation takes values in $[0,1]$ and converges to $\mathbf{1}_{\{S_T > K\}}$ as $h \downarrow 0$, so dominated convergence gives the right derivative. In the same way, for $0 < h < K$, $\big(C(K) - C(K-h)\big)/h = -B\,\mathbb{E}^{\mathbb{Q}}\big[\min\big((S_T - K + h)^+, h\big)/h\big]$, and the variable inside converges to $\mathbf{1}_{\{S_T \geq K\}}$.

(ii) With a density, $\mathbb{Q}(S_T = K) = 0$ for every K, so by (i) C is differentiable on $(0,\infty)$ with $C'(K) = -B\int_K^\infty q_T(y)\,\mathrm{d}y$. By the fundamental theorem of calculus, C' is differentiable at every point of I, where q_T is continuous, with derivative $B\,q_T(K)$, a continuous function on I.

(iii) By (i), $C'(K) = \partial^+ C(K) = -B\,\mathbb{Q}(S_T > K)$ for $K \in I$. For $a < b$ in I, $\mathbb{Q}(a < S_T \leq b) = \big(C'(b) - C'(a)\big)/B = B^{-1}\int_a^b C''(K)\,\mathrm{d}K$, by absolute continuity. □

So the curvature of the call price is the risk-neutral density, up to the factor B, and convexity in the strike (Theorem 3.12) is the statement that this density is nonnegative. A butterfly with strikes $K - h$, K, $K + h$ costs approximately $B\,h^2 q_T(K)$; divided by h, it is the price of a bet that S_T ends near K, the continuous counterpart of the weights in the proof of Theorem 3.14. The law of S_T under $\mathbb{Q}$ need not be its law under $\mathbb{P}$: the risk-neutral density says what the market charges for payoffs, not how likely they are. Part (iii) is the version used on market data. It needs a differentiable call curve, which raw quotes at finitely many strikes do not provide: second differences of quotes are only rough estimates of C'', and they amplify rounding errors (Exercise 3.7). In practice one first fits an arbitrage-free curve to the quotes.

Remark 3.19 (What a model adds). Everything in this chapter is static: portfolios are set up at time 0 and never touched. Static arguments give relations between prices, not prices (Theorem 3.14). A static portfolio of the bond and the stock pays an affine function of S_T, and the payoff of a call is not affine, so in general a call cannot be replicated in this way; the exception, when S_T takes only two values, is the one-period binomial model. There are two ways forward. One is a model for how S evolves, with trading at intermediate dates: a call can then be replicated *dynamically* by the bond and the stock, and its price is the cost of that strategy.

This is the binomial model of Chapter 4 and the Black–Scholes model of Chapter 6. The other is to take option prices from the market and use (3.9) to price other payoffs; Chapter 15 does this for variance swaps.

Further reading

Merton (1973) derives the relations of Section 3.3, and more, from dominance arguments; it is still a very readable paper. Hull (2021), Chapters 11 and 12, discusses the same relations and the classical option strategies (spreads, straddles, butterflies) from a market perspective. Föllmer and Schied (2016), Chapter 1, develops the one-period theory in full generality: the fundamental theorem of asset pricing, and the interval of arbitrage-free prices of a payoff that cannot be replicated. Carr and Madan (1998) state the spanning formula (3.5) and use it to trade volatility, and Breeden and Litzenberger (1978) is the original source of Theorem 3.18. The one-period binomial model, which is where Chapter 4 starts, is in Shreve (2004a), Chapter 1.

Chapter 4

The binomial model

Chapter 3 priced claims by static replication: a forward contract, or a combination of calls and puts, is matched by a portfolio bought today and held to maturity. A call option cannot be replicated in this way by the stock and the money market, because its payoff is not an affine function of the stock price. It can be replicated *dynamically*, by a portfolio that is rebalanced as prices move. The simplest model in which this can be done is the binomial model of Cox et al. (1979): in each period the stock price is multiplied by one of two factors.

Everything in this chapter is finite, and every statement comes down to a finite system of linear equations or inequalities. Yet the central ideas of the course all appear: replication, the risk-neutral measure, discounted prices as martingales, the fundamental theorem of asset pricing, optimal stopping for American options and, with the trinomial model, what goes wrong when replication fails. Chapter 6 lets the number of periods tend to infinity and recovers the Black–Scholes formula.

4.1 The model

Definition 4.1 (The binomial model)**.** Fix a number of periods $N \geq 1$, an initial stock price $S_0 > 0$, factors $0 < d < u$, an interest rate $r > -1$ per period, and $p \in (0,1)$.

(i) The outcomes are the sequences of N coin tosses, $\Omega = \{H,T\}^N$, written $\omega = \omega_1\omega_2\cdots\omega_N$. The tosses are independent, with probability p of heads: $\mathbb{P}(\{\omega\}) = p^{\#H(\omega)}(1-p)^{\#T(\omega)}$, where $\#H(\omega)$ and $\#T(\omega)$ count the heads and the tails in ω.

(ii) The stock price starts at S_0 and is multiplied by u after a head and by d after a tail: $S_n = uS_{n-1}$ if $\omega_n = H$ and $S_n = dS_{n-1}$ if $\omega_n = T$, for $1 \leq n \leq N$.

(iii) The money market turns one dollar at time 0 into $(1+r)^n$ at time n; money can be lent and borrowed at this rate.

(iv) The information at time n is $\mathcal{F}_n = \sigma(\omega_1,\ldots,\omega_n)$, generated by the first n tosses, with $\mathcal{F}_0 = \{\emptyset,\Omega\}$.

The *one-period model* is the case $N = 1$.

The filtration $(\mathcal{F}_n)$ (Theorem 1.7) is generated by the partition of Ω into the 2^n sets on which

$\omega_1 \cdots \omega_n$ is fixed, so a random variable is $\mathcal{F}_n$-measurable if and only if it depends on ω only through the first n tosses (Theorem 1.2); we then write it $Y(\omega_1 \cdots \omega_n)$. For instance $S_n = S_0 u^{H_n} d^{n-H_n}$, with H_n the number of heads among the first n tosses. Since $d < u$, the price moves reveal the tosses: $\mathcal{F}_n = \sigma(S_1, \ldots, S_n)$. Every outcome has positive probability, so "almost surely" means "for every ω", and conditional expectations given $\mathcal{F}_n$ are defined at every ω by the formula for finite partitions (Theorem 1.12).

A *claim* (or derivative) with maturity N is a random variable V_N on Ω: the amount it pays at time N. It may depend only on S_N, as the call $(S_N - K)^+$ and the put $(K - S_N)^+$, or on the whole path, as the lookback $(\max_{n \le N} S_n - K)^+$.

Definition 4.2 (Self-financing portfolio)**.** A *portfolio* consists of an initial capital $X_0 \in \mathbb{R}$ and a process $(\Delta_n)_{0 \le n \le N-1}$ adapted to $(\mathcal{F}_n)$: Δ_n is the number of shares held from time n to time $n+1$. Its value process is defined by

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)\,(X_n - \Delta_n S_n), \qquad 0 \le n \le N-1. \tag{4.1}$$

At time n the investor splits the wealth X_n into $\Delta_n S_n$ in the stock and $X_n - \Delta_n S_n$ in the money market; a negative Δ_n is a short sale and a negative cash position a loan. The equation says that one period later the wealth is the value of these two positions, with no money added or withdrawn: the portfolio finances itself. Dividing (4.1) by $(1+r)^{n+1}$,

$$(1+r)^{-(n+1)} X_{n+1} - (1+r)^{-n} X_n = \Delta_n \Big((1+r)^{-(n+1)} S_{n+1} - (1+r)^{-n} S_n \Big). \tag{4.2}$$

The discounted wealth is X_0 plus the gains of the strategy in the discounted stock: the martingale transform of Theorem 2.7, with the predictable position $H_{n+1} = \Delta_n$ held over the period from n to $n+1$.

An *arbitrage* is a portfolio with $X_0 = 0$, $X_N \ge 0$ almost surely and $\mathbb{P}(X_N > 0) > 0$: type (b) of Theorem 3.1, for a portfolio rebalanced over time (type (a) reduces to it by investing $-X_0 > 0$ in the money market). As every outcome has positive probability, this means $X_N \ge 0$ everywhere and $X_N > 0$ somewhere; p plays no role.

4.2 One period

In the one-period model, a portfolio is a pair of numbers (X_0, Δ_0) and its value at time 1 is

$$X_1 = (1+r)X_0 + \Delta_0(S_1 - (1+r)S_0). \tag{4.3}$$

Proposition 4.3 (No arbitrage in one period)**.** *The one-period binomial model admits no arbitrage if and only if $d < 1 + r < u$.*

Proof. With $X_0 = 0$, (4.3) gives $X_1(H) = \Delta_0(u - 1 - r)S_0$ and $X_1(T) = \Delta_0(d - 1 - r)S_0$.

Suppose $d < 1 + r < u$, so that the first bracket is positive and the second negative. If $\Delta_0 > 0$ then $X_1(T) < 0$; if $\Delta_0 < 0$ then $X_1(H) < 0$; if $\Delta_0 = 0$ then $X_1 = 0$. No choice is an arbitrage.

Conversely, if $d \ge 1+r$, buy one share with borrowed money, $\Delta_0 = 1$: then $X_1(T) = (d-1-r)S_0 \ge 0$ and $X_1(H) = (u - 1 - r)S_0 > 0$ because $u > d$. If $u \le 1 + r$, sell one share short and invest

the proceeds, $\Delta_0 = -1$: then $X_1(H) = (1+r-u)S_0 \geq 0$ and $X_1(T) = (1+r-d)S_0 > 0$. Both are arbitrages. □

The condition says that neither the stock nor the money market dominates the other in the sense of Theorem 3.3.

Theorem 4.4 (Replication in one period)**.** *Assume $d < 1+r < u$ and set*

$$q_u = \frac{1+r-d}{u-d}, \qquad q_d = \frac{u-(1+r)}{u-d}. \tag{4.4}$$

Then $q_u, q_d \in (0,1)$ and $q_u + q_d = 1$. Every claim V_1 is replicated by exactly one portfolio, namely

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}, \qquad X_0 = \frac{q_u V_1(H) + q_d V_1(T)}{1+r}. \tag{4.5}$$

The number $V_0 := X_0$ is the only price of the claim at time 0 that does not create an arbitrage when the claim is traded together with the stock and the money market.

Proof. $q_u > 0$ because $1+r > d$, $q_d > 0$ because $u > 1+r$, and $q_u + q_d = (u-d)/(u-d) = 1$.

By (4.3), (X_0, Δ_0) replicates V_1 if and only if

$$(1+r)X_0 + \Delta_0(u-1-r)S_0 = V_1(H), \qquad (1+r)X_0 + \Delta_0(d-1-r)S_0 = V_1(T).$$

Subtracting the second equation from the first gives $\Delta_0(u-d)S_0 = V_1(H) - V_1(T)$, which determines Δ_0 because $u \neq d$; this is the first formula, as $S_1(H) - S_1(T) = (u-d)S_0$. Substituting into the first equation and using $(u-1-r)/(u-d) = q_d$,

$$(1+r)X_0 = V_1(H) - q_d(V_1(H) - V_1(T)) = q_u V_1(H) + q_d V_1(T).$$

So there is at most one replicating portfolio. It does replicate: with $(d-1-r)/(u-d) = -q_u$, the left side of the second equation equals $q_u V_1(H) + q_d V_1(T) - q_u(V_1(H) - V_1(T)) = V_1(T)$.

Suppose the claim can be bought or sold at time 0 for a price c. If $c > V_0$, sell the claim, invest V_0 in the replicating portfolio and $c - V_0$ in the money market: at time 1 the portfolio pays V_1, which settles the claim, and $(1+r)(c-V_0) > 0$ is left in every state. If $c < V_0$, buy the claim, run the replicating portfolio in reverse ($-\Delta_0$ shares) and invest $V_0 - c$. Either way there is an arbitrage; this is the law of one price of Theorem 3.3. Conversely, the probability $\mathbb{Q}$ with $\mathbb{Q}(H) = q_u$ and $\mathbb{Q}(T) = q_d$ is a pricing measure (Theorem 3.5) for the money market and the stock, with $B = (1+r)^{-1}$: indeed $\mathbb{E}^{\mathbb{Q}}[S_1] = (q_u u + q_d d)S_0$ and $(1+r-d)u + (u-1-r)d = (1+r)(u-d)$, so $\mathbb{E}^{\mathbb{Q}}[S_1]/(1+r) = S_0$. Since $V_0 = \mathbb{E}^{\mathbb{Q}}[V_1]/(1+r)$, Theorem 3.6 shows that adding the claim at the price V_0 creates no arbitrage. □

4.2.1 The risk-neutral measure

The weights q_u and q_d are positive and add up to one: they are probabilities, and (4.5) says that the price is a discounted expectation under them. This is the central object of the chapter, and we define it directly for N periods.

Definition 4.5 (Risk-neutral measure)**.** A probability measure $\mathbb{Q}$ on Ω is *risk-neutral*, or an *equivalent martingale measure*, for the N-period binomial model if

(i) $\mathbb{Q}$ is equivalent to $\mathbb{P}$ (Theorem 1.22); since every outcome has positive $\mathbb{P}$-probability, this means $\mathbb{Q}(\omega) := \mathbb{Q}(\{\omega\}) > 0$ for every $\omega \in \Omega$;

(ii) the discounted stock price $\big((1+r)^{-n}S_n\big)_{0\le n\le N}$ is a $\mathbb{Q}$-martingale with respect to $(\mathcal{F}_n)$ (Theorem 2.1).

For $N = 1$ this is a pricing measure in the sense of Theorem 3.5 for the money market and the stock. Condition (ii) reads $\mathbb{E}^{\mathbb{Q}}[S_1] = (1+r)S_0$, that is, $\mathbb{Q}(H)\,uS_0 + (1-\mathbb{Q}(H))dS_0 = (1+r)S_0$, whose only solution is $\mathbb{Q}(H) = q_u$; and (i) holds if and only if $q_u \in (0,1)$, that is, $d < 1+r < u$. So under the no-arbitrage condition there is exactly one risk-neutral measure, $\mathbb{Q}(H) = q_u$, $\mathbb{Q}(T) = q_d$, and Theorem 4.4 reads

$$V_0 = \mathbb{E}^{\mathbb{Q}}\Big[\frac{V_1}{1+r}\Big], \qquad \Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}. \tag{4.6}$$

The price is the expected discounted payoff, computed with $\mathbb{Q}$ and not with $\mathbb{P}$, and the hedge ratio is a discrete derivative of the claim with respect to the stock. Under $\mathbb{Q}$ the stock earns on average the risk-free rate, the return that an investor indifferent to risk would accept; hence the name, which describes the measure, not the investors.

Remark 4.6 (Why p does not enter the price). The real-world probability p appears nowhere in (4.5): replication must work in every state, so only which states are possible matters. This does not make p irrelevant: with investors' attitudes to risk, it shapes the stock price S_0, and claims are priced *relative to* S_0. Traders who agree on S_0, u, d and r but not on p must agree on every price, or one offers the other an arbitrage. What p changes is expected returns: under $\mathbb{P}$ the stock's mean return $pu + (1-p)d - 1$ may exceed r (a risk premium), and a claim's in general differs from both (Exercise 4.1).

4.3 N periods

In N periods the one-period argument is applied at every node, from the last date backwards. The successors of a node $\omega_1 \cdots \omega_n$ are $\omega_1 \cdots \omega_n H$ and $\omega_1 \cdots \omega_n T$.

Theorem 4.7 (Replication by backward induction)**.** *Assume $d < 1 + r < u$ and let V_N be a claim. Define V_n and Δ_n for $n = N-1, \dots, 0$, node by node, by*

$$V_n(\omega_1 \cdots \omega_n) = \frac{q_u V_{n+1}(\omega_1 \cdots \omega_n H) + q_d V_{n+1}(\omega_1 \cdots \omega_n T)}{1+r}, \tag{4.7}$$

$$\Delta_n(\omega_1 \cdots \omega_n) = \frac{V_{n+1}(\omega_1 \cdots \omega_n H) - V_{n+1}(\omega_1 \cdots \omega_n T)}{S_{n+1}(\omega_1 \cdots \omega_n H) - S_{n+1}(\omega_1 \cdots \omega_n T)}. \tag{4.8}$$

(i) The portfolio with initial capital $X_0 = V_0$ and hedge (Δ_n) satisfies $X_n = V_n$ for every n; in particular $X_N = V_N$ in every state. It is the only portfolio that replicates V_N.

(ii) Let $\mathbb{Q}$ be the probability under which the tosses are independent with $\mathbb{Q}(\omega_n = H) = q_u$, that is, $\mathbb{Q}(\omega) = q_u^{\#H(\omega)} q_d^{\#T(\omega)}$. Then

$$V_n = \mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-(N-n)} V_N \,\big|\, \mathcal{F}_n\big], \quad 0 \le n \le N; \qquad V_0 = \mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-N} V_N\big]. \tag{4.9}$$

Proof. (i) The formulas define V_n and Δ_n as functions of $\omega_1 \cdots \omega_n$, so (Δ_n) is adapted. We show $X_n = V_n$ by induction on n; it holds for $n = 0$ by construction. Suppose $X_n = V_n$ at a node $\omega_1 \cdots \omega_n$. Seen from this node, the next period is a one-period model with initial price $S_n(\omega_1 \cdots \omega_n)$, the same factors u and d, and a claim paying $V_{n+1}(\omega_1 \cdots \omega_n H)$ or $V_{n+1}(\omega_1 \cdots \omega_n T)$. Formulas (4.7) and (4.8) are (4.5) for this model, so Theorem 4.4 gives $X_{n+1} = V_{n+1}$ at both successors.

Conversely, let (X'_0, Δ') replicate V_N. At each node at time $N-1$, the pair $(X'_{N-1}, \Delta'_{N-1})$ replicates V_N over the last period, so the uniqueness part of Theorem 4.4 forces $X'_{N-1} = V_{N-1}$ and $\Delta'_{N-1} = \Delta_{N-1}$. Repeating the argument at times $N-2, \ldots, 0$ gives $X'_0 = V_0$ and $\Delta' = \Delta$.

(ii) Let $Y = f(\omega_1 \cdots \omega_{n+1})$ be $\mathcal{F}_{n+1}$-measurable, and let A be the event that the first n tosses are $\omega_1 \cdots \omega_n$. Then $\mathbb{Q}(A) > 0$ and $\mathbb{Q}(A \cap \{\omega_{n+1} = H\}) = \mathbb{Q}(A)\, q_u$, so the formula for conditional expectations given a finite partition (Theorem 1.12) gives, on A,

$$\mathbb{E}^{\mathbb{Q}}[Y \mid \mathcal{F}_n] = \frac{\mathbb{E}^{\mathbb{Q}}[Y \mathbf{1}_A]}{\mathbb{Q}(A)} = q_u\, f(\omega_1 \cdots \omega_n H) + q_d\, f(\omega_1 \cdots \omega_n T). \tag{4.10}$$

(This is the independence lemma, Theorem 1.19, since ω_{n+1} is independent of $\mathcal{F}_n$ under $\mathbb{Q}$.) Thus (4.7) says $(1+r)^{-n} V_n = \mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-(n+1)} V_{n+1} \mid \mathcal{F}_n\big]$ for every n, and the tower property (Theorem 1.16) gives $(1+r)^{-n} V_n = \mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-N} V_N \mid \mathcal{F}_n\big]$. □

As in Theorem 4.4, a price $c \neq V_0$ is exploited by trading the claim against its replicating portfolio, so V_0 is the only candidate price, and it creates no arbitrage (Theorem 4.10); likewise V_n is the price at time n.

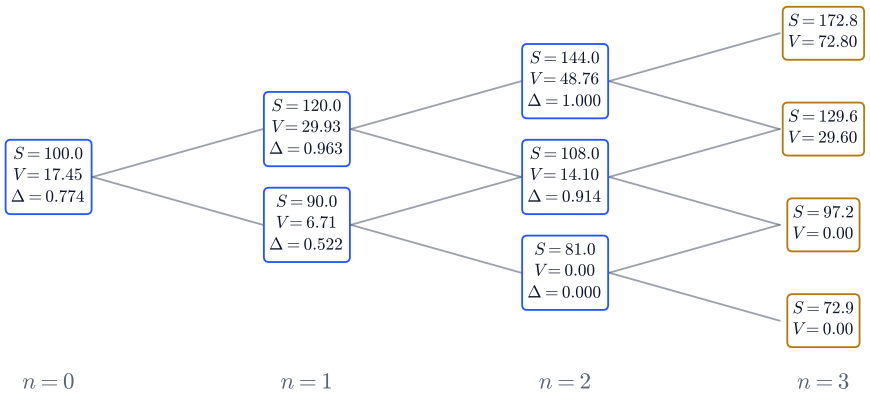

Figure 4.1: Three periods with $S_0 = 100$, $u = 1.2$, $d = 0.9$ and $r = 5\%$ (so $q_u = \frac{1}{2}$), and a call with strike 100. At each node: the stock price S_n, the price V_n of the call from (4.7), and the number of shares Δ_n held in the replicating portfolio over the next period.

Recombining trees. If $V_N = g(S_N)$, then by induction $V_n = v_n(S_n)$, with $v_N = g$ and $v_n(s) = (q_u v_{n+1}(us) + q_d v_{n+1}(ds))/(1+r)$, and the computation runs over the $n+1$ values of S_n rather than the 2^n nodes (Figure 4.1). Under $\mathbb{Q}$ the number of heads H_N is binomial with

parameters N and q_u, so (4.9) is the *Cox–Ross–Rubinstein formula*

$$V_0 = (1+r)^{-N} \sum_{k=0}^{N} \binom{N}{k} q_u^k \, q_d^{N-k} \, g(S_0 u^k d^{N-k}). \tag{4.11}$$

Its limit as $N \to \infty$ is the subject of Chapter 6. A path-dependent claim needs the whole tree (Exercise 4.2).

▶ Interactive demo: build the tree for any u, d, r and N, read off prices and hedge ratios at every node, check the replication along all 2^N paths, and see that p moves expected payoffs but not prices.

Proposition 4.8 (Discounted prices are martingales)**.** *Assume $d < 1+r < u$ and let $\mathbb{Q}$ be the measure of Theorem 4.7.*

(i) The discounted stock price $(1+r)^{-n} S_n$ is a $\mathbb{Q}$-martingale, so $\mathbb{Q}$ is a risk-neutral measure.

(ii) Let $\mathbb{Q}'$ be any probability measure under which the discounted stock price is a martingale. For every portfolio, the discounted value $(1+r)^{-n} X_n$ is a $\mathbb{Q}'$-martingale; in particular $\mathbb{E}^{\mathbb{Q}'}\left[(1+r)^{-N} X_N\right] = X_0$.

(iii) For every claim, the discounted price $(1+r)^{-n} V_n$ is a $\mathbb{Q}$-martingale.

Proof. (i) S_{n+1} equals uS_n or dS_n according to ω_{n+1}, so by (4.10)

$$\mathbb{E}^{\mathbb{Q}}[S_{n+1} \mid \mathcal{F}_n] = (q_u u + q_d d)\, S_n = \frac{(1+r-d)u + (u-1-r)d}{u-d}\, S_n = (1+r)\, S_n.$$

Dividing by $(1+r)^{n+1}$ gives the martingale property, and $\mathbb{Q}(\omega) > 0$ for every ω because $q_u, q_d > 0$.

(ii) On the finite space Ω every random variable is bounded. Write $D_n = (1+r)^{-n} S_n$. By (4.2), taking out what is known (Δ_n is $\mathcal{F}_n$-measurable),

$$\mathbb{E}^{\mathbb{Q}'}\left[(1+r)^{-(n+1)} X_{n+1} - (1+r)^{-n} X_n \,\middle|\, \mathcal{F}_n\right] = \Delta_n \, \mathbb{E}^{\mathbb{Q}'}[D_{n+1} - D_n \mid \mathcal{F}_n] = 0.$$

This is Theorem 2.8: Δ_n is chosen at time n and multiplies the increment from n to $n+1$. Taking expectations gives the last claim.

(iii) This is (4.9) with the tower property, or (ii) applied to the replicating portfolio. □

4.4 The fundamental theorem of asset pricing

Theorem 4.9 (Fundamental theorem of asset pricing, binomial model)**.** *In the N-period binomial model the following are equivalent:*

(a) there is no arbitrage;

(b) there is a risk-neutral measure;

(c) $d < 1+r < u$.

If they hold, the risk-neutral measure is unique, it is the measure $\mathbb{Q}$ of Theorem 4.7, and every claim can be replicated: the model is complete.

Proof. (b)⇒(a). Let $\mathbb{Q}'$ be risk-neutral and X a portfolio with $X_0 = 0$ and $X_N \geq 0$. By Theorem 4.8(ii), $\mathbb{E}^{\mathbb{Q}'}\left[(1+r)^{-N}X_N\right] = 0$. A nonnegative random variable with mean zero vanishes at every outcome of positive probability, and $\mathbb{Q}'(\omega) > 0$ for every ω; so $X_N = 0$ everywhere and X is not an arbitrage. (This is the argument of Theorem 3.6, with the martingale property of discounted wealth in place of the linearity of static prices.)

(a)⇒(c). Suppose $d \geq 1 + r$. Borrow S_0, buy one share at time 0, sell it at time 1 and keep the balance in the money market until time N: $X_0 = 0$, $\Delta_0 = 1$ and $\Delta_n = 0$ for $n \geq 1$. Then $X_1 = S_1 - (1 + r)S_0$ is nonnegative, and positive if $\omega_1 = H$, as in the proof of Theorem 4.3; and $X_N = (1 + r)^{N-1}X_1$. Since $\mathbb{P}(\omega_1 = H) = p > 0$, this is an arbitrage. If $u \leq 1 + r$, take $\Delta_0 = -1$ instead.

(c)⇒(b). This is Theorem 4.8(i).

Uniqueness and completeness. Under (c) every claim can be replicated, by Theorem 4.7. Let $\mathbb{Q}'$ be a risk-neutral measure and fix an outcome ϖ. The claim $\mathbf{1}_{\{\varpi\}}$, which pays one dollar if $\omega = \varpi$ and nothing otherwise, is replicated by a portfolio X with $X_0 = \mathbb{E}^{\mathbb{Q}}\left[(1+r)^{-N}\mathbf{1}_{\{\varpi\}}\right] = (1+r)^{-N}\mathbb{Q}(\varpi)$, by (4.9). By Theorem 4.8(ii) applied to $\mathbb{Q}'$, also $X_0 = \mathbb{E}^{\mathbb{Q}'}\left[(1 + r)^{-N}X_N\right] = (1 + r)^{-N}\mathbb{Q}'(\varpi)$. Hence $\mathbb{Q}'(\varpi) = \mathbb{Q}(\varpi)$ for every ϖ. □

The claims $\mathbf{1}_{\{\varpi\}}$ are the *Arrow–Debreu securities* of the model, and their prices $(1 + r)^{-N}\mathbb{Q}(\varpi)$ are the state prices of Chapter 3 (Exercise 4.3).

Remark 4.10 (Trading the claim creates no arbitrage). If the claim is traded at V_n, its discounted price is a $\mathbb{Q}$-martingale (Theorem 4.8(iii)), and the proof of (b)⇒(a) applies verbatim to portfolios in the three assets: V_n is a consistent price.

Remark 4.11 (The general theorem). The equivalence (a)⇔(b) holds in every market with finitely many trading dates and finitely many assets: Harrison and Pliska (1981) proved the multi-period version on a finite probability space, and Dalang et al. (1990) extended it to general probability spaces; see Föllmer and Schied (2016), Section 1.2 for one period and Section 5.2 for several. On a finite Ω, (b)⇒(a) has the proof above; in general the discounted wealth of an unbounded strategy need not be an integrable martingale, and one first shows that it is when its terminal value is nonnegative (Föllmer and Schied (2016), Chapter 5). The converse needs a separating hyperplane argument, which the binomial model replaces by the explicit formula (4.4). Completeness is also general: an arbitrage-free market of this kind is complete if and only if its risk-neutral measure is unique (Föllmer and Schied (2016), Section 5.4); the Arrow–Debreu argument above is one half of that statement, and Section 4.6 shows the other half at work.

4.5 American options

An American option can be exercised at any date up to its maturity N; exercise at time n pays G_n, where (G_n) is adapted: $G_n = (K - S_n)^+$ for the American put and $G_n = (S_n - K)^+$ for the American call. The holder decides when to exercise with the information available, so an exercise strategy is a stopping time (Theorem 2.11): a random time τ with values in $\{0, \ldots, N\}$ such that $\{\tau = n\} \in \mathcal{F}_n$ for every n. We write $\mathcal{T}_n$ for the set of stopping times with $n \leq \tau \leq N$.

The price is computed backwards, comparing at every node exercise with continuation:

$$U_N = G_N, \qquad U_n = \max\Big(G_n,\ \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[U_{n+1} \mid \mathcal{F}_n]\Big), \quad n = N-1, \dots, 0. \tag{4.12}$$

The second term is the one-period price of a claim paying U_{n+1} one period later: the value of keeping the option alive. (4.12) is a dynamic programming equation, and $(1+r)^{-n}U_n$ is the *Snell envelope* of $(1+r)^{-n}G_n$.

The proof rests on optional stopping, which we need conditionally on $\mathcal{F}_n$; for $n = 0$ the next lemma is Theorem 2.14(a).

Lemma 4.12 (Optional stopping on a finite horizon)**.** *On any probability space with a filtration $(\mathcal{F}_k)_{0\le k\le N}$, let $(Y_k)_{0\le k\le N}$ be a supermartingale and τ a stopping time with $n \le \tau \le N$. Then $\mathbb{E}[Y_\tau \mid \mathcal{F}_n] \le Y_n$, with equality if Y is a martingale.*

Proof. $Y_\tau = \sum_k \mathbf{1}_{\{\tau=k\}} Y_k$ is integrable. Since $n \le \tau \le N$, the following sum telescopes:

$$Y_\tau - Y_n = \sum_{k=n+1}^{N} \mathbf{1}_{\{\tau \ge k\}}\,(Y_k - Y_{k-1}).$$

The event $\{\tau \ge k\} = \Omega \setminus \{\tau \le k-1\}$ belongs to $\mathcal{F}_{k-1}$. For $k \ge n+1$, the tower property, taking out what is known and positivity give

$$\mathbb{E}\big[\mathbf{1}_{\{\tau\ge k\}}(Y_k - Y_{k-1}) \,\big|\, \mathcal{F}_n\big] = \mathbb{E}\big[\mathbf{1}_{\{\tau\ge k\}}\,\mathbb{E}[Y_k - Y_{k-1} \mid \mathcal{F}_{k-1}] \,\big|\, \mathcal{F}_n\big] \le 0,$$

with equality for a martingale. Summing over k proves the lemma. □

Theorem 4.13 (American options)**.** *Assume $d < 1+r < u$, let G be an adapted payoff process, define U by (4.12), and let $\tau_n^* = \min\{k \ge n : U_k = G_k\}$.*

(i) $U_n \ge G_n$ for every n, and $(1+r)^{-n}U_n$ is a $\mathbb{Q}$-supermartingale.

(ii) $\tau_n^* \in \mathcal{T}_n$, and for every $\tau \in \mathcal{T}_n$,

$$(1+r)^n\,\mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-\tau}G_\tau \,\big|\, \mathcal{F}_n\big] \le U_n = (1+r)^n\,\mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-\tau_n^*}G_{\tau_n^*} \,\big|\, \mathcal{F}_n\big].$$

In particular $U_0 = \max_{\tau\in\mathcal{T}_0}\mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-\tau}G_\tau\big]$, and the maximum is attained at $\tau^* := \tau_0^*$.

(iii) There is a portfolio with initial capital U_0 whose value satisfies $X_n \ge U_n \ge G_n$ for every n; and every portfolio with $X_n \ge G_n$ for all n has $X_0 \ge U_0$.

(iv) (American call) Let $r \ge 0$, $K \ge 0$ and $G_n = (S_n - K)^+$, and let V_n be the price (4.9) of the European call with payoff $(S_N - K)^+$. Then $U_n = V_n$ for every n, and $\tau = N$ is an optimal exercise time: nothing is gained by exercising early. If $r > 0$ and $K > 0$, exercise at a time $n < N$ in a state where $S_n > K$ is strictly suboptimal: $U_n > G_n$ there.

Proof. Write $Y_k = (1+r)^{-k}U_k$ and $c_k = \mathbb{E}^{\mathbb{Q}}[U_{k+1} \mid \mathcal{F}_k]/(1+r)$ for the continuation value.

(i) Both terms of the maximum in (4.12) are $\mathcal{F}_n$-measurable, so U is adapted. By definition $U_n \ge G_n$, and $U_n \ge c_n$ is, after division by $(1+r)^n$, the supermartingale inequality $Y_n \ge \mathbb{E}^{\mathbb{Q}}[Y_{n+1} \mid \mathcal{F}_n]$.

(ii) Since $U_N = G_N$, we have $n \le \tau_n^* \le N$, and for $n \le k \le N$, $\{\tau_n^* = k\} = \{U_n > G_n, \dots, U_{k-1} > G_{k-1}, U_k = G_k\} \in \mathcal{F}_k$; so $\tau_n^* \in \mathcal{T}_n$. Let $\tau \in \mathcal{T}_n$. By Theorem 4.12 for the supermartingale Y, and by (i),

$$Y_n \ge \mathbb{E}^{\mathbb{Q}}[Y_\tau \mid \mathcal{F}_n] \ge \mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-\tau} G_\tau \,\big|\, \mathcal{F}_n\big],$$

which is the inequality. For τ_n^*, consider the stopped process $M_k = Y_{k \wedge \tau_n^*}$, $n \le k \le N$. Its increments are $M_{k+1} - M_k = \mathbf{1}_{\{\tau_n^* > k\}}(Y_{k+1} - Y_k)$, and $\{\tau_n^* > k\} \in \mathcal{F}_k$. On this event $U_k > G_k$, so $U_k = c_k$, that is, $\mathbb{E}^{\mathbb{Q}}[Y_{k+1} \mid \mathcal{F}_k] = Y_k$. Taking out what is known,

$$\mathbb{E}^{\mathbb{Q}}[M_{k+1} - M_k \mid \mathcal{F}_k] = \mathbf{1}_{\{\tau_n^* > k\}}\big(\mathbb{E}^{\mathbb{Q}}[Y_{k+1} \mid \mathcal{F}_k] - Y_k\big) = 0,$$

so M is a $\mathbb{Q}$-martingale on $\{n, \dots, N\}$ and, by the tower property, $Y_n = M_n = \mathbb{E}^{\mathbb{Q}}[M_N \mid \mathcal{F}_n] = \mathbb{E}^{\mathbb{Q}}[Y_{\tau_n^*} \mid \mathcal{F}_n]$. Finally $Y_{\tau_n^*} = (1+r)^{-\tau_n^*} G_{\tau_n^*}$, because $U = G$ at time τ_n^*.

(iii) Define Δ_n by (4.8) with U in place of V, and start from $X_0 = U_0$. At each node, Theorem 4.4 says that the capital c_n with Δ_n shares replicates U_{n+1} one period later: $(1+r)c_n + \Delta_n(S_{n+1} - (1+r)S_n) = U_{n+1}$. Subtracting this from (4.1) written as $X_{n+1} = (1+r)X_n + \Delta_n(S_{n+1} - (1+r)S_n)$,

$$X_{n+1} = U_{n+1} + (1+r)(X_n - c_n).$$

If $X_n \ge U_n$, then $X_n \ge c_n$ and so $X_{n+1} \ge U_{n+1}$; by induction $X_n \ge U_n$ for every n. Conversely, let X be a portfolio with $X_n \ge G_n$ for all n. Its discounted value is a $\mathbb{Q}$-martingale (Theorem 4.8(ii)), so Theorem 4.12 with $\tau = \tau^*$ gives $X_0 = \mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-\tau^*} X_{\tau^*}\big] \ge \mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-\tau^*} G_{\tau^*}\big] = U_0$.

(iv) By the conditional Jensen inequality for the convex function $x \mapsto (x-K)^+$ (Theorem 1.16) and the martingale property $\mathbb{E}^{\mathbb{Q}}[S_N \mid \mathcal{F}_n] = (1+r)^{N-n} S_n$,

$$\begin{aligned} V_n = \frac{\mathbb{E}^{\mathbb{Q}}[(S_N - K)^+ \mid \mathcal{F}_n]}{(1+r)^{N-n}} &\ge \frac{\big((1+r)^{N-n} S_n - K\big)^+}{(1+r)^{N-n}} \\ &= \big(S_n - (1+r)^{-(N-n)} K\big)^+ \ge (S_n - K)^+ = G_n, \end{aligned} \tag{4.13}$$

where the last step uses $(1+r)^{-(N-n)} K \le K$, valid because $r \ge 0$ and $K \ge 0$. Now $U_N = G_N = V_N$, and if $U_{n+1} = V_{n+1}$, then (4.12) and (4.7) give $U_n = \max(G_n, V_n) = V_n$. By backward induction $U_n = V_n$ for all n, and $\tau = N$ attains $U_0 = V_0 = \mathbb{E}^{\mathbb{Q}}\big[(1+r)^{-N} G_N\big]$. If $r > 0$, $K > 0$, $n < N$ and $S_n > K$, (4.13) gives $U_n = V_n \ge S_n - (1+r)^{-(N-n)} K > S_n - K = G_n$. □

The arbitrage-free price. Suppose the option can be bought or sold at time 0 for c, and trading in the stock and the money market is allowed. The buyer chooses an exercise time $\tau \in \mathcal{T}_0$ and keeps G_τ in the money market; the seller pays G_τ at whatever τ the buyer chooses. Following Föllmer and Schied (2016), Chapter 6, call c *arbitrage-free* if (a) for no τ can the buyer, borrowing c, exercising at τ and trading, end at time N with a nonnegative and not identically zero amount; and (b) for some τ the seller, receiving c, paying G_τ and trading, cannot. Then U_0 is the only arbitrage-free price. If $c > U_0$, the seller invests U_0 in the portfolio of (iii) and $c - U_0$ in the money market; whatever τ, $X_\tau \ge G_\tau$ covers the payment, and (b) fails. If $c < U_0$, the buyer exercises at τ^* and sells short the portfolio replicating $(1+r)^{N-\tau^*} G_{\tau^*}$ paid at N, worth U_0 by (4.9) and (ii), and keeps $U_0 - c > 0$; (a) fails. If $c = U_0$, a buyer using τ and a portfolio started from 0 ends with an amount whose discounted $\mathbb{Q}$-mean is $\mathbb{E}^{\mathbb{Q}}[(1+r)^{-\tau} G_\tau] - U_0 \le 0$ by (ii) and Theorem 4.8(ii); a seller facing τ^* ends with discounted $\mathbb{Q}$-mean $U_0 - \mathbb{E}^{\mathbb{Q}}[(1+r)^{-\tau^*} G_{\tau^*}] = 0$.

As $\mathbb{Q}(\omega) > 0$ for every ω, a nonnegative amount with nonpositive mean is zero: neither has an arbitrage. The time τ^* is the smallest optimal exercise time; there may be others (Exercise 4.6).

In (iv), the bound (4.13) is the lower bound of Theorem 3.10 at time n, which holds in every arbitrage-free market, as Merton (1973) observed: a call on a stock that pays no dividend is worth more alive than exercised. (The time τ^* may still stop at nodes where the call is worthless, $U_n = G_n = 0$, which changes nothing.) For the put the argument fails: the corresponding bound $V_n \geq \left((1+r)^{-(N-n)}K - S_n\right)^+$ lies below the exercise value $K - S_n$ when $r > 0$, and deep in the money early exercise is strictly better (Figure 4.2). With $r = 0$ the American and European puts coincide (Exercise 4.4).

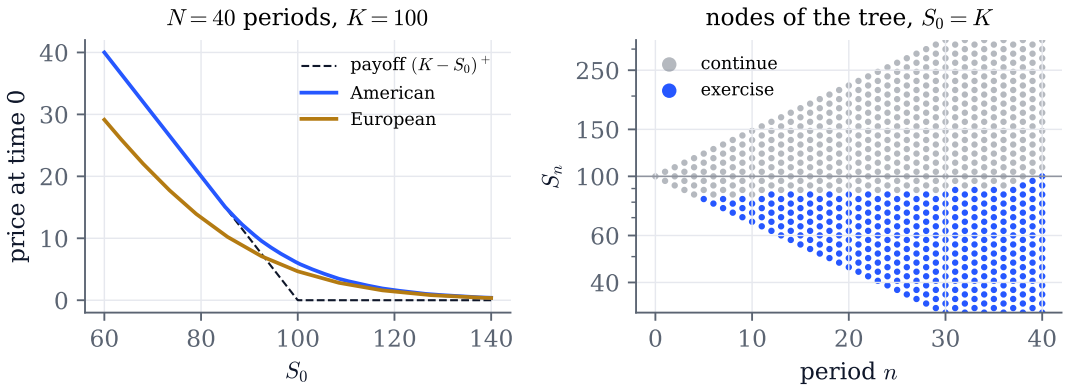

Figure 4.2: American and European puts with $K = 100$ on a tree with $N = 40$, $u = 1.04$, $d = 1/u$ and $r = 0.3\%$ per period. Left: prices at time 0 as functions of S_0; the American price meets the payoff where immediate exercise is optimal. Right: the nodes of the tree started at $S_0 = K$, marked where exercise is optimal ($G_n > 0$ and G_n at least the continuation value).

▶ Interactive demo: compare American and European puts and watch the early-exercise region as r, u, N and the strike move.

4.6 The trinomial model: a first incomplete market

In the one-period *trinomial model*, S_1 takes the values uS_0, mS_0, dS_0, with $0 < d < m < u$ and positive $\mathbb{P}$-probability for each, and the money market pays $R := 1 + r$ per dollar. A risk-neutral measure (q_u, q_m, q_d) must satisfy two linear equations in three unknowns. A portfolio is worth $RX_0 + \Delta_0(S_1 - RS_0)$, affine in S_1, so it replicates only claims whose three values lie on a line; a call with strike $dS_0 < K < uS_0$ is not one of them. We call c an *arbitrage-free price* for a claim $h(S_1)$ if the market in which the money market, the stock and the claim (at price c at time 0) are traded admits no arbitrage.

Proposition 4.14 (The one-period trinomial model)**.** *Let $0 < d < m < u$ and $R = 1 + r$.*

(i) The model has no arbitrage if and only if $d < R < u$.

(ii) In that case the risk-neutral measures are the measures $\mathbb{Q}_\theta$, $0 < \theta < \theta_{\max}$, given by

$$q_m = \theta, \quad q_u = \frac{R - d - \theta(m-d)}{u-d}, \quad q_d = \frac{u - R - \theta(u-m)}{u-d}, \quad \theta_{\max} = \frac{R-d}{m-d} \wedge \frac{u-R}{u-m}.$$

(iii) Assume $d < R < u$, let $\theta_{\max}$ be as in (ii), let $dS_0 < K < uS_0$ and $h(x) = (x-K)^+$, and write $h_x = h(xS_0)$ for $x \in \{d, m, u\}$. A price c for the call is arbitrage-free if and only if $C_{\text{low}} < c < C_{\text{high}}$, where

$$C_{\text{high}} = \frac{1}{R}\Big(\frac{R-d}{u-d}\,h_u + \frac{u-R}{u-d}\,h_d\Big), \quad C_{\text{low}} = \begin{cases} \dfrac{1}{R}\Big(\dfrac{R-m}{u-m}\,h_u + \dfrac{u-R}{u-m}\,h_m\Big) & \text{if } m \le R, \\ \dfrac{1}{R}\Big(\dfrac{R-d}{m-d}\,h_m + \dfrac{m-R}{m-d}\,h_d\Big) & \text{if } m \ge R. \end{cases}$$

The interval $(C_{\text{low}}, C_{\text{high}})$ is the set $\{\mathbb{E}^{\mathbb{Q}_\theta}[h(S_1)]/R : 0 < \theta < \theta_{\max}\}$.

Proof. (i) If $d < R < u$, then $\theta_{\max} > 0$ and the measures of (ii) exist. Each of them is a pricing measure (Theorem 3.5) for the money market and the stock, with $B = 1/R$, because $\mathbb{E}^{\mathbb{Q}}[S_1] = RS_0$; so there is no arbitrage, by Theorem 3.6. If $d \ge R$, buying a share with borrowed money gives $X_1 = S_1 - RS_0 \ge 0$, positive in the states m and u; if $u \le R$, selling one short gives $X_1 = RS_0 - S_1 \ge 0$, positive in the states d and m.

(ii) The conditions are $q_u + q_m + q_d = 1$ and $uq_u + mq_m + dq_d = R$. For $q_m = \theta$ fixed, this is a linear system for (q_u, q_d) with nonzero determinant $d - u$, and substitution checks that the displayed values solve it. They are positive exactly when $0 < \theta < \theta_{\max}$.

(iii) The price under $\mathbb{Q}_\theta$, $\pi(\theta) = (q_u h_u + \theta h_m + q_d h_d)/R$, is affine in θ:

$$R\,\pi(\theta) = R\,\pi(0) - \theta\,(\ell_m - h_m), \qquad \ell_m = \frac{(m-d)\,h_u + (u-m)\,h_d}{u-d},$$

where ℓ_m is the value at mS_0 of the line ℓ through (dS_0, h_d) and (uS_0, h_u). The payoff is convex, so $h_m \le \ell_m$, and here strictly: $h_d = 0 < h_u$, so if $mS_0 \le K$ then $h_m = 0 < \ell_m$, and if $mS_0 > K$ a short computation gives $\ell_m - h_m = (u-m)(K - dS_0)/(u-d) > 0$. Hence π is strictly decreasing and $\{\pi(\theta) : 0 < \theta < \theta_{\max}\} = (\pi(\theta_{\max}), \pi(0))$. The measure $\mathbb{Q}_0$ has $q_m = 0$ and is the binomial measure on $\{d, u\}$, so $\pi(0) = C_{\text{high}}$. One checks that $(u-R)/(u-m) \le (R-d)/(m-d)$ if and only if $m \le R$. If $m \le R$, then $\theta_{\max} = (u-R)/(u-m)$, where $q_d = 0$ and $q_u = (R-m)/(u-m)$, and $\pi(\theta_{\max})$ is the first formula for C_{low}; if $m \ge R$, then $\theta_{\max} = (R-d)/(m-d)$, where $q_u = 0$, and we get the second.

Arbitrage-free prices. If $C_{\text{low}} < c < C_{\text{high}}$, then $c = \pi(\theta)$ for some $\theta \in (0, \theta_{\max})$: the call is priced at $\mathbb{E}^{\mathbb{Q}_\theta}[h(S_1)]/R$, and Theorem 3.6 shows that adding it to the market creates no arbitrage. For the converse, note that an affine payoff $\alpha + \beta S_1$ costs $\alpha/R + \beta S_0$, which is $\mathbb{E}^{\mathbb{Q}^*}[\alpha + \beta S_1]/R$ for every probability $\mathbb{Q}^*$ on the three states with $\mathbb{E}^{\mathbb{Q}^*}[S_1] = RS_0$, including $\mathbb{Q}_0$ and $\mathbb{Q}_{\theta_{\max}}$.

Suppose $c \ge C_{\text{high}}$. The payoff $\ell(S_1)$ costs $\mathbb{E}^{\mathbb{Q}_0}[\ell(S_1)]/R = \mathbb{E}^{\mathbb{Q}_0}[h(S_1)]/R = C_{\text{high}}$, because $\ell = h$ at the two states that carry $\mathbb{Q}_0$. Sell the call at c, buy $\ell(S_1)$ and invest $c - C_{\text{high}}$: the position costs nothing, and at time 1 it is worth $\ell(S_1) - h(S_1) + R(c - C_{\text{high}})$, which is nonnegative in every state and at least $\ell_m - h_m > 0$ in the state m. This is an arbitrage.

Suppose $c \le C_{\text{low}}$ and $m \le R$. Let ℓ' be the affine function through (mS_0, h_m) and (uS_0, h_u); as above, $\ell'(S_1)$ costs $\mathbb{E}^{\mathbb{Q}_{\theta_{\max}}}[h(S_1)]/R = C_{\text{low}}$. Write $mS_0 = \lambda dS_0 + (1-\lambda)uS_0$ with $\lambda \in (0,1)$;

convexity of h and affinity of ℓ' give $\lambda\ell'(dS_0)+(1-\lambda)h_u = h_m \le \lambda h_d+(1-\lambda)h_u$, so $\ell'(dS_0) \le h_d$, and the inequality is strict, since otherwise the three points (xS_0, h_x) would lie on ℓ' and $\ell_m = h_m$. Buy the call at c, sell $\ell'(S_1)$ and invest $C_{\text{low}} - c$: at time 1 the position is worth $h(S_1) - \ell'(S_1) + R(C_{\text{low}} - c) \ge 0$, and positive in the state d. The case $m \ge R$ is symmetric, with the line through (dS_0, h_d) and (mS_0, h_m) and the state u. □

No arbitrage now pins down the price only up to an interval. C_{high} is the *superreplication price*: $\ell(S_1)$ dominates the call, and every dominating portfolio costs $\mathbb{E}^{\mathbb{Q}_\theta}[X_1]/R \ge \pi(\theta)$ for all θ; C_{low} is the subreplication price. Choosing a price inside needs more information, such as the price of another option; Exercise 4.7 shows this, and that a claim is replicable exactly when its price is the same under every risk-neutral measure (cf. Theorem 4.11).

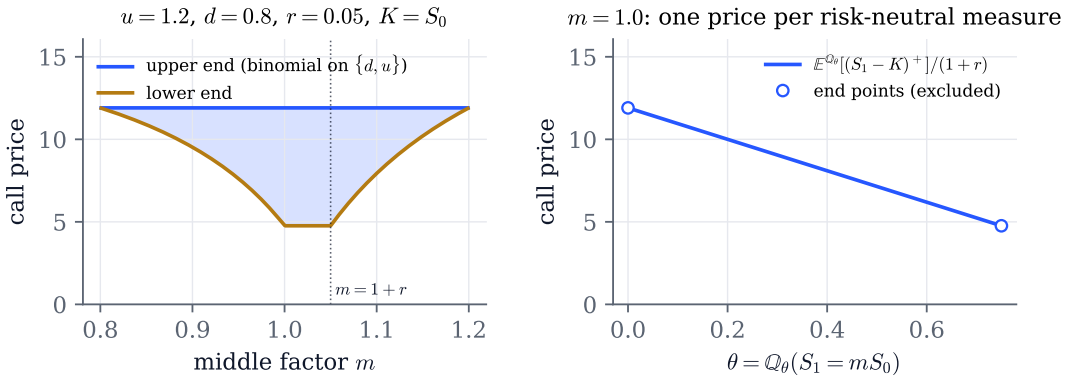

Figure 4.3: Arbitrage-free prices of a call with $K = S_0$ in the one-period trinomial model with $u = 1.2$, $d = 0.8$, $r = 5\%$. Left: the interval of Theorem 4.14 as m moves from d to u. For $K/S_0 \le m \le 1 + r$ the lower end is the static bound $S_0 - K/(1+r)$ of Theorem 3.10: the payoff is affine on the two states that carry $\mathbb{Q}_{\theta_{\max}}$. Right: for $m = 1$, the price is affine in $\theta = q_m$; the excluded end points give zero weight to one state.

▶ Interactive demo: move the middle factor m, the strike and the rate, and watch the interval of arbitrage-free call prices.

Remark 4.15 (Trinomial trees as numerical schemes). Trinomial trees used for computation (Chapter 16) choose one measure by matching the mean $\nu\,\delta t$ and second moment $s^2 := \sigma^2\delta t + \nu^2\delta t^2$ of a log-price step $\pm\delta x$ or 0 (the Black–Scholes values of Chapter 6): $p_{u,d} = \frac{1}{2}\left(s^2/\delta x^2 \pm \nu\,\delta t/\delta x\right)$, $p_m = 1 - s^2/\delta x^2$. $\delta x = \sigma\sqrt{\delta t}$ gives $p_m = -\nu^2\delta t/\sigma^2 < 0$ if $\nu \ne 0$; $\delta x = \sigma\sqrt{3\,\delta t}$ gives $p_m = \frac{2}{3} - \nu^2\delta t/(3\sigma^2) > 0$ for $\delta t < 2\sigma^2/\nu^2$ and $p_{u,d} = (1 + a^2 \pm \sqrt{3}\,a)/6 > 0$, $a = \nu\sqrt{\delta t}/\sigma$. Such a tree is only approximately risk-neutral: $\mathbb{E}[e^{\delta X}] \ne e^{r\delta t}$, with a per-step error of order δt^2 in general and δt^3 for $\delta x = \sigma\sqrt{3\,\delta t}$.

Further reading

Shreve (2004a), Chapters 1–4, covers this material at the same level; the model is due to Cox et al. (1979). Föllmer and Schied (2016) treat discrete-time markets with full rigour: Chapters 1 and 5 for arbitrage, completeness and incomplete markets, Chapter 6 for American options. Hull (2021) explains how trees are used in practice.

Part II

Brownian motion and the Black–Scholes formula

Chapter 5

Brownian motion

Every continuous-time model in these notes is driven by Brownian motion. We build it the way it arises in finance: as the limit of a random walk whose steps become small and frequent. A price that receives many small, independent shocks per unit of time moves, in the limit, like a Brownian path.

Two features of the limit run through the rest of the course. Brownian motion is a martingale, which is what makes risk-neutral pricing work. And its paths are rough: over an interval of length t, the sum of squared increments does not vanish as the mesh goes to zero, as it would for a smooth function, but converges to t. This *quadratic variation* is why stochastic calculus differs from ordinary calculus (Chapter 8).

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathcal{N}(m, v)$ is the normal law with mean m and variance v.

5.1 The symmetric random walk

Let $X_1, X_2, \dots$ be independent with $\mathbb{P}(X_j = 1) = \mathbb{P}(X_j = -1) = \frac{1}{2}$: the outcomes of a fair coin, with heads counted as $+1$ and tails as -1. The *symmetric random walk* is

$$M_0 = 0, \qquad M_k = \sum_{j=1}^{k} X_j, \quad k \geq 1,$$

and its natural filtration is $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$, the information carried by the first k tosses.

Proposition 5.1. *Let $0 = k_0 < k_1 < \cdots < k_m$ be integers.*

(i) The increments $M_{k_1} - M_{k_0}$, $M_{k_2} - M_{k_1}, \dots, M_{k_m} - M_{k_{m-1}}$ are independent, and $M_{k_{i+1}} - M_{k_i}$ has mean 0 and variance $k_{i+1} - k_i$.

(ii) $(M_k)_{k \geq 0}$ is a martingale with respect to $(\mathcal{F}_k)_{k \geq 0}$.

(iii) For every k and every outcome ω, the quadratic variation $[M, M]_k := \sum_{j=1}^{k} (M_j - M_{j-1})^2$ *equals k.*

Proof. (i) The increment $M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j$ is a function of the block $(X_{k_i+1}, \dots, X_{k_{i+1}})$.

Different increments use disjoint blocks of independent variables, so they are independent. Each X_j has mean 0 and variance $\mathbb{E}[X_j^2] = 1$, and variances of independent variables add, so the increment has mean 0 and variance $k_{i+1} - k_i$.

(ii) M_k is a function of $X_1, \dots, X_k$, so it is $\mathcal{F}_k$-measurable, and $|M_k| \le k$ is integrable. For $\ell \ge k$, the increment $M_\ell - M_k$ depends only on $X_{k+1}, \dots, X_\ell$, which are independent of $\mathcal{F}_k$. Taking out what is known and using independence (Theorem 1.16),

$$\mathbb{E}[M_\ell \mid \mathcal{F}_k] = M_k + \mathbb{E}[M_\ell - M_k \mid \mathcal{F}_k] = M_k + \mathbb{E}[M_\ell - M_k] = M_k.$$

(iii) $(M_j - M_{j-1})^2 = X_j^2 = 1$ for every j and every ω. □

Remark 5.2. $\mathrm{Var}(M_k)$ and $[M, M]_k$ are both equal to k, but they are different objects. The variance is an average over all possible paths. The quadratic variation is computed along a single path, and here it happens not to depend on the path at all. For Brownian motion we will find the same number t through a limit, not by exact counting.

5.2 Scaling the walk

To pass to continuous time, speed the walk up and shrink its steps. Fix $n \ge 1$ and take n steps per unit of time, each of size $1/\sqrt{n}$.

Definition 5.3 (Scaled random walk)**.** For $n \ge 1$, the *scaled random walk* is $W^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt}$ for $t \ge 0$ with $nt \in \mathbb{N}$, extended to all $t \ge 0$ by linear interpolation between consecutive grid points k/n.

The step size $1/\sqrt{n}$ is the only choice for which the limit is neither trivial nor explosive: with n steps of size h per unit of time, the variance at time t is $nt\,h^2$, which stays finite and nonzero only if $h \propto 1/\sqrt{n}$.

Proposition 5.4. *Let $0 \le s < t$ with $ns, nt \in \mathbb{N}$. Then $W^{(n)}(t) - W^{(n)}(s)$ has mean 0 and variance $t - s$; increments over disjoint grid intervals are independent; $(W^{(n)}(k/n))_{k \ge 0}$ is a martingale; and along every path*

$$[W^{(n)}, W^{(n)}](t) := \sum_{j=1}^{nt} \left(W^{(n)}\left(\tfrac{j}{n}\right) - W^{(n)}\left(\tfrac{j-1}{n}\right) \right)^2 = t.$$

Proof. Every statement is the corresponding part of Theorem 5.1 after dividing by $\sqrt{n}$. In particular, $\mathrm{Var}(W^{(n)}(t) - W^{(n)}(s)) = \frac{1}{n}(nt - ns) = t - s$, and each of the nt squared increments equals $\frac{1}{n}$. □

At a fixed time, the scaled walk is a normalized sum of independent coin tosses, so the central limit theorem applies. We prove the version we need directly.

Theorem 5.5 (Central limit theorem for the scaled walk)**.** *For every $t > 0$, $W^{(n)}(t)$ converges in distribution to $\mathcal{N}(0, t)$ as $n \to \infty$.*

Proof. Let $k_n = \lfloor nt \rfloor$, so that $k_n/n \le t < (k_n+1)/n$ and $k_n/n \to t$. The characteristic function of $X_1/\sqrt{n}$ is $\mathbb{E}[e^{iuX_1/\sqrt{n}}] = \cos(u/\sqrt{n})$, so by independence

$$\varphi_n(u) := \mathbb{E}\Big[e^{iuW^{(n)}(k_n/n)}\Big] = \cos\left(u/\sqrt{n}\right)^{k_n}.$$

Since $\log\cos x = -\frac{x^2}{2} + O(x^4)$ as $x \to 0$, for fixed u and n large,

$$\log\varphi_n(u) = k_n \log\cos\left(u/\sqrt{n}\right) = -\frac{u^2}{2}\,\frac{k_n}{n} + O\Big(\frac{k_n}{n^2}\Big) \longrightarrow -\frac{u^2 t}{2}.$$

The limit $e^{-u^2t/2}$ is the characteristic function of $\mathcal{N}(0,t)$, so by Lévy's continuity theorem $W^{(n)}(k_n/n)$ converges in distribution to $\mathcal{N}(0,t)$. Finally, $W^{(n)}(t)$ interpolates between the grid values at k_n/n and $(k_n+1)/n$, which differ by one step, so $|W^{(n)}(t) - W^{(n)}(k_n/n)| \le 1/\sqrt{n}$ for every outcome. A perturbation that goes to 0 surely does not change a limit in distribution (Slutsky's lemma), so $W^{(n)}(t)$ converges to $\mathcal{N}(0,t)$ as well. □

The same argument, applied to the vector of increments over disjoint intervals with the multivariate version of Lévy's continuity theorem, shows that $(W^{(n)}(t_1), \dots, W^{(n)}(t_m))$ converges in distribution to a vector with independent Gaussian increments of variances $t_1, t_2 - t_1, \dots, t_m - t_{m-1}$. *Donsker's invariance principle* strengthens this to the convergence in distribution of the whole path $t \mapsto W^{(n)}(t)$, as a random continuous function, to Brownian motion; see Billingsley (1999), Theorem 8.2. We do not need that strength here. What matters is the shape of the limit, which the next definition records.

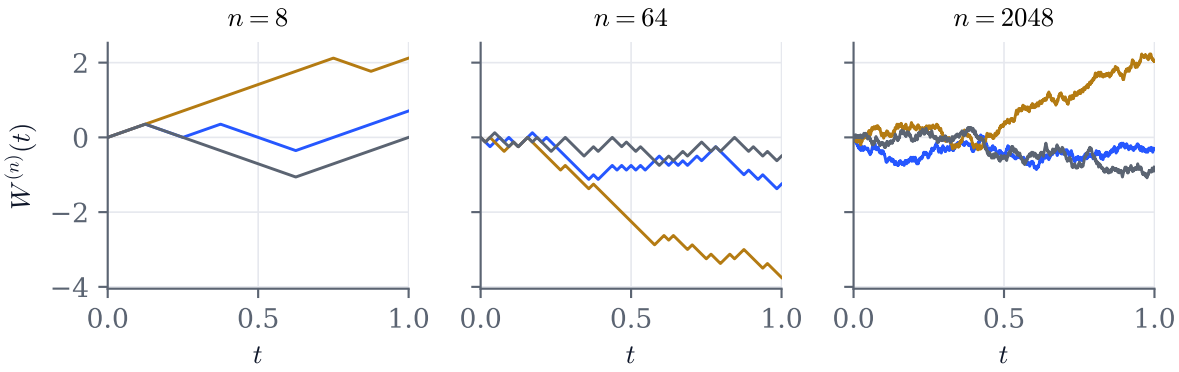

Figure 5.1: Three paths of the scaled random walk $W^{(n)}$ for $n = 8$, 64 and 2048 steps per unit of time. As n grows, the paths become indistinguishable from Brownian paths.

▶ Interactive demo: move the number of steps n and watch the scaled walk and the histogram of $W^{(n)}(1)$ approach Brownian motion and $\mathcal{N}(0,1)$.

5.3 Brownian motion

Definition 5.6 (Brownian motion)**.** A stochastic process $W = (W_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is a *(standard) Brownian motion* if

(i) $W_0 = 0$;

(ii) for all $0 = t_0 < t_1 < \cdots < t_m$, the increments $W_{t_1} - W_{t_0}, \ldots, W_{t_m} - W_{t_{m-1}}$ are independent;

(iii) for all $0 \leq s < t$, $W_t - W_s \sim \mathcal{N}(0, t-s)$;

(iv) for every ω, the path $t \mapsto W_t(\omega)$ is continuous.

That such a process exists is a theorem, due to Wiener (1923). The difficulty is property (iv): it is easy to produce Gaussian variables with the right covariances at countably many times, and it takes work to show that they can be chosen to fit together into continuous paths. A short construction, due to Lévy, is in Mörters and Peres (2010), Theorem 1.3; we take existence for granted.

The next result is the most convenient way to check that a process is a Brownian motion. Recall that a random vector is *Gaussian* if every linear combination of its entries is normally distributed (a constant counts as a normal variable with variance 0), and that the law of a Gaussian vector is determined by its means and covariances.

Proposition 5.7 (Gaussian characterization)**.** *Let W be a process with continuous paths and $W_0 = 0$. Then W is a Brownian motion if and only if, for all $0 \leq t_1 < \cdots < t_m$, the vector $(W_{t_1}, \ldots, W_{t_m})$ is Gaussian with mean zero and*

$$\operatorname{Cov}(W_s, W_t) = \min(s,t) \qquad \text{for all } s, t \geq 0. \tag{5.1}$$

Proof. ($\Rightarrow$) The increments over $0 < t_1 < \cdots < t_m$ are independent Gaussian variables, so they form a Gaussian vector, and $(W_{t_1}, \ldots, W_{t_m})$ is a linear transformation of it (partial sums), hence Gaussian with mean zero. For $s \leq t$,

$$\operatorname{Cov}(W_s, W_t) = \operatorname{Cov}\left(W_s,\, W_s + (W_t - W_s)\right) = \operatorname{Var}(W_s) + \operatorname{Cov}(W_s - W_0,\, W_t - W_s) = s,$$

because the increments $W_s - W_0$ and $W_t - W_s$ are independent.

($\Leftarrow$) The increments over $0 = t_0 < t_1 < \cdots < t_m$ are linear functions of a Gaussian vector, so they are jointly Gaussian. For $i < j$, using (5.1) and $t_i < t_{i+1} \leq t_j < t_{j+1}$,

$$\operatorname{Cov}\left(W_{t_{i+1}} - W_{t_i},\, W_{t_{j+1}} - W_{t_j}\right) = t_{i+1} - t_{i+1} - t_i + t_i = 0.$$

Uncorrelated entries of a Gaussian vector are independent, which gives property (ii). Finally $W_t - W_s$ is Gaussian with mean zero and variance $t - 2\min(s,t) + s = t - s$ for $s < t$, which is property (iii). □

Remark 5.8 (Independence of processes). A process X is *independent of a σ-algebra $\mathcal{G}$* if every vector $(X_{t_1}, \ldots, X_{t_m})$ is independent of $\mathcal{G}$, and two processes X and Y are *independent* if every vector $(X_{s_1}, \ldots, X_{s_k})$ is independent of every vector $(Y_{t_1}, \ldots, Y_{t_m})$. By Dynkin's π–λ theorem, these finite-dimensional conditions imply the independence of the σ-algebras $\sigma(X_t : t \geq 0)$ and $\mathcal{G}$, respectively $\sigma(Y_t : t \geq 0)$; see Durrett (2019), Section 2.1. We use the two forms interchangeably.

Proposition 5.9 (Symmetry, scaling, restart)**.** *Let W be a Brownian motion, and fix $c > 0$ and $s \geq 0$. Each of the following processes is again a Brownian motion:*

(a) $(-W_t)_{t\geq 0}$*;*

(b) $\left(c^{-1/2}W_{ct}\right)_{t\geq 0}$*;*

(c) $(W_{s+t} - W_s)_{t\geq 0}$*, which is moreover independent of* $\mathcal{F}_s^W = \sigma(W_u : u \leq s)$*.*

Proof. Each process has continuous paths, starts at 0, and is a linear transformation of W, so its finite-dimensional vectors are centered Gaussian. By Theorem 5.7 it is enough to compute covariances: (a) $\mathrm{Cov}(-W_u, -W_t) = \min(u,t)$; (b) $\mathrm{Cov}\left(c^{-1/2}W_{cu}, c^{-1/2}W_{ct}\right) = c^{-1}\min(cu, ct) = \min(u,t)$; (c) $\mathrm{Cov}(W_{s+u} - W_s, W_{s+t} - W_s) = \min(s+u, s+t) - s - s + s = \min(u,t)$. For the independence in (c): every vector $(W_{s+t_1} - W_s, \ldots, W_{s+t_m} - W_s)$ is built from increments after time s, and every vector $(W_{u_1}, \ldots, W_{u_k})$ with $u_i \leq s$ from increments before s. These vectors are independent by property (ii), and Theorem 5.8 upgrades this to independence from $\mathcal{F}_s^W$. □

Property (b) is *Brownian scaling*: zooming in on a Brownian path by a factor c in time and $\sqrt{c}$ in space produces another Brownian path. No matter how closely you look, the path never straightens out. Section 5.5 turns this picture into a theorem. A fourth invariance, time inversion, is Exercise 5.5.

Remark 5.10 (Laws of path functionals). Because the paths are continuous, quantities such as $\sup_{u\leq t} W_u$ or the first time W leaves an interval are determined by the values of W at rational times: for instance $\sup_{u\leq t} W_u = \sup\left(\{W_u : u \in [0,t] \cap \mathbb{Q}\} \cup \{W_t\}\right)$. Their laws are therefore the same for every Brownian motion, on whatever probability space it is built. Combined with Theorem 5.9, this lets us move statements about such laws from one Brownian motion to another, for example from W to $c^{-1/2}W_{c\cdot}$.

5.4 Filtrations and martingales

In finance the information available at time t is usually more than the path of one Brownian motion: other assets, news, and so on. We therefore allow filtrations larger than the one generated by W, as long as they do not see the future increments of W.

Definition 5.11 (Filtration for a Brownian motion)**.** A *filtration for the Brownian motion W* is a family $(\mathcal{F}_t)_{t\geq 0}$ of σ-algebras such that

(i) $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$ (information accumulates);

(ii) W_t is $\mathcal{F}_t$-measurable for every t (the present is known);

(iii) $W_t - W_s$ is independent of $\mathcal{F}_s$ for every $s < t$ (future increments are unpredictable).

The smallest example is the natural filtration $\mathcal{F}_t^W = \sigma(W_u : u \leq t)$, for which (iii) is Theorem 5.9(c). A process (X_t) is a *martingale* with respect to $(\mathcal{F}_t)$ if X_t is $\mathcal{F}_t$-measurable and integrable for every t, and $\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$ for all $s \leq t$.

Theorem 5.12. *Let W be a Brownian motion with filtration $(\mathcal{F}_t)$, and let $\sigma \in \mathbb{R}$. The*

following processes are martingales:

$$\text{(i) } W_t, \qquad \text{(ii) } W_t^2 - t, \qquad \text{(iii) } Z_t = \exp\Big(\sigma W_t - \frac{\sigma^2}{2} t\Big).$$

Proof. All three are functions of (t, W_t), hence adapted. They are integrable because W_t is Gaussian: $\mathbb{E}[W_t^2] = t$, and the Gaussian moment generating function gives $\mathbb{E}[e^{\sigma W_t}] = e^{\sigma^2 t/2}$. Fix $s < t$ and write $W_t = W_s + D$ with $D = W_t - W_s \sim \mathcal{N}(0, t-s)$. Here W_s is $\mathcal{F}_s$-measurable and D is independent of $\mathcal{F}_s$, so for any functions f and g with the needed integrability, $\mathbb{E}[f(W_s)g(D) \mid \mathcal{F}_s] = f(W_s)\,\mathbb{E}[g(D)]$ by Theorem 1.16 (or, more generally, the independence lemma, Theorem 1.19).

(i) $\mathbb{E}[W_t \mid \mathcal{F}_s] = W_s + \mathbb{E}[D] = W_s$.

(ii) $W_t^2 = W_s^2 + 2W_s D + D^2$, so $\mathbb{E}[W_t^2 \mid \mathcal{F}_s] = W_s^2 + 2W_s\mathbb{E}[D] + \mathbb{E}[D^2] = W_s^2 + (t-s)$. Subtracting t from both sides gives $\mathbb{E}[W_t^2 - t \mid \mathcal{F}_s] = W_s^2 - s$.

(iii) $Z_t = Z_s \exp\left(\sigma D - \frac{\sigma^2}{2}(t-s)\right)$, and $\mathbb{E}[e^{\sigma D}] = e^{\sigma^2(t-s)/2}$, so $\mathbb{E}[Z_t \mid \mathcal{F}_s] = Z_s$. □

Remark 5.13 (The exponential martingale and stock prices). The martingale Z in (iii) is the template for asset prices. In the Black–Scholes model of Chapter 6, the stock under the risk-neutral measure is $S_t = S_0 \exp\left(\sigma W_t + (r - \frac{\sigma^2}{2})t\right)$, so the discounted price is $e^{-rt}S_t = S_0 Z_t$, a martingale. The correction $-\frac{\sigma^2}{2}t$ is exactly what cancels the convexity of the exponential: without it, $\mathbb{E}[e^{\sigma W_t}] = e^{\sigma^2 t/2}$ would grow.

Remark 5.14 (A martingale that dies). Let $\sigma \neq 0$. Then $\mathbb{E}[Z_t] = 1$ for all t, and yet $Z_t \to 0$ almost surely as $t \to \infty$, because $W_t/t \to 0$ (Exercise 5.5) makes the exponent $t\,(\sigma W_t/t - \sigma^2/2)$ tend to $-\infty$. The mean is carried by rarer and rarer paths that become very large. A martingale need not describe what a typical path does.

▶ Interactive demo: sample means of W_t, $W_t^2 - t$ and Z_t stay flat while those of W_t^2 and $e^{\sigma W_t}$ do not; and individual paths of Z_t go to zero while their exact mean stays at one.

5.5 Quadratic variation

Let $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ be a partition of $[0, t]$, with mesh $\|\Pi\| = \max_j (t_{j+1} - t_j)$. For a function f and $p > 0$, the *p-variation of f along* Π is

$$V_p(f; \Pi) = \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|^p.$$

The first variation ($p = 1$) measures how far the graph travels up and down. The quadratic variation ($p = 2$) is the case that matters for stochastic calculus.

Lemma 5.15 (Smooth functions have zero quadratic variation)**.** *If f is continuously differentiable on $[0, t]$, then $V_2(f; \Pi) \le \left(\max_{[0,t]} |f'|\right)^2 t\,\|\Pi\|$, which tends to 0 as $\|\Pi\| \to 0$.*

Proof. By the mean value theorem, $|f(t_{j+1}) - f(t_j)| \le \max |f'|\,(t_{j+1} - t_j)$. Squaring and using

$(t_{j+1}-t_j)^2 \le \|\Pi\|\,(t_{j+1}-t_j)$,

$$V_2(f;\Pi) \le (\max|f'|)^2 \sum_j (t_{j+1}-t_j)^2 \le (\max|f'|)^2\,\|\Pi\|\,t. \qquad \square$$

Brownian paths behave in the opposite way.

Theorem 5.16 (Quadratic variation of Brownian motion)**.** *Let W be a Brownian motion, $t>0$, and (Π_n) partitions of $[0,t]$ with $\|\Pi_n\|\to 0$. Then*

$$\mathbb{E}\Big[\big(V_2(W;\Pi_n)-t\big)^2\Big] = 2\sum_j \big(t_{j+1}^{(n)}-t_j^{(n)}\big)^2 \le 2t\,\|\Pi_n\| \xrightarrow[n\to\infty]{} 0, \tag{5.2}$$

so $V_2(W;\Pi_n)\to t$ in L^2. If moreover $\sum_n \|\Pi_n\|<\infty$ (for example, for the dyadic partitions $t_j = jt2^{-n}$), then $V_2(W;\Pi_n)\to t$ almost surely.

We write this as $[W,W]_t = t$.

Proof. Fix n and drop it from the notation. Write $\Delta_j = W_{t_{j+1}} - W_{t_j}$ and $h_j = t_{j+1}-t_j$, so that $\Delta_j \sim \mathcal{N}(0,h_j)$ independently, and

$$V_2(W;\Pi) - t = \sum_j (\Delta_j^2 - h_j).$$

The terms $\Delta_j^2 - h_j$ are independent with mean 0. Since $\mathbb{E}[\Delta_j^4] = 3h_j^2$ (the fourth moment of $\mathcal{N}(0,h)$ is $3h^2$), each has variance $\mathbb{E}[\Delta_j^4] - h_j^2 = 2h_j^2$. The variance of a sum of independent terms is the sum of the variances, so

$$\mathbb{E}\big[(V_2(W;\Pi)-t)^2\big] = \sum_j 2h_j^2 \le 2\|\Pi\| \sum_j h_j = 2t\|\Pi\|.$$

This is (5.2).

For the almost sure statement, fix $\varepsilon>0$. By Chebyshev's inequality and (5.2), $\mathbb{P}(|V_2(W;\Pi_n) - t| > \varepsilon) \le 2t\|\Pi_n\|/\varepsilon^2$, which is summable in n. By the Borel–Cantelli lemma,[1] almost surely $|V_2(W;\Pi_n)-t| \le \varepsilon$ for all n large enough. Applying this to $\varepsilon = 1/k$ for every integer $k\ge 1$ (a countable intersection of almost sure events is almost sure) gives $V_2(W;\Pi_n)\to t$ almost surely. $\square$

Note the mechanism. Each squared increment Δ_j^2 is random, but there are many of them and they are independent, so their sum concentrates around its mean t: a law of large numbers along a single path. For the scaled random walk of Section 5.2 there was nothing to average: every squared step was exactly $1/n$.

Corollary 5.17 (Infinite first variation)**.** *In the setting of Theorem 5.16, with $\sum_n \|\Pi_n\|<\infty$, $V_1(W;\Pi_n)\to\infty$ almost surely. In particular, almost every Brownian path has infinite total variation on $[0,t]$: it is not the difference of two increasing functions. Riemann–Stieltjes sums $\sum_j g(t_j)(W_{t_{j+1}}-W_{t_j})$ still converge when the integrand g itself has bounded variation (sum by parts), but they cannot be made to converge for every continuous g, and $\int_0^t W\,\mathrm{d}W$ is the first*

[1] If events A_n satisfy $\sum_n \mathbb{P}(A_n)<\infty$, then almost surely only finitely many of them occur.

casualty (Chapter 7).

Proof. For any partition,

$$V_2(W;\Pi) = \sum_j |\Delta_j|^2 \le \Big(\max_j |\Delta_j|\Big) \sum_j |\Delta_j| = \Big(\max_j |\Delta_j|\Big)\, V_1(W;\Pi).$$

A continuous path is uniformly continuous on $[0,t]$, so $\max_j |\Delta_j| \to 0$ as $\|\Pi_n\| \to 0$, for every ω. On the almost sure event where $V_2(W;\Pi_n) \to t > 0$, the inequality forces $V_1(W;\Pi_n) \ge V_2(W;\Pi_n)/\max_j |\Delta_j| \to \infty$. The total variation, the supremum of $V_1(W;\Pi)$ over all partitions, is therefore infinite. □

The roughness goes further: almost every Brownian path is nowhere differentiable (Paley, Wiener and Zygmund, 1933; see Mörters and Peres (2010), Theorem 1.30).

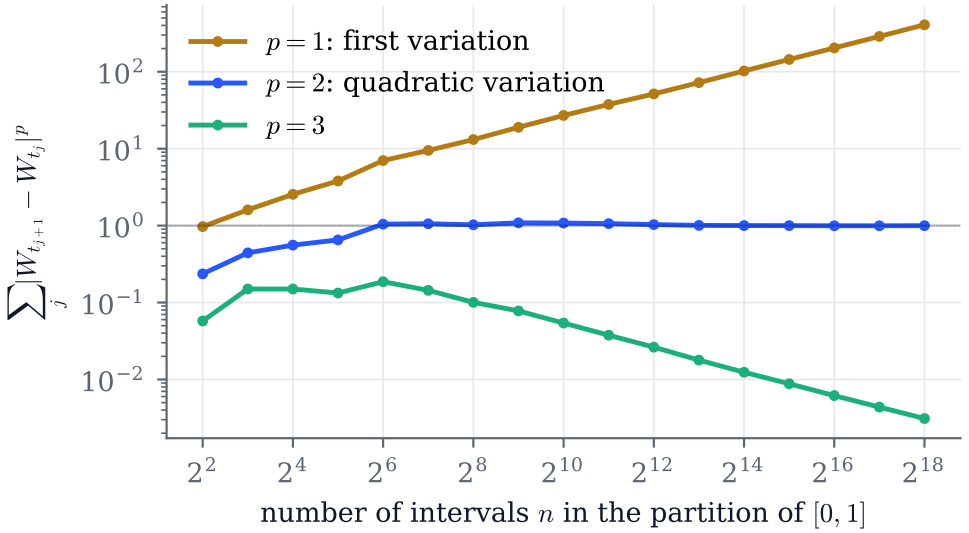

Figure 5.2: p-variations of one Brownian path on $[0,1]$ along dyadic partitions with $n = 2^2, \dots, 2^{18}$ intervals. The first variation grows like $\sqrt{n}$, the quadratic variation settles at $t = 1$ (Theorem 5.16), and the cubic variation goes to 0 (Exercise 5.7).

The rules $dW\,dW = dt$, $dW\,dt = 0$, $dt\,dt = 0$

The same arguments compute the cross terms. Along partitions with $\|\Pi\| \to 0$:

- $\sum_j \Delta_j^2 \to t$: this is Theorem 5.16, written $dW\,dW = dt$;
- $|\sum_j \Delta_j h_j| \le \max_j |\Delta_j| \sum_j h_j = t \max_j |\Delta_j| \to 0$, written $dW\,dt = 0$;
- $\sum_j h_j^2 \le t\|\Pi\| \to 0$ (Theorem 5.15 for $f(s) = s$), written $dt\,dt = 0$.

These three rules are all the bookkeeping that Itô's formula needs (Chapter 8). The first is the one that has no counterpart in ordinary calculus: in a second-order Taylor expansion of $f(W_t)$, the term $\frac{1}{2}f''(W_t)(dW_t)^2$ does not vanish but contributes $\frac{1}{2}f''(W_t)\,dt$.

▶ Interactive demo: the p-variations of a Brownian path on finer and finer partitions, for any p you choose.

Further reading

Shreve (2004b), Chapter 3, covers the same ground at the same level and with the same notation. Mörters and Peres (2010) is a complete and readable account of Brownian motion itself: construction, path properties and much more. Karatzas and Shreve (1991), Chapter 2, and Le Gall (2016), Chapter 2, give the measure-theoretic details that we have skipped, including the existence theorem.

Chapter 6

From the binomial model to Black–Scholes

Chapter 4 priced every claim in the binomial model by a finite computation, and Chapter 5 obtained Brownian motion as the limit of a random walk. This chapter puts the two together. Let the number of periods grow while the up and down factors shrink to one at the right rate: the log-price becomes a scaled random walk with a small drift, its terminal law becomes normal, and the binomial prices of calls and puts converge to explicit limits. These limits are the formula of Black and Scholes (1973) and Merton (1973).

The limit also suggests how to price in continuous time: the price of a claim is its expected discounted payoff under a measure $\mathbb{Q}$ that makes the discounted stock price a martingale. That this is the only price consistent with no arbitrage, because the claim can be replicated by trading continuously, needs stochastic integrals; it is the subject of Chapters 10 and 11. Here we compute: the formula at every date, the prices of digital options, the sensitivities of the price (the Greeks) and the approximations that traders use at the money.

6.1 Geometric Brownian motion

Definition 6.1 (Geometric Brownian motion)**.** Let W be a Brownian motion with a filtration $(\mathcal{F}_t)$ (Theorems 5.6 and 5.11), and let $S_0 > 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$. The process

$$S_t = S_0 \exp\Big(\sigma W_t + \Big(\mu - \frac{\sigma^2}{2}\Big)t\Big), \qquad t \geq 0,$$

is a *geometric Brownian motion* with *drift* μ and *volatility* σ.

The log-price $\log S_t = \log S_0 + (\mu - \frac{\sigma^2}{2})t + \sigma W_t$ is a Brownian motion with drift, so the log-returns over disjoint time intervals are independent and normal: the continuous-time counterpart of independent coin tosses. The next proposition shows that μ is the expected rate of return; the correction $-\frac{\sigma^2}{2}$ compensates the convexity of the exponential, as in Theorem 5.13. In Chapter 8 we will see that S solves the stochastic differential equation $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$ (Theorem 8.8).

Proposition 6.2. *Let S be a geometric Brownian motion with drift μ and volatility σ, and let $0 \le s < t$.*

(i) The ratio S_t/S_s is independent of $\mathcal{F}_s$, and $\log(S_t/S_s) \sim \mathcal{N}((\mu - \frac{\sigma^2}{2})(t-s),\, \sigma^2(t-s))$. In particular $S_t > 0$, and for $t > 0$ the law of S_t is lognormal, *with density*

$$f_t(y) = \frac{1}{y\sigma\sqrt{t}}\, N'\!\left(\frac{\log(y/S_0) - (\mu - \sigma^2/2)t}{\sigma\sqrt{t}}\right), \qquad y > 0.$$

(ii) For every $p \in \mathbb{R}$, S_t^p is integrable and $\mathbb{E}[S_t^p \mid \mathcal{F}_s] = S_s^p \exp\left(p(\mu - \frac{\sigma^2}{2})(t-s) + \frac{p^2\sigma^2}{2}(t-s)\right)$. In particular $\mathbb{E}[S_t \mid \mathcal{F}_s] = e^{\mu(t-s)} S_s$.

(iii) Let $r \in \mathbb{R}$. The discounted price $(e^{-rt}S_t)_{t\ge 0}$ is a martingale if and only if $\mu = r$.

Proof. (i) $S_t/S_s = \exp\left(\sigma D + (\mu - \frac{\sigma^2}{2})(t-s)\right)$ is a function of $D = W_t - W_s \sim \mathcal{N}(0, t-s)$, which is independent of $\mathcal{F}_s$ (Theorem 5.11). With $s = 0$, $\log S_t$ is normal with mean $m = \log S_0 + (\mu - \frac{\sigma^2}{2})t$ and variance $\sigma^2 t$, so $\mathbb{P}(S_t \le y) = N\big((\log y - m)/(\sigma\sqrt{t})\big)$ for $y > 0$; differentiating in y gives f_t.

(ii) By the Gaussian moment generating function, $\mathbb{E}[e^{p\sigma D}] = e^{p^2\sigma^2(t-s)/2}$, so $(S_t/S_s)^p$ and S_t^p are integrable. Taking out what is known (S_s^p is $\mathcal{F}_s$-measurable and $S_t^p = S_s^p (S_t/S_s)^p$ is integrable) and using independence (Theorem 1.16), $\mathbb{E}[S_t^p \mid \mathcal{F}_s] = S_s^p\, \mathbb{E}[(S_t/S_s)^p]$, which is the formula.

(iii) $e^{-rt}S_t$ is adapted and integrable, and by (ii) $\mathbb{E}[e^{-rt}S_t \mid \mathcal{F}_s] = e^{(\mu-r)(t-s)}\, e^{-rs} S_s$. If $\mu = r$ this is the martingale property. If $\mu \ne r$, take $s = 0$ and $t > 0$: $\mathbb{E}[e^{-rt}S_t] = e^{(\mu-r)t} S_0 \ne S_0$. □

The mean of S_t grows like $e^{\mu t}$, but its median is $S_0 e^{(\mu - \sigma^2/2)t}$, and if $\mu < \sigma^2/2$ then $\log S_t = t\left(\sigma W_t/t + \mu - \frac{\sigma^2}{2}\right) \to -\infty$ almost surely, because $W_t/t \to 0$ (Exercise 5.5): the phenomenon of Theorem 5.14, for a stock.

6.2 The binomial model in the limit

Fix a maturity $T > 0$, an initial price $S_0 > 0$, a continuously compounded interest rate $r \in \mathbb{R}$ and a volatility $\sigma > 0$. For $n \ge 1$, divide $[0, T]$ into n periods of length $\delta_n = T/n$, and consider the n-period binomial model of Theorem 4.1 with

$$u_n = e^{\sigma\sqrt{\delta_n}}, \qquad d_n = e^{-\sigma\sqrt{\delta_n}}, \qquad 1 + r_n = e^{r\delta_n}, \tag{6.1}$$

where r_n is the interest rate per period. Period k ends at time $k\delta_n$, and a dollar in the money market grows to $(1+r_n)^n = e^{rT}$ at time T, whatever n. The log-price moves by $\pm\sigma\sqrt{\delta_n}$ in each period: after k periods, $\log(S_k/S_0) = \sigma\sqrt{\delta_n}\, M_k = \sigma\sqrt{T}\, W^{(n)}(k/n)$, where M_k counts heads minus tails and $W^{(n)}$ is the scaled random walk of Theorem 5.3. The condition $d_n < 1 + r_n < u_n$ of Theorem 4.9 reads $|r|\delta_n < \sigma\sqrt{\delta_n}$, that is,

$$n > r^2 T/\sigma^2;$$

for these n the model has no arbitrage, and its risk-neutral measure $\mathbb{Q}_n$ makes the tosses independent with $\mathbb{Q}_n(H) = q_n = (e^{r\delta_n} - d_n)/(u_n - d_n)$ (Theorem 4.7). This parametrization is due to Cox et al. (1979).

Recall that random variables X_n, possibly defined on different probability spaces, *converge in distribution* to X if $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for every bounded continuous $f : \mathbb{R} \to \mathbb{R}$; equivalently, if the distribution functions converge at every point where the limit is continuous (Durrett (2019), Section 3.2). If $X_n \to X$ in distribution and g is continuous, then $g(X_n) \to g(X)$ in distribution, since $f \circ g$ is bounded and continuous.

The central limit theorem we need is for a *triangular array*: in the n-th tree the steps are independent and identically distributed, but their law changes with n. With bounded steps it has a short proof along the lines of Theorem 5.5. When the limiting variance is positive it also follows, after centring the steps, from the Lindeberg–Feller theorem (Durrett (2019), Section 3.4), whose Lindeberg condition holds trivially for steps bounded by $\varepsilon_n \to 0$; when it is zero, Chebyshev's inequality suffices.

Lemma 6.3 (Central limit theorem for small bounded steps)**.** *For each $n \geq 1$, let $k_n \geq 1$ be an integer and $\xi_{n,1}, \dots, \xi_{n,k_n}$ independent random variables with a common law, such that $|\xi_{n,j}| \leq \varepsilon_n$. Suppose that $k_n \to \infty$, $\varepsilon_n \to 0$, and*

$$k_n \, \mathbb{E}[\xi_{n,1}] \to m, \qquad k_n \, \mathbb{E}[\xi_{n,1}^2] \to v.$$

Then $X_n = \sum_{j=1}^{k_n} \xi_{n,j}$ converges in distribution to $\mathcal{N}(m, v)$ (the point mass at m if $v = 0$).

Proof. Fix $u \in \mathbb{R}$ and write $\xi = \xi_{n,1}$, $m_n = k_n \mathbb{E}[\xi]$, $v_n = k_n \mathbb{E}[\xi^2]$ and $z_n = ium_n - \frac{u^2}{2} v_n$, so that $z_n \to z = ium - \frac{u^2}{2} v$. Taylor's formula with integral remainder gives $e^{iy} - 1 - iy + \frac{y^2}{2} = \frac{-i}{2} \int_0^y (y-s)^2 e^{is} \, \mathrm{d}s$, so $|e^{iy} - 1 - iy + \frac{y^2}{2}| \leq |y|^3/6$ for real y. Taking expectations, and using $\mathbb{E}|\xi|^3 \leq \varepsilon_n \mathbb{E}[\xi^2]$,

$$\psi_n := \mathbb{E}\big[e^{iu\xi}\big] = 1 + \frac{z_n}{k_n} + \rho_n, \qquad |\rho_n| \leq \frac{|u|^3 \varepsilon_n \mathbb{E}[\xi^2]}{6} = \frac{|u|^3 \varepsilon_n v_n}{6 k_n}.$$

By independence, $\mathbb{E}[e^{iuX_n}] = \psi_n^{k_n}$. For complex a, b and $k \geq 1$, $a^k - b^k = (a-b) \sum_{j<k} a^j b^{k-1-j}$, so $|a^k - b^k| \leq k |a-b| \max(|a|, |b|)^{k-1}$. With $a = \psi_n$ (so $|a| \leq 1$) and $b = 1 + z_n/k_n$ (so $|b| \leq e^{|z_n|/k_n}$),

$$\Big| \psi_n^{k_n} - \Big(1 + \frac{z_n}{k_n}\Big)^{k_n} \Big| \leq k_n |\rho_n| \, e^{|z_n|} \leq \frac{|u|^3 \varepsilon_n v_n}{6} \, e^{|z_n|} \longrightarrow 0.$$

With $a = 1 + z_n/k_n$ and $b = e^{z_n/k_n}$, the bound $|e^y - 1 - y| \leq \frac{|y|^2}{2} e^{|y|}$ (from the power series) gives $|a - b| \leq \frac{|z_n|^2}{2k_n^2} e^{|z_n|/k_n}$, hence $|(1 + z_n/k_n)^{k_n} - e^{z_n}| \leq \frac{|z_n|^2}{2k_n} e^{2|z_n|} \to 0$. Since $e^{z_n} \to e^z$, we get $\mathbb{E}[e^{iuX_n}] \to e^{ium - u^2 v/2}$, the characteristic function of $\mathcal{N}(m, v)$. By Lévy's continuity theorem (Durrett (2019), Section 3.3), X_n converges in distribution to $\mathcal{N}(m, v)$. □

Theorem 6.4 (From the binomial model to Black–Scholes)**.** *Let $T > 0$, $S_0 > 0$, $r \in \mathbb{R}$ and $\sigma > 0$. For $n > r^2 T/\sigma^2$, let $S_T^{(n)}$ be the stock price after n periods (at time T) in the binomial model (6.1), with its risk-neutral measure $\mathbb{Q}_n$.*

(i) $q_n \to \frac{1}{2}$. Under $\mathbb{Q}_n$, $\log(S_T^{(n)}/S_0)$ converges in distribution to $\mathcal{N}((r - \frac{\sigma^2}{2})T, \sigma^2 T)$. Consequently, $S_T^{(n)}$ converges in distribution to

$$S_T = S_0 \exp\Big(\sigma \sqrt{T}\, Y + \Big(r - \frac{\sigma^2}{2}\Big) T\Big), \qquad Y \sim \mathcal{N}(0, 1), \tag{6.2}$$

the value at time T of a geometric Brownian motion with drift r.

(ii) For every strike $K \geq 0$, the binomial prices (4.11) of the European call and put with strike K and maturity T converge:

$$C_0^{(n)} \longrightarrow e^{-rT}\mathbb{E}[(S_T - K)^+], \qquad P_0^{(n)} \longrightarrow e^{-rT}\mathbb{E}[(K - S_T)^+].$$

The limits are computed explicitly in Theorem 6.11.

Proof. (i) Write $s_n = \sigma\sqrt{\delta_n}$ and $\delta = \delta_n$. Under $\mathbb{Q}_n$ the tosses are independent with $\mathbb{Q}_n(H) = q_n$, so $\log(S_T^{(n)}/S_0) = \sum_{j=1}^n \xi_{n,j}$, where $\xi_{n,j} = s_n$ if the j-th toss is a head and $-s_n$ otherwise. These steps are independent, identically distributed and bounded by $\varepsilon_n = s_n \to 0$, and $n\mathbb{E}[\xi_{n,1}^2] = ns_n^2 = \sigma^2 T$. For the mean, $q_n(e^{s_n} - e^{-s_n}) = e^{r\delta} - e^{-s_n}$ gives

$$2q_n - 1 = \frac{2e^{r\delta} - e^{s_n} - e^{-s_n}}{e^{s_n} - e^{-s_n}} = \frac{e^{r\delta} - \cosh s_n}{\sinh s_n},$$

so that

$$n\mathbb{E}[\xi_{n,1}] = ns_n(2q_n - 1) = \Big(n(e^{r\delta} - 1) - n\big(\cosh s_n - 1\big)\Big)\frac{s_n}{\sinh s_n}.$$

As $n \to \infty$: $n(e^{r\delta} - 1) = T(e^{r\delta} - 1)/\delta \to rT$; $n(\cosh s_n - 1) = \sigma^2 T\,(\cosh s_n - 1)/s_n^2 \to \sigma^2 T/2$; and $s_n/\sinh s_n \to 1$. Hence $n\mathbb{E}[\xi_{n,1}] \to (r - \frac{\sigma^2}{2})T$, and also $2q_n - 1 = n\mathbb{E}[\xi_{n,1}]/(ns_n) \to 0$, because $ns_n = \sigma\sqrt{nT} \to \infty$. Theorem 6.3 with $k_n = n$ gives the convergence of $\log(S_T^{(n)}/S_0)$, and that of $S_T^{(n)} = S_0 \exp\big(\log(S_T^{(n)}/S_0)\big)$ follows because the exponential is continuous.

(ii) In the n-period model the discount factor is $(1 + r_n)^{-n} = e^{-rT}$, so by (4.11), $P_0^{(n)} = e^{-rT}\mathbb{E}^{\mathbb{Q}_n}[f(S_T^{(n)})]$ with $f(x) = \min\big((K - x)^+, K\big)$, a bounded continuous function equal to the put payoff on $(0, \infty)$. By (i), $P_0^{(n)} \to e^{-rT}\mathbb{E}[f(S_T)] = e^{-rT}\mathbb{E}[(K - S_T)^+]$. The call payoff is unbounded, and convergence in distribution alone does not control the expectation of an unbounded function; the martingale property does. Since $(x - K)^+ - (K - x)^+ = x - K$ and $\mathbb{E}^{\mathbb{Q}_n}[S_T^{(n)}] = (1 + r_n)^n S_0 = e^{rT}S_0$ (Theorem 4.8), put–call parity holds in every tree: $C_0^{(n)} = P_0^{(n)} + S_0 - Ke^{-rT}$. The same identity holds for the limit, because $\mathbb{E}[S_T] = S_0 e^{rT}$ by the Gaussian moment generating function. Hence $C_0^{(n)} \to e^{-rT}\mathbb{E}[(K - S_T)^+] + S_0 - Ke^{-rT} = e^{-rT}\mathbb{E}[(S_T - K)^+]$. □

Remark 6.5 (The scaling). The proof uses four properties of the trees: steps bounded by $\varepsilon_n \to 0$, $n\mathbb{E}^{\mathbb{Q}_n}[\xi] \to (r - \frac{\sigma^2}{2})T$, $n\mathbb{E}^{\mathbb{Q}_n}[\xi^2] \to \sigma^2 T$, and the martingale identity $\mathbb{E}^{\mathbb{Q}_n}[S_T^{(n)}] = S_0 e^{rT}$. Any family of trees with these properties has the same limit; one checks in the same way that $u_n = 1 + \sigma\sqrt{\delta_n}$, $d_n = 1 - \sigma\sqrt{\delta_n}$ is one. The exponent $\frac{1}{2}$ in $\sqrt{\delta_n}$ is what keeps the limit both random and finite, as in Section 5.2: with $u_n = 1/d_n = e^{\sigma\delta_n^\alpha}$ and $0 < \alpha < 1$, the variance of the log-price under $\mathbb{Q}_n$ is $n\sigma^2\delta_n^{2\alpha}(1 - (2q_n - 1)^2)$, which for large n is of order $n^{1-2\alpha}$: it tends to infinity if $\alpha < \frac{1}{2}$ and to zero if $\frac{1}{2} < \alpha < 1$. In the second case Theorem 6.3 with $v = 0$ gives $S_T^{(n)} \to S_0 e^{rT}$ in distribution, and the argument of (ii) gives $C_0^{(n)} \to (S_0 - Ke^{-rT})^+$, the lower bound of Theorem 3.10: without randomness in the limit, the option is worth only its discounted intrinsic value.

Remark 6.6 (The real-world limit). Theorem 6.4 is a statement under $\mathbb{Q}_n$. Under $\mathbb{P}$, let the tosses have probability $p_n = \frac{1}{2} + \frac{(\mu - \sigma^2/2)\sqrt{\delta_n}}{2\sigma}$ of heads, for some $\mu \in \mathbb{R}$ and n large enough that

$p_n \in (0,1)$. Then $n\mathbb{E}^{\mathbb{P}}[\xi] = ns_n(2p_n - 1) = (\mu - \frac{\sigma^2}{2})T$ exactly, and Theorem 6.3 shows that under $\mathbb{P}$ the terminal price converges to the lognormal law of a geometric Brownian motion with drift μ. The prices depend only on $\mathbb{Q}_n$, so μ does not enter their limit: this is Theorem 4.6, in the limit.

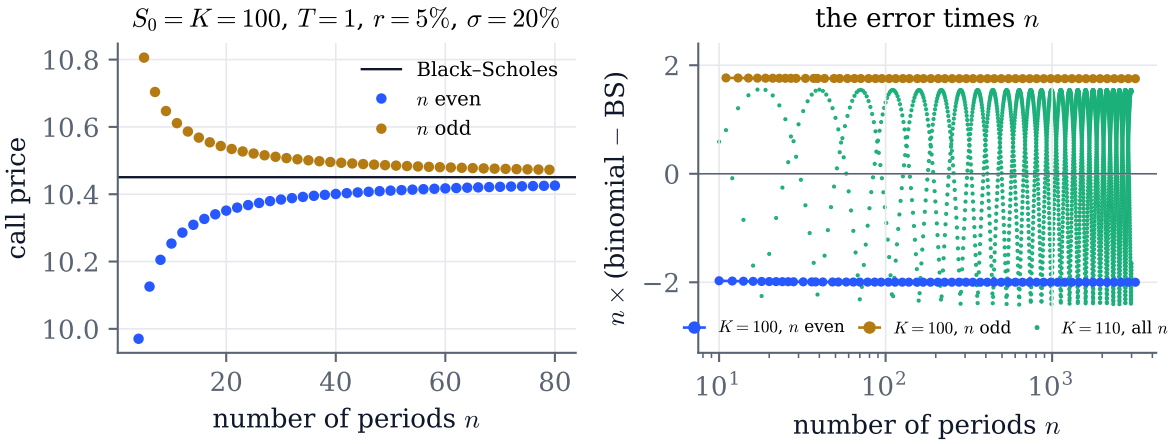

Figure 6.1: Binomial prices of a call converging to the Black–Scholes price. Left: with $K = S_0$, the strike is a node of the terminal lattice when n is even and lies halfway between two nodes when n is odd; the prices approach the limit from below and from above. Right: the error times n. For $K = S_0$ it settles near -2 (n even) and 1.75 (n odd); for $K = 110$ it keeps oscillating as the lattice slides past the strike.

Remark 6.7 (Odd and even). The convergence is not monotone (Figure 6.1). The terminal prices $S_0 u_n^{2j-n}$, $0 \le j \le n$, form a lattice whose spacing in log is $2\sigma\sqrt{\delta_n}$, and the call payoff has a kink at K; the error depends on where $\log(K/S_0)$ falls in the lattice. Numerically, n times the error stays bounded; for $K = S_0$ it settles along the even and along the odd n, where the strike is a node or a mid-point, while for $K = 110$ it keeps oscillating. A precise expansion of the error is given by F. Diener and M. Diener (*Mathematical Finance*, 2004). Remedies, among them the smoothed tree of M. Broadie and J. Detemple (*Review of Financial Studies*, 1996), are explored in Exercise 6.8.

▶ Interactive demo: watch the law of the log-price in the tree approach the normal density, and the binomial prices approach the Black–Scholes price, for any strike, rate, volatility and maturity.

6.3 Risk-neutral pricing in continuous time

Theorem 6.4 suggests a model and a pricing rule. The model: under the real-world measure $\mathbb{P}$, the stock is a geometric Brownian motion with some drift μ (Theorem 6.6). The pricing rule: prices are expected discounted payoffs under a measure $\mathbb{Q}$ under which the stock is a geometric Brownian motion with drift r, that is, by Theorem 6.2(iii), under which $e^{-rt}S_t$ is a martingale. We now fix this setting for the rest of the chapter.

Setting. Let $T > 0$, $r \in \mathbb{R}$, $\sigma > 0$ and $S_0 > 0$. Let $\mathbb{Q}$ be a probability measure on $(\Omega, \mathcal{F})$ and $\widetilde{W}$ a Brownian motion under $\mathbb{Q}$, with filtration $(\mathcal{F}_t)_{t \ge 0}$; only times $t \le T$ will matter. The money

market is worth e^{rt} at time t and the stock

$$S_t = S_0 \exp\Big(\sigma \widetilde{W}_t + \Big(r - \frac{\sigma^2}{2}\Big)t\Big), \qquad t \geq 0. \tag{6.3}$$

By Theorem 6.2(iii), applied under $\mathbb{Q}$, $(e^{-rt}S_t)$ is a $\mathbb{Q}$-martingale. We *define* the *price* at time $t \in [0,T]$ of a claim paying $g(S_T)$ at time T as

$$V_t = e^{-r(T-t)}\,\mathbb{E}^{\mathbb{Q}}[g(S_T)\,|\,\mathcal{F}_t], \tag{6.4}$$

for every Borel $g : (0,\infty) \to \mathbb{R}$ such that $g \geq 0$ or $g(S_T)$ is $\mathbb{Q}$-integrable. In this chapter (6.4) is a definition; Chapters 10 and 11 justify it by replication and no arbitrage.

Remark 6.8 (Why this price). Three facts support (6.4). (a) For calls and puts at time 0, V_0 is the limit of the unique arbitrage-free prices of the binomial trees (Theorem 6.4). (b) The discounted price $e^{-rt}V_t$ is a $\mathbb{Q}$-martingale (Theorem 6.9). In the binomial model this is what rules out arbitrage (Theorems 4.9 and 4.10); the continuous-time version needs stochastic integrals. (c) In Chapter 10 the claim is replicated by a self-financing portfolio in the stock and the money market, and in Chapter 11 the measure $\mathbb{Q}$ is built from $\mathbb{P}$: if S is a geometric Brownian motion with drift μ driven by a $\mathbb{P}$-Brownian motion W, then $\widetilde{W}_t = W_t + \frac{\mu - r}{\sigma}t$, $t \leq T$, is a Brownian motion on $[0,T]$ under the measure $\mathbb{Q}$ with $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = \exp\big(\theta W_T - \frac{\theta^2}{2}T\big)$, $\theta = -\frac{\mu - r}{\sigma}$ (Girsanov's theorem; for the single time T this is Theorem 1.27). The number $(\mu - r)/\sigma$ is the *market price of risk*. In this chapter neither μ nor $\mathbb{P}$ plays any role.

To compute (6.4) at $t > 0$ we condition on $\mathcal{F}_t$. The tool is the independence lemma of Chapter 1: the stock price at T is the price at t times a factor that is independent of $\mathcal{F}_t$.

Proposition 6.9 (Markov property and prices at intermediate times)**.** *In the setting* (6.3)*, let* $0 \leq t \leq T$*,* $\tau = T - t$*, and let* $g : (0,\infty) \to \mathbb{R}$ *be Borel with* $|g(y)| \leq a(y^p + y^{-p})$ *for all* $y > 0$*, for some constants* $a, p \geq 0$*. Then, almost surely,*

$$\mathbb{E}^{\mathbb{Q}}[g(S_T)\,|\,\mathcal{F}_t] = v(t,S_t), \qquad v(t,x) = \mathbb{E}\Big[g\Big(x\exp\Big(\sigma\sqrt{\tau}\,Y + \Big(r - \frac{\sigma^2}{2}\Big)\tau\Big)\Big)\Big],$$

where $Y \sim \mathcal{N}(0,1)$*. So the price* $V_t = e^{-r\tau}v(t,S_t)$ *depends on the past only through the current stock price, and* $(e^{-rt}V_t)_{0\leq t\leq T}$ *is a* $\mathbb{Q}$*-martingale.*

Proof. For $t = T$ both sides equal $g(S_T)$. Let $t < T$ and $Y = (\widetilde{W}_T - \widetilde{W}_t)/\sqrt{\tau}$, which under $\mathbb{Q}$ is $\mathcal{N}(0,1)$ and independent of $\mathcal{F}_t$ (Theorem 5.11). Then

$$S_T = S_t \exp\Big(\sigma\sqrt{\tau}\,Y + \Big(r - \frac{\sigma^2}{2}\Big)\tau\Big), \qquad \text{so} \qquad g(S_T) = f(S_t, Y),$$

with $f(x,y) = g(x\exp(\sigma\sqrt{\tau}\,y + (r - \sigma^2/2)\tau))$ Borel, and S_t is $\mathcal{F}_t$-measurable. Write $Z = \sigma\sqrt{\tau}\,Y + (r - \frac{\sigma^2}{2})\tau$. The growth condition gives $|f(x,Y)| \leq a(x^p + x^{-p})\,e^{p|Z|} \leq a(x^p + x^{-p})(e^{pZ} + e^{-pZ})$, which is integrable for every $x > 0$ by the Gaussian moment generating function; and $g(S_T)$ is integrable by Theorem 6.2(ii). The independence lemma, Theorem 1.19 under $\mathbb{Q}$, gives $\mathbb{E}^{\mathbb{Q}}[g(S_T) \mid \mathcal{F}_t] = v(t,S_t)$. Finally $e^{-rt}V_t = e^{-rT}\mathbb{E}^{\mathbb{Q}}[g(S_T) \mid \mathcal{F}_t]$ is a martingale by the tower property. □

In the binomial model the same fact appeared as the recombining-tree recursion $V_n = v_n(S_n)$ of Section 4.3. Conditional expectations of functions of S_T given $\mathcal{F}_t$ depend on $\mathcal{F}_t$ only through S_t: this is the *Markov property*, which Chapter 9 extends to solutions of stochastic differential equations (Theorem 9.9).

6.4 The Black–Scholes formula

For $\tau > 0$, $x > 0$ and a strike $K > 0$, set

$$d_\pm(\tau, x) = \frac{\log(x/K) + (r \pm \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}}, \qquad d_-(\tau, x) = d_+(\tau, x) - \sigma\sqrt{\tau}. \tag{6.5}$$

The whole computation rests on one Gaussian identity, which is the Gaussian shift of Theorem 1.26 in disguise.

Lemma 6.10. *Let $Y \sim \mathcal{N}(0,1)$ and $a, d \in \mathbb{R}$. Then $\mathbb{P}(Y > -d) = N(d)$ and*

$$\mathbb{E}\big[e^{aY - a^2/2}\,\mathbf{1}_{\{Y > -d\}}\big] = N(d + a).$$

Proof. By symmetry of the normal law, $\mathbb{P}(Y > -d) = \mathbb{P}(-Y < d) = N(d)$. The left side is $\mathbb{E}[Z\,h(Y)]$ with $Z = e^{aY - a^2/2}$ and $h = \mathbf{1}_{(-d,\infty)}$; by Theorem 1.26 with $\theta = a$, it equals $\mathbb{E}[h(Y + a)] = \mathbb{P}(Y > -d - a) = N(d + a)$. □

Theorem 6.11 (The Black–Scholes formula)**.** *In the setting (6.3), let $K > 0$, $0 \le t < T$ and $\tau = T - t$. The prices at time t of the European call and put with strike K and maturity T are*

$$C_t = e^{-r\tau}\mathbb{E}^{\mathbb{Q}}\big[(S_T - K)^+ \,\big|\, \mathcal{F}_t\big] = c(t, S_t), \qquad P_t = e^{-r\tau}\mathbb{E}^{\mathbb{Q}}\big[(K - S_T)^+ \,\big|\, \mathcal{F}_t\big] = p(t, S_t),$$

where, for $x > 0$ and with $d_\pm = d_\pm(\tau, x)$ as in (6.5),

$$c(t, x) = x\,N(d_+) - Ke^{-r\tau}N(d_-), \qquad p(t, x) = Ke^{-r\tau}N(-d_-) - x\,N(-d_+).$$

In particular $C_0 = S_0 N(d_+) - Ke^{-rT}N(d_-)$ with $d_\pm = \big(\log(S_0/K) + (r \pm \frac{\sigma^2}{2})T\big)/(\sigma\sqrt{T})$, and $c(t, x) - p(t, x) = x - Ke^{-r\tau}$ for all $t < T$ and $x > 0$.

Proof. The payoff $g(y) = (y - K)^+ \le y$ satisfies the hypothesis of Theorem 6.9, so $C_t = e^{-r\tau}v(t, S_t)$ with $v(t, x) = \mathbb{E}\big[(xe^{\sigma\sqrt{\tau}Y + (r - \sigma^2/2)\tau} - K)^+\big]$. Fix $x > 0$. The payoff is positive exactly when $\sigma\sqrt{\tau}\,Y + (r - \frac{\sigma^2}{2})\tau > \log(K/x)$, that is, when $Y > -d_-$. Hence

$$e^{-r\tau}v(t, x) = x\,\mathbb{E}\big[e^{\sigma\sqrt{\tau}\,Y - \sigma^2\tau/2}\mathbf{1}_{\{Y > -d_-\}}\big] - Ke^{-r\tau}\,\mathbb{P}(Y > -d_-) = x\,N(d_- + \sigma\sqrt{\tau}) - Ke^{-r\tau}N(d_-),$$

by Theorem 6.10 with $a = \sigma\sqrt{\tau}$ and $d = d_-$; and $d_- + \sigma\sqrt{\tau} = d_+$. For the put, $(y - K)^+ - (K - y)^+ = y - K$ and $\mathbb{E}\big[xe^{\sigma\sqrt{\tau}Y + (r - \sigma^2/2)\tau}\big] = xe^{r\tau}$, so the put function is $c(t, x) - x + Ke^{-r\tau} = -x\big(1 - N(d_+)\big) + Ke^{-r\tau}\big(1 - N(d_-)\big)$, which is $p(t, x)$ because $1 - N(y) = N(-y)$. □

Remark 6.12 (Reading the formula). (a) The term $Ke^{-r\tau}N(d_-)$ is the discounted strike times the risk-neutral probability that the call is exercised, and $xN(d_+)$ is the value of receiving the stock in that event (Theorem 6.13). Neither is a real-world quantity: μ does not appear.

(b) The price is a function of x, the discounted strike $Ke^{-r\tau}$ and the *total volatility* $\sigma\sqrt{\tau}$ only: writing $m = \log\left(x/(Ke^{-r\tau})\right)$ and $w = \sigma\sqrt{\tau}$, we have $d_\pm = m/w \pm w/2$.

(c) The prices satisfy the parity of Theorem 3.9 and the bounds of Theorem 3.10, with $B = e^{-r\tau}$, as they must, being expectations under a pricing measure.

(d) Boundary cases. As $\tau \downarrow 0$, $c(t,x) \to (x-K)^+$ for every $x > 0$ (at $x = K$, $d_\pm \to 0$ and both terms tend to $K/2$). For $K = 0$ the call pays S_T and $C_t = e^{-r\tau}\mathbb{E}^{\mathbb{Q}}[S_T \mid \mathcal{F}_t] = S_t$ by the martingale property; the formula gives the same value with the convention $d_\pm = +\infty$, $N(+\infty) = 1$.

Proposition 6.13 (Digital options)**.** *In the setting of Theorem 6.11, with $d_\pm = d_\pm(\tau, S_t)$, the prices at time t of the* digital call *paying $\mathbf{1}_{\{S_T>K\}}$, the* digital put *paying $\mathbf{1}_{\{S_T\le K\}}$, and the* asset-or-nothing call *paying $S_T\mathbf{1}_{\{S_T>K\}}$ are*

$$e^{-r\tau}N(d_-), \qquad e^{-r\tau}N(-d_-), \qquad S_t\,N(d_+).$$

In particular $\mathbb{Q}(S_T > K \mid \mathcal{F}_t) = N\big(d_-(\tau, S_t)\big)$ and $\mathbb{Q}(S_T > K) = N\big(d_-(T, S_0)\big)$, and the call is one asset-or-nothing call minus K digital calls.

Proof. Apply Theorem 6.9 to the payoffs $\mathbf{1}_{(K,\infty)}(y) \le 1$ and $y\mathbf{1}_{(K,\infty)}(y) \le y$. As in the proof of Theorem 6.11, $S_T > K$ corresponds to $Y > -d_-$, so $v(t,x) = \mathbb{P}(Y > -d_-) = N(d_-)$ for the first and $v(t,x) = xe^{r\tau}N(d_+)$ for the second, by Theorem 6.10. The two digital payoffs add up to 1, whose price is $e^{-r\tau}$, and $1 - N(d_-) = N(-d_-)$. The conditional probability is $\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{S_T>K\}} \mid \mathcal{F}_t]$, and $(y-K)^+ = y\mathbf{1}_{\{y>K\}} - K\mathbf{1}_{\{y>K\}}$. □

Since S_T has a density, $\mathbb{Q}(S_T = K) = 0$, and it does not matter whether the digitals pay at $S_T = K$ or not. For the same reason the digital prices of the trees converge too, although the payoff is discontinuous: by Theorem 6.4(i), $\mathbb{Q}_n(S_T^{(n)} > K) \to \mathbb{Q}(S_T > K)$, because the limit distribution function is continuous at K.

6.5 The Greeks

The *Greeks* are the partial derivatives of a price with respect to the stock price, time and the parameters. They measure how a position reacts when the market moves, and in Chapter 10 the delta becomes the number of shares held in the replicating portfolio, as the discrete derivative (4.8) was in the binomial model.

Proposition 6.14 (The Greeks of calls and puts)**.** *Let $K > 0$, $0 \le t < T$, $\tau = T - t$ and $x > 0$, write $d_\pm = d_\pm(\tau, x)$, and regard c, p and $d_\pm$ as functions of (t,x) and of the parameters σ, r and K. Then*

$$Ke^{-r\tau}N'(d_-) = x\,N'(d_+), \tag{6.6}$$

and the derivatives of the call and put prices are as follows.

(i) Delta: $\partial_x c = N(d_+)$ and $\partial_x p = N(d_+) - 1 = -N(-d_+)$.

(ii) Gamma: $\partial_{xx}c = \partial_{xx}p = \dfrac{N'(d_+)}{x\sigma\sqrt{\tau}}$.

(iii) Vega: $\partial_\sigma c = \partial_\sigma p = x\,N'(d_+)\sqrt{\tau}$.

(iv) Theta: $\partial_t c = -\dfrac{x\sigma N'(d_+)}{2\sqrt{\tau}} - rKe^{-r\tau}N(d_-)$ *and* $\partial_t p = -\dfrac{x\sigma N'(d_+)}{2\sqrt{\tau}} + rKe^{-r\tau}N(-d_-)$.

(v) Rho: $\partial_r c = K\tau e^{-r\tau}N(d_-)$ *and* $\partial_r p = -K\tau e^{-r\tau}N(-d_-)$.

(vi) Strike: $\partial_K c = -e^{-r\tau}N(d_-)$ *and* $\partial_K p = e^{-r\tau}N(-d_-)$.

Proof. Since $d_+ - d_- = \sigma\sqrt{\tau}$ and $d_+ + d_- = 2\big(\log(x/K) + r\tau\big)/(\sigma\sqrt{\tau})$,

$$\frac{N'(d_-)}{N'(d_+)} = \exp\Big(\frac{d_+^2 - d_-^2}{2}\Big) = \exp\Big(\frac{(d_+ - d_-)(d_+ + d_-)}{2}\Big) = \frac{x}{K}\,e^{r\tau},$$

which is (6.6). Now let θ be any of the variables x, t, σ, r, K. By the chain rule,

$$\partial_\theta c = (\partial_\theta x)\,N(d_+) - \partial_\theta(Ke^{-r\tau})N(d_-) + x\,N'(d_+)\,\partial_\theta d_+ - Ke^{-r\tau}N'(d_-)\,\partial_\theta d_-,$$

and by (6.6) the last two terms add up to $xN'(d_+)\,\partial_\theta(d_+ - d_-) = xN'(d_+)\,\partial_\theta(\sigma\sqrt{\tau})$. Only this combination of σ and τ survives. For $\theta = x$, neither $Ke^{-r\tau}$ nor $\sigma\sqrt{\tau}$ depends on x, which gives $N(d_+)$; $\theta = \sigma$ gives $xN'(d_+)\sqrt{\tau}$; $\theta = r$ gives $K\tau e^{-r\tau}N(d_-)$; $\theta = K$ gives $-e^{-r\tau}N(d_-)$; and $\theta = t$, with $\partial_t\tau = -1$, gives $-rKe^{-r\tau}N(d_-) - xN'(d_+)\sigma/(2\sqrt{\tau})$. The gamma is $\partial_x N(d_+) = N'(d_+)\,\partial_x d_+ = N'(d_+)/(x\sigma\sqrt{\tau})$. For the put, differentiate the parity $p = c - x + Ke^{-r\tau}$ of Theorem 6.11: the delta loses 1, gamma and vega are unchanged, theta gains $rKe^{-r\tau}$, rho loses $\tau Ke^{-r\tau}$, ∂_K gains $e^{-r\tau}$, and $1 - N(y) = N(-y)$ gives the put formulas. □

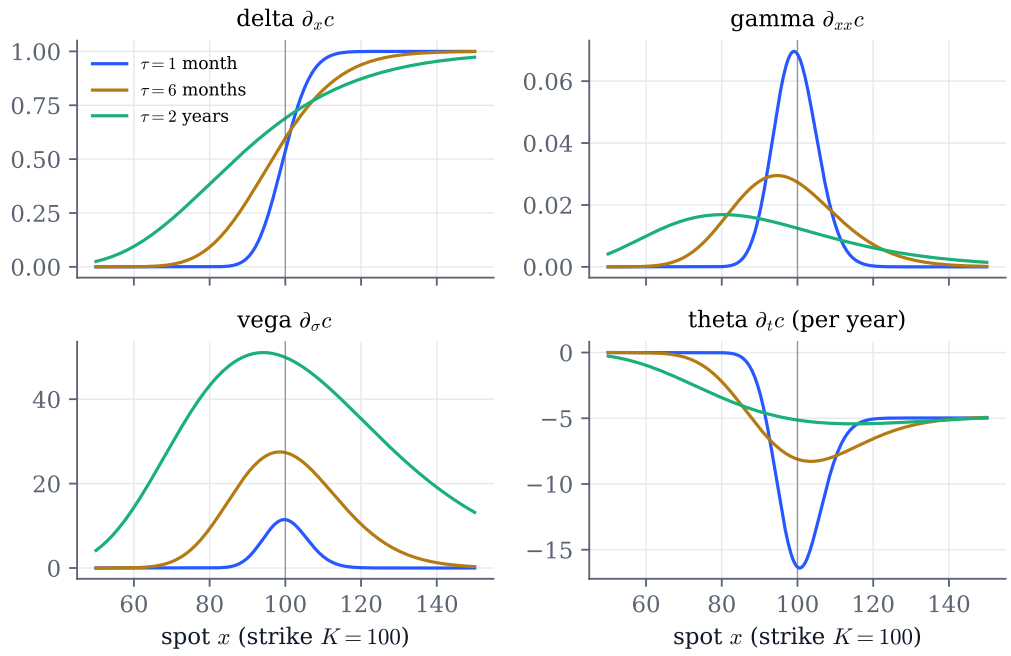

Figure 6.2: Greeks of a call with $K = 100$, $r = 5\%$, $\sigma = 20\%$, against the spot x, for three times to maturity. Near maturity the delta approaches the step $\mathbf{1}_{\{x>K\}}$, and the gamma and the theta concentrate around the strike.

Remark 6.15 (Reading the Greeks). (a) For every $r \in \mathbb{R}$: $0 < \partial_x c < 1$, the gamma is positive (the price is strictly convex in x), the vega is positive (the price strictly increases with σ), the call rho is positive and the put rho negative. If $r \geq 0$, the call theta is negative: with x fixed, the call loses value as time passes. If $r \leq 0$ the put theta is negative; if $r > 0$ it takes both signs: it tends to $rKe^{-r\tau} > 0$ as $x \downarrow 0$, and it is negative for x large, since by (6.6) it equals $Ke^{-r\tau}(rN(-d_-) - \sigma N'(d_-)/(2\sqrt{\tau}))$ and $N(-y) \leq N'(y)/y$ for $y > 0$ (because $\int_y^\infty N'(u)\,\mathrm{d}u \leq \int_y^\infty \frac{u}{y} N'(u)\,\mathrm{d}u$).

(b) At time 0, Theorem 3.18 says that in any model with a pricing measure the right derivative of the call price in the strike is $-B\,\mathbb{Q}(S_T > K)$, and the left one $-B\,\mathbb{Q}(S_T \geq K)$; here they agree because $\mathbb{Q}(S_T = K) = 0$. With $B = e^{-rT}$, $\partial_K c(0, S_0) = -e^{-rT} N(d_-)$ is indeed minus the digital price of Theorem 6.13. Differentiating once more, $\partial_{KK} c(0, S_0) = e^{-rT} N'(d_-)/(K\sigma\sqrt{T})$, which is e^{-rT} times the lognormal density of S_T at K (Theorem 6.2(i) with $\mu = r$), as in Theorem 3.18(ii).

(c) Two identities link the Greeks: $\partial_\sigma c = \sigma\tau x^2\, \partial_{xx} c$, and

$$\partial_t c + rx\, \partial_x c + \frac{\sigma^2 x^2}{2}\, \partial_{xx} c = r\, c, \tag{6.7}$$

both checked by substituting Theorem 6.14 (in the second, the terms in $N'(d_+)$ cancel and $rxN(d_+) - rKe^{-r\tau} N(d_-) = rc$ remains). The same holds for p, since $x - Ke^{-r\tau}$ satisfies (6.7). This is the Black–Scholes equation; in Chapter 10 it is derived from hedging, without knowing the formula.

(d) Near maturity (Figure 6.2), for $x \neq K$ the delta tends to $\mathbf{1}_{\{x>K\}}$ and the gamma to 0 as $\tau \downarrow 0$, while at $x = K$ the gamma behaves like $1/(K\sigma\sqrt{2\pi\tau})$ and blows up: close to maturity and near the strike, small moves of the stock change the hedge a lot.

(e) Units: t and τ are in years, so theta is a loss per year (divide by about 365 for a calendar day), and vega is per unit of σ (divide by 100 for one volatility point, from 20% to 21%).

▶ Interactive demo: the Greeks as surfaces over spot and maturity, for any rate and volatility; and the time-t price $c(t, S_t)$ along simulated paths, with the discounted price keeping a constant mean.

6.6 At the money

Options with strikes near the current price are among the most liquid, and there the formula can be done in one's head. The convenient reference is the *at-the-money-forward* strike $K = S_0 e^{rT}$, the forward price of Theorem 3.7. At this strike the call and the put have the same price, by parity: $C_0 - P_0 = S_0 - Ke^{-rT} = 0$.

Proposition 6.16 (At-the-money approximations). *Let $K = S_0 e^{rT}$ and write $v = \sigma\sqrt{T}$ for the total volatility. Then $d_\pm = \pm v/2$ and $C_0 = P_0 = S_0(2N(v/2) - 1)$. Moreover, with $1/\sqrt{2\pi} = 0.3989\ldots$,*

(i) price: $0 \leq \dfrac{S_0 v}{\sqrt{2\pi}} - C_0 \leq \dfrac{S_0 v^3}{24\sqrt{2\pi}}$;

(ii) delta: $0 \le \frac{1}{2} + \frac{v}{2\sqrt{2\pi}} - \partial_x c(0,S_0) \le \frac{v^3}{48\sqrt{2\pi}}$, *and the put delta is* $\partial_x c(0,S_0) - 1$*;*

(iii) vega: $0 \le \frac{S_0\sqrt{T}}{\sqrt{2\pi}} - \partial_\sigma c(0,S_0) \le \frac{S_0\sqrt{T}}{\sqrt{2\pi}}\,\frac{v^2}{8}$.

Hence $C_0 = P_0 \approx 0.4\,S_0\sigma\sqrt{T}$*, the call delta is about* $0.5 + 0.2\,\sigma\sqrt{T}$ *and the vega about* $0.4\,S_0\sqrt{T}$*, with relative errors of order* v^2 *in the price and the vega and an absolute error of order* v^3 *in the delta. Rounding* $1/\sqrt{2\pi}$ *to* 0.4 *adds a relative error of* 0.27% *to the terms in* v*.*

Proof. With $\log(S_0/K) = -rT$, (6.5) gives $d_\pm = \pm\frac{\sigma^2 T/2}{\sigma\sqrt{T}} = \pm v/2$, and Theorem 6.11 gives $C_0 = S_0(N(v/2) - N(-v/2)) = S_0(2N(v/2) - 1)$; the put has the same price by parity. All three bounds come from $1 - y^2/2 \le e^{-y^2/2} \le 1$, that is, $\frac{1}{\sqrt{2\pi}}(1 - \frac{y^2}{2}) \le N'(y) \le \frac{1}{\sqrt{2\pi}}$. For (i), $C_0/S_0 = \int_{-v/2}^{v/2} N'(y)\,\mathrm{d}y$ and $\int_{-v/2}^{v/2} \frac{y^2}{2}\,\mathrm{d}y = \frac{v^3}{24}$. For (ii), the delta is $N(v/2) = \frac{1}{2} + \int_0^{v/2} N'(y)\,\mathrm{d}y$ by Theorem 6.14, and $\int_0^{v/2} \frac{y^2}{2}\,\mathrm{d}y = \frac{v^3}{48}$. For (iii), the vega is $S_0 N'(v/2)\sqrt{T} = \frac{S_0\sqrt{T}}{\sqrt{2\pi}} e^{-v^2/8}$, and $0 \le 1 - e^{-v^2/8} \le v^2/8$. □

The approximation is due to Brenner and Subrahmanyam (1988). For a three-month option on a stock at $S_0 = 50$ with $\sigma = 30\%$, $v = 0.15$: the exact price is 2.9893, $S_0 v/\sqrt{2\pi}$ gives 2.9921 and the rule of thumb $0.4\,S_0 v$ gives 3.00; the delta is 0.52989 against 0.52992. The order v^3 in (i) cannot be improved: since $1 - e^{-a} \ge a - a^2/2$, the same computation bounds the error from below by $S_0(v^3/24 - v^5/640)/\sqrt{2\pi}$, and Figure 6.3 shows the error following the bound. For v below about 0.25 the rounding to 0.4 is the larger error.

Read backwards, (i) gives a volatility from a price: $\sigma \approx C_0/(0.4\,S_0\sqrt{T})$. This is a first approximation to the *implied volatility* of Chapter 15, which is what Brenner and Subrahmanyam (1988) proposed it for. The forward strike matters: at the spot strike $K = S_0$, parity gives $C_0 - P_0 = S_0(1 - e^{-rT}) \approx S_0 rT$, so the rule $0.4\,S_0\sigma\sqrt{T}$ misses at least one of the two prices by at least $\frac{1}{2}S_0|1 - e^{-rT}|$ (Exercise 6.1).

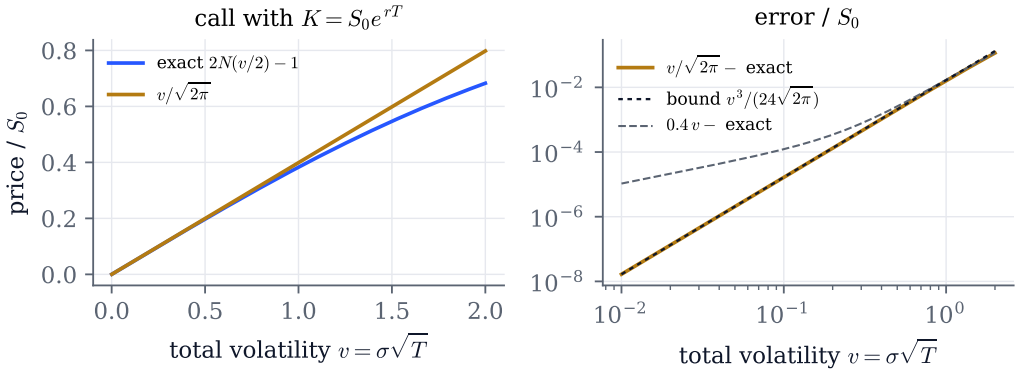

Figure 6.3: The call with strike $S_0 e^{rT}$. Left: exact price and the approximation $S_0 v/\sqrt{2\pi}$ against the total volatility $v = \sigma\sqrt{T}$. Right, on logarithmic scales: the error of $S_0 v/\sqrt{2\pi}$, the bound of Theorem 6.16(i), and the error of the rounded rule $0.4\,S_0 v$.

▶ Interactive demo: compare the exact at-the-money price, delta and vega with the rules of

thumb as the volatility and the maturity grow.

Further reading

The formula is due to Black and Scholes (1973) and Merton (1973); the binomial parametrization and its limit to Cox et al. (1979). Shreve (2004b) proves the lognormal limit for $r = 0$ in Section 3.2.7, with the trees $u_n = 1 + \sigma/\sqrt{n}$, $d_n = 1 - \sigma/\sqrt{n}$ of Theorem 6.5 (Exercise 3.8 adds a rate), derives the formula by the independence lemma in Section 5.2.5, and lists the Greeks in Section 4.5.5. Hull (2021) discusses the Greeks and their use on a trading desk. The rate of convergence of binomial prices, with its oscillations, is analysed by J. B. Walsh (*Finance and Stochastics*, 2003) and by F. Diener and M. Diener (*Mathematical Finance*, 2004).

Part III

Stochastic calculus

Chapter 7

The Itô integral

In discrete time, the gains of a trading strategy are a martingale transform (Chapter 2): a sum of positions times price changes, each position chosen before the price moves. This chapter takes the limit of such sums when the strategy is rebalanced continuously and the price is driven by a Brownian motion. The limit is the *Itô integral* $\int_0^T H_t \, \mathrm{d}W_t$. Brownian paths have infinite first variation (Theorem 5.17), so the integral cannot be defined path by path as a Riemann–Stieltjes integral. Itô's construction evaluates the integrand at the left end of each interval, as a trader must, and takes limits in mean square, where an identity between second moments, the *Itô isometry*, controls everything. We build the integral for simple integrands, then for square-integrable ones by approximation, and finally, by localization, for locally square-integrable ones.

Standing assumptions. W is a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{t\geq 0}$ is a filtration for W (Theorem 5.11), and $T > 0$ is a fixed horizon. We also assume that every event of probability zero belongs to $\mathcal{F}_0$.[1] For a partition $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ of $[0, t]$ we write $\|\Pi\| = \max_j (t_{j+1} - t_j)$ and $V_2(X; \Pi) = \sum_j (X_{t_{j+1}} - X_{t_j})^2$, as in Section 5.5.

7.1 Gains from trading

Consider a stock with price S_t and a money market that pays no interest; with a constant rate r, everything below applies to the discounted price $e^{-rt} S_t$, as in (4.2). A trader rebalances at the dates $0 = t_0 < t_1 < \cdots < t_n = T$. At t_j she trades so as to hold H_{t_j} shares until t_{j+1}, using the information available at t_j: H_{t_j} is $\mathcal{F}_{t_j}$-measurable. If no money is added or withdrawn after time 0, her wealth at the trading dates is

$$V_{t_k} = V_0 + \sum_{j=0}^{k-1} H_{t_j} (S_{t_{j+1}} - S_{t_j}),$$

the initial capital plus the martingale transform (Theorem 2.7) of the sequence $(S_{t_k})_k$ by the predictable sequence of positions. When $(S_{t_k})_k$ is a martingale and the positions are bounded,

[1] This costs nothing: the events that differ from an event of $\mathcal{F}_t$ by a null event form a σ-algebra, and replacing $\mathcal{F}_t$ by it changes no probability. It lets us change a process on a null event without losing adaptedness.

the gains form a martingale (Theorem 2.8): no strategy beats a fair game.

As the trader rebalances more and more often, the gains should converge to an integral $\int_0^T H_t \, \mathrm{d}S_t$. In the Bachelier model $S_t = S_0 + \sigma W_t$ the gains are $\sigma \sum_j H_{t_j}(W_{t_{j+1}} - W_{t_j})$; for geometric Brownian motion (Theorem 6.1), $\mathrm{d}S_t = \mu S_t \, \mathrm{d}t + \sigma S_t \, \mathrm{d}W_t$ (Theorem 8.8). Either way the new object is the limit of the *Itô sums*

$$\sum_{j=0}^{n-1} H_{t_j}(W_{t_{j+1}} - W_{t_j}), \qquad H_{t_j} \text{ known at time } t_j, \tag{7.1}$$

as the mesh of the partition goes to zero.

Two features of Brownian motion stand in the way of an ordinary integral: its paths have infinite first variation, so the sums need not converge path by path; and the evaluation point matters, since replacing H_{t_j} by $H_{t_{j+1}}$ changes the limit already for $H = W$ (Theorem 2.10, Theorem 7.11). Both come from the quadratic variation $[W, W]_t = t$. Itô's way around them is probabilistic. When H_{t_j} is known at time t_j and square integrable, the terms of (7.1) are orthogonal in L^2, and

$$\mathbb{E}\Big[\Big(\sum_j H_{t_j}(W_{t_{j+1}} - W_{t_j})\Big)^2\Big] = \mathbb{E}\Big[\sum_j H_{t_j}^2 \, (t_{j+1} - t_j)\Big], \tag{7.2}$$

as in the discrete isometry of Exercise 2.2. The right side is a Riemann sum of $\mathbb{E}\int_0^T H_t^2 \, \mathrm{d}t$, which does not see the roughness of the paths, and this identity lets us pass to the limit in mean square.

7.2 Simple integrands

Definition 7.1 (Simple integrands and their integral)**.** A process $H = (H_t)_{0 \le t \le T}$ is *simple* if there are a partition $0 = t_0 < t_1 < \cdots < t_n = T$ and random variables $\xi_0, \dots, \xi_{n-1}$ such that each ξ_j is $\mathcal{F}_{t_j}$-measurable with $\mathbb{E}[\xi_j^2] < \infty$, and

$$H_t = \xi_j \quad \text{for } t_j \le t < t_{j+1},\ 0 \le j \le n-1, \qquad H_T = \xi_{n-1}.$$

The *Itô integral* of H is the process

$$I_t = \int_0^t H_s \, \mathrm{d}W_s := \sum_{j=0}^{n-1} \xi_j (W_{t_{j+1} \wedge t} - W_{t_j \wedge t}), \qquad 0 \le t \le T. \tag{7.3}$$

For $t_k \le t \le t_{k+1}$ the sum reads $I_t = \sum_{j<k} \xi_j (W_{t_{j+1}} - W_{t_j}) + \xi_k (W_t - W_{t_k})$: the gains of a trader who holds ξ_j units of an asset with price W from t_j to t_{j+1}, with the open position marked to market at time t. At the dates t_k, I is the martingale transform of $(W_{t_k})_k$ by the predictable sequence of positions.

A simple process has many representations. Inserting a point s into (t_j, t_{j+1}), with the same ξ_j on both sides, does not change (7.3), because $W_{t_{j+1} \wedge t} - W_{t_j \wedge t} = (W_{s \wedge t} - W_{t_j \wedge t}) + (W_{t_{j+1} \wedge t} - W_{s \wedge t})$. Any two representations have a common refinement, so the integral does not depend on the representation. In the same way two simple processes H and K can be written on a common partition, so $aH + bK$ is simple for $a, b \in \mathbb{R}$, and $\int_0^t (aH_s + bK_s) \, \mathrm{d}W_s = a \int_0^t H_s \, \mathrm{d}W_s + b \int_0^t K_s \, \mathrm{d}W_s$.

Proposition 7.2 (Integrals of simple processes)**.** *Let H be simple and $I_t = \int_0^t H_s \, dW_s$.*

(i) $I_0 = 0$, the paths of I are continuous, and I is adapted with $\mathbb{E}[I_t^2] < \infty$ for every t.

(ii) I is a martingale. In particular $\mathbb{E}[I_t] = 0$.

(iii) (Itô isometry) *For $0 \le s \le t \le T$,*

$$\mathbb{E}\big[(I_t - I_s)^2 \,\big|\, \mathcal{F}_s\big] = \mathbb{E}\Big[\int_s^t H_u^2 \, du \,\Big|\, \mathcal{F}_s\Big]. \tag{7.4}$$

In particular $\mathbb{E}[I_t^2] = \mathbb{E}\int_0^t H_u^2 \, du$, and $I_t^2 - \int_0^t H_u^2 \, du$ is a martingale.

(iv) (Quadratic variation) *Let $0 < t \le T$, and let Π_m be partitions of $[0,t]$ with $\|\Pi_m\| \to 0$. Then $V_2(I;\Pi_m) \to \int_0^t H_u^2 \, du$ in probability.*

Proof. (i) Each term of (7.3) is continuous in t and vanishes at $t = 0$. Fix t and let k be the largest index with $t_k \le t$. The terms with $j > k$ vanish, and the others involve only $\xi_0, \dots, \xi_k$ and values of W at times at most t, all $\mathcal{F}_t$-measurable. For $j \le k$ the increment $D = W_{t_{j+1}\wedge t} - W_{t_j}$ is independent of $\mathcal{F}_{t_j}$, so $\mathbb{E}[\xi_j^2 D^2] = \mathbb{E}[\xi_j^2]\,\mathbb{E}[D^2] < \infty$, and I_t is square integrable.

(ii) and (iii). Fix $s \le t$. After refining the partition we may assume that $s = t_k$ and $t = t_m$ with $k \le m$, so that $I_t - I_s = \sum_{j=k}^{m-1} \xi_j D_j$ with $D_j = W_{t_{j+1}} - W_{t_j}$. Here ξ_j is $\mathcal{F}_{t_j}$-measurable, and D_j is independent of $\mathcal{F}_{t_j}$ with mean 0 and variance $h_j = t_{j+1} - t_j$. Taking out what is known and using independence (Theorem 1.16),

$$\mathbb{E}[\xi_j D_j \mid \mathcal{F}_{t_j}] = \xi_j \,\mathbb{E}[D_j] = 0, \qquad \mathbb{E}[\xi_j^2 D_j^2 \mid \mathcal{F}_{t_j}] = \xi_j^2 \, h_j,$$

and for $j < l$, since $\xi_j D_j \xi_l$ is $\mathcal{F}_{t_l}$-measurable, $\mathbb{E}[\xi_j D_j \, \xi_l D_l \mid \mathcal{F}_{t_l}] = \xi_j D_j \xi_l \,\mathbb{E}[D_l] = 0$. All these products are integrable: $\xi_j D_j \xi_l$ is a product of two square-integrable variables, and $|D_l|$ is independent of it. Conditioning further on $\mathcal{F}_s \subseteq \mathcal{F}_{t_j}$ (tower property), we get $\mathbb{E}[I_t - I_s \mid \mathcal{F}_s] = 0$, which is (ii), and

$$\mathbb{E}\big[(I_t - I_s)^2 \,\big|\, \mathcal{F}_s\big] = \mathbb{E}\Big[\sum_{j=k}^{m-1} \xi_j^2 h_j \,\Big|\, \mathcal{F}_s\Big] = \mathbb{E}\Big[\int_s^t H_u^2 \, du \,\Big|\, \mathcal{F}_s\Big],$$

which is (7.4). With $s = 0$, taking expectations gives $\mathbb{E}[I_t^2] = \mathbb{E}\int_0^t H_u^2 \, du$. Finally, since I is a martingale, $\mathbb{E}[(I_t - I_s)^2 \mid \mathcal{F}_s] = \mathbb{E}[I_t^2 \mid \mathcal{F}_s] - 2I_s\mathbb{E}[I_t \mid \mathcal{F}_s] + I_s^2 = \mathbb{E}[I_t^2 \mid \mathcal{F}_s] - I_s^2$, and $\int_0^s H_u^2 \, du = \sum_j \xi_j^2 (t_{j+1}\wedge s - t_j \wedge s)$ is $\mathcal{F}_s$-measurable, so (7.4) rearranges to $\mathbb{E}\big[I_t^2 - \int_0^t H_u^2 \, du \,\big|\, \mathcal{F}_s\big] = I_s^2 - \int_0^s H_u^2 \, du$.

(iv) Let $\Pi = \{0 = s_0 < \cdots < s_K = t\}$ be a partition of $[0,t]$ with mesh smaller than every $t_{j+1} - t_j$, and let Π' be Π together with the points $t_j < t$. An interval of Π then contains at most one t_j in its interior; if it does, splitting it at t_j changes the sum of squares by $2(I_{t_j} - I_{s_i})(I_{s_{i+1}} - I_{t_j})$. There are at most n such intervals, so

$$\big|V_2(I;\Pi) - V_2(I;\Pi')\big| \le 2n\,\omega(\|\Pi\|)^2, \qquad \omega(\delta) = \max\big\{|I_u - I_v| : u, v \in [0,t],\ |u - v| \le \delta\big\},$$

which tends to 0 with $\|\Pi\|$ for every outcome, because the paths of I are uniformly continuous on $[0,t]$. On an interval of Π' inside $[t_j, t_{j+1}]$ the increment of I is ξ_j times the increment of W, so

$$V_2(I;\Pi') = \sum_{j:\, t_j < t} \xi_j^2 \, V_2\big(W; \Pi' \cap [t_j, t_{j+1} \wedge t]\big).$$

By Theorem 5.16, applied to the Brownian motion $(W_{t_j+u} - W_{t_j})_{u\ge 0}$ (Theorem 5.9), each $V_2(W;\Pi' \cap [t_j, t_{j+1}\wedge t])$ converges in L^2, hence in probability, to $t_{j+1}\wedge t - t_j$. Multiplying by the finite random variable ξ_j^2 preserves convergence in probability,[2] and so does adding finitely many terms. Hence $V_2(I;\Pi_m) \to \sum_j \xi_j^2 (t_{j+1}\wedge t - t_j \wedge t) = \int_0^t H_u^2\,\mathrm{d}u$ in probability. □

7.3 Square-integrable integrands

Definition 7.3 (The spaces $\mathbb{H}^2$ and $\mathbb{H}^2_{\mathrm{loc}}$)**.** A process $H = (H_t)_{0\le t\le T}$ is *progressively measurable* if, for every $t \in [0,T]$, the map $(s,\omega) \mapsto H_s(\omega)$ on $[0,t]\times\Omega$ is measurable with respect to $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$. We write $\mathbb{H}^2 = \mathbb{H}^2[0,T]$ for the set of progressively measurable processes H with

$$\|H\|_{\mathbb{H}^2}^2 := \mathbb{E}\Big[\int_0^T H_t^2\,\mathrm{d}t\Big] < \infty,$$

and $\mathbb{H}^2_{\mathrm{loc}} = \mathbb{H}^2_{\mathrm{loc}}[0,T]$ for the set of progressively measurable processes with $\int_0^T H_t^2\,\mathrm{d}t < \infty$ almost surely.

Progressive measurability implies adaptedness (take $s = t$) and adds just enough joint measurability for integrals in time to be adapted: by Tonelli's theorem, $\int_0^t H_s^2\,\mathrm{d}s$ is then $\mathcal{F}_t$-measurable for every t, and so is $\int_0^t H_s\,\mathrm{d}s$ when $\int_0^t |H_s|\,\mathrm{d}s < \infty$. Every adapted process whose paths are all right-continuous, or all left-continuous, is progressively measurable (Karatzas and Shreve (1991), Proposition 1.1.13). So are simple processes, and $f(t, X_t)$ for a Borel function f and a progressively measurable X. $\|\cdot\|_{\mathbb{H}^2}$ is the norm of $L^2(\mathrm{d}t\otimes\mathrm{d}\mathbb{P})$ on $[0,T]\times\Omega$; processes H and K with $\|H-K\|_{\mathbb{H}^2} = 0$ are identified, and they will have the same integral. A simple process has $\|H\|_{\mathbb{H}^2}^2 = \sum_j \mathbb{E}[\xi_j^2](t_{j+1}-t_j)$, so it is in $\mathbb{H}^2$, and $\|W\|_{\mathbb{H}^2}^2 = \int_0^T t\,\mathrm{d}t = T^2/2$. Since $\mathbb{E}[e^{2W_t^2}] = (1-4t)^{-1/2}$ for $t < \frac14$ and $= \infty$ for $t \ge \frac14$, $e^{W_t^2}$ is in $\mathbb{H}^2[0,T]$ exactly when $T \le \frac14$, but in $\mathbb{H}^2_{\mathrm{loc}}[0,T]$ for every T (its paths are continuous).

Lemma 7.4 (Left-point approximation)**.** *Let H be progressively measurable, with $\mathbb{E}[H_t^2] < \infty$ for every $t \in [0,T]$ and continuous in mean square: $\mathbb{E}[(H_u - H_t)^2] \to 0$ as $u \to t$, for every $t \in [0,T]$. For a partition $\Pi = \{t_j\}$ of $[0,T]$, let H^Π be the simple process equal to H_{t_j} on $[t_j, t_{j+1})$, with $H_T^\Pi = H_{t_{n-1}}$. Then $H \in \mathbb{H}^2$, and $\|H^\Pi - H\|_{\mathbb{H}^2} \to 0$ as $\|\Pi\| \to 0$.*

Proof. The map $t \mapsto H_t$ from $[0,T]$ to L^2 is continuous on a compact interval, hence bounded, $c = \sup_t \mathbb{E}[H_t^2] < \infty$, and uniformly continuous: $\rho(\delta) = \sup\{\mathbb{E}[(H_u - H_v)^2] : |u-v| \le \delta\} \to 0$ as $\delta \to 0$. By Tonelli's theorem $\|H\|_{\mathbb{H}^2}^2 = \int_0^T \mathbb{E}[H_t^2]\,\mathrm{d}t \le cT$. Each H_{t_j} is $\mathcal{F}_{t_j}$-measurable and square integrable, so H^Π is simple, and by Tonelli again

$$\|H^\Pi - H\|_{\mathbb{H}^2}^2 = \sum_j \int_{t_j}^{t_{j+1}} \mathbb{E}\big[(H_{t_j} - H_t)^2\big]\,\mathrm{d}t \le T\rho(\|\Pi\|) \longrightarrow 0. \qquad \square$$

The lemma applies to W, with $\mathbb{E}[(W_u - W_t)^2] = |u-t|$ and $\|W^\Pi - W\|_{\mathbb{H}^2}^2 = \sum_j \int_{t_j}^{t_{j+1}} (t - t_j)\,\mathrm{d}t = \frac12 \sum_j (t_{j+1}-t_j)^2$; to deterministic continuous functions; and, by dominated convergence, to adapted processes with continuous paths and $\mathbb{E}[\sup_{t\le T} H_t^2] < \infty$.

[2] If $X_m \to c$ in probability and ξ is a finite random variable, then $\mathbb{P}(|\xi X_m - \xi c| > \varepsilon) \le \mathbb{P}(|\xi| > K) + \mathbb{P}(|X_m - c| > \varepsilon/K)$ for all $K, \varepsilon > 0$; let $m \to \infty$, then $K \to \infty$.

Lemma 7.5 (Simple processes are dense)**.** *For every $H \in \mathbb{H}^2$ there are simple processes H^n with $\|H^n - H\|_{\mathbb{H}^2} \to 0$.*

This is Øksendal (2003), Section 3.1, Steps 1–3, for integrands that are adapted and jointly measurable, as progressively measurable ones are. The proof truncates H, replaces it by its averages $n\int_{(t-1/n)\vee 0}^{t} H_s\,\mathrm{d}s$ over the recent past, which are continuous and adapted (by progressive measurability), and applies Theorem 7.4; Exercise 7.6 gives the details.

Lemma 7.6 (Doob's maximal inequality)**.** *If M is a martingale with continuous paths on $[0,T]$ and $\mathbb{E}[M_T^2] < \infty$, then $\mathbb{E}\big[\sup_{0\le t\le T} M_t^2\big] \le 4\,\mathbb{E}[M_T^2]$.*

This is Karatzas and Shreve (1991), Theorem 1.3.8(iv), applied to the nonnegative submartingale $|M|$ (a submartingale by conditional Jensen, Theorem 1.16). Processes X and Y are *indistinguishable* if $\mathbb{P}(X_t = Y_t \text{ for all } t) = 1$; for continuous processes it suffices that $\mathbb{P}(X_t = Y_t) = 1$ for every rational t.

Theorem 7.7 (The Itô integral)**.** *Let $H \in \mathbb{H}^2$. There is a process $I = (I_t)_{0\le t\le T}$ with continuous paths such that, for every simple process G,*

$$\mathbb{E}\Big[\sup_{0\le t\le T}\Big(I_t - \int_0^t G_s\,\mathrm{d}W_s\Big)^2\Big] \le 4\,\|H - G\|_{\mathbb{H}^2}^2. \tag{7.5}$$

It is unique up to indistinguishability, and for simple H it is the integral of Theorem 7.1. It is called the Itô integral *of H and written $I_t = \int_0^t H_s\,\mathrm{d}W_s$. In particular, if simple processes H^n converge to H in $\mathbb{H}^2$ (Theorem 7.5), their integrals converge to I in mean square, uniformly in t. Moreover, for $H, K \in \mathbb{H}^2$, $a, b \in \mathbb{R}$ and $0 \le s \le t \le T$:*

(i) (continuity) *$I_0 = 0$ and the paths of I are continuous;*

(ii) (adaptedness) *I_t is $\mathcal{F}_t$-measurable;*

(iii) (linearity) *$\int_0^t (aH_u + bK_u)\,\mathrm{d}W_u = a\int_0^t H_u\,\mathrm{d}W_u + b\int_0^t K_u\,\mathrm{d}W_u$;*

(iv) (martingale) *I is a martingale, so is $I_t^2 - \int_0^t H_u^2\,\mathrm{d}u$, and the conditional isometry (7.4) holds;*

(v) (Itô isometry) *$\mathbb{E}[I_t^2] = \mathbb{E}\int_0^t H_u^2\,\mathrm{d}u$, and $\mathbb{E}\big[\int_0^t H_u\,\mathrm{d}W_u \int_0^t K_u\,\mathrm{d}W_u\big] = \mathbb{E}\int_0^t H_u K_u\,\mathrm{d}u$;*

(vi) (quadratic variation) *if $t > 0$ and Π_m are partitions of $[0,t]$ with $\|\Pi_m\| \to 0$, then $V_2(I;\Pi_m) \to \int_0^t H_u^2\,\mathrm{d}u$ in probability. In the notation of Theorem 7.12, $[I,I]_t = \int_0^t H_u^2\,\mathrm{d}u$.*

Proof. Construction. By Theorem 7.5 there are simple H^n with $\|H^n - H\|_{\mathbb{H}^2} \to 0$; let $I_t^n = \int_0^t H_u^n\,\mathrm{d}W_u$. For simple G and G', the difference of their integrals is the integral of $G - G'$, a martingale with continuous paths, so Theorem 7.6 and the isometry of Theorem 7.2 give

$$\mathbb{E}\Big[\sup_{t\le T}\Big(\int_0^t G_u\,\mathrm{d}W_u - \int_0^t G'_u\,\mathrm{d}W_u\Big)^2\Big] \le 4\,\|G - G'\|_{\mathbb{H}^2}^2. \tag{7.6}$$

The sequence (H^n) is Cauchy in $\mathbb{H}^2$, so there are $n_1 < n_2 < \cdots$ with $\|H^n - H^m\| \le 4^{-k}$ for $n, m \ge n_k$. By (7.6) and Chebyshev's inequality, $\mathbb{P}\big(\sup_t |I_t^{n_{k+1}} - I_t^{n_k}| > 2^{-k}\big) \le 4\cdot 16^{-k}\cdot 4^k = 4^{1-k}$, which is summable. By the Borel–Cantelli lemma, almost surely $\sup_t |I_t^{n_{k+1}} - I_t^{n_k}| \le 2^{-k}$ for all large k, and then I^{n_k} converges uniformly on $[0,T]$. Let Ω_0 be the event where it does, and set $I = \lim_k I^{n_k}$ on Ω_0 and $I = 0$ off Ω_0. Uniform limits of continuous functions are continuous,

so I has continuous paths and $I_0 = 0$: this is (i). Since Ω_0^c is a null event, it is in $\mathcal{F}_0$, and $I_t = \mathbf{1}_{\Omega_0} \limsup_k I_t^{n_k}$ is $\mathcal{F}_t$-measurable: this is (ii).

The bound (7.5), uniqueness. Fix n. On Ω_0, $\sup_t (I_t - I_t^n)^2 = \lim_k \sup_t (I_t^{n_k} - I_t^n)^2$, so Fatou's lemma and (7.6) give $\mathbb{E}[\sup_t (I_t - I_t^n)^2] \le 4 \liminf_k \|H^{n_k} - H^n\|^2 = 4\|H - H^n\|^2$. For a simple G, the triangle inequality in L^2 and (7.6) give

$$\mathbb{E}\Big[\sup_t \Big(I_t - \int_0^t G_u \, dW_u\Big)^2\Big]^{1/2} \le 2\|H - H^n\| + 2\|H^n - G\| \xrightarrow[n\to\infty]{} 2\|H - G\|,$$

which is (7.5). If $\tilde{I}$ also satisfies (7.5), the same argument gives $\mathbb{E}[\sup_t (I_t - \tilde{I}_t)^2]^{1/2} \le 4\|H - H^n\| \to 0$, so I and $\tilde{I}$ are indistinguishable. For simple H, (7.6) shows that the integral of Theorem 7.1 satisfies (7.5).

(iii) Let K^n be simple with $K^n \to K$, and J the integral of K. The integrals of the simple processes $aH^n + bK^n$ are $aI^n + bJ^n$. They converge to the integral of $aH + bK$ in the sense of (7.5), and to $aI + bJ$; the two limits are indistinguishable.

(iv) By (7.5), $I_t^n \to I_t$ in L^2 for every t. Conditional expectation does not increase the L^2 norm (Theorem 1.21), so $I_s^n = \mathbb{E}[I_t^n \mid \mathcal{F}_s] \to \mathbb{E}[I_t \mid \mathcal{F}_s]$ in L^2, while $I_s^n \to I_s$. Hence $\mathbb{E}[I_t \mid \mathcal{F}_s] = I_s$, and I_t is square integrable. Next, by the Cauchy–Schwarz inequality,

$$\mathbb{E}|(I_t^n)^2 - I_t^2| \le \|I_t^n - I_t\|_{L^2} \, \|I_t^n + I_t\|_{L^2} \to 0,$$
$$\mathbb{E}\Big|\int_0^t ((H_u^n)^2 - H_u^2)\, du\Big| \le \|H^n - H\|_{\mathbb{H}^2} \, \|H^n + H\|_{\mathbb{H}^2} \to 0.$$

Conditional expectation does not increase the L^1 norm (Theorem 1.15), and $\int_0^t H_u^2 \, du$ is $\mathcal{F}_t$-measurable, so the martingale property of $(I_t^n)^2 - \int_0^t (H_u^n)^2 \, du$ (Theorem 7.2) passes to the limit. Expanding $(I_t - I_s)^2$ as in the proof of Theorem 7.2 gives (7.4).

(v) The martingale $I_t^2 - \int_0^t H_u^2 \, du$ starts at 0, so its mean is 0. For the second identity, write $4IJ = (I + J)^2 - (I - J)^2$, where $I \pm J$ is the integral of $H \pm K$ by (iii), and apply the first one.

(vi) Write $Q_m(X) = V_2(X; \Pi_m)$. If M is a martingale with $M_0 = 0$ and $\mathbb{E}[M_t^2] < \infty$, its increments over disjoint intervals are orthogonal: for $a < b \le c < d$, $\mathbb{E}[(M_b - M_a)(M_d - M_c)] = \mathbb{E}\big[(M_b - M_a)\,\mathbb{E}[M_d - M_c \mid \mathcal{F}_c]\big] = 0$. Hence $\mathbb{E}[Q_m(M)] = \mathbb{E}[M_t^2]$ for every partition. Apply this to $M = I - I^n$, the integral of $H - H^n$, and to I^n: with $\varepsilon_n = \|H - H^n\|$, $\mathbb{E}[Q_m(I - I^n)] \le \varepsilon_n^2$ and $\mathbb{E}[Q_m(I^n)] \le \|H^n\|^2$. If a_i and b_i denote the increments of I and I^n over the intervals of Π_m, the Cauchy–Schwarz and triangle inequalities in $\mathbb{R}^K$ give

$$|Q_m(I) - Q_m(I^n)| = \Big|\sum_i (a_i - b_i)(a_i + b_i)\Big| \le Q_m(I - I^n)^{1/2}\big(Q_m(I - I^n)^{1/2} + 2Q_m(I^n)^{1/2}\big),$$

and taking expectations, with Cauchy–Schwarz once more, $\mathbb{E}|Q_m(I) - Q_m(I^n)| \le \varepsilon_n^2 + 2\varepsilon_n \|H^n\|$, for every m. Also $\mathbb{E}\big|\int_0^t (H_u^n)^2 \, du - \int_0^t H_u^2 \, du\big| \le \varepsilon_n \|H^n + H\|$. For $\delta > 0$, Markov's inequality gives

$$\mathbb{P}\Big(\Big|Q_m(I) - \int_0^t H_u^2 \, du\Big| > 3\delta\Big) \le \frac{\varepsilon_n^2 + 2\varepsilon_n\|H^n\| + \varepsilon_n\|H^n + H\|}{\delta} + \mathbb{P}\Big(\Big|Q_m(I^n) - \int_0^t (H_u^n)^2 \, du\Big| > \delta\Big).$$

As $m \to \infty$ the last term tends to 0 by Theorem 7.2(iv). Then let $n \to \infty$. □

Remark 7.8 (Notation). For $0 \le s \le t \le T$ we write $\int_s^t H_u \, \mathrm{d}W_u = I_t - I_s$. This is also the integral over $[0,T]$ of $H\mathbf{1}_{[s,t)}$: the identity holds for simple H, after adding s and t to the partition, and passes to the limit by (7.5), since $\|H^n\mathbf{1}_{[s,t)} - H\mathbf{1}_{[s,t)}\| \le \|H^n - H\|$. In the same way the horizon does not matter: for $t' \le T$, the integral of H restricted to $[0,t']$, constructed on the horizon t', is indistinguishable from $(I_s)_{s \le t'}$. The *differential notation* $\mathrm{d}I_t = H_t \, \mathrm{d}W_t$ is shorthand for $I_t = I_0 + \int_0^t H_u \, \mathrm{d}W_u$: over a short interval, I moves by H, taken at the start, times the move of W.

For integrands that are continuous in mean square, the integral is the limit of the Itô sums (7.1), as the trading story suggests.

Corollary 7.9 (Itô sums converge)**.** *Let H be as in Theorem 7.4. Along partitions $\Pi = \{t_j\}$ of $[0,T]$ with $\|\Pi\| \to 0$,*

$$\mathbb{E}\Big[\sup_{0 \le t \le T} \Big(\sum_j H_{t_j}\big(W_{t_{j+1}\wedge t} - W_{t_j \wedge t}\big) - \int_0^t H_u \, \mathrm{d}W_u\Big)^2\Big] \le 4\,\|H^\Pi - H\|_{\mathbb{H}^2}^2 \longrightarrow 0.$$

Proof. The sum is $\int_0^t H_u^\Pi \, \mathrm{d}W_u$; apply (7.5) with $G = H^\Pi$, and Theorem 7.4. □

Some regularity is needed: $H_t = \mathbf{1}_{\mathbb{Q}}(t)$ has $\|H\|_{\mathbb{H}^2} = 0$ and integral 0, yet its Itô sums along partitions with rational points equal W_T.

7.4 Deterministic integrands

When the integrand is a deterministic function of time, the integral inherits the Gaussian structure of W. Such integrals are sometimes called *Wiener integrals*.

Proposition 7.10 (Deterministic integrands)**.** *Let $h : [0,T] \to \mathbb{R}$ be Borel measurable with $\int_0^T h(u)^2 \, \mathrm{d}u < \infty$, and $I_t = \int_0^t h(u) \, \mathrm{d}W_u$.*

(i) For $0 \le s < t \le T$, the increment $I_t - I_s$ is independent of $\mathcal{F}_s$ and has law $\mathcal{N}(0, \int_s^t h(u)^2 \, \mathrm{d}u)$. In particular, I has independent increments.

(ii) For $0 \le t_1 < \cdots < t_m \le T$, the vector $(I_{t_1}, \dots, I_{t_m})$ is centred Gaussian, with $\mathrm{Cov}(I_s, I_t) = \int_0^{s\wedge t} h(u)^2 \, \mathrm{d}u$.

(iii) $[I,I]_t = \int_0^t h(u)^2 \, \mathrm{d}u = \mathrm{Var}(I_t)$: the quadratic variation is deterministic and equal to the variance.

Proof. As a process, h does not depend on ω, so it is progressively measurable, and it is in $\mathbb{H}^2$.

Step functions. By Theorem 7.5 there are simple K^n with $K^n \to h$ in $\mathbb{H}^2$. Then $h^n(t) = \mathbb{E}[K_t^n]$ is a deterministic step function, and by Jensen's inequality $\int_0^T (h^n - h)^2 \, \mathrm{d}t \le \|K^n - h\|_{\mathbb{H}^2}^2 \to 0$. Let $I^n = \int h^n \, \mathrm{d}W$. Fix $s < t$, refine the partition of h^n to contain s and t, and write $I_t^n - I_s^n = \sum_{j=1}^k c_j D_j$, with $D_j = W_{u_j} - W_{u_{j-1}}$ for $s = u_0 < u_1 < \cdots < u_k = t$. Let $A \in \mathcal{F}_s$ and $\theta \in \mathbb{R}$. Conditioning on $\mathcal{F}_{u_{k-1}}$, of which $\mathbf{1}_A$ and $D_1, \dots, D_{k-1}$ are measurable while D_k is independent of it, gives $\mathbb{E}\big[\mathbf{1}_A e^{i\theta \sum_{j\le k} c_j D_j}\big] = \mathbb{E}\big[\mathbf{1}_A e^{i\theta \sum_{j<k} c_j D_j}\big]\, e^{-\theta^2 c_k^2 (u_k - u_{k-1})/2}$, and repeating,

$$\mathbb{E}\big[\mathbf{1}_A e^{i\theta(I_t^n - I_s^n)}\big] = \mathbb{P}(A)\, e^{-\theta^2 v_n/2}, \qquad v_n = \int_s^t h^n(u)^2 \, \mathrm{d}u. \tag{7.7}$$

The limit. By (7.5), $I_t^n - I_s^n \to I_t - I_s$ in L^2, and since $|e^{ix} - e^{iy}| \le |x - y|$ the left side of (7.7) converges to $\mathbb{E}[\mathbf{1}_A e^{i\theta(I_t - I_s)}]$; on the right, $v_n \to v = \int_s^t h^2 \, \mathrm{d}u$. With $A = \Omega$, this shows that $I_t - I_s$ has the characteristic function of $\mathcal{N}(0, v)$. For A with $\mathbb{P}(A) > 0$, it shows that $I_t - I_s$ has the same characteristic function, hence the same law, under $\mathbb{P}(\cdot \mid A)$ as under $\mathbb{P}$: $I_t - I_s$ is independent of $\mathcal{F}_s$. Given $0 = t_0 < t_1 < \cdots < t_m$, each increment $I_{t_k} - I_{t_{k-1}}$ is independent of $\mathcal{F}_{t_{k-1}}$, which contains the earlier increments, so the increments are independent. This proves (i). Independent Gaussian variables form a Gaussian vector, and $(I_{t_1}, \dots, I_{t_m})$ is a linear transformation of the vector of increments, hence centred Gaussian. For $s \le t$, by (i), $\mathrm{Cov}(I_s, I_t) = \mathrm{Var}(I_s) + \mathrm{Cov}(I_s, I_t - I_s) = \int_0^s h^2 \, \mathrm{d}u$. This proves (ii), and (iii) is Theorem 7.7(vi) with (i) for $s = 0$. □

In the Bachelier model with zero interest rate and volatility $\sigma(t)$, the price under the pricing measure is $S_T = S_0 + \int_0^T \sigma(u) \, \mathrm{d}W_u$, normal with variance $\int_0^T \sigma(u)^2 \, \mathrm{d}u$: European prices depend on $\sigma(\cdot)$ only through this total variance. Vasicek's short-rate model is built from integrals $\int_0^t e^{-a(t-u)} \, \mathrm{d}W_u$ (Chapter 9). A random integrand need not give a Gaussian integral (Theorem 7.11), although it may (Exercise 7.5).

7.5 The integral of W against itself

For a continuously differentiable g with $g(0) = 0$, ordinary calculus gives $\int_0^T g(t) \, \mathrm{d}g(t) = \int_0^T g(t) g'(t) \, \mathrm{d}t = \frac{1}{2} g(T)^2$. For Brownian motion the answer depends on where the integrand is evaluated, and the Itô integral picks up a correction.

Proposition 7.11 ($\int_0^T W \, \mathrm{d}W$)**.**

(i) $W \in \mathbb{H}^2$ *and, almost surely,* $\displaystyle\int_0^t W_s \, \mathrm{d}W_s = \frac{1}{2}(W_t^2 - t)$ *for every* $t \in [0, T]$.

(ii) Let $\lambda \in [0, 1]$. *For a partition* $\Pi = \{t_j\}$ *of* $[0, T]$, *let* $S_\lambda(\Pi) = \sum_j \big((1 - \lambda) W_{t_j} + \lambda W_{t_{j+1}}\big)(W_{t_{j+1}} - W_{t_j})$. *Then, for every partition,*

$$S_\lambda(\Pi) = \frac{1}{2} W_T^2 + \Big(\lambda - \frac{1}{2}\Big) V_2(W; \Pi), \tag{7.8}$$

and as $\|\Pi\| \to 0$, $S_\lambda(\Pi) \to \frac{1}{2} W_T^2 + (\lambda - \frac{1}{2}) T$ *in* L^2, *with* $\mathbb{E}\big[\big(S_\lambda(\Pi) - \frac{1}{2} W_T^2 - (\lambda - \frac{1}{2}) T\big)^2\big] = 2(\lambda - \frac{1}{2})^2 \sum_j (t_{j+1} - t_j)^2$.

(iii) The midpoint sums $M(\Pi) = \sum_j W_{(t_j + t_{j+1})/2} (W_{t_{j+1}} - W_{t_j})$ *converge in* L^2 *to* $\frac{1}{2} W_T^2$.

Proof. (ii) For real numbers a and b, $((1 - \lambda)a + \lambda b)(b - a) = \frac{1}{2}(b^2 - a^2) + (\lambda - \frac{1}{2})(b - a)^2$. Summing over the intervals of Π, the first terms telescope to $\frac{1}{2}(W_T^2 - W_0^2) = \frac{1}{2} W_T^2$, which gives (7.8). The error is $(\lambda - \frac{1}{2})(V_2(W; \Pi) - T)$, and (5.2) computes its second moment.

(i) $W \in \mathbb{H}^2$ by Theorem 7.4. Fix $t \in (0, T]$ and apply (ii) on $[0, t]$: along partitions of $[0, t]$, the left-point sums S_0 converge in L^2 to $\frac{1}{2}(W_t^2 - t)$. Extend each partition of $[0, t]$ to one of $[0, T]$ by points of $[t, T]$ with gaps at most its mesh; at time t the same sums are the Itô sums of Theorem 7.9, which converge in L^2 to $\int_0^t W_s \, \mathrm{d}W_s$. Hence $\int_0^t W_s \, \mathrm{d}W_s = \frac{1}{2}(W_t^2 - t)$ almost surely, for each t. Both sides are continuous in t, so they are indistinguishable.

(iii) Let $m_j = (t_j + t_{j+1})/2$, $A_j = W_{m_j} - W_{t_j}$ and $B_j = W_{t_{j+1}} - W_{m_j}$, so that $W_{m_j}(W_{t_{j+1}} - W_{t_j}) = W_{t_j}(W_{t_{j+1}} - W_{t_j}) + A_j^2 + A_j B_j$ and $M(\Pi) = S_0(\Pi) + \sum_j A_j^2 + \sum_j A_j B_j$. The variables A_j, B_j are independent $\mathcal{N}(0, h_j/2)$, with $h_j = t_{j+1} - t_j$. The sum $\sum_j A_j^2$ has mean $T/2$ and variance $\sum_j 2(h_j/2)^2$, and the terms $A_j B_j$ are orthogonal with mean 0 and variance $h_j^2/4$, so in L^2, $\sum_j A_j^2 \to T/2$ and $\sum_j A_j B_j \to 0$. With (ii) for $\lambda = 0$, $M(\Pi) \to \frac{1}{2}W_T^2 - \frac{T}{2} + \frac{T}{2} = \frac{1}{2}W_T^2$. □

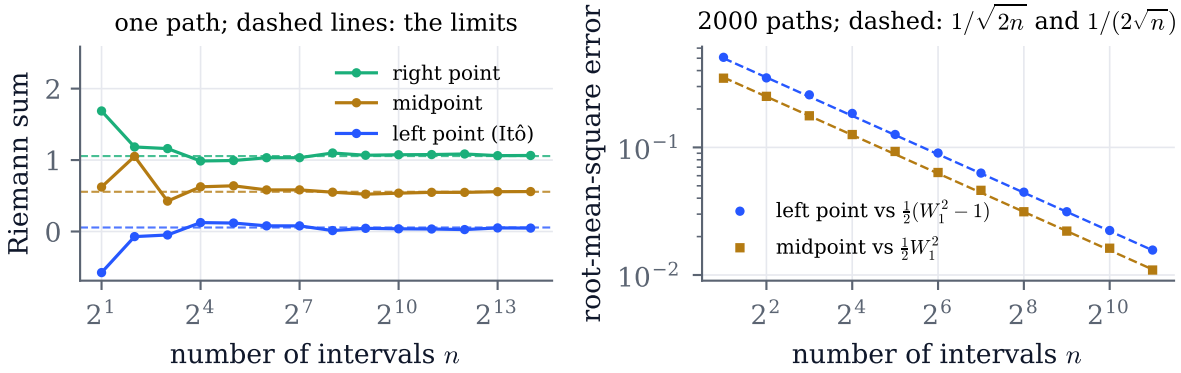

Figure 7.1: Riemann sums for $\int_0^1 W \, \mathrm{d}W$ on uniform partitions with n intervals. Left: one path; the left-point, midpoint and right-point sums settle at $\frac{1}{2}(W_1^2 - 1)$, $\frac{1}{2}W_1^2$ and $\frac{1}{2}(W_1^2 + 1)$ (Theorem 7.11). Right: the root-mean-square errors of the left-point and midpoint sums over 2000 paths, against the exact values $1/\sqrt{2n}$ and $1/(2\sqrt{n})$ from the proof. Computed with `code/src/ieor222/itointegral.py`.

Three limits from the same path (Figure 7.1). The Itô integral ($\lambda = 0$), $\frac{1}{2}(W_T^2 - T)$, is a martingale with mean 0, in line with Theorem 5.12. It is the gain of the strategy that holds W_t units of an asset with price W_t, and it is not Gaussian: it is bounded below by $-T/2$. The right-point limit $\frac{1}{2}(W_T^2 + T)$ has mean $T > 0$: it is the gain of a trader who knows the price at the end of each period, the clairvoyance that Theorem 2.8 excludes. The trapezoid sums equal $\frac{1}{2}W_T^2$ exactly, and the midpoint sums have the same limit, the answer of ordinary calculus: the *Stratonovich integral* $\int_0^T W \circ \mathrm{d}W$, with mean $T/2$, not a martingale. The left point is the one that matches trading, and the one used throughout these notes.

The gap between the right-point and left-point limits is $[W, W]_T = T$, as for the random walk in Theorem 2.10. Rearranged, (i) reads $W_T^2 = 2\int_0^T W_t \, \mathrm{d}W_t + [W, W]_T$, while for a continuously differentiable g with $g(0) = 0$, $g(T)^2 = 2\int_0^T g \, \mathrm{d}g$ because $[g, g]_T = 0$ (Theorem 5.15). The extra term is the first instance of Itô's formula (Chapter 8).

▶ Interactive demo: left-point, midpoint and right-point sums for $\int W \, \mathrm{d}W$, for any evaluation point λ, on one path and on average; only the left point gives a martingale.

7.6 Quadratic variation and covariation

Theorem 5.16 computed the quadratic variation of W, and Theorem 7.7 that of an Itô integral. We now fix the general definition, which also covers pairs of processes.

Definition 7.12 (Quadratic variation and covariation)**.** Let X and Y be processes with continuous paths and $t > 0$. For a partition $\Pi = \{0 = s_0 < \cdots < s_K = t\}$ of $[0, t]$, let

$$C(X, Y; \Pi) = \sum_{i=0}^{K-1} (X_{s_{i+1}} - X_{s_i})(Y_{s_{i+1}} - Y_{s_i}).$$

If $C(X, Y; \Pi_m)$ converges in probability for every sequence of partitions Π_m of $[0, t]$ with $\|\Pi_m\| \to 0$, the limit is the *covariation* $[X, Y]_t$ of X and Y on $[0, t]$. (It does not depend on the sequence: two sequences can be merged into one by alternating between them.) The *quadratic variation* is $[X, X]_t$, the limit of $C(X, X; \Pi_m) = V_2(X; \Pi_m)$. We set $[X, Y]_0 = 0$.

By Theorem 5.16, $[W, W]_t = t$, and by Theorem 7.7(vi), $[I, I]_t = \int_0^t H_u^2 \, du$ for $I = \int H \, dW$ with $H \in \mathbb{H}^2$. The sums $C(X, Y; \Pi)$ are symmetric and bilinear in (X, Y), and limits in probability are linear, so the covariation is symmetric and bilinear wherever it exists. In particular, polarization gives

$$[X, Y]_t = \tfrac{1}{4}([X + Y, X + Y]_t - [X - Y, X - Y]_t)$$

whenever the two quadratic variations on the right exist.

Lemma 7.13 (Finite variation)**.** *Let A be a process with continuous paths of finite total variation on $[0, t]$ (for instance $A_s = \int_0^s \Theta_u \, du$ with $\int_0^t |\Theta_u| \, du < \infty$), and X a process with continuous paths. Then $[X, A]_t = 0$, with almost sure convergence; in particular $[A, A]_t = 0$. Consequently, if $[X, X]_t$ exists, then $[X + A, X + A]_t = [X, X]_t$.*

Proof. For every outcome, $|C(X, A; \Pi)| \le \max_i |X_{s_{i+1}} - X_{s_i}| \sum_i |A_{s_{i+1}} - A_{s_i}|$. The sum is at most the total variation of the path of A (at most $\int_0^t |\Theta_u| \, du$ in the example), and the maximum tends to 0 with $\|\Pi\|$ by uniform continuity. The last claim follows from $V_2(X + A; \Pi) = V_2(X; \Pi) + 2C(X, A; \Pi) + V_2(A; \Pi)$. □

With $X = W$ and $A_s = s$, this gives the rules $dW \, dt = dt \, dt = 0$ of Section 5.5.

Proposition 7.14 (Covariations of Itô integrals)**.**

(i) If $H, K \in \mathbb{H}^2$, $I = \int H \, dW$ and $J = \int K \, dW$, then $[I, J]_t = \int_0^t H_u K_u \, du$ for $t \in [0, T]$.

(ii) If W and B are independent Brownian motions, then $[W, B]_t = 0$ for every t, and $\mathbb{E}[C(W, B; \Pi)^2] = \sum_i (s_{i+1} - s_i)^2 \le t\|\Pi\|$.

Proof. (i) By linearity, $I \pm J = \int (H \pm K) \, dW$, whose quadratic variations are $\int_0^t (H \pm K)^2 \, du$ by Theorem 7.7(vi). Polarize: $\frac{1}{4}((H + K)^2 - (H - K)^2) = HK$.

(ii) The increments $(W_{s_{i+1}} - W_{s_i})_i$ are independent of each other and of the increments $(B_{s_{i+1}} - B_{s_i})_i$ (Theorem 5.8), so the products $(W_{s_{i+1}} - W_{s_i})(B_{s_{i+1}} - B_{s_i})$ are independent, with mean 0 and variance $(s_{i+1} - s_i)^2$. The variance of their sum is the sum of the variances. □

In differential notation, $dI_t \, dJ_t = H_t K_t \, dt$ and $dW_t \, dB_t = 0$.

Quadratic variation is not variance. The isometry says $\mathbb{E}[I_t^2] = \mathbb{E}[[I, I]_t]$. For a deterministic integrand the two coincide (Theorem 7.10); for $H = W$, $[I, I]_t = \int_0^t W_s^2 \, ds$ varies from path to path around its mean $t^2/2$ (Figure 7.2, Exercise 7.1). The quadratic variation can be estimated

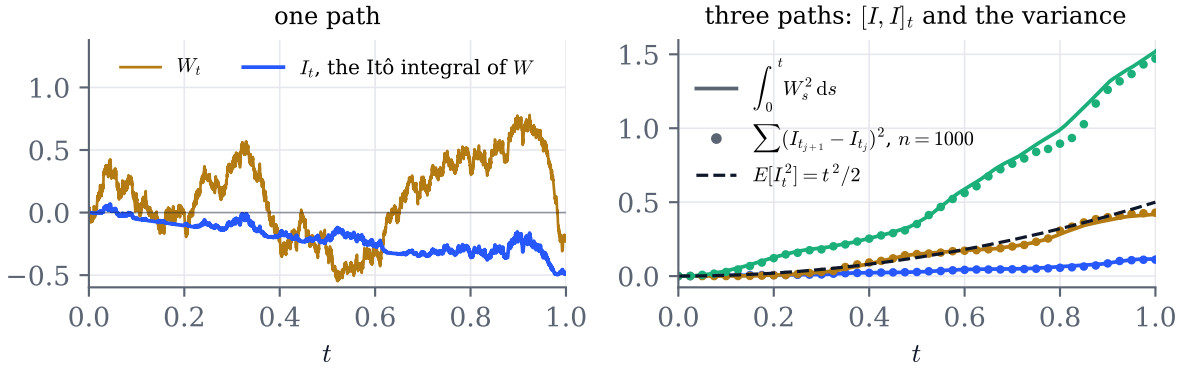

Figure 7.2: Left: a path of W and of $I_t = \int_0^t W_s\,\mathrm{d}W_s = \frac{1}{2}(W_t^2 - t)$. Right: for three paths, the quadratic variation $[I, I]_t = \int_0^t W_s^2\,\mathrm{d}s$ (lines), its estimate from 1000 increments (dots), and the variance $\mathbb{E}[I_t^2] = t^2/2$ (dashed): the mean of $[I, I]_t$, not the quadratic variation of a path.

from a single price history (the realized variance of variance swaps, Chapter 15); the variance is a property of the law.

▶ Interactive demo: the Itô isometry checked by simulation for several integrands, with the histograms of $\int_0^T H\,\mathrm{d}W$ (Gaussian or not); and the realized quadratic variation of one integral path converging to $\int_0^t H_s^2\,\mathrm{d}s$ as the grid is refined.

7.7 Stopping and localization

A *stopping time* is a random time τ with values in $[0, T]$ such that $\{\tau \le t\} \in \mathcal{F}_t$ for every t. The process $\mathbf{1}_{[0,\tau]}(s) = \mathbf{1}_{\{s \le \tau\}}$ is adapted with left-continuous paths, so $H\mathbf{1}_{[0,\tau]} \in \mathbb{H}^2$ whenever $H \in \mathbb{H}^2$.

Proposition 7.15 (Stopped integrals)**.** *Let $H \in \mathbb{H}^2$, $I = \int H\,\mathrm{d}W$, and τ a stopping time. Almost surely, $\int_0^u H_s \mathbf{1}_{[0,\tau]}(s)\,\mathrm{d}W_s = I_{u\wedge\tau}$ for every $u \in [0, T]$.*

Proof. Both sides are continuous in u, so it suffices to prove the identity almost surely for each u. For $H, H' \in \mathbb{H}^2$, $\int H\,\mathrm{d}W - \int H'\,\mathrm{d}W = \int (H - H')\,\mathrm{d}W$ is a continuous martingale, so by Theorem 7.6 and the isometry, $\mathbb{E}\big[\sup_u \big(\int_0^u H\,\mathrm{d}W - \int_0^u H'\,\mathrm{d}W\big)^2\big] \le 4\|H - H'\|^2$; we use this bound twice. *Simple H, τ with finitely many values.* Add the values of τ to the partition of H. Then $\mathbf{1}_{[0,\tau]}$ and $\mathbf{1}_{[0,\tau)}$ differ only at $s = \tau$, and for $s \in [t_j, t_{j+1})$, $s < \tau$ if and only if $\tau > t_j$. So $H\mathbf{1}_{[0,\tau)}$ is simple, with $\mathcal{F}_{t_j}$-measurable values $\xi_j \mathbf{1}_{\{\tau > t_j\}}$, and its integral at time u is $\sum_{j:\, t_j < \tau} \xi_j (W_{t_{j+1}\wedge u} - W_{t_j \wedge u}) = I_{u\wedge\tau}$, since $t_j < \tau$ implies $t_{j+1} \le \tau$. *General H.* For simple $H^n \to H$, $\|H^n\mathbf{1}_{[0,\tau]} - H\mathbf{1}_{[0,\tau]}\| \le \|H^n - H\|$ and $\sup_u |I^n_{u\wedge\tau} - I_{u\wedge\tau}| \le \sup_u |I^n_u - I_u|$, so both sides of the identity for H^n converge in L^2 to those for H. *General τ.* Let τ_n be the smallest point of $\{kT2^{-n} : 0 \le k \le 2^n\}$ that is at least τ. Then $\{\tau_n \le t\} = \{\tau \le d\} \in \mathcal{F}_t$, with d the largest grid point not after t, so τ_n is a stopping time with finitely many values, and $\tau_n \downarrow \tau$. By dominated convergence, $\|H\mathbf{1}_{[0,\tau_n]} - H\mathbf{1}_{[0,\tau]}\|^2 = \mathbb{E}\int_0^T H_s^2 \mathbf{1}_{\{\tau < s \le \tau_n\}}\,\mathrm{d}s \to 0$, so the left sides for τ_n converge in L^2 to that for τ, while $I_{u\wedge\tau_n} \to I_{u\wedge\tau}$ by continuity of the paths. Limits in L^2 and almost sure limits agree almost surely (along a subsequence), which concludes. □

Remark 7.16 (Integrands in $\mathbb{H}^2_{\mathrm{loc}}$ and local martingales). The condition $\mathbb{E}\int_0^T H_t^2\,\mathrm{d}t < \infty$ is not always available; $e^{W_t^2}$ fails it for $T > \frac{1}{4}$ (Section 7.3). The integral extends to every $H \in \mathbb{H}^2_{\mathrm{loc}}$ by *localization*. Since $t \mapsto \int_0^t H_s^2\,\mathrm{d}s$ is continuous, nondecreasing and adapted, the times $\tau_n = \inf\{t : \int_0^t H_s^2\,\mathrm{d}s \geq n\}$ ($\inf\emptyset = T$) are stopping times; they increase with n, and almost surely $\tau_n = T$ for all large n. The process $H\mathbf{1}_{[0,\tau_n]}$ is in $\mathbb{H}^2$, with norm at most $\sqrt{n}$; let $I^{(n)}$ be its integral. By Theorem 7.15, applied to $H\mathbf{1}_{[0,\tau_{n+1}]}$ and τ_n, $I^{(n+1)}_{t\wedge\tau_n} = I^{(n)}_t$. Hence $I_t = I^{(n)}_t$ on $\{t \leq \tau_n\}$ defines a continuous adapted process with $I_{t\wedge\tau_n} = I^{(n)}_t$: the Itô integral of H (for $H \in \mathbb{H}^2$, the old one, by Theorem 7.15). It is linear in H, and $[I,I]_t = \int_0^t H_s^2\,\mathrm{d}s$ still holds, because $I = I^{(n)}$ on the event $\{\tau_n = T\}$, whose probability tends to one. It is a *local martingale*: each stopped process $(I_{t\wedge\tau_n})_{t\leq T}$ is a martingale. But I itself need not be a martingale, and the isometry need not hold (Theorem 7.17). A sufficient condition for an integral to be a martingale is that its integrand be in $\mathbb{H}^2$ (more in Karatzas and Shreve (1991), Section 3.2.D).

Example 7.17 (A doubling strategy in continuous time)**.** Take $T = 1$ and $u_k = 1 - 1/k$. By Theorem 7.10 on each $[0,u_k]$, glued along k (Theorem 7.8), $X_u = \int_0^u (1-s)^{-1/2}\,\mathrm{d}W_s$, $0 \leq u < 1$, is a centred Gaussian process with continuous paths, independent increments and $\mathrm{Var}(X_u) = \int_0^u \frac{\mathrm{d}s}{1-s} = -\log(1-u)$. In the clock $v = -\log(1-u)$, the process $B_v = X_{1-e^{-v}}$, $v \geq 0$, has continuous paths and $\mathrm{Cov}(B_v, B_w) = \min(v,w)$: it is a Brownian motion (Theorem 5.7). Brownian motion reaches every level almost surely (let $a \to -\infty$ in Exercise 5.6(b)), so the stopping time $\tau = \inf\{u \in [0,1) : X_u = 1\}$ is less than 1 almost surely (set $\tau = 1$ on the null event where X never reaches 1). Let $H_s = (1-s)^{-1/2}\mathbf{1}_{\{s\leq\tau\}}\mathbf{1}_{\{\tau<1\}}$, progressively measurable because $\{\tau < 1\}$ has probability one and so belongs to $\mathcal{F}_0$. Then $\int_0^1 H_s^2\,\mathrm{d}s = -\log(1-\tau) < \infty$ almost surely, so $H \in \mathbb{H}^2_{\mathrm{loc}}[0,1]$. By Theorem 7.15 on each $[0,u_k]$, its gains process is $I_u = X_{u\wedge\tau}$ for $u < 1$, and since H vanishes after $\tau < 1$, $I_1 = X_\tau = 1$.

Starting with no capital, this strategy ends with one dollar for sure: an arbitrage. Its gains process is a local martingale but not a martingale ($\mathbb{E}[I_1] = 1$), and $H \notin \mathbb{H}^2$: $\int_0^1 H_s^2\,\mathrm{d}s$ is the time B needs to reach 1, of infinite mean (Exercise 7.7). As for the doubling strategy of Theorem 2.13, the catch is unlimited credit: for every $a < 0$, the gains reach a before 1 with probability $1/(1-a)$ (Exercise 5.6(b) for B on $(a,1)$). Continuous-time models exclude such strategies by restricting the admissible ones, for instance to those whose wealth is bounded below (Harrison and Pliska (1981); Karatzas and Shreve (1991), Section 5.8, where such strategies are called *tame*).

Further reading

Shreve (2004b), Sections 4.2 and 4.3, follows the same order, with integrands written Δ. Øksendal (2003), Chapter 3, and Steele (2001), Chapters 6 and 7, give more detail, including Stratonovich integrals and localization; Karatzas and Shreve (1991), Chapter 3, and Revuz and Yor (1999), Chapter IV, integrate against continuous martingales.

Chapter 8

Itô's formula

The chain rule of ordinary calculus says that $\mathrm{d}f(g(t)) = f'(g(t))\,\mathrm{d}g(t)$ for a differentiable function g. Chapter 7 found that Brownian motion breaks this rule: $W_t^2 = 2\int_0^t W_s\,\mathrm{d}W_s + t$ (Theorem 7.11), and the chain rule misses the term t. This chapter explains where the extra term comes from and computes it for every smooth function. Expand $f(W)$ to second order along a partition: the second-order term is $\frac{1}{2}f''(W)(\delta W)^2$, and since the sum of the squared increments $(\delta W)^2$ does not vanish but converges to t (Theorem 5.16), this term survives in the limit. Every other term of the expansion disappears. The result is Itô's formula.

Itô's formula is the main computational tool of the rest of the course: it gives the dynamics of smooth functions of prices, tells when they are martingales, and handles several correlated assets.

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, W a Brownian motion and $(\mathcal{F}_t)$ a filtration for W (Theorem 5.11); as in Chapter 7, every event of probability zero belongs to $\mathcal{F}_0$. For a partition $\Pi = \{0 = t_0 < t_1 < \cdots < t_m = t\}$ of $[0,t]$ we write $h_j = t_{j+1} - t_j$, $\|\Pi\| = \max_j h_j$, and $\delta_j X = X_{t_{j+1}} - X_{t_j}$ for the increments of a process X. For a process X with continuous paths and $r > 0$, the modulus of continuity $\operatorname{osc}_X(r) = \sup\{|X_u - X_v| : u, v \in [0,t],\ |u - v| \le r\}$ tends to 0 as $r \to 0$, for every outcome, because a continuous function is uniformly continuous on $[0,t]$. Identities between random variables hold almost surely.

8.1 Variations, and what survives a Taylor expansion

Recall the p-variation $V_p(g;\Pi) = \sum_j |\delta_j g|^p$ of a function g along a partition (Section 5.5). A function g on $[0,t]$ has *finite variation* if $\sup_\Pi V_1(g;\Pi) < \infty$. The basic example in this chapter is $A_t = \int_0^t \Theta_s\,\mathrm{d}s$ with $\int_0^t |\Theta_s|\,\mathrm{d}s < \infty$: since $|\delta_j A| \le \int_{t_j}^{t_{j+1}} |\Theta_s|\,\mathrm{d}s$, we have $V_1(A;\Pi) \le \int_0^t |\Theta_s|\,\mathrm{d}s$ for every partition.

Lemma 8.1 (Higher variations vanish)**.** *Let g be continuous on $[0,t]$, let $0 < p < q$, and let (Π_n) be partitions of $[0,t]$ with $\|\Pi_n\| \to 0$. Then*

$$V_q(g;\Pi_n) \le \Big(\max_j |\delta_j g|\Big)^{q-p} V_p(g;\Pi_n), \qquad \text{and} \qquad \max_j |\delta_j g| \to 0.$$

In particular, if $\sup_n V_p(g;\Pi_n) < \infty$, then $V_q(g;\Pi_n) \to 0$; a continuous function of finite variation has zero quadratic variation.

Proof. Write $|\delta_j g|^q = |\delta_j g|^{q-p}|\delta_j g|^p$ and bound the first factor by its maximum over j. The maximum is at most $\mathrm{osc}_g(\|\Pi_n\|)$, which tends to 0. □

For Brownian motion, $V_2(W;\Pi_n) \to t$ in probability, so the inequality of the lemma gives $V_q(W;\Pi_n) \to 0$ in probability for every $q > 2$ (Exercise 5.7); for $p < 2$ it reads $V_p(W;\Pi_n) \geq V_2(W;\Pi_n)/(\max_j |\delta_j W|)^{2-p}$, so $V_p(W;\Pi_n) \to \infty$ in probability (for $p = 1$ and $\sum_n \|\Pi_n\| < \infty$, almost surely: Theorem 5.17). Among the p-variations of Brownian motion, only the quadratic one has a finite nonzero limit, and this decides which terms of a Taylor expansion survive. Let f be smooth and g a continuous path. Telescoping over a partition of $[0,t]$ and expanding each term to second order,

$$f(g_t) - f(g_0) = \sum_j f'(g_{t_j})\,\delta_j g + \frac{1}{2}\sum_j f''(g_{t_j})\,(\delta_j g)^2 + \sum_j R_j, \qquad |R_j| \leq C|\delta_j g|^3, \tag{8.1}$$

with C depending on f and on the range of the path. If g is continuously differentiable, the second and third sums are bounded by $\max|f''|\,V_2(g;\Pi)$ and $C\,V_3(g;\Pi)$, which tend to 0: only the first sum remains, and its limit $\int_0^t f'(g_s)g'_s\,\mathrm{d}s$ gives the chain rule. If $g = W$, the third sum still vanishes, but the second does not: it is a weighted sum of squared increments, and the next section shows that it converges to $\frac{1}{2}\int_0^t f''(W_s)\,\mathrm{d}s$. The bookkeeping rules $\mathrm{d}W\,\mathrm{d}W = \mathrm{d}t$, $\mathrm{d}W\,\mathrm{d}t = 0$, $\mathrm{d}t\,\mathrm{d}t = 0$ of Section 5.5 are the short form of this statement.

8.2 Itô's formula for Brownian motion

Remark 8.2 (What we use from Chapter 7). For $H \in \mathbb{H}^2$ (Theorem 7.3), the integral $I_t = \int_0^t H_s\,\mathrm{d}W_s$ of Theorem 7.7 is continuous, adapted, linear in H, a martingale with $\mathbb{E}[I_t^2] = \mathbb{E}\int_0^t H_s^2\,\mathrm{d}s$, and $[I,I]_t = \int_0^t H_s^2\,\mathrm{d}s$; for simple H it is the sum of Theorem 7.1. For $H \in \mathbb{H}^2_{\mathrm{loc}}$ it is defined by localization (Theorem 7.16): it is continuous, adapted and linear, $[I,I]_t = \int_0^t H_s^2\,\mathrm{d}s$ still holds, and it is in general only a local martingale. Chapter 7 works on a fixed horizon $[0,T]$, which plays no role: if $T < T'$ and simple $H^n \to H$ in $\mathbb{H}^2[0,T']$, then $H^n \to H$ in $\mathbb{H}^2[0,T]$ and the integrals of the H^n over the two horizons agree on $[0,T]$, so by the uniqueness in Theorem 7.7 the two integrals of H are indistinguishable on $[0,T]$ (for $\mathbb{H}^2_{\mathrm{loc}}$, by localization). Taking $T = 1, 2, \ldots$, statements on every $[0,T]$ combine into statements for all $t \geq 0$. Continuous adapted processes are progressively measurable (Karatzas and Shreve (1991), Proposition 1.1.13), hence in $\mathbb{H}^2_{\mathrm{loc}}$, and in $\mathbb{H}^2$ if bounded. Quadratic variations and covariations are those of Theorem 7.12, limits in probability.

> **Lemma 8.3** (Locality)**.** *Let $H, G \in \mathbb{H}^2_{\mathrm{loc}}$, $t \in (0,T]$, and E an event such that $H_s(\omega) = G_s(\omega)$ for all $s \leq t$ and $\omega \in E$. Then, almost surely on E, $\int_0^s H\,\mathrm{d}W = \int_0^s G\,\mathrm{d}W$ for every $s \leq t$.*

Proof. Let $K = H - G$, $\tau_n = \inf\{s : \int_0^s K_u^2\,\mathrm{d}u \geq 1/n\} \wedge t$ and $\tau = \inf\{s : \int_0^s K_u^2\,\mathrm{d}u \geq 1\} \wedge T$, stopping times with $\tau_n \leq \tau$. By Theorem 7.16, $\int K\,\mathrm{d}W$ coincides up to time τ with the integral of $K\mathbf{1}_{[0,\tau]} \in \mathbb{H}^2$, so Theorem 7.15 gives $\int_0^{s\wedge\tau_n} K\,\mathrm{d}W = \int_0^s K\mathbf{1}_{[0,\tau_n]}\,\mathrm{d}W$ for $s \leq t$. Since $\int_0^t K_u^2\mathbf{1}_{[0,\tau_n]}(u)\,\mathrm{d}u \leq 1/n$, Doob's inequality (Theorem 7.6) and the isometry give $\mathbb{E}\big[\sup_{s\leq t}\big(\int_0^{s\wedge\tau_n} K\,\mathrm{d}W\big)^2\big] \leq 4/n$. On E, $\int_0^t K_u^2\,\mathrm{d}u = 0$, so $\tau_n = t$ for every n, and $\mathbb{P}(E \cap \{\sup_{s\leq t}|\int_0^s K\,\mathrm{d}W| > \varepsilon\}) \leq 4/(n\varepsilon^2)$ for all n and $\varepsilon > 0$. So $\int_0^s K\,\mathrm{d}W = 0$ for all $s \leq t$, almost surely on E, and linearity concludes. □

The first step is to identify the limit of the first-order sum in (8.1): the gains of a strategy that holds H_{t_j} units over $[t_j, t_{j+1})$.

Lemma 8.4 (Riemann sums with continuous integrands)**.** *Let $\Delta \in \mathbb{H}^2_{\mathrm{loc}}$, $M_t = \int_0^t \Delta_s \, \mathrm{d}W_s$, and let H be a continuous adapted process. For $t > 0$ and partitions Π of $[0,t]$ with $\|\Pi\| \to 0$,*

$$\sum_j H_{t_j} \, \delta_j M \longrightarrow \int_0^t H_s \Delta_s \, \mathrm{d}W_s \quad \textit{in probability.}$$

If H is bounded and $\Delta \in \mathbb{H}^2$, the convergence holds in L^2. With $\Delta \equiv 1$: $\sum_j H_{t_j} \delta_j W \to \int_0^t H_s \, \mathrm{d}W_s$.

Proof. The product $H\Delta$ is progressively measurable and $\int_0^t H_s^2 \Delta_s^2 \, \mathrm{d}s \le \max_{s \le t} H_s^2 \int_0^t \Delta_s^2 \, \mathrm{d}s < \infty$, so $H\Delta \in \mathbb{H}^2_{\mathrm{loc}}$. Write S_Π for the sum.

Bounded case: $|H| \le c$ and $\Delta \in \mathbb{H}^2$. For $0 \le a < b \le t$ and ξ bounded and $\mathcal{F}_a$-measurable,

$$\int_0^t \xi \mathbf{1}_{[a,b)}(s) \, \Delta_s \, \mathrm{d}W_s = \xi \, (M_b - M_a). \tag{8.2}$$

For $\xi = 1$ this is Theorem 7.8; for simple Δ it is Theorem 7.1, after adding a and b to the grid of Δ. In general, take simple $\Delta^n \to \Delta$ in $\mathbb{H}^2$; then $\xi \mathbf{1}_{[a,b)} \Delta^n \to \xi \mathbf{1}_{[a,b)} \Delta$ in $\mathbb{H}^2$, and by the isometry both sides of (8.2) for Δ^n converge in L^2 to those for Δ. Summing (8.2) over the intervals of Π gives $S_\Pi = \int_0^t H_s^\Pi \Delta_s \, \mathrm{d}W_s$, where $H_s^\Pi = H_{t_j}$ for $s \in [t_j, t_{j+1})$. By linearity and the isometry,

$$\mathbb{E}\Big[\Big(S_\Pi - \int_0^t H_s \Delta_s \, \mathrm{d}W_s\Big)^2\Big] = \mathbb{E} \int_0^t (H_s^\Pi - H_s)^2 \Delta_s^2 \, \mathrm{d}s.$$

For every (s, ω), $|H_s^\Pi - H_s| \le \mathrm{osc}_H(\|\Pi\|) \to 0$, and the integrand is at most $4c^2\Delta_s^2$, which is integrable on $[0,t] \times \Omega$. By dominated convergence, the right side tends to 0.

General case. For $K > 0$ let $H^K = \max(-K, \min(H, K))$, a bounded continuous adapted process, and $\Delta_s^K = \Delta_s \mathbf{1}\{\int_0^s \Delta_u^2 \, \mathrm{d}u \le K\}$, which is in $\mathbb{H}^2$ since $\int_0^t (\Delta_s^K)^2 \, \mathrm{d}s \le K$. On the event $E_K = \{\max_{s \le t} |H_s| \le K, \ \int_0^t \Delta_s^2 \, \mathrm{d}s \le K\}$ we have $H^K = H$ and $\Delta^K = \Delta$ on $[0,t]$, so by Theorem 8.3, almost surely on E_K, M coincides on $[0,t]$ with $\int \Delta^K \mathrm{d}W$ and $\int_0^t H\Delta \, \mathrm{d}W = \int_0^t H^K \Delta^K \mathrm{d}W$. Hence, with S_Π^K the sum for (H^K, Δ^K),

$$\mathbb{P}\Big(\Big|S_\Pi - \int_0^t H\Delta \, \mathrm{d}W\Big| > \varepsilon\Big) \le \mathbb{P}(E_K^c) + \mathbb{P}\Big(\Big|S_\Pi^K - \int_0^t H^K \Delta^K \mathrm{d}W\Big| > \varepsilon\Big).$$

The last term tends to 0 as $\|\Pi\| \to 0$ by the bounded case, and $\mathbb{P}(E_K^c) \to 0$ as $K \to \infty$ because $\max_{s \le t} |H_s|$ and $\int_0^t \Delta_s^2 \, \mathrm{d}s$ are finite. □

We write $f \in C^{1,2}([0,\infty) \times \mathbb{R})$ when $f(t,x)$ has partial derivatives $\partial_t f$, $\partial_x f$ and $\partial_{xx} f$ that are continuous on $[0,\infty) \times \mathbb{R}$ (one-sided in t at $t = 0$); $C^{1,2}([0,T] \times \mathbb{R})$ is defined in the same way.

Theorem 8.5 (Itô's formula for Brownian motion)**.** *Let $f \in C^{1,2}([0,\infty) \times \mathbb{R})$. Almost surely, for every $t \ge 0$,*

$$f(t, W_t) = f(0,0) + \int_0^t \partial_t f(s, W_s) \, \mathrm{d}s + \int_0^t \partial_x f(s, W_s) \, \mathrm{d}W_s + \frac{1}{2} \int_0^t \partial_{xx} f(s, W_s) \, \mathrm{d}s. \tag{8.3}$$

The integrand $\partial_x f(s, W_s)$ is continuous and adapted, hence in $\mathbb{H}^2_{\text{loc}}$. If $\mathbb{E}\int_0^T \partial_x f(s, W_s)^2 \, \mathrm{d}s < \infty$, it is in $\mathbb{H}^2$ on $[0, T]$, and the stochastic integral in (8.3) is a martingale on $[0, T]$. The same holds on $[0, T]$ for $f \in C^{1,2}([0, T] \times \mathbb{R})$.

In differential notation, $\mathrm{d}f(t, W_t) = (\partial_t f + \frac{1}{2}\partial_{xx} f)(t, W_t)\,\mathrm{d}t + \partial_x f(t, W_t)\,\mathrm{d}W_t$: expand f to first order in t and to second order in x, and replace $(\mathrm{d}W_t)^2$ by $\mathrm{d}t$.

Proof. Fix $t > 0$ and partitions Π of $[0, t]$ with $\|\Pi\| \to 0$, and write $\delta_j = \delta_j W$. Telescoping,

$$f(t, W_t) - f(0, 0) = \underbrace{\sum_j \left[f(t_{j+1}, W_{t_{j+1}}) - f(t_j, W_{t_{j+1}})\right]}_{A_\Pi} + \underbrace{\sum_j \left[f(t_j, W_{t_{j+1}}) - f(t_j, W_{t_j})\right]}_{B_\Pi}.$$

Fix an outcome ω. Its path is bounded on $[0, t]$, by k_ω say, and the three derivatives of f are uniformly continuous on the compact set $[0, t] \times [-k_\omega, k_\omega]$. Let $\varpi(r)$ be the largest of their moduli of continuity there, so that $|g(s, x) - g(s', x')| \le \varpi(r)$ whenever $|s - s'| \le r$ and $|x - x'| \le r$, for $g \in \{\partial_t f, \partial_x f, \partial_{xx} f\}$; then $\varpi(r) \to 0$ as $r \to 0$. Let $r_\Pi = \|\Pi\| + \mathrm{osc}_W(\|\Pi\|)$, which tends to 0.

Time increments. By the mean value theorem, the j-th term of A_Π is $\partial_t f(\tau_j, W_{t_{j+1}})\, h_j$ for some $\tau_j \in [t_j, t_{j+1}]$, and for every $s \in [t_j, t_{j+1}]$ the point $(\tau_j, W_{t_{j+1}})$ is within r_Π of (s, W_s) in each coordinate. Hence $\left|A_\Pi - \int_0^t \partial_t f(s, W_s)\,\mathrm{d}s\right| \le t\,\varpi(r_\Pi) \to 0$, for every outcome.

Space increments. By Taylor's formula with the Lagrange remainder, $f(t_j, W_{t_{j+1}}) - f(t_j, W_{t_j}) = \partial_x f(t_j, W_{t_j})\,\delta_j + \frac{1}{2}\partial_{xx} f(t_j, \xi_j)\,\delta_j^2$ for some ξ_j between W_{t_j} and $W_{t_{j+1}}$. Hence

$$B_\Pi = \sum_j \partial_x f(t_j, W_{t_j})\,\delta_j + \frac{1}{2}\sum_j g_j\,\delta_j^2 + \frac{1}{2}R_\Pi, \qquad g_j = \partial_{xx} f(t_j, W_{t_j}),$$

where R_Π, defined by this identity, is a random variable with $|R_\Pi| \le \varpi(r_\Pi)\, V_2(W; \Pi)$ (since $|\xi_j - W_{t_j}| \le r_\Pi$). The first sum converges in probability to $\int_0^t \partial_x f(s, W_s)\,\mathrm{d}W_s$ by Theorem 8.4. The remainder tends to 0 in probability, because $\varpi(r_\Pi) \to 0$ for every outcome and $V_2(W; \Pi) \to t$ in probability (Theorem 5.16).

Second-order term. Write $\sum_j g_j \delta_j^2 = \sum_j g_j h_j + \sum_j g_j(\delta_j^2 - h_j)$. The first sum is a Riemann sum, within $t\,\varpi(r_\Pi)$ of $\int_0^t \partial_{xx} f(s, W_s)\,\mathrm{d}s$. The second sum tends to 0 in probability: this is the law of large numbers along a path announced in Section 5.5. Let $K > 0$, let $g_j^K = \partial_{xx} f(t_j, \max(-K, \min(W_{t_j}, K)))$, which is $\mathcal{F}_{t_j}$-measurable and bounded by $c_K = \max_{[0,t]\times[-K,K]} |\partial_{xx} f|$, and let $D_j = \delta_j^2 - h_j$. As in the proof of Theorem 5.16, D_j is independent of $\mathcal{F}_{t_j}$, with $\mathbb{E}[D_j] = 0$ and $\mathbb{E}[D_j^2] = 2h_j^2$. For $i < j$, $g_i^K D_i g_j^K$ is $\mathcal{F}_{t_j}$-measurable and integrable, so $\mathbb{E}[g_i^K D_i g_j^K D_j] = \mathbb{E}[g_i^K D_i g_j^K]\,\mathbb{E}[D_j] = 0$: the cross terms vanish, and

$$\mathbb{E}\Big[\Big(\sum_j g_j^K D_j\Big)^2\Big] = \sum_j \mathbb{E}[(g_j^K)^2]\,\mathbb{E}[D_j^2] \le 2c_K^2 \sum_j h_j^2 \le 2c_K^2\, t\,\|\Pi\|.$$

On the event $E_K = \{\max_{s \le t} |W_s| \le K\}$, $g_j^K = g_j$ for every j. By Chebyshev's inequality, $\mathbb{P}(|\sum_j g_j D_j| > \varepsilon) \le \mathbb{P}(E_K^c) + 2c_K^2 t\|\Pi\|/\varepsilon^2$. Let $\|\Pi\| \to 0$ and then $K \to \infty$; $\mathbb{P}(E_K^c) \to 0$ because the path is bounded on $[0, t]$.

Conclusion. $A_\Pi + B_\Pi = f(t, W_t) - f(0, 0)$ does not depend on Π, and we have shown that it converges in probability to the right side of (8.3). Limits in probability are unique almost

surely, so (8.3) holds almost surely for each fixed t. Both sides are continuous in t, so equality almost surely at all rational t at once gives equality for all t. The statements about $\mathbb{H}^2$ are Theorem 8.2. □

Example 8.6. (a) For $f(x) = x^2$: $W_t^2 = \int_0^t 2W_s\,\mathrm{d}W_s + t$. The integrand is in $\mathbb{H}^2$, since $\mathbb{E}\int_0^T 4W_s^2\,\mathrm{d}s = 2T^2$, so $W_t^2 - t$ is a martingale, as in Theorem 5.12.

(b) For $f(t,x) = \exp(\sigma x - \frac{\sigma^2}{2}t)$: $\partial_t f = -\frac{\sigma^2}{2}f$, $\partial_x f = \sigma f$ and $\partial_{xx} f = \sigma^2 f$, so the $\mathrm{d}t$-terms cancel and $Z_t = f(t, W_t)$ satisfies $Z_t = 1 + \int_0^t \sigma Z_s\,\mathrm{d}W_s$, or $\mathrm{d}Z_t = \sigma Z_t\,\mathrm{d}W_t$. Since $\mathbb{E}[Z_s^2] = e^{\sigma^2 s}$, $\mathbb{E}\int_0^T \sigma^2 Z_s^2\,\mathrm{d}s < \infty$ and Z is a martingale, as in Theorem 5.12(iii).

Both examples are instances of one mechanism: when the $\mathrm{d}t$-terms of Itô's formula cancel, what remains is a stochastic integral.

Corollary 8.7 (Martingales by drift cancellation)**.** *Let $T > 0$ and $f \in C^{1,2}([0,T]\times\mathbb{R})$ with*

$$\partial_t f + \tfrac{1}{2}\partial_{xx} f = 0 \quad \text{on } [0,T]\times\mathbb{R}. \tag{8.4}$$

Then $f(t,W_t) = f(0,0) + \int_0^t \partial_x f(s,W_s)\,\mathrm{d}W_s$ for $t \le T$. If moreover $\mathbb{E}\int_0^T \partial_x f(s,W_s)^2\,\mathrm{d}s < \infty$, then $(f(t,W_t))_{t\le T}$ is a martingale.

Proof. The $\mathrm{d}s$-integrals in (8.3) cancel by (8.4); the rest is Theorem 8.5. □

Equation (8.4) is the backward heat equation. It is also solved by $x^3 - 3tx$ and $x^4 - 6tx^2 + 3t^2$, which recovers Exercise 5.2 (the integrability condition holds since W_s has moments of all orders). Without that condition, $f(t,W_t)$ is in general only a local martingale (Theorem 7.16); a converse is Exercise 8.3, and Chapter 12 turns the corollary around.

Proposition 8.8 (Geometric Brownian motion)**.** *Let $S_0 > 0$, $\mu,\sigma \in \mathbb{R}$, and $S_t = S_0\exp\left(\sigma W_t + (\mu - \frac{\sigma^2}{2})t\right)$ (Theorem 6.1, where $\sigma > 0$; here any $\sigma \in \mathbb{R}$ is allowed). For every $r \in \mathbb{R}$,*

(i) $S_t = S_0 + \int_0^t \mu S_s\,\mathrm{d}s + \int_0^t \sigma S_s\,\mathrm{d}W_s$, that is, $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$;

(ii) $\mathrm{d}(e^{-rt}S_t) = (\mu - r)\,e^{-rt}S_t\,\mathrm{d}t + \sigma\,e^{-rt}S_t\,\mathrm{d}W_t$;

(iii) the integrands σS and $\sigma e^{-rt}S$ are in $\mathbb{H}^2$ on every $[0,T]$, and $(e^{-rt}S_t)_{t\ge 0}$ is a martingale if and only if $\mu = r$.

Proof. (i) Apply Theorem 8.5 to $f(t,x) = S_0\exp\left(\sigma x + (\mu - \frac{\sigma^2}{2})t\right)$: $\partial_t f = (\mu - \frac{\sigma^2}{2})f$, $\partial_x f = \sigma f$, $\partial_{xx} f = \sigma^2 f$, and the $\mathrm{d}t$-coefficient is $(\mu - \frac{\sigma^2}{2} + \frac{\sigma^2}{2})f = \mu f$. (ii) The same computation with $\mu - r$ in place of μ, since $e^{-rt}S_t = S_0\exp\left(\sigma W_t + (\mu - r - \frac{\sigma^2}{2})t\right)$. (iii) The Gaussian moment generating function gives $\mathbb{E}[S_s^2] = S_0^2 e^{(2\mu+\sigma^2)s}$, which is bounded on $[0,T]$. If $\mu = r$, (ii) has no $\mathrm{d}t$-term and the integrand is in $\mathbb{H}^2$, so $e^{-rt}S_t$ is a martingale. If $\mu \ne r$, $\mathbb{E}[e^{-rt}S_t] = S_0 e^{(\mu-r)t} \ne S_0$ for $t > 0$. □

In (i), $\mathrm{d}S_t/S_t = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, while $\log S_t$ grows at the smaller rate $\mu - \frac{\sigma^2}{2}$: the difference is the volatility drag of Theorems 2.6 and 5.14, produced by the second-order term. Uniqueness for the equation in (i) is proved in Chapter 9, and the measure under which $\mu = r$ is constructed in Chapter 11.

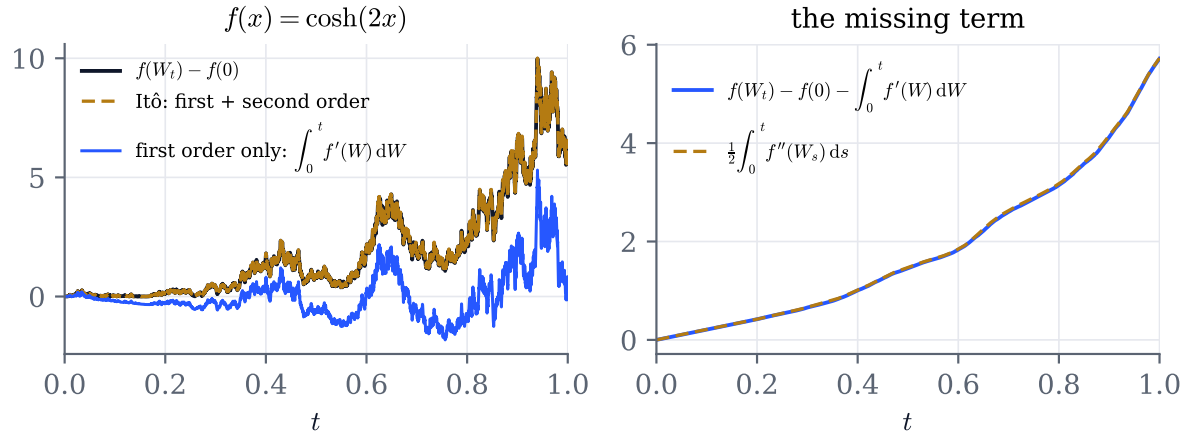

Figure 8.1: Itô's formula along one Brownian path on a grid of 2^{16} steps, for $f(x) = \cosh(2x)$, with the integrals computed by left-point sums (code/src/ieor222/itoformula.py). Left: $f(W_t) - f(0)$, which is nonnegative, and the right side of (8.3) lie on top of each other; the chain rule of ordinary calculus, $\int_0^t f'(W_s)\,\mathrm{d}W_s$, drifts away and even turns negative. Right: the gap is the second-order term $\frac{1}{2}\int_0^t f''(W_s)\,\mathrm{d}s$.

▶ Interactive demo: Itô's formula checked along a path for the function, grid and seed of your choice, with and without the second-order term.

8.3 Itô processes

Prices are rarely functions of (t, W_t) alone. The class of processes that Itô's formula maps into itself is the following.

Definition 8.9 (Itô process)**.** An *Itô process* is a process of the form

$$X_t = x_0 + \int_0^t \Theta_s\,\mathrm{d}s + \int_0^t \Delta_s\,\mathrm{d}W_s, \qquad t \geq 0,$$

where x_0 is $\mathcal{F}_0$-measurable (often a constant), $\Delta \in \mathbb{H}^2_{\mathrm{loc}}$, and Θ is progressively measurable with $\int_0^T |\Theta_s|\,\mathrm{d}s < \infty$ almost surely for every T. Θ is the *drift* and Δ the *diffusion coefficient*; in differential notation, $\mathrm{d}X_t = \Theta_t\,\mathrm{d}t + \Delta_t\,\mathrm{d}W_t$, $X_0 = x_0$.

An Itô process has continuous paths and is adapted, and linear combinations of Itô processes are Itô processes. The proofs below (of Theorems 8.11, 8.13, 8.17 and 8.18) use x_0 only through $X_0 = x_0$ and work path by path, so they hold verbatim for random x_0. By Theorem 8.5, $f(t, W_t)$ is an Itô process with $\Theta = (\partial_t f + \frac{1}{2}\partial_{xx} f)(t, W_t)$ and $\Delta = \partial_x f(t, W_t)$. If $\Delta \in \mathbb{H}^2$, then $X - \int \Theta\,\mathrm{d}s$ is a martingale; with $\Delta \in \mathbb{H}^2_{\mathrm{loc}}$ only, it is a local martingale. The larger class is needed: $e^{W_t^2}$ is an Itô process whose diffusion coefficient is not in $\mathbb{H}^2$ on $[0, T]$ for $T \geq \frac{1}{4}$ (Exercise 8.2).

Definition 8.10 (Integral with respect to an Itô process)**.** Let X be an Itô process as in Theorem 8.9, and H a progressively measurable process with $\int_0^T |H_s \Theta_s|\,\mathrm{d}s < \infty$ and

$\int_0^T H_s^2\Delta_s^2\,\mathrm{d}s<\infty$ almost surely for every T (for instance, H continuous and adapted). Then

$$\int_0^t H_s\,\mathrm{d}X_s=\int_0^t H_s\Theta_s\,\mathrm{d}s+\int_0^t H_s\Delta_s\,\mathrm{d}W_s.$$

If H is continuous and adapted, $\sum_j H_{t_j}\delta_j X\to\int_0^t H_s\,\mathrm{d}X_s$ in probability: the $\mathrm{d}W$ part is Theorem 8.4, and the $\mathrm{d}s$ part converges for every outcome, since $\bigl|\sum_j H_{t_j}\int_{t_j}^{t_{j+1}}\Theta_s\,\mathrm{d}s-\int_0^t H_s\Theta_s\,\mathrm{d}s\bigr|\le \operatorname{osc}_H(\|\Pi\|)\int_0^t|\Theta_s|\,\mathrm{d}s$. So $\int H\,\mathrm{d}X$ is the gain of holding H_t units of an asset with price X, the continuous-time form of the martingale transform of Theorem 2.7.

Proposition 8.11 (Covariation and uniqueness of the decomposition)**.** *Let X be as in Theorem 8.9 and $Y_t=y_0+\int_0^t\tilde\Theta_s\,\mathrm{d}s+\int_0^t\tilde\Delta_s\,\mathrm{d}W_s$ another Itô process.*

(i) For every $t\ge0$, $[X,X]_t=\int_0^t\Delta_s^2\,\mathrm{d}s$ and $[X,Y]_t=\int_0^t\Delta_s\tilde\Delta_s\,\mathrm{d}s$.

(ii) If $X_t=Y_t$ for all $t\in[0,T]$ almost surely, then $x_0=y_0$, and $\Theta=\tilde\Theta$ and $\Delta=\tilde\Delta$ for $\mathrm{d}t\otimes\mathrm{d}\mathbb{P}$-almost every $(t,\omega)\in[0,T]\times\Omega$.

Proof. (i) Write $X=x_0+A+M$ and $Y=y_0+\tilde A+\tilde M$ with $A=\int\Theta\,\mathrm{d}s$, $M=\int\Delta\,\mathrm{d}W$, and similarly for Y. The paths of A and $\tilde A$ have finite variation, so by bilinearity and Theorem 7.13, $[X,Y]_t=[M,\tilde M]_t$. This is $\int_0^t\Delta_s\tilde\Delta_s\,\mathrm{d}s$ by Theorem 7.14(i) when $\Delta,\tilde\Delta\in\mathbb{H}^2$, and in general by polarization, since $[M\pm\tilde M,M\pm\tilde M]_t=\int_0^t(\Delta_s\pm\tilde\Delta_s)^2\,\mathrm{d}s$ (Theorem 7.16).

(ii) $D=X-Y$ is an Itô process with $D_0=x_0-y_0$, drift $\Theta-\tilde\Theta$ and diffusion coefficient $\Delta-\tilde\Delta$, and $D\equiv0$ on $[0,T]$. So $x_0=y_0$ almost surely, and all quadratic variation sums of D vanish: by (i), $\int_0^T(\Delta_s-\tilde\Delta_s)^2\,\mathrm{d}s=0$ almost surely. By Tonelli's theorem, $\Delta=\tilde\Delta$ for $\mathrm{d}t\otimes\mathrm{d}\mathbb{P}$-almost every (t,ω). Then $\mathbb{E}\int_0^T(\Delta-\tilde\Delta)^2\mathrm{d}s=0$, so $\int(\Delta-\tilde\Delta)\,\mathrm{d}W=0$ on $[0,T]$ by the isometry, and $\int_0^t(\Theta_s-\tilde\Theta_s)\,\mathrm{d}s=0$ for all $t\le T$, almost surely. Fix such an outcome and let $\phi=\Theta-\tilde\Theta$. The finite measures $\int_B\phi^+\mathrm{d}s$ and $\int_B\phi^-\mathrm{d}s$ on $[0,T]$ agree on the intervals $[0,t]$, which form a π-system generating the Borel sets, so they agree on every Borel set B by Dynkin's π–λ theorem. With $B=\{\phi>0\}$ this gives $\int_0^T\phi^+\mathrm{d}s=0$, and likewise $\int_0^T\phi^-\mathrm{d}s=0$. So $\phi=0$ for almost every s, and Tonelli's theorem concludes again. □

In differential form, part (i) is a multiplication table:

	$\mathrm{d}t$	$\mathrm{d}W_t$	
$\mathrm{d}t$	0	0	$\mathrm{d}[X,X]_t=\Delta_t^2\,\mathrm{d}t$
$\mathrm{d}W_t$	0	$\mathrm{d}t$	$\mathrm{d}[X,Y]_t=\Delta_t\tilde\Delta_t\,\mathrm{d}t$

The drift does not contribute to the quadratic variation: $[X,X]$ can be read off a single path on any time interval, whatever the drift. Part (ii) says that the drift and the diffusion coefficient are determined by the process; in Chapter 10, matching the $\mathrm{d}W$-terms and the $\mathrm{d}t$-terms of two expressions for a hedged portfolio gives the hedge ratio and the Black–Scholes equation.

For Itô processes, the martingale argument for the second-order term is replaced by a lemma that uses only the quadratic variation.

Lemma 8.12 (Weighted quadratic variation)**.** *Let X be a continuous process and A a continuous nondecreasing process with $A_0=0$, such that $V_2(X;\Pi)\to A_s$ in probability for every $s\in(0,t]$ and partitions Π of $[0,s]$ with $\|\Pi\|\to0$. Let G be a process with continuous*

paths. Then, for partitions Π *of* $[0,t]$ *with* $\|\Pi\| \to 0$,

$$\sum_j G_{t_j}(\delta_j X)^2 \longrightarrow \int_0^t G_s \,\mathrm{d}A_s \quad \textit{in probability,}$$

the right side being a Riemann–Stieltjes integral along each path. For an Itô process, $A = [X,X]$ *and the limit is* $\int_0^t G_s \Delta_s^2 \,\mathrm{d}s$.

Proof. Write $S(\Pi)$ for the sum and $\|G\| = \max_{s\le t}|G_s|$. Fix an integer $m \ge 1$, let $r_k = kt/m$, and let Π' be Π with the points $r_1, \dots, r_{m-1}$ added. Inserting a point u into an interval $[t_j, t_{j+1}]$ replaces $G_{t_j}(a+b)^2$ by $G_{t_j}a^2 + G_u b^2$, with $a = X_u - X_{t_j}$ and $b = X_{t_{j+1}} - X_u$; the change $2G_{t_j}ab + (G_{t_j} - G_u)b^2$ is at most $4\|G\|\,\mathrm{osc}_X(\|\Pi\|)^2$. Hence $|S(\Pi) - S(\Pi')| \le 4m\|G\|\,\mathrm{osc}_X(\|\Pi\|)^2$, which tends to 0 for every outcome. Next, group the intervals of Π' according to the block $[r_k, r_{k+1}]$ that contains them:

$$S(\Pi') = \sum_{k=0}^{m-1} G_{r_k} Q_k + R, \qquad |R| \le \mathrm{osc}_G(t/m)\, V_2(X;\Pi'),$$

where Q_k is the sum of $(\delta X)^2$ over the intervals of Π' in $[r_k, r_{k+1}]$. Since Q_k is the difference of the quadratic variation sums of X along the partitions $\Pi' \cap [0, r_{k+1}]$ and $\Pi' \cap [0, r_k]$, whose meshes are at most $\|\Pi\|$, $Q_k \to A_{r_{k+1}} - A_{r_k}$ in probability, and $V_2(X;\Pi') \to A_t$. Finally, for every outcome, $\big|\sum_k G_{r_k}(A_{r_{k+1}} - A_{r_k}) - \int_0^t G\,\mathrm{d}A\big| \le \mathrm{osc}_G(t/m)\, A_t$. Altogether,

$$\begin{aligned}\Big|S(\Pi) - \int_0^t G\,\mathrm{d}A\Big| &\le 4m\|G\|\,\mathrm{osc}_X(\|\Pi\|)^2 + \Big|\sum_k G_{r_k}(Q_k - A_{r_{k+1}} + A_{r_k})\Big| \\ &\quad + \mathrm{osc}_G(t/m)(V_2(X;\Pi') + A_t).\end{aligned}$$

Given $\varepsilon, \eta > 0$, choose m with $\mathbb{P}\big(\mathrm{osc}_G(t/m)(2A_t + 1) > \varepsilon\big) < \eta$, which is possible because $\mathrm{osc}_G(t/m) \to 0$ for every outcome. As $\|\Pi\| \to 0$, the first term tends to 0, the second tends to 0 in probability (a finite sum of finite random variables times terms that tend to 0 in probability), and $\mathbb{P}(V_2(X;\Pi') > A_t + 1) \to 0$. So $\limsup \mathbb{P}(|S(\Pi) - \int_0^t G\,\mathrm{d}A| > 3\varepsilon) \le \eta$. □

By polarization, $\sum_j G_{t_j}\delta_j X \delta_j Y \to \int_0^t G_s \,\mathrm{d}[X,Y]_s$ when $X \pm Y$ satisfy the hypothesis; for Itô processes the limit is $\int_0^t G_s \Delta_s \tilde{\Delta}_s \,\mathrm{d}s$.

Theorem 8.13 (Itô's formula for Itô processes)**.** *Let* X *be an Itô process as in Theorem 8.9 and* $f \in C^{1,2}([0,\infty)\times\mathbb{R})$. *Almost surely, for every* $t \ge 0$,

$$\begin{aligned} f(t, X_t) &= f(0,x_0) + \int_0^t \partial_t f(s, X_s)\,\mathrm{d}s + \int_0^t \partial_x f(s, X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t \partial_{xx} f(s, X_s)\,\mathrm{d}[X,X]_s \\ &= f(0,x_0) + \int_0^t \Big(\partial_t f + \Theta_s\,\partial_x f + \frac{1}{2}\Delta_s^2\,\partial_{xx} f\Big)(s, X_s)\,\mathrm{d}s + \int_0^t \partial_x f(s, X_s)\,\Delta_s\,\mathrm{d}W_s. \end{aligned}$$

In particular, $f(t, X_t)$ *is an Itô process.*

Proof. Repeat the proof of Theorem 8.5 with X in place of W. The time increments and the bound $|R_\Pi| \le \varpi(r_\Pi)V_2(X;\Pi)$ use only the continuity of the paths, and $V_2(X;\Pi) \to [X,X]_t$ (Theorem 8.11) replaces Theorem 5.16. The first-order sum converges to $\int_0^t \partial_x f(s, X_s)\,\mathrm{d}X_s$ by

the remark after Theorem 8.10, and the second-order sum $\sum_j \partial_{xx} f(t_j, X_{t_j})(\delta_j X)^2$ converges to $\int_0^t \partial_{xx} f(s, X_s)\Delta_s^2 \, \mathrm{d}s$ by Theorem 8.12. The result is an Itô process: $\partial_x f(s, X_s)\Delta_s$ is in $\mathbb{H}^2_{\mathrm{loc}}$, and the $\mathrm{d}s$-integrand is integrable on $[0, T]$, because $\partial_t f$, $\partial_x f$, $\partial_{xx} f$ are bounded along the path on $[0, T]$ and $\int_0^T (1 + |\Theta_s| + \Delta_s^2) \, \mathrm{d}s < \infty$. □

Remark 8.14 (Functions on an open set). If $D \subseteq \mathbb{R}$ is an open interval, X takes values in D, and $f \in C^{1,2}([0, \infty) \times D)$ (for instance $\log x$ or $1/x$ for a positive X), the formula still holds. Along each path, the values X_s, $s \le t$, form a compact subset of D at a positive distance d_ω from its complement; once $\mathrm{osc}_X(\|\Pi\|) < d_\omega/2$, every segment between X_{t_j} and $X_{t_{j+1}}$ lies in a compact subset of D, and the proof goes through unchanged.

Example 8.15 (Exponentials of Itô integrals)**.** Let $\theta \in \mathbb{H}^2_{\mathrm{loc}}$ and $Z_t = \exp\big(\int_0^t \theta_s \, \mathrm{d}W_s - \frac{1}{2}\int_0^t \theta_s^2 \, \mathrm{d}s\big)$, which is e^{X_t} for the Itô process $\mathrm{d}X_t = -\frac{1}{2}\theta_t^2 \, \mathrm{d}t + \theta_t \, \mathrm{d}W_t$. By Theorem 8.13 with $f(x) = e^x$ and $[X, X]_t = \int_0^t \theta_s^2 \, \mathrm{d}s$,

$$\mathrm{d}Z_t = Z_t\Big(-\frac{1}{2}\theta_t^2 \, \mathrm{d}t + \theta_t \, \mathrm{d}W_t\Big) + \frac{1}{2} Z_t \theta_t^2 \, \mathrm{d}t = Z_t \theta_t \, \mathrm{d}W_t.$$

So Z has no drift. If θ is deterministic, $\int_0^t \theta \, \mathrm{d}W \sim \mathcal{N}(0, \int_0^t \theta^2 \mathrm{d}s)$ (Theorem 7.10), so $\mathbb{E}[Z_s^2] = \exp(\int_0^s \theta_u^2 \mathrm{d}u)$, the integrand $Z\theta$ is in $\mathbb{H}^2$, and Z is a martingale. For random θ, conditions are given in Chapter 11, where Z is the density of Girsanov's change of measure.

8.4 Several Brownian motions

A market with several assets is driven by several sources of noise. A *d-dimensional Brownian motion* is a vector $W = (W^1, \dots, W^d)$ of independent Brownian motions, and $(\mathcal{F}_t)$ is a filtration for W if W_t is $\mathcal{F}_t$-measurable and the vector $W_t - W_s$ is independent of $\mathcal{F}_s$ for $s < t$. Then $(\mathcal{F}_t)$ is a filtration for each W^k, and the integrals $\int H \, \mathrm{d}W^k$, $H \in \mathbb{H}^2_{\mathrm{loc}}$, are those of Chapter 7. An *Itô process driven by* W is $X_t = x_0 + \int_0^t \Theta_s \, \mathrm{d}s + \sum_{k=1}^d \int_0^t \Delta_s^k \, \mathrm{d}W_s^k$, with Θ as in Theorem 8.9 and $\Delta^1, \dots, \Delta^d \in \mathbb{H}^2_{\mathrm{loc}}$; integrals with respect to X are defined as in Theorem 8.10.

> **Lemma 8.16** (Independent noises do not interact)**.** *For $H, G \in \mathbb{H}^2_{\mathrm{loc}}$ and $k, l \in \{1, \dots, d\}$,*
>
> $$\Big[\int H \, \mathrm{d}W^k, \int G \, \mathrm{d}W^l\Big]_t = \mathbf{1}_{\{k=l\}} \int_0^t H_s G_s \, \mathrm{d}s.$$
>
> *In particular $[W^k, W^l]_t = \mathbf{1}_{\{k=l\}} \, t$.*

Proof. For $k = l$ this is Theorem 8.11(i). Let $k \neq l$, $I = \int H \, \mathrm{d}W^k$, $J = \int G \, \mathrm{d}W^l$, and $C_\Pi(I, J) = \sum_j \delta_j I \, \delta_j J$.

Simple integrands. Let $H = \xi_i$ and $G = \zeta_i$ on $[s_i, s_{i+1})$ for a common grid $0 = s_0 < \cdots < s_p = t$, with ξ_i, ζ_i $\mathcal{F}_{s_i}$-measurable as in Theorem 7.1. Adding the points s_i to Π changes C_Π by at most $2p \, \mathrm{osc}_I(\|\Pi\|) \, \mathrm{osc}_J(\|\Pi\|)$, which tends to 0. Once they are in Π, $C_\Pi = \sum_i \xi_i \zeta_i \, c_i$, where c_i is the sum of $\delta_j W^k \delta_j W^l$ over the intervals of Π inside $[s_i, s_{i+1}]$: the difference of the covariation sums of W^k and W^l along $\Pi \cap [0, s_{i+1}]$ and $\Pi \cap [0, s_i]$. By Theorem 7.14(ii), $c_i \to 0$ in probability, and so does C_Π.

Integrands in $\mathbb{H}^2$. The increments of a square-integrable martingale N over a partition are orthogonal, so $\mathbb{E}[V_2(N;\Pi)] = \mathbb{E}[(N_t - N_0)^2]$ for every Π. Let H', G' be simple, with integrals I', J'. By the Cauchy–Schwarz inequality, first for sums and then for expectations, and the isometry,

$$\mathbb{E}|C_\Pi(I,J) - C_\Pi(I',J')| \le \mathbb{E}|C_\Pi(I-I',J)| + \mathbb{E}|C_\Pi(I',J-J')| \le \|H-H'\|\,\|G\| + \|H'\|\,\|G-G'\|,$$

with $\|H\|^2 = \mathbb{E}\int_0^t H_s^2\,\mathrm{d}s$. The bound does not depend on Π and is as small as we wish for suitable simple H', G'; with the simple case and Markov's inequality, $C_\Pi(I,J) \to 0$ in probability. For $H, G \in \mathbb{H}^2_{\mathrm{loc}}$, localize with Theorem 8.3 as in the proof of Theorem 8.4. □

By bilinearity and the argument of Theorem 8.11, Itô processes $X^i = x_0^i + \int \Theta^i\,\mathrm{d}s + \sum_k \int \Delta^{ik}\mathrm{d}W^k$ satisfy

$$[X^i, X^j]_t = \int_0^t \sum_{k=1}^{d} \Delta_s^{ik}\Delta_s^{jk}\,\mathrm{d}s, \tag{8.5}$$

which is the rule $\mathrm{d}W^k\mathrm{d}W^l = \mathbf{1}_{\{k=l\}}\mathrm{d}t$, and the decomposition is unique: the proof of Theorem 8.11(ii) applies with $[D,D]_t = \int_0^t \sum_k (\Delta_s^k - \tilde\Delta_s^k)^2\mathrm{d}s$.

Theorem 8.17 (Multidimensional Itô formula)**.** *Let $X = (X^1,\dots,X^n)$ be Itô processes driven by W, with $X_0 = x_0$, and let $f \in C^{1,2}([0,\infty)\times\mathbb{R}^n)$ (continuous $\partial_t f$, $\partial_i f$, $\partial_{ij} f$). Almost surely, for every $t \ge 0$,*

$$f(t,X_t) = f(0,x_0) + \int_0^t \partial_t f(s,X_s)\,\mathrm{d}s + \sum_{i=1}^n \int_0^t \partial_i f(s,X_s)\,\mathrm{d}X_s^i + \frac{1}{2}\sum_{i,j=1}^n \int_0^t \partial_{ij} f(s,X_s)\,\mathrm{d}[X^i,X^j]_s,$$

with $[X^i,X^j]$ given by (8.5). As in Theorem 8.14, $\mathbb{R}^n$ may be replaced by an open convex set in which X takes its values.

Proof. Repeat the proof of Theorem 8.13 with the second-order Taylor formula in $\mathbb{R}^n$, $f(t_j,y) - f(t_j,x) = \sum_i \partial_i f(t_j,x)(y-x)_i + \frac{1}{2}\sum_{i,l}\partial_{il} f(t_j,\xi)(y-x)_i(y-x)_l$ with ξ on the segment $[x,y]$. The first-order sums converge by Theorem 8.4, applied to each W^k, and the second-order sums $\sum_j \partial_{il} f(t_j,X_{t_j})\,\delta_j X^i\,\delta_j X^l$ converge to $\int_0^t \partial_{il} f(s,X_s)\,\mathrm{d}[X^i,X^l]_s$ by Theorem 8.12 and polarization, and the remainder is at most $\varpi(r_\Pi)\sum_{i,l}\sum_j |\delta_j X^i||\delta_j X^l| \le n\,\varpi(r_\Pi)\sum_i V_2(X^i;\Pi)$, which tends to 0 in probability. □

Correlated Brownian motions. Let $C = (\rho_{ij})$ be a correlation matrix and L a matrix with $LL^\top = C$ (a Cholesky factor). Then $B = LW$ has components $B^i = \sum_k L_{ik}W^k$ that are Brownian motions for $(\mathcal{F}_t)$: they are continuous, start at 0, and $B_t - B_s = L(W_t - W_s)$ is independent of $\mathcal{F}_s$, with $B_t^i - B_s^i \sim \mathcal{N}(0,(t-s)\rho_{ii}) = \mathcal{N}(0,t-s)$. By Theorem 8.16, $[B^i,B^j]_t = \sum_k L_{ik}L_{jk}t = \rho_{ij}t$, which is the rule $\mathrm{d}B^i\mathrm{d}B^j = \rho_{ij}\,\mathrm{d}t$. Since $\int H\,\mathrm{d}B^i = \sum_k L_{ik}\int H\,\mathrm{d}W^k$ (for simple H by definition, in general by density and the isometry), an Itô process driven by B is an Itô process driven by W, and Theorem 8.17 applies. For two stocks $\mathrm{d}S_t^i = \mu_i S_t^i\,\mathrm{d}t + \sigma_i S_t^i\,\mathrm{d}B_t^i$ with $[B^1,B^2]_t = \rho t$, it gives $\mathrm{d}[S^1,S^2]_t = \rho\sigma_1\sigma_2 S_t^1 S_t^2\,\mathrm{d}t$. Conversely, by Theorem 8.20 below, two Brownian motions for the same filtration with $[B^1,B^2]_t = \rho t$, $|\rho| < 1$, are of this form, with $W = (B^1, (B^2 - \rho B^1)/\sqrt{1-\rho^2})$. A random covariation is a different matter (Exercise 8.5).

Corollary 8.18 (Product rule)**.** *For Itô processes X and Y driven by W, almost surely, for every $t \geq 0$,*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s \, \mathrm{d}Y_s + \int_0^t Y_s \, \mathrm{d}X_s + [X, Y]_t,$$

that is, $\mathrm{d}(XY) = X \, \mathrm{d}Y + Y \, \mathrm{d}X + \mathrm{d}[X, Y]$.

Proof. Theorem 8.17 with $n = 2$ and $f(x, y) = xy$: $\partial_x f = y$, $\partial_y f = x$, $\partial_{xy} f = \partial_{yx} f = 1$ and $\partial_{xx} f = \partial_{yy} f = 0$. □

The covariation is the term that ordinary integration by parts misses. It vanishes when one of the processes has no $\mathrm{d}W$-part, for instance a deterministic discount factor: $\mathrm{d}(e^{-rt} S_t) = e^{-rt} \, \mathrm{d}S_t - r e^{-rt} S_t \, \mathrm{d}t$, which gives Theorem 8.8(ii) again.

Example 8.19 (The time integral of Brownian motion)**.** With $X_t = t$ and $Y = W$, $[X, W] = 0$ and $tW_t = \int_0^t s \, \mathrm{d}W_s + \int_0^t W_s \, \mathrm{d}s$. Writing $tW_t = \int_0^t t \, \mathrm{d}W_s$,

$$\int_0^t W_s \, \mathrm{d}s = \int_0^t (t - s) \, \mathrm{d}W_s \;\sim\; \mathcal{N}\Big(0, \int_0^t (t - s)^2 \mathrm{d}s\Big) = \mathcal{N}(0, t^3/3)$$

by Theorem 7.10. Early increments weigh more, having been part of the integral for longer. The same computation gives the integrated short rate of Gaussian models (Chapter 14).

The last result of the chapter identifies Brownian motion by two properties that Itô calculus can check: being a martingale, and having quadratic variation t.

Theorem 8.20 (Lévy's characterization of Brownian motion)**.** *Let $M = (M^1, \dots, M^d)$ be a process with continuous paths and $M_0 = 0$, such that each M^i is a martingale with respect to $(\mathcal{F}_t)$ and $[M^i, M^j]_t = \mathbf{1}_{\{i=j\}} \, t$ for all $t \geq 0$ and all i, j. Then M is a d-dimensional Brownian motion and $(\mathcal{F}_t)$ is a filtration for it.*

This is Shreve (2004b), Theorems 4.6.4 ($d = 1$) and 4.6.5 ($d = 2$), and Karatzas and Shreve (1991), Theorem 3.3.16, for every d and for continuous local martingales; in the latter the covariation is defined through the Doob–Meyer decomposition, which for continuous local martingales agrees with the limit in probability of sums used here (Revuz and Yor (1999), Chapter IV, Theorem 1.8). We prove the case in which the M^i are Itô integrals against W, which is the one we can reach with the tools of this chapter.

Proof when $M^i = \sum_k \int \Delta^{ik} \mathrm{d}W^k$ with $\Delta^{ik} \in \mathbb{H}^2_{\mathrm{loc}}$. Fix $u \in \mathbb{R}^d$ and let $Y = u \cdot M = \sum_i u_i M^i$, an Itô process driven by W with no drift and diffusion coefficients $\Gamma^k = \sum_i u_i \Delta^{ik}$. By bilinearity, $[Y, Y]_t = \sum_{i,j} u_i u_j [M^i, M^j]_t = |u|^2 t$, and by (8.5), $\int_0^t \sum_k (\Gamma^k_s)^2 \mathrm{d}s = |u|^2 t$. Apply Theorem 8.17 to the real and imaginary parts of $f(t, y) = \exp\left(iy + \tfrac{1}{2}|u|^2 t\right)$: since $\partial_t f = \tfrac{1}{2}|u|^2 f$, $\partial_{yy} f = -f$ and $\mathrm{d}[Y, Y]_t = |u|^2 \mathrm{d}t$, the $\mathrm{d}t$-terms cancel, and

$$f(t, Y_t) = 1 + \sum_k \int_0^t i \, f(s, Y_s) \, \Gamma^k_s \, \mathrm{d}W^k_s.$$

On $[0, T]$, $|f(s, Y_s)| \leq e^{|u|^2 T/2}$ and $\mathbb{E} \int_0^T \sum_k (\Gamma^k_s)^2 \mathrm{d}s = |u|^2 T$, so the integrands are in $\mathbb{H}^2$ and the real and imaginary parts of $f(t, Y_t)$ are martingales. For $s < t$, multiplying $\mathbb{E}[f(t, Y_t) \mid \mathcal{F}_s] = f(s, Y_s)$ by the bounded $\mathcal{F}_s$-measurable $e^{-iu \cdot M_s - |u|^2 t/2}$ gives

$$\mathbb{E}\big[e^{iu \cdot (M_t - M_s)} \,\big|\, \mathcal{F}_s\big] = e^{-|u|^2 (t-s)/2} \qquad \text{for every } u \in \mathbb{R}^d.$$

For $A \in \mathcal{F}_s$ with $\mathbb{P}(A) > 0$, the characteristic function of $M_t - M_s$ under $\mathbb{P}(\cdot \mid A)$ is therefore that of $\mathcal{N}(0, (t-s)I_d)$; characteristic functions determine laws, so $\mathbb{P}(\{M_t - M_s \in B\} \cap A) = \mathbb{P}(A)\,\mathcal{N}(0, (t-s)I_d)(B)$ for Borel $B \subseteq \mathbb{R}^d$: the increment is independent of $\mathcal{F}_s$ with law $\mathcal{N}(0, (t-s)I_d)$. Applying this to successive intervals, the increments of M over $0 \le t_1 < \cdots < t_m$ are independent with laws $\mathcal{N}(0, (t_{k+1} - t_k)I_d)$. So all the $M^i_{t_k}$ together form a centred Gaussian vector with $\mathrm{Cov}(M^i_s, M^j_t) = \mathbf{1}_{\{i=j\}} \min(s,t)$: each M^i is a Brownian motion by Theorem 5.7, and the M^i are independent (Theorem 5.8). □

For example, let $\mathrm{sgn}(x) = 1$ for $x > 0$ and -1 for $x \le 0$, and $B_t = \int_0^t \mathrm{sgn}(W_s)\,\mathrm{d}W_s$. Then $[B,B]_t = \int_0^t \mathrm{sgn}(W_s)^2 \mathrm{d}s = t$, so B is a Brownian motion: it moves with W when W is positive and against it when W is not (Figure 8.2); its relation to $|W|$ is Exercise 8.8. In Chapter 11, Lévy's theorem identifies the Brownian motion of Girsanov's theorem.

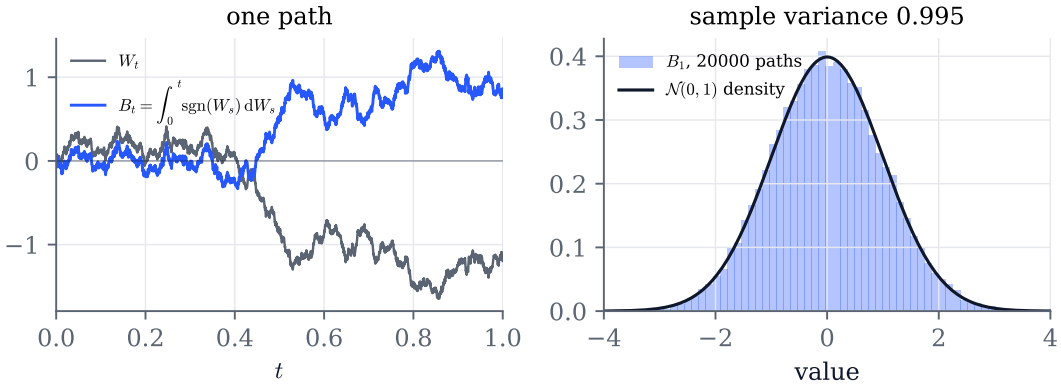

Figure 8.2: $B_t = \int_0^t \mathrm{sgn}(W_s)\,\mathrm{d}W_s$ by left-point sums. Left: one path of W and of B; during the long negative excursion of W, B moves against it. Right: the law of B_1 over 20000 paths, against the $\mathcal{N}(0,1)$ density. On the grid the increments $\mathrm{sgn}(W_{t_j})\,\delta_j W$ are exactly independent $\mathcal{N}(0, h_j)$, so this panel checks the computation; Theorem 8.20 is the statement that the limit is a Brownian motion.

▶ Interactive demo: covariations and the product rule for correlated Brownian motions with the correlation of your choice; and Lévy's theorem for integrands with $|\Delta| = 1$ that you select, against one with $|\Delta| \neq 1$.

Further reading

Shreve (2004b), Sections 4.4 and 4.6, covers the same ground with the same notation and more financial examples. Karatzas and Shreve (1991), Section 3.3, and Revuz and Yor (1999), Chapter IV, prove Itô's formula for continuous semimartingales, of which Itô processes are a special case, with the localization done in full. Øksendal (2003), Chapter 4, is a short and readable alternative, and Steele (2001), Chapter 8, develops Itô's formula and its first applications with much intuition. Le Gall (2016), Chapter 5, is a concise modern account that includes Lévy's characterization.

Chapter 9

Stochastic differential equations

The Black–Scholes stock satisfies $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$: over a short interval its relative change is a deterministic drift plus a Gaussian shock (Theorem 8.8). Most continuous-time models in finance have this shape, with coefficients that depend on the current state: a short rate is pulled back towards a long-run level, the volatility of a stock rises when its price falls. Such a model is a *stochastic differential equation* $\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t$. When does it have a solution, and only one? Which equations can be solved in closed form? How do we simulate the others? On the way we meet the Markov property and the generator, which lead to the pricing equations of Chapters 10 and 12.

Throughout, $T > 0$ is a fixed horizon, W is a Brownian motion and $(\mathcal{F}_t)$ a filtration for W (Theorem 5.11). We assume that every event of probability zero belongs to $\mathcal{F}_0$, so that changing a process on a null event does not affect adaptedness. Stochastic integrals are those of Chapter 7, for integrands in $\mathbb{H}^2$ (Theorem 7.3) or in $\mathbb{H}^2_{\mathrm{loc}}$ (Theorem 7.16); Itô processes and Itô's formula are those of Chapter 8, where the starting value of an Itô process may be any $\mathcal{F}_0$-measurable random variable (Theorem 8.9).

9.1 Stochastic differential equations

Let $b, \sigma : [0,T] \times \mathbb{R} \to \mathbb{R}$ be Borel functions, the *drift* and the *diffusion coefficient*, and let ξ be an $\mathcal{F}_0$-measurable random variable, the initial condition.

Definition 9.1 (Strong solution)**.** A *(strong) solution* of

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \qquad X_0 = \xi, \tag{9.1}$$

on $[0,T]$ is a process $X = (X_t)_{t\in[0,T]}$ such that

(i) X is adapted to $(\mathcal{F}_t)$ and every path $t \mapsto X_t(\omega)$ is continuous;

(ii) almost surely, $\int_0^T \left(|b(s, X_s)| + \sigma(s, X_s)^2\right) \mathrm{d}s < \infty$;

(iii) almost surely, for every $t \in [0,T]$, $X_t = \xi + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s$.

Condition (ii) is what (iii) needs. Since X is continuous and adapted, it is progressively measurable

(Karatzas and Shreve (1991), Proposition 1.1.13), hence so are $b(s, X_s)$ and $\sigma(s, X_s)$; by (ii), the first is integrable along almost every path and the second lies in $\mathbb{H}^2_{\mathrm{loc}}$. A solution is thus an Itô process (Theorem 8.9) whose coefficients at time t depend only on t and X_t. *Strong* means that W and $(\mathcal{F}_t)$ are given in advance; in a *weak* solution they are part of the answer (see the further reading).

The equation is a recipe for the next move. Over a short interval $[t, t+h]$,

$$X_{t+h} - X_t \approx b(t, X_t)\, h + \sigma(t, X_t)\,(W_{t+h} - W_t)\,:$$

given $\mathcal{F}_t$, the increment has mean about $b(t, X_t)h$ and variance about $\sigma(t, X_t)^2 h$. Iterating this approximation on a grid is the Euler–Maruyama scheme of Section 9.7.

Geometric Brownian motion is the case $b(t, x) = \mu x$, $\sigma(t, x) = \sigma x$; the Ornstein–Uhlenbeck and CEV models of Sections 9.4 and 9.6 are two more.

Even without noise, an equation of this kind can fail to have a solution, or have many.

Example 9.2 (What can go wrong)**.** Take $\sigma = 0$, so that (9.1) is an ordinary differential equation.

(a) *Explosion.* $x' = x^2$, $x(0) = 1$, is solved by $x(t) = 1/(1-t)$, which explodes at $t = 1$; the drift grows faster than linearly.

(b) *Branching.* $x' = 3|x|^{2/3}$, $x(0) = 0$, is solved by $x \equiv 0$ and, for every $a \geq 0$, by $x(t) = ((t-a)^+)^3$; the drift is not Lipschitz at 0.

The next theorem rules these out with a linear-growth bound and a Lipschitz condition.

9.2 Existence and uniqueness

The proofs below bound a mean square distance $\varphi(t)$ by an integral inequality, which Gronwall's lemma turns into a bound.

Lemma 9.3 (Gronwall)**.** *Let $\varphi : [0, T] \to [0, \infty)$ be Borel and integrable, and let $a, c \geq 0$ be such that $\varphi(t) \leq a + c \int_0^t \varphi(s)\, \mathrm{d}s$ for every $t \in [0, T]$. Then $\varphi(t) \leq a\, e^{ct}$ for every $t \in [0, T]$. In particular, if $a = 0$ then $\varphi \equiv 0$.*

Proof. We show by induction that, for every $n \geq 1$ and $t \in [0, T]$,

$$\varphi(t) \leq a \sum_{k=0}^{n-1} \frac{(ct)^k}{k!} + c^n \int_0^t \frac{(t-s)^{n-1}}{(n-1)!}\, \varphi(s)\, \mathrm{d}s. \tag{9.2}$$

For $n = 1$ this is the hypothesis. If it holds for n, bound $\varphi(s)$ in the last integral by $a + c\int_0^s \varphi(u)\, \mathrm{d}u$, and exchange the order of integration (Tonelli):

$$c^n \int_0^t \frac{(t-s)^{n-1}}{(n-1)!} \Big(a + c \int_0^s \varphi(u)\, \mathrm{d}u\Big) \mathrm{d}s = a \frac{(ct)^n}{n!} + c^{n+1} \int_0^t \frac{(t-u)^n}{n!}\, \varphi(u)\, \mathrm{d}u,$$

which gives (9.2) for $n+1$. The last term of (9.2) is at most $c^n T^{n-1} \int_0^T \varphi/(n-1)!$, which tends to 0, and the sum tends to ae^{ct}. □

Lemma 9.4 (Maximal estimates). *Let $t \in [0,T]$ and let H be progressively measurable with $\mathbb{E}\int_0^t H_s^2\,\mathrm{d}s < \infty$. Then*

$$\mathbb{E}\Big[\sup_{u\le t}\Big|\int_0^u H_s\,\mathrm{d}s\Big|^2\Big] \le t\,\mathbb{E}\int_0^t H_s^2\,\mathrm{d}s, \qquad \mathbb{E}\Big[\sup_{u\le t}\Big|\int_0^u H_s\,\mathrm{d}W_s\Big|^2\Big] \le 4\,\mathbb{E}\int_0^t H_s^2\,\mathrm{d}s.$$

Proof. The first bound is the Cauchy–Schwarz inequality, $|\int_0^u H_s\,\mathrm{d}s|^2 \le u\int_0^u H_s^2\,\mathrm{d}s \le t\int_0^t H_s^2\,\mathrm{d}s$. For the second, $I_u = \int_0^u H_s\,\mathrm{d}W_s$ is a continuous martingale on $[0,t]$ with $\mathbb{E}[I_t^2] = \mathbb{E}\int_0^t H_s^2\,\mathrm{d}s$ (Theorem 7.7). Doob's maximal inequality for continuous square-integrable martingales gives $\mathbb{E}[\sup_{u\le t} I_u^2] \le 4\,\mathbb{E}[I_t^2]$ (Karatzas and Shreve (1991), Theorem 1.3.8(iv)). □

Theorem 9.5 (Existence and uniqueness). *Suppose that there is a constant K such that, for all $t \in [0,T]$ and $x, y \in \mathbb{R}$,*

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \le K|x-y|, \tag{9.3}$$
$$|b(t,x)| + |\sigma(t,x)| \le K(1+|x|). \tag{9.4}$$

(a) Pathwise uniqueness. *If X and Y are solutions of (9.1) with the same initial condition ξ, then $\mathbb{P}(X_t = Y_t \text{ for all } t \in [0,T]) = 1$. Only (9.3) is needed for this.*

(b) Existence. *If $\mathbb{E}[\xi^2] < \infty$, then (9.1) has a solution X, and $\mathbb{E}[\sup_{t\le T} X_t^2] \le C(1+\mathbb{E}[\xi^2])$ for a constant C that depends only on K and T.*

Proof. (a) Let $D = X - Y$. Subtracting the two integral equations, $D_t = \int_0^t \beta_s\,\mathrm{d}s + \int_0^t \gamma_s\,\mathrm{d}W_s$ with $\beta_s = b(s,X_s) - b(s,Y_s)$ and $\gamma_s = \sigma(s,X_s) - \sigma(s,Y_s)$, and (9.3) gives $|\beta_s| \le K|D_s|$ and $|\gamma_s| \le K|D_s|$. We know nothing about the moments of D, so we do not apply Itô's formula to D^2 but to a bounded function of D: $f(x) = x^2/(1+x^2)$, which takes values in $[0,1)$, vanishes only at 0, and has $f'(x) = 2x/(1+x^2)^2$ and $f''(x) = (2-6x^2)/(1+x^2)^3$, so that $|xf'(x)| \le \min(2f(x), \frac{1}{2})$ and $x^2|f''(x)| \le 6f(x)$. By Itô's formula (Theorem 8.13, with the stochastic integral understood as in Theorem 7.16),

$$f(D_t) = \int_0^t \Big(f'(D_s)\beta_s + \tfrac{1}{2}f''(D_s)\gamma_s^2\Big)\mathrm{d}s + \int_0^t f'(D_s)\gamma_s\,\mathrm{d}W_s.$$

The integrand of the stochastic integral is bounded by $K|D_s f'(D_s)| \le K/2$, so the integral is a martingale with mean zero (Theorem 7.7). The $\mathrm{d}s$ integrand is bounded by $(2K+3K^2)f(D_s) \le 2K + 3K^2$. Taking expectations and using Fubini, $\varphi(t) = \mathbb{E}[f(D_t)]$ satisfies $\varphi(t) \le (2K + 3K^2)\int_0^t \varphi(s)\,\mathrm{d}s$. The function φ is bounded by 1 and continuous (bounded convergence and continuity of D), so Gronwall's lemma with $a = 0$ gives $\varphi \equiv 0$. Hence, for each t, $D_t = 0$ almost surely. A countable intersection of almost sure events is almost sure, so almost surely $D_t = 0$ for every rational t, and then for every t by continuity.

(b) We construct the solution by *Picard iteration*, as for ordinary differential equations. For a process Z write $\|Z\|_t^2 = \mathbb{E}[\sup_{u\le t} Z_u^2]$. Set $X_t^0 = \xi$ and

$$X_t^{n+1} = \xi + \int_0^t b(s,X_s^n)\,\mathrm{d}s + \int_0^t \sigma(s,X_s^n)\,\mathrm{d}W_s, \qquad t \in [0,T].$$

Step 1: the iterates are well defined. Suppose X^n is continuous, adapted and $\|X^n\|_T < \infty$ (true for $n = 0$). By (9.4), $b^2 + \sigma^2 \le (|b| + |\sigma|)^2 \le 2K^2(1+x^2)$, so $\mathbb{E}\int_0^T (b^2+\sigma^2)(s,X_s^n)\,\mathrm{d}s \le$

$2K^2T(1+\|X^n\|_T^2)<\infty$. Hence X^{n+1} is defined, continuous and adapted, and by Theorem 9.4 and $(u+v+w)^2\le 3(u^2+v^2+w^2)$, $\|X^{n+1}\|_T^2\le 3\mathbb{E}[\xi^2]+3(T+4)\mathbb{E}\int_0^T(b^2+\sigma^2)(s,X_s^n)\,\mathrm{d}s<\infty$.

Step 2: contraction. Let $\Delta_n(t)=\|X^{n+1}-X^n\|_t^2$. The difference $X^{n+1}-X^n$ has drift $b(s,X_s^n)-b(s,X_s^{n-1})$ and diffusion $\sigma(s,X_s^n)-\sigma(s,X_s^{n-1})$, both bounded by $K|X_s^n-X_s^{n-1}|$. By $(u+v)^2\le 2u^2+2v^2$ and Theorem 9.4,

$$\Delta_n(t)\le 2(t+4)K^2\,\mathbb{E}\int_0^t|X_s^n-X_s^{n-1}|^2\,\mathrm{d}s\le L\int_0^t\Delta_{n-1}(s)\,\mathrm{d}s,\qquad L=2K^2(T+4).$$

(Δ_{n-1} is nondecreasing, hence Borel.) Since $\Delta_0(t)\le M:=\Delta_0(T)<\infty$, induction gives $\Delta_n(t)\le M(Lt)^n/n!$.

Step 3: convergence. By Chebyshev's inequality, $\mathbb{P}\big(\sup_{t\le T}|X_t^{n+1}-X_t^n|>2^{-n}\big)\le 4^nM(LT)^n/n!$, which is summable. By the Borel–Cantelli lemma, almost surely $\sup_t|X_t^{n+1}-X_t^n|\le 2^{-n}$ for all large n, and on this event X^n converges uniformly on $[0,T]$ to a continuous limit X. Set $X=0$ on the exceptional null event, which lies in $\mathcal{F}_0$; then X is adapted. By Fatou's lemma and the triangle inequality for $\|\cdot\|_T$, $\|X-X^n\|_T\le\sum_{k\ge n}\sqrt{\Delta_k(T)}\to 0$; in particular $\|X\|_T<\infty$.

Step 4: X solves the equation. By (9.3) and Theorem 9.4, as in Step 2,

$$\mathbb{E}\Big[\sup_{t\le T}\Big|\int_0^t\big(b(s,X_s^n)-b(s,X_s)\big)\mathrm{d}s+\int_0^t\big(\sigma(s,X_s^n)-\sigma(s,X_s)\big)\mathrm{d}W_s\Big|^2\Big]\le LT\,\|X^n-X\|_T^2\to 0.$$

So the right-hand side of the recursion converges, uniformly in t in mean square, to $\xi+\int_0^t b(s,X_s)\,\mathrm{d}s+\int_0^t\sigma(s,X_s)\,\mathrm{d}W_s$, while the left-hand side X^{n+1} converges to X. The two limits agree almost surely for each t, and, both being continuous in t, for all t at once. Condition (ii) of Theorem 9.1 holds because $\mathbb{E}\int_0^T(b^2+\sigma^2)(s,X_s)\,\mathrm{d}s<\infty$.

Step 5: the moment bound. As in Step 1, $\varphi(t)=\|X\|_t^2$ satisfies $\varphi(t)\le 3\mathbb{E}[\xi^2]+6K^2T(T+4)+6K^2(T+4)\int_0^t\varphi(s)\,\mathrm{d}s$, and Gronwall's lemma gives $\varphi(T)\le\big(3\mathbb{E}[\xi^2]+6K^2T(T+4)\big)e^{6K^2T(T+4)}$. □

Part (a) is a special case of Karatzas and Shreve (1991), Theorem 5.2.5, and part (b) is their Theorem 5.2.9 (which bounds $\sup_t\mathbb{E}[X_t^2]$; the bound on $\mathbb{E}[\sup_t X_t^2]$ is Step 5). Their theorem also shows that X_t is measurable with respect to $\sigma(\xi,W_s:s\le t)$ augmented by the null events.

Remark 9.6 (On the hypotheses).

(a) The Lipschitz constant must be uniform in $t\in[0,T]$. Given (9.3), the growth bound (9.4) holds, with a possibly larger constant, if and only if $|b(t,0)|+|\sigma(t,0)|$ is bounded on $[0,T]$, because $|b(t,x)|\le|b(t,0)|+K|x|$. The proof uses (9.4) only in Steps 1, 4 and 5, and with small changes there it goes through under the weaker condition $\int_0^T\big(|b(t,0)|+\sigma(t,0)^2\big)\,\mathrm{d}t<\infty$; the constant C then also depends on this integral.

(b) The Lipschitz condition can be weakened in two useful ways. If b and σ are only *locally* Lipschitz in x (Lipschitz on every bounded set, with constants uniform in t), pathwise uniqueness still holds (Karatzas and Shreve (1991), Theorem 5.2.5), although a solution on $[0,T]$ need not exist (Theorem 9.2(a)). In dimension one, it is enough that b be Lipschitz and that $|\sigma(t,x)-\sigma(t,y)|\le h(|x-y|)$ for a strictly increasing h with $h(0)=0$ and $\int_0^\varepsilon h(u)^{-2}\,\mathrm{d}u=\infty$ for every $\varepsilon>0$, for instance $h(u)=c\sqrt{u}$, for pathwise uniqueness (not existence): this is the Yamada–Watanabe criterion (Karatzas and Shreve (1991), Proposition 5.2.13), which covers the square-root diffusions of Section 9.6 and Chapter 14.

9.3 Linear equations

When the coefficients are affine in x, the equation can be solved explicitly, as a linear ordinary differential equation is solved with an integrating factor.

Proposition 9.7 (Linear SDEs)**.** *Let $\mu_0, \mu_1, \sigma_0, \sigma_1 : [0,T] \to \mathbb{R}$ be bounded Borel functions and ξ an $\mathcal{F}_0$-measurable random variable. Let*

$$Z_t = \exp\Big(\int_0^t \sigma_1(s)\,\mathrm{d}W_s + \int_0^t (\mu_1(s) - \tfrac{1}{2}\sigma_1(s)^2)\mathrm{d}s\Big).$$

The equation

$$\mathrm{d}X_t = (\mu_0(t) + \mu_1(t)X_t)\mathrm{d}t + (\sigma_0(t) + \sigma_1(t)X_t)\mathrm{d}W_t, \qquad X_0 = \xi, \tag{9.5}$$

has the unique solution

$$X_t = Z_t\Big(\xi + \int_0^t \frac{\mu_0(s) - \sigma_0(s)\sigma_1(s)}{Z_s}\,\mathrm{d}s + \int_0^t \frac{\sigma_0(s)}{Z_s}\,\mathrm{d}W_s\Big). \tag{9.6}$$

In particular:

(a) ($\mu_0 = \sigma_0 = 0$, $\mu_1 = \mu$, $\sigma_1 = \sigma$ constant) geometric Brownian motion $S_t = S_0 \exp\big(\sigma W_t + (\mu - \frac{\sigma^2}{2})t\big)$ is the unique solution of $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$;

(b) ($\sigma_1 = 0$, additive noise) with $B(t) = \int_0^t \mu_1(s)\,\mathrm{d}s$,

$$X_t = e^{B(t)}\xi + \int_0^t e^{B(t)-B(s)}\mu_0(s)\,\mathrm{d}s + \int_0^t e^{B(t)-B(s)}\sigma_0(s)\,\mathrm{d}W_s;$$

if $\xi = x_0$ is a constant, X_t is normal with mean $e^{B(t)}x_0 + \int_0^t e^{B(t)-B(s)}\mu_0(s)\,\mathrm{d}s$ and variance $\int_0^t e^{2(B(t)-B(s))}\sigma_0(s)^2\,\mathrm{d}s$.

Proof. Uniqueness. The coefficients $(t,x) \mapsto \mu_0(t) + \mu_1(t)x$ and $\sigma_0(t) + \sigma_1(t)x$ satisfy (9.3) with $K = \sup|\mu_1| + \sup|\sigma_1|$, so Theorem 9.5(a) applies. It requires no moment of ξ.

The formula is a solution. By Itô's formula for the exponential of the Itô process $\int_0^t \sigma_1\,\mathrm{d}W + \int_0^t(\mu_1 - \frac{1}{2}\sigma_1^2)\,\mathrm{d}s$ (Theorem 8.13), $\mathrm{d}Z_t = Z_t(\mu_1(t)\,\mathrm{d}t + \sigma_1(t)\,\mathrm{d}W_t)$ with $Z_0 = 1$. Let U_t be the bracket in (9.6), so that $\mathrm{d}U_t = Z_t^{-1}\big((\mu_0 - \sigma_0\sigma_1)\,\mathrm{d}t + \sigma_0\,\mathrm{d}W_t\big)$. Since $\int_0^s \sigma_1\,\mathrm{d}W$ is normal (Theorem 7.10) and the coefficients are bounded, $\mathbb{E}[Z_s^2]$ and $\mathbb{E}[Z_s^{-2}]$ are bounded on $[0,T]$; so Z and U are Itô processes with integrands in $\mathbb{H}^2$. By the product rule (Theorem 8.18), with $\mathrm{d}[Z,U]_t = Z_t\sigma_1(t)\cdot Z_t^{-1}\sigma_0(t)\,\mathrm{d}t$ (Theorem 8.11),

$$\mathrm{d}(Z_tU_t) = Z_t\,\mathrm{d}U_t + U_t\,\mathrm{d}Z_t + \mathrm{d}[Z,U]_t = (\mu_0 + \mu_1 Z_tU_t)\mathrm{d}t + (\sigma_0 + \sigma_1 Z_tU_t)\mathrm{d}W_t,$$

and $Z_0U_0 = \xi$. So $X = ZU$ solves (9.5).

Special cases. (a) is the case $Z_t = S_t/S_0$, $U \equiv S_0$. In (b), $Z_t = e^{B(t)}$ is deterministic and can be moved inside the integrals. The stochastic integral of the deterministic integrand $e^{B(t)-B(s)}\sigma_0(s)$, $s \le t$, is normal with mean 0 and variance $\int_0^t e^{2(B(t)-B(s))}\sigma_0(s)^2\,\mathrm{d}s$ (Theorem 7.10); the other two terms are constants. □

The formula comes from the integrating factor Z^{-1}: $\mathrm{d}(Z_t^{-1}X_t) = Z_t^{-1}((\mu_0 - \sigma_0\sigma_1)\,\mathrm{d}t + \sigma_0\,\mathrm{d}W_t)$ no longer involves X; the term $-\sigma_0\sigma_1$ comes from the covariation of Z^{-1} and X.

9.4 The Ornstein–Uhlenbeck process and the Vasicek model

Proposition 9.8 (Ornstein–Uhlenbeck process)**.** *Let $\kappa > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$, let ξ be $\mathcal{F}_0$-measurable, and write $v(t) = \sigma^2(1 - e^{-2\kappa t})/(2\kappa)$. The equation $\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, $X_0 = \xi$, has the unique solution*

$$X_t = \theta + (\xi - \theta)e^{-\kappa t} + \sigma \int_0^t e^{-\kappa(t-s)}\,\mathrm{d}W_s. \tag{9.7}$$

(i) For $0 \le t \le u$, $X_u = \theta + (X_t - \theta)e^{-\kappa(u-t)} + \sigma\int_t^u e^{-\kappa(u-s)}\,\mathrm{d}W_s$, and the last term is normal with mean 0 and variance $v(u-t)$, and independent of $\mathcal{F}_t$.

(ii) If $\xi = x_0$ is a constant, X_t is normal with mean $\theta + (x_0 - \theta)e^{-\kappa t}$ and variance $v(t)$. As $t \to \infty$, its law converges to $\mathcal{N}(\theta, \frac{\sigma^2}{2\kappa})$, whatever x_0.

(iii) If $\xi \sim \mathcal{N}(\theta, \frac{\sigma^2}{2\kappa})$, then $X_t \sim \mathcal{N}(\theta, \frac{\sigma^2}{2\kappa})$ for every $t \ge 0$: this is the stationary law.

Formula (9.7) gives the solution on every horizon at once, which gives a meaning to $t \to \infty$ in (ii).

Proof. This is Theorem 9.7 with $\mu_0 = \kappa\theta$, $\mu_1 = -\kappa$, $\sigma_0 = \sigma$, $\sigma_1 = 0$: then $Z_t = e^{-\kappa t}$ and (9.6) becomes (9.7), since $e^{-\kappa t}\int_0^t e^{\kappa s}\kappa\theta\,\mathrm{d}s = \theta(1 - e^{-\kappa t})$.

(i) Write $e^{-\kappa(u-s)} = e^{-\kappa(u-t)}e^{-\kappa(t-s)}$ in (9.7) at time u and split $\int_0^u = \int_0^t + \int_t^u$; the terms up to time t recombine into $e^{-\kappa(u-t)}(X_t - \theta)$. Let $J = \int_t^u e^{-\kappa(u-s)}\,\mathrm{d}W_s$ and, for a partition $t = s_0 < \cdots < s_m = u$, $J_m = \sum_j e^{-\kappa(u-s_j)}(W_{s_{j+1}} - W_{s_j})$, the integral of a simple integrand (Theorem 7.1). Each increment $W_{s_{j+1}} - W_{s_j}$ is normal and independent of $\mathcal{F}_{s_j}$, which contains $\mathcal{F}_t$ and the earlier increments (Theorem 5.11); so J_m is normal and independent of $\mathcal{F}_t$. As the mesh goes to 0, $J_m \to J$ in L^2 by the Itô isometry (the integrands converge uniformly on $[t,u]$), so for $A \in \mathcal{F}_t$, $\lambda \in \mathbb{R}$, the identity $\mathbb{E}[\mathbf{1}_A e^{i\lambda J_m}] = \mathbb{P}(A)e^{-\lambda^2\operatorname{Var}(J_m)/2}$ passes to the limit: $\mathbb{E}[\mathbf{1}_A e^{i\lambda J}] = \mathbb{P}(A)e^{-\lambda^2 w/2}$ with $w = \int_t^u e^{-2\kappa(u-s)}\,\mathrm{d}s$. Taking $A = \Omega$ gives $J \sim \mathcal{N}(0, w)$, and $\sigma^2 w = v(u-t)$; for $\mathbb{P}(A) > 0$ the law of J given A is that of J, so J is independent of $\mathcal{F}_t$.

(ii) Take $t = 0$ in (i). As $t \to \infty$, the mean tends to θ and $v(t)$ to $\sigma^2/(2\kappa)$; the characteristic functions of X_t converge to that of $\mathcal{N}(\theta, \sigma^2/(2\kappa))$, which is convergence in law by Lévy's continuity theorem.

(iii) By (i) with $t = 0$, $X_t = \theta + (\xi - \theta)e^{-\kappa t} + \sigma J$ with σJ independent of $\mathcal{F}_0$, hence of ξ. A sum of independent normal variables is normal, with mean θ and variance $e^{-2\kappa t}\frac{\sigma^2}{2\kappa} + v(t) = \frac{\sigma^2}{2\kappa}$. □

The parameters read directly off the solution. The mean forgets the starting point at rate κ: the gap $\mathbb{E}[X_t] - \theta$ halves every $\log 2/\kappa$ units of time. The variance does not grow without bound, as it would for $x_0 + \sigma W_t$, but levels off at $\sigma^2/(2\kappa)$, where the pull towards θ balances the noise. Both features need $\kappa > 0$: for $\kappa \le 0$ the same computation writes X_t as a function of ξ plus an independent normal variable whose variance tends to infinity, so the characteristic function of X_t tends to 0 at every $\lambda \neq 0$, and no law stays invariant.

The Vasicek model. Vasicek (1977) models the short rate as an Ornstein–Uhlenbeck process $\mathrm{d}r_t = \kappa(\theta - r_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$: rates are pulled towards a long-run level, and Gaussian rates make bond prices explicit (Chapter 14). The price of this tractability is that $\mathbb{P}(r_t < 0) > 0$ for every $t > 0$, by (ii). With the parameters of Figure 9.1 ($\theta = 4\%$, $\kappa = 1.5$, $\sigma = 2\%$), the stationary probability of a negative rate is $N\big(-\theta\sqrt{2\kappa}/\sigma\big) \approx 2.7 \times 10^{-4}$.

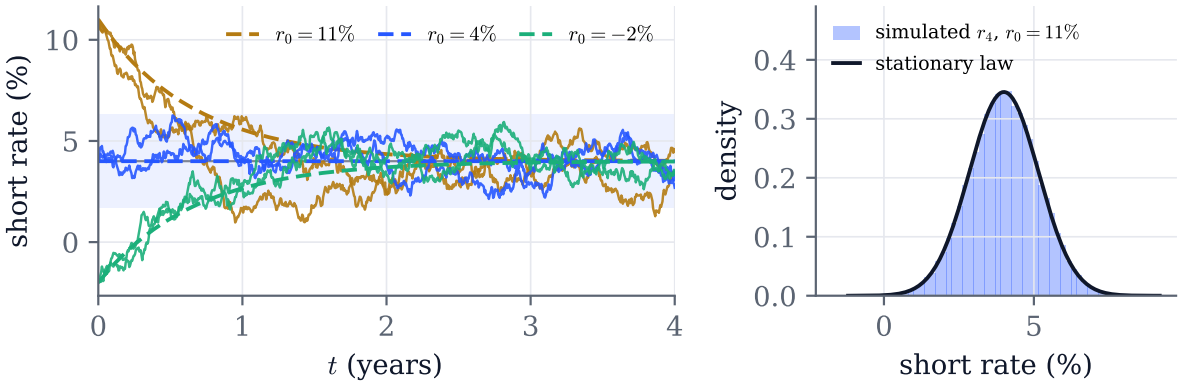

Figure 9.1: Left: Vasicek short rates with $\kappa = 1.5$, $\theta = 4\%$, $\sigma = 2\%$, two paths from each of three starting points, simulated exactly with the transition law of Theorem 9.8(i). Dashed: the means $\mathbb{E}[r_t]$; shaded: $\theta \pm 2$ stationary standard deviations. Right: the simulated law of r_4 started from 11%, against the stationary law $\mathcal{N}(\theta, \sigma^2/(2\kappa))$.

▶ Interactive demo: Ornstein–Uhlenbeck paths and their law at any time, for any κ, θ, σ and starting point, against the stationary law.

9.5 The Markov property and the generator

The coefficients of (9.1) depend on the present state only, not on the past of the path. Since the future increments of W are independent of the past, the future of X should depend on the past only through the present value X_t.

Proposition 9.9 (Markov property)**.** *Assume the hypotheses of Theorem 9.5, and let X be the solution of (9.1) with $\mathbb{E}[\xi^2] < \infty$. For $t \in [0,T]$ and $x \in \mathbb{R}$, let $X^{t,x}$ be the solution on $[t,T]$ of the same equation started at time t from x: $X^{t,x}_u = x + \int_t^u b(s, X^{t,x}_s)\,\mathrm{d}s + \int_t^u \sigma(s, X^{t,x}_s)\,\mathrm{d}W_s$. Then for every bounded Borel $h : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}\big[h(X_T) \,\big|\, \mathcal{F}_t\big] = g(t, X_t), \qquad \text{where } g(t,x) = \mathbb{E}\big[h(X^{t,x}_T)\big].$$

In particular $\mathbb{E}[h(X_T) \mid \mathcal{F}_t] = \mathbb{E}[h(X_T) \mid X_t]$: given the present state, the rest of the information available at time t does not help to predict $h(X_T)$.

The process $\tilde{W}_s = W_{t+s} - W_t$ is a Brownian motion (Theorem 5.9(c)), and $(\mathcal{F}_{t+s})_{s \ge 0}$ is a filtration for it by Theorem 5.11(iii). An integral $\int_t^u H\,\mathrm{d}W$ equals $\int_0^{u-t} H_{t+s}\,\mathrm{d}\tilde{W}_s$: both are limits of the same sums for simple integrands. So $X^{t,x}$ exists by Theorem 9.5 on $[t,T]$. The last statement follows from the first by the tower property. For the general case we cite the result: it is stated without proof in Shreve (2004b), Theorem 6.3.1, and proved for coefficients that do not depend

on t, with the Brownian filtration, in Øksendal (2003), Theorem 7.1.2. The difficulty is to show that $X_T^{t,x}$ depends measurably on x, so that x can be replaced by X_t. For the equations solved in closed form, the solution shows the structure directly.

Proof for geometric Brownian motion and the Ornstein–Uhlenbeck process. For geometric Brownian motion, $S_T = S_t \exp\big(\sigma(W_T - W_t) + (\mu - \frac{\sigma^2}{2})(T-t)\big)$, where S_t is $\mathcal{F}_t$-measurable and $W_T - W_t$ is independent of $\mathcal{F}_t$. The independence lemma (Theorem 1.19) gives $\mathbb{E}[h(S_T) \mid \mathcal{F}_t] = g(t, S_t)$ with $g(t,x) = \mathbb{E}\big[h\big(x \exp(\sigma(W_T - W_t) + (\mu - \frac{\sigma^2}{2})(T-t))\big)\big]$. The variable inside h is $S_T^{t,x}$, by Theorem 9.7(a) applied on $[t, T]$.

For the Ornstein–Uhlenbeck process, Theorem 9.8(i) gives $X_T = \theta + (X_t - \theta)e^{-\kappa(T-t)} + \sigma J$ with $J = \int_t^T e^{-\kappa(T-s)}\,\mathrm{d}W_s$ independent of $\mathcal{F}_t$, and the same formula with X_t replaced by x is $X_T^{t,x}$. The independence lemma concludes as before. □

By the Markov property, prices at time t of payoffs $h(S_T)$ are functions $v(t, S_t)$ (Chapter 10 onwards), which solve equations built from the following operator.

Definition 9.10 (Generator)**.** The *generator* of (9.1) at time t is the operator that maps $f \in C^2(\mathbb{R})$ to the function

$$(\mathcal{L}_t f)(x) = \tfrac{1}{2}\sigma(t,x)^2 f''(x) + b(t,x) f'(x).$$

When b and σ do not depend on t we write $\mathcal{L}$. The process $M_t^f = f(X_t) - f(X_0) - \int_0^t (\mathcal{L}_s f)(X_s)\,\mathrm{d}s$ is the *local martingale associated with* f; it is a martingale under the integrability condition below.

Proposition 9.11 (Martingales of a diffusion)**.** *Let X be a solution of (9.1) and $f \in C^2(\mathbb{R})$.*

(a) $M_t^f = \int_0^t f'(X_s)\sigma(s, X_s)\,\mathrm{d}W_s$, a local martingale, and a martingale if $\mathbb{E}\int_0^T f'(X_s)^2\sigma(s,X_s)^2\,\mathrm{d}s < \infty$. This holds, for instance, under the hypotheses of Theorem 9.5 with $\mathbb{E}[\xi^2] < \infty$ when f' is bounded.

*(b) If M^f is a martingale and $\mathbb{E}|f(X_0)| < \infty$, then $\mathbb{E}[f(X_t)] = \mathbb{E}[f(X_0)] + \int_0^t \mathbb{E}[(\mathcal{L}_s f)(X_s)]\,\mathrm{d}s$ whenever $\int_0^T \mathbb{E}|(\mathcal{L}_s f)(X_s)|\,\mathrm{d}s < \infty$ (*Dynkin's formula*).*

Proof. (a) Since $\mathrm{d}[X,X]_t = \sigma(t,X_t)^2\,\mathrm{d}t$ (Theorem 8.11), Itô's formula (Theorem 8.13) gives $\mathrm{d}f(X_t) = f'(X_t)(b\,\mathrm{d}t + \sigma\,\mathrm{d}W_t) + \frac{1}{2}f''(X_t)\sigma^2\,\mathrm{d}t = (\mathcal{L}_t f)(X_t)\,\mathrm{d}t + f'(X_t)\sigma(t,X_t)\,\mathrm{d}W_t$. The integrand $f'(X)\sigma(\cdot, X)$ is progressively measurable and, since $f'(X)$ is continuous, square integrable along almost every path by Theorem 9.1(ii); so M^f is a local martingale (Theorem 7.16), and a martingale when the integrand is in $\mathbb{H}^2$ (Theorem 7.7). If $|f'| \le c$, then by (9.4) $\mathbb{E}\int_0^T f'(X_s)^2\sigma(s,X_s)^2\,\mathrm{d}s \le 2c^2K^2(T + \int_0^T \mathbb{E}[X_s^2]\,\mathrm{d}s) < \infty$ by Theorem 9.5(b).

(b) Take expectations in the definition of M^f, using $\mathbb{E}[M_t^f] = 0$ and Fubini. □

For example, with $f(x) = x$ the Ornstein–Uhlenbeck generator gives $\mathcal{L}f(x) = \kappa(\theta - x)$, so $m(t) = \mathbb{E}[X_t]$ satisfies $m(t) = x_0 + \int_0^t \kappa(\theta - m(s))\,\mathrm{d}s$, that is $m' = \kappa(\theta - m)$: the mean found in Theorem 9.8 without solving the equation. Dynkin's formula reads: the drift moves the mean of $f(X_t)$ at rate bf', and the noise adds $\frac{1}{2}\sigma^2 f''$, through the convexity of f.

Remark 9.12 (Towards the pricing equations). Let $u \in C^{1,2}([0,T]\times\mathbb{R})$ solve $\partial_t u + \mathcal{L}_t u = 0$ with $u(T,x) = h(x)$. By Itô's formula, $\mathrm{d}u(t,X_t) = \partial_x u\,\sigma(t,X_t)\,\mathrm{d}W_t$; if this local martingale is a martingale, $u(t,X_t) = \mathbb{E}[h(X_T)\mid\mathcal{F}_t] = g(t,X_t)$ in the notation of Theorem 9.9. This is the Feynman–Kac formula of Chapter 12; with discounting and the Black–Scholes generator it becomes the Black–Scholes equation of Chapter 10.

9.6 Local volatility and the CEV model

In the Black–Scholes model the volatility of returns is constant, but implied volatilities depend on the strike (Chapter 15). The simplest models that reproduce such prices let the volatility depend on time and on the current price.

Definition 9.13 (Local volatility model)**.** Let $\mu\in\mathbb{R}$ and let $\sigma:[0,T]\times(0,\infty)\to[0,\infty)$ be Borel. A *local volatility model* for a stock price is a positive solution S, with $S_0>0$, of

$$\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma(t,S_t)S_t\,\mathrm{d}W_t.$$

In this section $\sigma(t,x)$ is a *relative* volatility, the volatility of returns $\mathrm{d}S_t/S_t$; the diffusion coefficient in the sense of Theorem 9.1 is $x\sigma(t,x)$. The *CEV model* is the case $\sigma(t,x)=\sigma x^{\gamma-1}$, with constants $\sigma>0$ and $\gamma\ge 0$.

Under a risk-neutral measure μ becomes the interest rate r (Chapter 11), and $\sigma(t,x)$ can be read off call prices by Dupire's formula (Chapter 15). If σ is bounded and Lipschitz in $\log x$, the model is well posed.

Proposition 9.14 (Bounded local volatility)**.** *Suppose that there are constants M and K such that $\sigma(t,x)\le M$ and $|\sigma(t,x)-\sigma(t,x')|\le K|\log x-\log x'|$ for all $t\in[0,T]$ and $x,x'>0$. Then for every $S_0>0$ there is exactly one positive solution, and it satisfies*

$$S_t = S_0\exp\Big(\int_0^t \sigma(u,S_u)\,\mathrm{d}W_u + \int_0^t (\mu - \tfrac12\sigma(u,S_u)^2)\mathrm{d}u\Big). \tag{9.8}$$

Proof. Uniqueness. Set $a(t,x)=x\sigma(t,x)$ for $x>0$ and $a(t,x)=0$ for $x\le 0$. For $0<x'\le x$, $|a(t,x)-a(t,x')|\le (x-x')\sigma(t,x)+x'K\log(x/x')\le (M+K)(x-x')$, since $\log y\le y-1$; and $|a(t,x)|\le M|x|$. So the equation with coefficients μx and a on $\mathbb{R}$, which every positive solution solves, satisfies (9.3) and (9.4), and has at most one solution by Theorem 9.5(a). *Existence.* The coefficients $\mu-\frac12\sigma(t,e^y)^2$ and $\sigma(t,e^y)$ are bounded and Lipschitz in y (constants MK and K), so Theorem 9.5 gives a solution Y of $\mathrm{d}Y_t=(\mu-\frac12\sigma(t,e^{Y_t})^2)\,\mathrm{d}t+\sigma(t,e^{Y_t})\,\mathrm{d}W_t$, $Y_0=\log S_0$. By Itô's formula, $S=e^Y$ satisfies $\mathrm{d}S_t = S_t\,\mathrm{d}Y_t+\frac12 S_t\,\mathrm{d}[Y,Y]_t=\mu S_t\,\mathrm{d}t+\sigma(t,S_t)S_t\,\mathrm{d}W_t$, and (9.8) is $S=e^Y$ written out. □

Formula (9.8) is not a closed-form solution, since S appears on both sides; it shows that S stays positive. In practice such models are simulated or solved numerically (Section 9.7 and Chapter 16).

Example 9.15 (The CEV model)**.** In the CEV model, $\mathrm{d}S_t=\mu S_t\,\mathrm{d}t+\sigma S_t^{\gamma}\,\mathrm{d}W_t$, introduced by Cox (1975), the volatility of returns $\sigma S^{\gamma-1}$ rises when the price falls if $\gamma<1$. This is the *leverage effect* observed in equity markets. The volatility function is unbounded for $\gamma\neq 1$, so

Theorem 9.14 does not apply, and Theorem 9.5 applies only in two cases. Write the diffusion coefficient as $a(x) = \sigma(x^+)^\gamma$ for $\gamma > 0$.

(a) $\gamma = 1$ (geometric Brownian motion) and $\gamma = 0$ (the linear equation $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$) satisfy both hypotheses. For $\gamma = 0$ the solution is normal (Theorem 9.7(b)), hence negative with positive probability: not a stock price in the sense of Theorem 9.13.

(b) For $0 < \gamma < 1$, a satisfies the growth bound ($x^\gamma \le 1 + x$) but is not Lipschitz at 0: $a(x)/x = \sigma x^{\gamma-1} \to \infty$ as $x \downarrow 0$, so the theorem does not apply. For $\frac{1}{2} \le \gamma < 1$, $|a(x) - a(y)| \le \sigma|x - y|^\gamma$, and the Yamada–Watanabe criterion (Theorem 9.6(b)) gives pathwise uniqueness. A solution exists. For $n \ge 1$, the truncated coefficients $\mu(x^+ \wedge n)$ and $\sigma(x^+ \wedge n)^\gamma$ are bounded and continuous, so the truncated equation has a weak solution (Karatzas and Shreve (1991), Theorem 5.4.22); they satisfy the same Lipschitz and Hölder bounds, so pathwise uniqueness holds and gives a strong solution S^n (their Proposition 5.2.13 and Corollary 5.3.23). For $m \ge n$ the coefficients agree on $(-\infty, n]$, so pathwise uniqueness, applied to the processes stopped at their first passage at n (Theorem 7.15), gives $S^m = S^n$ up to that time. The growth bounds are uniform in n, so the moment bound of Step 5 gives $\mathbb{E}[\sup_{t\le T}(S^n_t)^2] \le C$ and $\mathbb{P}(\sup_{t\le T} S^n_t \ge n) \le C/n^2 \to 0$; hence $S = \lim_n S^n$ exists on $[0, T]$ and solves the equation. Paths reach 0 with positive probability, by Feller's test (Karatzas and Shreve (1991), Section 5.5.C; Exercise 9.6 computes the probability for $\gamma = \frac{1}{2}$, $\mu = 0$). They stay there: if τ is the first time S hits 0, then, since both coefficients vanish at 0 and $\int_0^{t\wedge\tau} H\,\mathrm{d}W = \int_0^t \mathbf{1}_{\{s\le\tau\}} H_s\,\mathrm{d}W_s$ (Theorem 7.15), the stopped process $S_{t\wedge\tau}$ solves the same equation, and pathwise uniqueness gives $S_t = S_{t\wedge\tau}$: the firm has defaulted. For $0 < \gamma < \frac{1}{2}$ the criterion fails, and the model is completed by declaring 0 absorbing, as in the figure.

(c) For $\gamma > 1$, a is continuously differentiable, hence locally Lipschitz, and pathwise uniqueness holds (Theorem 9.6(b)); but a grows faster than linearly, so the theorem does not apply. For $\gamma = 2$ and $\mu = 0$, Exercise 9.7 constructs a positive process that solves the equation with respect to a suitable Brownian motion, and shows that it is not a martingale, although it has no drift.

▶ Interactive demo: CEV paths near zero for any elasticity γ, and how the probability of reaching zero depends on γ and σ.

9.7 The Euler–Maruyama scheme

Most equations have no closed-form solution, and are solved by simulation. Fix $n \ge 1$, the step $h = T/n$ and the grid $t_k = kh$. The *Euler–Maruyama scheme* iterates the short-time approximation of Section 9.1:

$$X^h_0 = \xi, \qquad X^h_{k+1} = X^h_k + b(t_k, X^h_k)\,h + \sigma(t_k, X^h_k)\,(W_{t_{k+1}} - W_{t_k}), \quad k < n. \tag{9.9}$$

To simulate it, the increments $W_{t_{k+1}} - W_{t_k}$ are drawn as independent $\mathcal{N}(0, h)$ variables, which is their exact joint law. Two notions of error measure how well X^h_n approximates X_T. The *strong error* $\mathbb{E}|X_T - X^h_n|$ compares the two on the same Brownian path; it matters when the path itself is used, for example to follow a hedge along a scenario. The *weak error* $|\mathbb{E}f(X_T) - \mathbb{E}f(X^h_n)|$, for a function f, compares only laws; it is the bias of a Monte Carlo price.

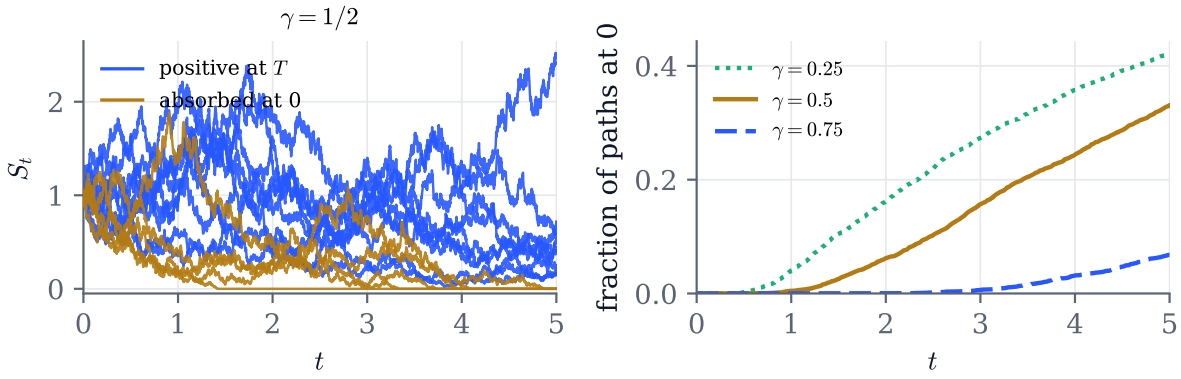

Figure 9.2: The CEV model with $S_0 = 1$, $\mu = 0$, $\sigma = 0.6$ (so that the volatility of returns is 60% at $S = 1$), simulated by the Euler scheme with step 10^{-3} and absorption at 0. Left: 14 paths for $\gamma = \frac{1}{2}$; five reach 0 before $T = 5$. Right: the fraction of 4000 paths absorbed by time t. The smaller γ, the faster the volatility of returns rises as the price falls, and the more paths default.

Theorem 9.16 (Convergence of the Euler–Maruyama scheme)**.**

(a) Strong order $\frac{1}{2}$. *Assume* (9.3), (9.4), $\mathbb{E}[\xi^2] < \infty$, *and* $|b(s,x)-b(t,x)|+|\sigma(s,x)-\sigma(t,x)| \le K(1+|x|)|s-t|^{1/2}$ *for* $s,t \in [0,T]$, $x \in \mathbb{R}$. *Then there is a constant* C *such that* $\mathbb{E}|X_T - X_n^h| \le C\,h^{1/2}$ *for every* n.

(b) Weak order 1. *Assume that* b *and* σ *do not depend on* t, *are four times continuously differentiable with bounded first derivatives and derivatives up to order four of polynomial growth, and that* $\xi = x_0$ *is a constant. Let* $f \in C^4(\mathbb{R})$, *with* f *and its first four derivatives of polynomial growth. Then there is a constant* C_f *such that* $|\mathbb{E}f(X_T) - \mathbb{E}f(X_n^h)| \le C_f\,h$ *for every* n.

Part (a) is Kloeden and Platen (1992), Theorem 10.2.2, and part (b) is a special case of their Theorem 14.1.5. The proof of (a) follows the pattern of Theorem 9.5: compare X with the continuous-time interpolation $Y_t = \xi + \int_0^t b(\eta_s, Y_{\eta_s})\,\mathrm{d}s + \int_0^t \sigma(\eta_s, Y_{\eta_s})\,\mathrm{d}W_s$ of the scheme, where $\eta_s = t_k$ for $s \in [t_k, t_{k+1})$, and apply Gronwall's lemma to $\mathbb{E}[\sup_{u\le t}|X_u - Y_u|^2]$. The order comes from the one-step bound $\mathbb{E}|Y_s - Y_{\eta_s}|^2 \le Ch$: over a step the path moves by about $\sqrt{h}$.

Example 9.17 (Euler–Maruyama for geometric Brownian motion)**.** For $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$, the scheme multiplies the price at each step by $1 + \mu h + \sigma(W_{t_{k+1}} - W_{t_k})$, so $S_n^h = S_0 \prod_{k<n}(1 + \mu h + \sigma(W_{t_{k+1}} - W_{t_k}))$. The factors are independent with mean $1+\mu h$, hence $\mathbb{E}[S_n^h] = S_0(1+\mu h)^n$, while $\mathbb{E}[S_T] = S_0 e^{\mu T}$. With $n = T/h$ and $\log(1+\mu h) = \mu h - \frac{1}{2}\mu^2 h^2 + O(h^3)$,

$$\mathbb{E}[S_T] - \mathbb{E}[S_n^h] = S_0 e^{\mu T}\left(1 - e^{-\mu^2 T h/2 + O(h^2)}\right) = \tfrac{1}{2} S_0 e^{\mu T}\mu^2 T\,h + O(h^2).$$

For $\mu \neq 0$ the weak error for $f(x) = x$ is of order exactly h; for $\mu = 0$ it vanishes, since both means equal S_0. Note also that a factor $1 + \mu h + \sigma(W_{t_{k+1}} - W_{t_k})$ is negative with positive probability: the scheme can produce negative prices, which the exact solution never does. Applying the scheme to $\log S$, which is a Brownian motion with drift, avoids both problems for this model.

Heuristically, the one-step error has a part of root-mean-square size h and mean close to 0, and a part of size h^2: path by path, $n = T/h$ roughly independent errors of size h add up to $\sqrt{Th}$,

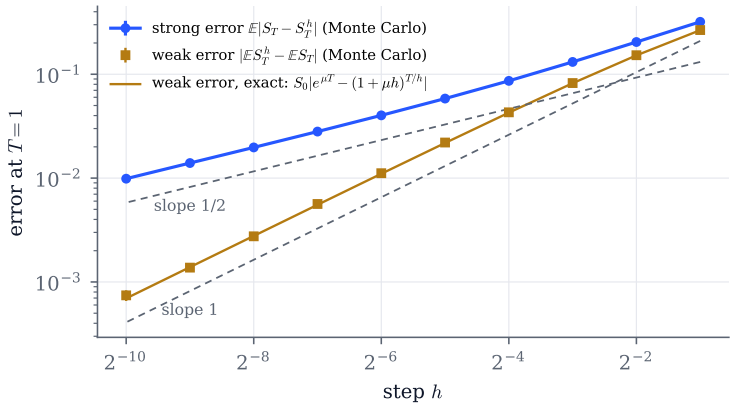

Figure 9.3: Errors of the Euler–Maruyama scheme for geometric Brownian motion with $S_0 = 1$, $\mu = 0.8$, $\sigma = 0.5$, $T = 1$, for $h = 2^{-1}, \dots, 2^{-10}$. Each scheme runs on the same 10^5 Brownian paths as the exact solution; bars are 95% confidence intervals. The weak error, estimated as the mean of $S_T^h - S_T$ over these coupled paths, agrees with the exact value of Theorem 9.17. On log–log axes the strong error has slope $\frac{1}{2}$ and the weak error slope 1 (Theorem 9.16).

while in the mean only the n terms of size h^2 survive. For pricing, the weak order is the relevant one: a Monte Carlo price has a bias of order h and a statistical error of order $M^{-1/2}$ for M paths (Chapter 16). The strong order matters for path-dependent uses and for multilevel Monte Carlo (Giles, 2008), which couples simulations with different steps. Higher-order schemes add terms from the Itô–Taylor expansion of the coefficients; the first of them, the Milstein scheme, has strong order 1 (Exercise 9.8).

▶ Interactive demo: Picard iterates converging along one Brownian path, and the Euler–Maruyama scheme against the exact solution of geometric Brownian motion, with the strong and weak errors measured for the step sizes and parameters you choose.

Further reading

Shreve (2004b), Sections 6.2–6.3, is at the level of these notes, without the existence proof. Øksendal (2003), Chapters 5 and 7, proves it and develops the Markov property and the generator; Steele (2001), Chapter 9, is a readable alternative. Karatzas and Shreve (1991), Chapter 5, covers strong solutions (Section 5.2), weak solutions and Tanaka's equation, which has a weak solution but no strong one (Section 5.3), and linear equations (Section 5.6); Revuz and Yor (1999), Chapter IX, goes further. Kloeden and Platen (1992) is the standard reference on the numerical solution of stochastic differential equations: Chapters 9–10 for strong and Chapter 14 for weak approximation.

Part IV

Pricing and hedging

Chapter 10

The Black–Scholes equation by hedging

Chapter 4 replicated every claim of the binomial model by a portfolio rebalanced at each node, with a hedge ratio given by a discrete derivative of the price. Chapter 6 let the tree become fine and obtained the Black–Scholes prices as expectations under a measure $\mathbb{Q}$, but nothing in that chapter replicates a claim in continuous time. With Itô's formula (Chapter 8) we can now trade continuously.

The argument is that of Black and Scholes (1973) and Merton (1973). If a claim can be replicated and its price is a smooth function $v(t, S_t)$ of time and the stock price, then two expressions for the same process must agree: the wealth equation of the replicating portfolio and Itô's formula for $v(t, S_t)$. Matching their $\mathrm{d}W$-terms gives the hedge, $\Delta_t = \partial_x v(t, S_t)$; matching their $\mathrm{d}t$-terms gives a partial differential equation for v in which the expected return μ of the stock does not appear. Conversely, every solution of that equation with moderate growth gives a hedge, and the Black–Scholes formula is one. The chapter ends with what happens when the hedge is rebalanced at finitely many dates, as it must be in practice: the hedging error no longer vanishes, but it is of the order of the square root of the time between rebalancing dates, and it is driven by the gamma of the claim.

Setting. Throughout, W is a Brownian motion with a filtration $(\mathcal{F}_t)_{t\ge 0}$ (Theorem 5.11) under the real-world measure $\mathbb{P}$, and $T > 0$ is a horizon. The market consists of a money market, worth e^{rt} at time t for some $r \in \mathbb{R}$, and a stock whose price is the geometric Brownian motion (Theorem 6.1)

$$S_t = S_0 \exp\Big(\sigma W_t + \Big(\mu - \frac{\sigma^2}{2}\Big)t\Big), \qquad S_0 > 0,\ \mu \in \mathbb{R},\ \sigma > 0. \tag{10.1}$$

By Theorem 8.8,

$$\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t, \qquad \mathrm{d}(e^{-rt}S_t) = (\mu - r)\,e^{-rt}S_t\,\mathrm{d}t + \sigma\, e^{-rt}S_t\,\mathrm{d}W_t. \tag{10.2}$$

Two facts about the law of S are used repeatedly. For $u > 0$, $\log S_u$ is normal with variance $\sigma^2 u > 0$, so S_u has a density that is positive on $(0, \infty)$. And for every $q \in \mathbb{R}$, by the Gaussian

moment generating function,

$$\mathbb{E}[S_u^q] = S_0^q \exp\Big(q\Big(\mu - \frac{\sigma^2}{2}\Big)u + \frac{q^2\sigma^2}{2}\,u\Big), \tag{10.3}$$

which is bounded for $u \in [0,T]$. As in Chapter 8, we write $v \in C^{1,2}([0,T)\times(0,\infty))$ when v has partial derivatives $\partial_t v$, $\partial_x v$ and $\partial_{xx} v$ that are continuous on $[0,T)\times(0,\infty)$ (one-sided in t at $t=0$).

The results of Chapter 8 are stated for processes indexed by $t \geq 0$ and functions on $[0,\infty)\times\mathbb{R}$. We use them on intervals $[0,t']$, for processes defined only up to time T or before T, and functions defined for $t < T$. This is legitimate. Multiplying the drift and the diffusion coefficient of an Itô process by $\mathbf{1}_{[0,t']}$ gives an Itô process on $[0,\infty)$ that agrees with the given one on $[0,t']$ and is constant afterwards. And if $t' < t''$ and $f \in C^{1,2}([0,t'']\times\mathbb{R})$, then $f(\phi(t),x)$ is in $C^{1,2}([0,\infty)\times\mathbb{R})$ and agrees with f on $[0,t']\times\mathbb{R}$, where ϕ is continuously differentiable, equal to the identity on $[0,t']$ and with values in $[0,t'']$, for instance $\phi(t) = t' + (t''-t')(1-e^{-(t-t')/(t''-t')})$ for $t > t'$.

10.1 Self-financing strategies

In the binomial model, a portfolio that holds Δ_n shares from time n to $n+1$ and keeps the rest of its value in the money market evolves by (4.1). Over an interval $[t,t+h]$ of continuous time, with continuous compounding, the same bookkeeping reads

$$X_{t+h} = \Delta_t S_{t+h} + e^{rh}(X_t - \Delta_t S_t), \quad \text{that is,} \quad e^{-r(t+h)}X_{t+h} - e^{-rt}X_t = \Delta_t\big(e^{-r(t+h)}S_{t+h} - e^{-rt}S_t\big),$$

the analogue of (4.2). Along a partition of $[0,t]$, the discounted wealth is X_0 plus a martingale transform (Theorem 2.7) of the discounted stock price, and as the partition is refined such sums converge to integrals with respect to the Itô process $e^{-rt}S_t$ (Chapter 8). We take the limit as the definition.

Definition 10.1 (Self-financing strategy)**.** A *self-financing strategy* is a pair (X_0,Δ) of an initial capital $X_0 \in \mathbb{R}$ and a progressively measurable process $\Delta = (\Delta_t)_{0\leq t\leq T}$ (Theorem 7.3) such that

$$\int_0^T \Delta_t^2 \,\mathrm{d}t < \infty \quad \text{almost surely.} \tag{10.4}$$

Δ_t is the number of shares held at time t. The *value* of the strategy is the process X defined, for $0 \leq t \leq T$, by

$$e^{-rt}X_t = X_0 + \int_0^t \Delta_u \,\mathrm{d}(e^{-ru}S_u) := X_0 + \int_0^t (\mu - r)\Delta_u\, e^{-ru}S_u \,\mathrm{d}u + \int_0^t \sigma\Delta_u\, e^{-ru}S_u \,\mathrm{d}W_u; \tag{10.5}$$

the amount $X_t - \Delta_t S_t$ is held in the money market.

The integrals exist. Along each path, S is continuous and positive on $[0,T]$, so $0 < m_\omega \leq S_u \leq M_\omega$ there. Hence $\int_0^T (\sigma\Delta_u e^{-ru}S_u)^2\,\mathrm{d}u \leq \sigma^2 e^{2|r|T}M_\omega^2 \int_0^T \Delta_u^2\,\mathrm{d}u < \infty$, so the integrand of $\mathrm{d}W$ is in $\mathbb{H}^2_{\mathrm{loc}}$, and $\int_0^T |\Delta_u| e^{-ru}S_u\,\mathrm{d}u < \infty$ by the Cauchy–Schwarz inequality: $e^{-rt}X_t$ is an Itô process (Theorem 8.9). For the same reason, (10.4) is equivalent to $\int_0^T \Delta_t^2 S_t^2\,\mathrm{d}t < \infty$: it says that the dollar amount held in the stock is square-integrable in time, which is what the stochastic integral needs.

Lemma 10.2 (Wealth equation)**.** *Let (X_0, Δ) be a self-financing strategy. Its value X is the only Itô process such that, almost surely, for all $t \in [0,T]$,*

$$X_t = X_0 + \int_0^t \Delta_u \, \mathrm{d}S_u + \int_0^t r(X_u - \Delta_u S_u) \, \mathrm{d}u, \tag{10.6}$$

where $\int_0^t \Delta_u \, \mathrm{d}S_u = \int_0^t \mu \Delta_u S_u \, \mathrm{d}u + \int_0^t \sigma \Delta_u S_u \, \mathrm{d}W_u$. Equivalently,

$$\mathrm{d}X_t = rX_t \, \mathrm{d}t + (\mu - r)\Delta_t S_t \, \mathrm{d}t + \sigma \Delta_t S_t \, \mathrm{d}W_t. \tag{10.7}$$

Proof. $Y_t = e^{-rt}X_t$ is an Itô process with drift $(\mu - r)\Delta_t e^{-rt} S_t$ and diffusion coefficient $\sigma \Delta_t e^{-rt} S_t$ (frozen after T, as explained in the setting), and e^{rt} is an Itô process with no $\mathrm{d}W$-part, so $[e^{r\cdot}, Y] = 0$ (Theorem 8.11). The product rule (Theorem 8.18) gives

$$\mathrm{d}X_t = \mathrm{d}(e^{rt}Y_t) = re^{rt}Y_t \, \mathrm{d}t + e^{rt} \, \mathrm{d}Y_t = rX_t \, \mathrm{d}t + (\mu - r)\Delta_t S_t \, \mathrm{d}t + \sigma \Delta_t S_t \, \mathrm{d}W_t,$$

which is (10.7), and (10.6) follows by writing $rX + (\mu - r)\Delta S = \mu \Delta S + r(X - \Delta S)$. If X' is another Itô process satisfying (10.6), then $D = X - X'$ satisfies $D_t = \int_0^t rD_u \, \mathrm{d}u$ for all t, almost surely. Along such a path D is continuous, hence continuously differentiable with $D' = rD$ and $D_0 = 0$; so $(e^{-rt}D_t)' = 0$ and $D \equiv 0$. □

The three terms of (10.7) are the return $rX_t \, \mathrm{d}t$ that the whole wealth would earn in the money market, the risk premium $(\mu - r)\Delta_t S_t \, \mathrm{d}t$ earned on the $\Delta_t S_t$ dollars held in the stock, and the risk $\sigma \Delta_t S_t \, \mathrm{d}W_t$, also proportional to the stock position. Holding no shares gives $X_t = X_0 e^{rt}$; holding one share with $X_0 = S_0$ gives $e^{-rt}X_t = S_0 + (e^{-rt}S_t - S_0)$ by (10.2), that is, $X = S$.

Remark 10.3 (Which strategies are allowed). Adaptedness, which progressive measurability includes, says that the position at time t uses only the information in $\mathcal{F}_t$. Condition (10.4) is what the integral needs, but it allows strategies that nobody could run. With $T = 1$, Theorem 7.17 constructs an integrand $H \in \mathbb{H}^2_{\mathrm{loc}}$ with $\int_0^1 H_u \, \mathrm{d}W_u = 1$ almost surely. If $\mu = r$, the strategy with $X_0 = 0$ and $\Delta_t = H_t/(\sigma e^{-rt} S_t)$ satisfies (10.4), because S is bounded away from 0 along each path, and (10.5) gives $X_1 = e^r$ almost surely: something for nothing. Like the doubling strategy of Theorem 2.13, it needs unlimited credit, since its value is not bounded below. Statements that rule out arbitrage must therefore restrict the class of strategies (Chapter 11). The results of this chapter need no such restriction: they construct a strategy that replicates a claim, rather than rule strategies out.

10.2 The Black–Scholes equation

We need Itô's formula for $v(t, S_t)$ when v is defined only for $x > 0$. Composing with the exponential reduces it to Itô's formula for Brownian motion.

Lemma 10.4 (Itô's formula for functions of the stock)**.** *Let $v \in C^{1,2}([0,T) \times (0,\infty))$ and write $\mathcal{L}^\mu v = \partial_t v + \mu x \, \partial_x v + \frac{1}{2}\sigma^2 x^2 \, \partial_{xx} v$. Almost surely, for every $t \in [0,T)$,*

$$e^{-rt}v(t, S_t) = v(0, S_0) + \int_0^t e^{-ru}(\mathcal{L}^\mu v - rv)(u, S_u) \, \mathrm{d}u + \int_0^t \sigma e^{-ru} S_u \, \partial_x v(u, S_u) \, \mathrm{d}W_u. \tag{10.8}$$

The integrand of $\mathrm{d}W$ is continuous and adapted on $[0,t]$, hence in $\mathbb{H}^2_{\mathrm{loc}}$.

Proof. Fix $t' < T$, let $s(u,w) = S_0 \exp\left(\sigma w + (\mu - \frac{\sigma^2}{2})u\right)$, so that $S_u = s(u, W_u)$, and let $f(u,w) = e^{-ru} v(u, s(u,w))$, which is in $C^{1,2}([0,t'] \times \mathbb{R})$ by the chain rule. Since $\partial_u s = (\mu - \frac{\sigma^2}{2})s$, $\partial_w s = \sigma s$ and $\partial_{ww} s = \sigma^2 s$, with the derivatives of v evaluated at $(u, s(u,w))$,

$$\partial_u f = e^{-ru}\Big(\partial_t v - rv + \Big(\mu - \frac{\sigma^2}{2}\Big) s\, \partial_x v\Big), \quad \partial_w f = e^{-ru}\sigma s\, \partial_x v, \quad \partial_{ww} f = e^{-ru}\big(\sigma^2 s^2\, \partial_{xx} v + \sigma^2 s\, \partial_x v\big).$$

So $\partial_u f + \frac{1}{2}\partial_{ww} f = e^{-ru}(\mathcal{L}^\mu v - rv)$, and Theorem 8.5 on $[0,t']$ gives (10.8) for $t \le t'$. Taking $t' = T - 1/m$, $m = 1, 2, \ldots$, covers $[0,T)$. □

This is the formula $\mathrm{d}v = \partial_t v\, \mathrm{d}t + \partial_x v\, \mathrm{d}S + \frac{1}{2}\partial_{xx} v\, \mathrm{d}[S,S]$ of Theorem 8.13, with $\mathrm{d}[S,S]_t = \sigma^2 S_t^2\, \mathrm{d}t$ (Theorem 8.11), followed by the product rule for the discount factor.

Consider a European claim that pays $g(S_T)$ at time T, for a Borel function $g : (0,\infty) \to \mathbb{R}$, and suppose that it is *replicated*: some self-financing strategy has $X_T = g(S_T)$ almost surely. A seller who receives X_0 and follows the strategy can pay $g(S_T)$ whatever happens, so X_0 is the cost of the claim, and X_t its value at time t. Following Black, Scholes and Merton, suppose further that this value depends on the past only through the current stock price, $X_t = v(t, S_t)$, as in the recombining trees of Section 4.3.

Theorem 10.5 (The Black–Scholes equation and the delta hedge)**.** *Let $g : (0,\infty) \to \mathbb{R}$ be Borel, let (X_0, Δ) be a self-financing strategy with $X_T = g(S_T)$ almost surely, and let $v \in C^{1,2}([0,T) \times (0,\infty))$ be such that $X_t = v(t, S_t)$ almost surely, for every $t \in [0,T)$. Then:*

(i) v solves the Black–Scholes equation

$$\partial_t v(t,x) + rx\, \partial_x v(t,x) + \frac{\sigma^2 x^2}{2}\, \partial_{xx} v(t,x) = r\, v(t,x), \qquad (t,x) \in [0,T) \times (0,\infty); \quad (10.9)$$

(ii) the strategy is the delta hedge*: $\Delta_t = \partial_x v(t, S_t)$ for $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost every $(t,\omega) \in [0,T) \times \Omega$;*

(iii) if moreover v is continuous on $[0,T] \times (0,\infty)$, then $v(T,x) = g(x)$ for Lebesgue-almost every $x > 0$, and for every $x > 0$ if g is continuous.

Neither the equation nor the hedge involves μ.

Proof. Fix $t' < T$. The processes X and $v(t, S_t)$ have continuous paths on $[0,t']$ and agree almost surely at each rational time, hence at all times in $[0,t']$, almost surely. So $e^{-rt}X_t$, given by (10.5), and $e^{-rt}v(t,S_t)$, given by (10.8), are two decompositions of one Itô process on $[0,t']$; after freezing both at t' as explained in the setting, Theorem 8.11(ii) applies with $T = t'$. For $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost every $(u,\omega) \in [0,t'] \times \Omega$ their diffusion coefficients agree,

$$\sigma \Delta_u\, e^{-ru} S_u = \sigma e^{-ru} S_u\, \partial_x v(u, S_u),$$

and so do their drifts,

$$(\mu - r)\Delta_u\, e^{-ru} S_u = e^{-ru}\Big(\partial_t v + \mu S_u\, \partial_x v + \frac{\sigma^2 S_u^2}{2}\, \partial_{xx} v - rv\Big)(u, S_u).$$

As $\sigma e^{-ru} S_u > 0$, the first identity gives $\Delta_u = \partial_x v(u, S_u)$. Substituting it into the second, the terms $\mu S_u \partial_x v(u, S_u)$ on the two sides cancel, and what remains is $F(u, S_u) = 0$, where

$F = \partial_t v + rx\,\partial_x v + \frac{1}{2}\sigma^2 x^2\,\partial_{xx} v - rv$ is continuous on $[0,T) \times (0,\infty)$. Letting $t' \uparrow T$ along a sequence proves (ii), and shows that $F(u, S_u) = 0$ for $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost every $(u,\omega) \in [0,T) \times \Omega$.

To pass from the path to the whole half-plane, note that by Tonelli's theorem $\int_0^T \mathbb{P}(F(u,S_u) \neq 0)\,\mathrm{d}u = \mathbb{E}\int_0^T \mathbf{1}_{\{F(u,S_u)\neq 0\}}\,\mathrm{d}u = 0$, so $\mathbb{P}(F(u,S_u) \neq 0) = 0$ for almost every $u \in (0,T)$. Fix such a u. The set $O_u = \{x > 0 : F(u,x) \neq 0\}$ is open, because $F(u,\cdot)$ is continuous, and S_u falls in it with probability zero. Since S_u has a density that is positive on $(0,\infty)$, O_u has Lebesgue measure zero, and an open set of measure zero is empty. Hence $F(u,\cdot) = 0$ for almost every $u \in [0,T)$, and by continuity of F for every u: this is (i).

(iii) X has continuous paths on $[0,T]$, and so has $v(t,S_t)$ under the extra hypothesis, so $X_T = \lim_{t\uparrow T} v(t,S_t) = v(T,S_T)$ almost surely. Hence $v(T,S_T) = g(S_T)$ almost surely, so the Borel set $\{x > 0 : v(T,x) \neq g(x)\}$ has probability zero under the law of S_T, and therefore Lebesgue measure zero. If g is continuous the set is open, hence empty. □

The proof is short because the work is done by Theorem 8.11: the hedge is read off the $\mathrm{d}W$-terms, and the equation off the $\mathrm{d}t$-terms. The original argument of Black and Scholes (1973) reads the same computation differently. The position long one claim and short $\partial_x v(t,S_t)$ shares has no $\mathrm{d}W$-term, so over a short interval it is riskless, and a riskless position must earn the riskless rate; written out, this is (10.9). The equation is linear, so a portfolio of claims is hedged by the sum of the hedges.

Remark 10.6 (The drift does not matter). (a) *Why μ drops out.* In the proof, μ enters twice: through the risk premium $(\mu - r)\Delta_t S_t$ earned by the shares in the portfolio (10.7), and through the drift $\mu S_t\,\partial_x v$ of the value of the claim (10.8). With $\Delta_t = \partial_x v(t,S_t)$ the two are equal: a portfolio that tracks the claim carries exactly the claim's exposure to the stock, and with it exactly the claim's share of the risk premium. Traders who agree on r and σ but not on μ agree on prices and hedges; this is Theorem 4.6 in continuous time.

(b) *What μ does affect.* By (10.8) and (10.9), the drift of $v(t,S_t)$ is $rv + (\mu - r)S_t\,\partial_x v$: over a short interval the claim's expected return is r plus the stock's excess return $\mu - r$ times the *elasticity* $S_t\,\partial_x v / v$. For a call, $c < xN(d_+) = x\,\partial_x c$ (Theorems 6.11 and 6.14), so the elasticity exceeds one: a call is a leveraged position in the stock, with a larger risk premium, not none. The drift also determines the probability that the claim ends in the money, and, in Section 10.4, the size of the error of a hedge rebalanced at discrete dates.

(c) *Why this is fortunate.* Since $\log S_u = \log S_0 + (\mu - \frac{\sigma^2}{2})u + \sigma W_u$, along dyadic partitions of $[0,t]$ the sum of squared log-returns converges almost surely to $\sigma^2 t$: the drift contributes nothing, by the rules $\mathrm{d}W\,\mathrm{d}t = 0$ and $\mathrm{d}t\,\mathrm{d}t = 0$ of Section 5.5, and $\sum(\sigma\,\delta W)^2 \to \sigma^2 t$ by Theorem 5.16. One path determines σ. The natural estimate of μ, $\log(S_t/S_0)/t + \frac{\sigma^2}{2}$, has standard deviation $\sigma/\sqrt{t}$: about 6% per year with $\sigma = 20\%$ and ten years of data. Prices depend on the parameter that is easy to estimate, not on the one that is hard.

(d) *The measure $\mathbb{Q}$.* In Chapter 6, prices were expectations under a measure under which the stock has drift r. Theorem 10.7(iii) below shows that the two routes give the same prices, and Chapter 11 constructs that measure from $\mathbb{P}$ (Theorem 11.2).

10.3 Verification and the Black–Scholes formula

Theorem 10.5 says what the price and the hedge must be if a smooth price exists. The converse is a *verification* theorem: a solution of the equation, with moderate growth, is a price, and its derivative is the hedge.

Theorem 10.7 (Verification)**.** *Let $g : (0,\infty) \to \mathbb{R}$ be continuous, and let $v \in C^{1,2}([0,T) \times (0,\infty))$ be continuous on $[0,T] \times (0,\infty)$, solve (10.9), and satisfy $v(T,x) = g(x)$ for $x > 0$. Assume the growth condition: for some constants $C, \kappa \geq 0$,*

$$|\partial_x v(t,x)| \leq C(1 + x^{\kappa} + x^{-\kappa}), \qquad (t,x) \in [0,T) \times (0,\infty). \tag{10.10}$$

Let $X_0 = v(0,S_0)$, $\Delta_t = \partial_x v(t,S_t)$ for $t < T$, and $\Delta_T = 0$.

(i) (X_0, Δ) is a self-financing strategy, and its value satisfies $X_t = v(t,S_t)$ for all $t \in [0,T]$, almost surely. In particular $X_T = g(S_T)$: the strategy replicates the claim.

(ii) $\mathbb{E}\int_0^T \Delta_t^2 S_t^2 \,\mathrm{d}t < \infty$.

(iii) If $\mu = r$, then $(e^{-rt}X_t)_{0 \leq t \leq T}$ is a martingale, $g(S_T)$ is integrable, and $v(t,S_t) = e^{-r(T-t)}\mathbb{E}[g(S_T) \,|\, \mathcal{F}_t]$ for every $t \in [0,T]$.

Proof. (i) Δ is continuous and adapted on $[0,T)$, so it is progressively measurable (Theorem 7.3), and by (10.10), $\int_0^T \Delta_t^2 \,\mathrm{d}t \leq C^2 \int_0^T (1 + S_t^{\kappa} + S_t^{-\kappa})^2 \,\mathrm{d}t < \infty$ along every path, since S is continuous and positive. By (10.9), $\mathcal{L}^{\mu} v - rv = (\mu - r)x\,\partial_x v$, so Theorem 10.4 reads, for $t < T$,

$$e^{-rt}v(t,S_t) = v(0,S_0) + \int_0^t (\mu - r)\,\partial_x v(u,S_u)\, e^{-ru} S_u \,\mathrm{d}u + \int_0^t \sigma\,\partial_x v(u,S_u)\, e^{-ru} S_u \,\mathrm{d}W_u,$$

which is the right side of (10.5). So $X_t = v(t,S_t)$ for all $t < T$, almost surely. As $t \uparrow T$, $X_t \to X_T$ because X is continuous, and $v(t,S_t) \to v(T,S_T) = g(S_T)$ because v is continuous on $[0,T] \times (0,\infty)$.

(ii) $\Delta_t^2 S_t^2 \leq 3C^2(S_t^2 + S_t^{2\kappa+2} + S_t^{2-2\kappa})$, whose expectation is bounded on $[0,T]$ by (10.3); Tonelli's theorem concludes.

(iii) If $\mu = r$, the $\mathrm{d}u$-integral in (10.5) vanishes, and by (ii) the integrand $\sigma\Delta_u e^{-ru}S_u$ is in $\mathbb{H}^2$, so $e^{-rt}X_t$ is a martingale on $[0,T]$ (Theorem 7.7). Its terminal value $e^{-rT}g(S_T)$ is therefore integrable, and $e^{-rt}X_t = \mathbb{E}[e^{-rT}g(S_T) \mid \mathcal{F}_t]$; multiply by e^{rt} and use (i). □

Remark 10.8 (What is proved about prices). (a) *One price.* Under the hypotheses of Theorem 10.7, suppose the claim can be traded at time 0 at a price $c > v(0,S_0)$. Sell it, run the strategy of the theorem from $v(0,S_0)$, and put $c - v(0,S_0)$ in the money market. At time T the strategy pays $g(S_T)$, which settles the claim, and $(c - v(0,S_0))e^{rT} > 0$ is left, almost surely. If $c < v(0,S_0)$, reverse every position. This is the argument of Theorem 4.4, and it uses one explicit strategy, whose value is $\pm v(t,S_t)$. For a call, the hedge of a sale is worth $c(t,S_t) \geq 0$; the reversed hedge is worth $-c(t,S_t) \geq -S_t$, which is not bounded below by a constant, although the call held alongside covers it. So $v(0,S_0)$ is the only price at which the claim can trade without offering a riskless profit of this kind.

(b) *No arbitrage at that price.* That no other strategy makes a riskless profit when the claim trades at $v(0,S_0)$ is a statement about all strategies, and by Theorem 10.3 it needs a

restriction on them. It is the fundamental theorem of Chapter 11 (Theorem 11.11), extended to a market that also trades the claim in Theorem 11.12.

(c) *One solution.* Could two solutions of (10.9) with the same terminal condition, both satisfying (10.10), give two hedges with different costs? No. The equation does not involve μ or S_0, so we may apply Theorem 10.7(iii) in the model with $\mu = r$ and any initial price $x > 0$, and to the function $(s, y) \mapsto v(t + s, y)$ on the horizon $T - t$, which satisfies the same hypotheses. At time 0 this gives

$$v(t,x) = e^{-r(T-t)}\,\mathbb{E}\Big[g\Big(x\exp\Big(\sigma W_{T-t} + \Big(r - \frac{\sigma^2}{2}\Big)(T-t)\Big)\Big)\Big], \qquad (t,x) \in [0,T]\times(0,\infty), \tag{10.11}$$

and the right side does not depend on the solution. This is the Feynman–Kac formula for (10.9); Chapter 12 proves it for general diffusions (Theorem 12.1). Uniqueness is a statement about a class of solutions: here, those satisfying (10.10).

It remains to find solutions. Chapter 6 computed the call and put prices as expectations; the next proposition checks that they solve the equation, with the right terminal condition and growth, by calculus alone.

Proposition 10.9 (The Black–Scholes formula solves the equation)**.** *Let $K > 0$, let c and p be the call and put functions of Theorem 6.11, and set $c(T,x) = (x-K)^+$ and $p(T,x) = (K-x)^+$.*

(i) $c, p \in C^{1,2}([0,T)\times(0,\infty))$, and both solve (10.9).

(ii) c and p are continuous on $[0,T]\times(0,\infty)$.

(iii) $0 < \partial_x c < 1$ and $-1 < \partial_x p < 0$ on $[0,T)\times(0,\infty)$.

Consequently, the call is replicated by the strategy with $X_0 = c(0,S_0)$ and $\Delta_t = N\big(d_+(T-t,S_t)\big)$, whose value is $X_t = c(t,S_t)$, and the put by the strategy with $X_0 = p(0,S_0)$ and $\Delta_t = -N\big(-d_+(T-t,S_t)\big)$, whose value is $p(t,S_t)$.

Proof. (i) For $t < T$, $\tau = T - t > 0$ and $d_\pm(\tau, x)$ are smooth functions of (t,x), and so are c and p. By Theorem 6.14(i), (ii) and (iv),

$$\partial_t c + rx\,\partial_x c + \frac{\sigma^2x^2}{2}\,\partial_{xx}c = -\frac{x\sigma N'(d_+)}{2\sqrt{\tau}} - rKe^{-r\tau}N(d_-) + rxN(d_+) + \frac{\sigma^2x^2}{2}\,\frac{N'(d_+)}{x\sigma\sqrt{\tau}} = r\,c.$$

The terms in $N'(d_+)$ cancel; this is where the identity $Ke^{-r\tau}N'(d_-) = xN'(d_+)$ of Theorem 6.14, used to compute the theta, does its work. For the put, $p = c - w$ with $w(t,x) = x - Ke^{-r\tau}$ by the parity of Theorem 6.11, and w solves (10.9): $\partial_t w + rx\,\partial_x w = -rKe^{-r\tau} + rx = rw$. The equation is linear.

(iii) By Theorem 6.14(i), $\partial_x c = N(d_+) \in (0,1)$ and $\partial_x p = N(d_+) - 1$.

(ii) Continuity at points with $t < T$ is part of (i). Fix $x > 0$ and let $t \uparrow T$: if $x > K$, then $d_\pm \to +\infty$ and $c(t,x) \to x - K$; if $x < K$, then $d_\pm \to -\infty$ and $c(t,x) \to 0$; if $x = K$, then $d_\pm = (r \pm \frac{\sigma^2}{2})\sqrt{\tau}/\sigma \to 0$ and $c(t,x) \to \frac{K}{2} - \frac{K}{2} = 0$. So $c(t,x) \to (x-K)^+$. By (iii) and the mean value theorem, $|c(t,y) - c(t,x)| \le |y - x|$ for $t < T$, and also for $t = T$. Hence $|c(t,y) - c(T,x)| \le |y-x| + |c(t,x) - c(T,x)| \to 0$ as $(t,y) \to (T,x)$. For the put, use $p = c - w$.

Finally, by (iii) the growth condition (10.10) holds with $C = 1$ and $\kappa = 0$, the payoffs are continuous, and Theorem 10.7 applies. □

By Theorem 10.7(iii) in the model with $\mu = r$, $c(t, S_t)$ is also $e^{-r(T-t)}\mathbb{E}[(S_T - K)^+ \mid \mathcal{F}_t]$, the expectation computed in Theorem 6.11: the hedging route of this chapter and the expectation route of Chapter 6 lead to the same function. The delta of the call, $N(d_+)$, is the number of shares in the replicating portfolio, and the rest of its value, $-Ke^{-r\tau}N(d_-)$, is borrowed: the hedge of a call is a leveraged long position in the stock.

Example 10.10 (Power claims and forwards)**.** For $a \in \mathbb{R}$, the claim paying S_T^a has the price $v(t,x) = e^{-\alpha(T-t)}x^a$ with $\alpha = (1-a)(r + \frac{a\sigma^2}{2})$, found by an expectation in Exercise 6.2. The equation is checked directly: $\partial_t v = \alpha v$, $x\,\partial_x v = av$ and $x^2\,\partial_{xx} v = a(a-1)v$, so the left side of (10.9) is $(\alpha + ra + \frac{\sigma^2}{2}a(a-1))v = rv$. Here v is smooth on $[0,T]\times(0,\infty)$, $|\partial_x v| = |a|e^{-\alpha(T-t)}x^{a-1}$ satisfies (10.10), and the hedge $\Delta_t = av(t,S_t)/S_t$ keeps the fixed fraction a of the claim's value in the stock. For $a = 1$ the claim is the stock and $\Delta = 1$; for $a = 0$ it is a zero-coupon bond, $v = e^{-r(T-t)}$ and $\Delta = 0$; for $a = 2$, $\Gamma = \partial_{xx} v = 2e^{(r+\sigma^2)(T-t)}$. Similarly, the forward contract $g(x) = x - K$ has $v(t,x) = x - Ke^{-r(T-t)}$ and $\Delta = 1$: the static hedge of Theorem 3.7.

▶ Interactive demo: follow the delta hedge of a call along a simulated path: the stock, the number of shares, and the value of the portfolio tracking the price of the call.

10.4 Hedging in discrete time

The delta hedge trades at every instant, which no one can do. Fix $n \geq 1$, let $h = T/n$ and $t_k = kh$, and consider a seller who receives $v(0,S_0)$, holds $\delta_k = \partial_x v(t_k, S_{t_k})$ shares from each date t_k, $k < n$, until t_{k+1}, and keeps the rest of her wealth in the money market. As in Chapter 4, her wealth satisfies

$$X_0^n = v(0,S_0), \qquad X_{t_{k+1}}^n = \delta_k S_{t_{k+1}} + e^{rh}\big(X_{t_k}^n - \delta_k S_{t_k}\big), \quad 0 \leq k \leq n-1,$$

and her *discounted hedging error* is $e_n = e^{-rT}\big(X_T^n - g(S_T)\big)$. Write

$$\Gamma_k = \partial_{xx} v(t_k, S_{t_k}), \qquad R_k = \frac{e^{-rt_{k+1}}S_{t_{k+1}}}{e^{-rt_k}S_{t_k}} - 1$$

for the gamma at the rebalancing date and the discounted return over the following period. We assume:

(H) v is as in Theorem 10.7, its partial derivatives $\partial_t^i \partial_x^j v$ with $2i + j \leq 4$ exist and are continuous on $[0,T]\times(0,\infty)$ (one-sided in t at 0 and T), and each of them is bounded by $C(1 + x^\kappa + x^{-\kappa})$ for some constants $C, \kappa \geq 0$.

The power claims of Theorem 10.10 satisfy (H). The call does not: its gamma $N'(d_+)/(x\sigma\sqrt{T-t})$ (Theorem 6.14) is unbounded near (T,K).

Proposition 10.11 (Discrete delta hedging)**.** *Assume (H). There is a constant C', depending on v, S_0, μ, r, σ and T but not on n, such that:*

(i) (gamma–theta decomposition)

$$e_n = -\sum_{k=0}^{n-1} \frac{1}{2}\, e^{-rt_k}\,\Gamma_k S_{t_k}^2 (R_k^2 - \sigma^2 h) + \rho_n, \qquad \mathbb{E}[\rho_n^2] \leq C' h^2; \tag{10.12}$$

(ii) $|\mathbb{E}[e_n]| \le C'h$, *and* $\mathbb{E}[e_n] = 0$ *if* $\mu = r$;

(iii) as $n \to \infty$, *with* $\Gamma = \partial_{xx} v$,

$$\mathbb{E}[e_n^2] = \frac{\sigma^4 h}{2} \int_0^T \mathbb{E}\Big[e^{-2ru}\,\Gamma(u, S_u)^2 S_u^4\Big]\,\mathrm{d}u + o(h). \tag{10.13}$$

Proof. Discounted coordinates. Let $Y_t = e^{-rt} S_t$ and $\bar v(t,y) = e^{-rt} v(t, e^{rt} y)$, so that $\bar v(t, Y_t) = e^{-rt} v(t, S_t)$. By the chain rule, $\partial_y \bar v(t,y) = \partial_x v(t, e^{rt} y)$ and $\partial_{yy} \bar v(t,y) = e^{rt} \partial_{xx} v(t, e^{rt} y)$, and (10.9), which extends to $t = T$ by continuity, becomes

$$\partial_t \bar v + \frac{\sigma^2 y^2}{2}\,\partial_{yy} \bar v = 0 \qquad \text{on } [0,T] \times (0, \infty). \tag{10.14}$$

Time decay of the discounted price is set by its convexity: *theta pays for gamma.* By (H) and the chain rule, the derivatives $\partial_t^i \partial_y^j \bar v$ with $2i + j \le 4$ are continuous and bounded by $C_1(1 + y^q + y^{-q})$ for some constants C_1, q.

The error as a sum. The discounted wealth changes by $\delta_k(Y_{t_{k+1}} - Y_{t_k})$ over $[t_k, t_{k+1}]$, and $\bar v(0, Y_0) = X_0^n$, $\bar v(T, Y_T) = e^{-rT} g(S_T)$. Telescoping,

$$e_n = -\sum_{k=0}^{n-1} P_k, \qquad P_k = \bar v(t_{k+1}, Y_{t_{k+1}}) - \bar v(t_k, Y_{t_k}) - \delta_k\big(Y_{t_{k+1}} - Y_{t_k}\big), \tag{10.15}$$

where P_k is the discounted profit over $[t_k, t_{k+1}]$ of a long position in the claim, marked at v, hedged with δ_k shares.

Taylor expansion. Fix k, write $y = Y_{t_k}$ and $\eta = Y_{t_{k+1}} - y = yR_k$, and evaluate the derivatives of $\bar v$ at (t_k, y) unless indicated. Expanding first in time at $y + \eta$, then $\partial_t \bar v(t_k, \cdot)$ and $\bar v(t_k, \cdot)$ in space around y, and using $\partial_y \bar v = \delta_k$ and (10.14),

$$P_k = \frac{1}{2}\,\partial_{yy} \bar v\,(\eta^2 - \sigma^2 y^2 h) + A_k + B_k, \qquad A_k = h\,\partial_{ty} \bar v\,\eta + \frac{1}{6}\,\partial_{yyy} \bar v\,\eta^3,$$

where the Taylor remainders satisfy $|B_k| \le \frac{1}{2} h^2 \sup|\partial_{tt} \bar v| + \frac{1}{2} h \eta^2 \sup|\partial_{tyy} \bar v| + \frac{1}{24} \eta^4 \sup|\partial_y^4 \bar v|$, the suprema being over $[t_k, t_{k+1}] \times \{y + \eta\}$ for the first and over $\{t_k\} \times [y, y + \eta]$ for the others. Since $\partial_{yy} \bar v\, y^2 = e^{-rt_k} \Gamma_k S_{t_k}^2$ and $\eta^2 = y^2 R_k^2$, the first term is the summand of (10.12), and $\rho_n = -\sum_k (A_k + B_k)$.

Moments. $1 + R_k = \exp\big(\sigma(W_{t_{k+1}} - W_{t_k}) + (\mu - r - \frac{\sigma^2}{2})h\big)$ is independent of $\mathcal{F}_{t_k}$, with a law that does not depend on k. Expanding $\mathbb{E}[(1 + R_k)^m] = \exp\big(m(\mu - r - \frac{\sigma^2}{2})h + \frac{m^2 \sigma^2}{2} h\big)$ in powers of h gives, for $h \le T$ and constants c_m not depending on h,

$$\mathbb{E}|R_k|^m \le c_m h^{m/2}, \qquad |\mathbb{E} R_k| \le c_1 |\mu - r| h, \qquad |\mathbb{E} R_k^3| \le c_3 h^2,$$
$$|\mathbb{E} R_k^2 - \sigma^2 h| \le c_2 h^2, \qquad \mathrm{Var}(R_k^2) = 2\sigma^4 h^2 + O(h^3);$$

for the first, use $|e^z - 1| \le |z| e^{|z|}$ and the Cauchy–Schwarz inequality. The points between y and $y + \eta = y(1 + R_k)$ lie in $[y \min(1, 1 + R_k),\, y \max(1, 1 + R_k)]$, so by the growth bounds and (10.3) all the coefficients and suprema above have moments of every order, bounded uniformly in k and n. With the Cauchy–Schwarz inequality this gives $\mathbb{E}[B_k^2] \le C_2 h^4$ and

$\mathbb{E}[A_k^2] \le C_2 h^3$. Moreover the coefficients of A_k are $\mathcal{F}_{t_k}$-measurable and R_k is independent of $\mathcal{F}_{t_k}$, so $\mathbb{E}[A_k \mid \mathcal{F}_{t_k}] = h\,\partial_{ty}\bar v\, y\,\mathbb{E}R_k + \frac{1}{6}\partial_{yyy}\bar v\, y^3\,\mathbb{E}R_k^3$ (Theorem 1.19), and $\mathbb{E}\big[\mathbb{E}[A_k \mid \mathcal{F}_{t_k}]^2\big] \le C_2 h^4$.

Proof of (i). Let $a_k = A_k - \mathbb{E}[A_k \mid \mathcal{F}_{t_k}]$. For $j < k$, a_j is $\mathcal{F}_{t_k}$-measurable, so $\mathbb{E}[a_j a_k] = \mathbb{E}\big[a_j\,\mathbb{E}[a_k \mid \mathcal{F}_{t_k}]\big] = 0$, and $\mathbb{E}\big[(\sum_k a_k)^2\big] = \sum_k \mathbb{E}[a_k^2] \le \sum_k \mathbb{E}[A_k^2] \le n C_2 h^3 = C_2 T h^2$. By the Cauchy–Schwarz inequality for sums, $\mathbb{E}\big[(\sum_k \mathbb{E}[A_k \mid \mathcal{F}_{t_k}])^2\big] \le n \sum_k C_2 h^4 = C_2 T^2 h^2$ and $\mathbb{E}\big[(\sum_k B_k)^2\big] \le C_2 T^2 h^2$. So $\mathbb{E}[\rho_n^2] \le C' h^2$.

Proof of (ii) and (iii). Write $G_k = e^{-rt_k}\Gamma_k S_{t_k}^2$, which is $\mathcal{F}_{t_k}$-measurable with $\mathbb{E}[G_k^2]$ bounded, and $R_k^2 - \sigma^2 h = \xi_k + m_h$ with $\xi_k = R_k^2 - \mathbb{E}R_k^2$ and $m_h = \mathbb{E}R_k^2 - \sigma^2 h = O(h^2)$. The main term of (10.12) is $N_n - \frac{1}{2} m_h \sum_k G_k$ with $N_n = -\frac{1}{2}\sum_k G_k \xi_k$. Since ξ_k is independent of $\mathcal{F}_{t_k}$, $\mathbb{E}[N_n] = 0$, and the second part has L^2-norm at most $\frac{1}{2}|m_h|\, n \max_k \mathbb{E}[G_k^2]^{1/2} = O(h)$. With (i), $\|e_n - N_n\|_2 = O(h)$, which gives $|\mathbb{E}e_n| \le C' h$. If $\mu = r$, then $\mathbb{E}R_k = 0$, so $\mathbb{E}[\delta_k(Y_{t_{k+1}} - Y_{t_k})] = \mathbb{E}[\delta_k Y_{t_k}]\,\mathbb{E}R_k = 0$ and $\mathbb{E}[e^{-rT} X_T^n] = X_0^n$, while $\mathbb{E}[e^{-rT} g(S_T)] = v(0, S_0) = X_0^n$ by Theorem 10.7(iii); so $\mathbb{E}[e_n] = 0$.

For (iii), the cross terms of $\mathbb{E}[N_n^2]$ vanish: for $j < k$, $G_j \xi_j G_k$ is $\mathcal{F}_{t_k}$-measurable and ξ_k is independent of $\mathcal{F}_{t_k}$ with mean 0. Hence

$$\mathbb{E}[N_n^2] = \frac{1}{4}\sum_k \mathbb{E}[G_k^2]\,\mathbb{E}[\xi_k^2] = \frac{\sigma^4 h}{2}(1 + O(h))\, h \sum_{k=0}^{n-1} \varphi(t_k), \qquad \varphi(u) = \mathbb{E}\big[e^{-2ru}\Gamma(u, S_u)^2 S_u^4\big].$$

Writing $S_u = S_0 \exp(\sigma\sqrt{u}\, Z + (\mu - \frac{\sigma^2}{2})u)$ with $Z \sim \mathcal{N}(0,1)$, the continuity and growth of Γ and dominated convergence show that φ is continuous on $[0, T]$, so the Riemann sums $h\sum_k \varphi(t_k)$ converge to $\int_0^T \varphi(u)\,\mathrm{d}u$ and $\mathbb{E}[N_n^2] = \frac{\sigma^4 h}{2}\int_0^T \varphi + o(h)$. Finally $\big|\|e_n\|_2^2 - \|N_n\|_2^2\big| \le \|e_n - N_n\|_2\,(\|e_n\|_2 + \|N_n\|_2) = O(h)\,O(\sqrt{h}) = o(h)$. □

Reading the proposition. Over each period, a delta-hedged long position in the claim gains $\frac{1}{2}\Gamma S^2 R^2$ from the move of the stock and loses $\frac{1}{2}\Gamma S^2 \sigma^2 h$ to the passage of time, by (10.14). For a claim with positive gamma it profits when the stock moves by more than $\sigma\sqrt{h}$ in either direction and loses otherwise (Figure 10.2, left); the seller who hedges has the opposite profit. In continuous time these gains and losses cancel exactly. In discrete time they cancel on average, by (ii), but not path by path: up to their small means, the terms of (10.12) are uncorrelated, each of variance about $\frac{1}{2}\sigma^4 h^2\,\mathbb{E}[G_k^2]$, so their sum has a standard deviation of order $\sqrt{h}$. To halve the error, rebalance four times as often.

Figure 10.1 shows the hedge of a call, which (H) does not cover. The errors are centred and shrink like $n^{-1/2}$: with $\mu = 15\%$, the root-mean-square error is 1.83 for monthly, 0.90 for weekly and 0.41 for daily rebalancing, against a premium of 10.45. The leading term of (10.13) predicts 1.91, 0.92 and 0.42. For the call the integral in (10.13) is still finite, because the gamma is large only within a distance of order $\sqrt{T-u}$ of the strike. But the proposition does not apply, and part of it fails: for the call, the remainder in (i) is only of order $h^{3/4}$ in L^2 ($\mathbb{E}[\rho_n^2]$ falls by a factor of about 8, not 16, when n is multiplied by 4). What survives is known. Zhang (1999), Theorem 2.4.1, showed that $\sqrt{n}\,\|e_n\|_2$ converges for Lipschitz payoffs, so $\mathbb{E}[e_n^2]$ is still of order h; and Bertsimas et al. (2000) showed that for calls and puts $\sqrt{n}\,e_n$ converges in law to a mixture of centred normal laws, which the leading term of (10.13) describes. Convergence in law alone does not give the limit of $n\,\mathbb{E}[e_n^2]$; numerically, it is within 3% of the leading term at $n = 1024$ (Figure 10.1, right). For less regular payoffs the rate changes: for a digital option $\mathbb{E}[e_n^2]$ is only of order $\sqrt{h}$ (Gobet and Temam (2001), Theorem 1, with the correction of Gobet (2014), which

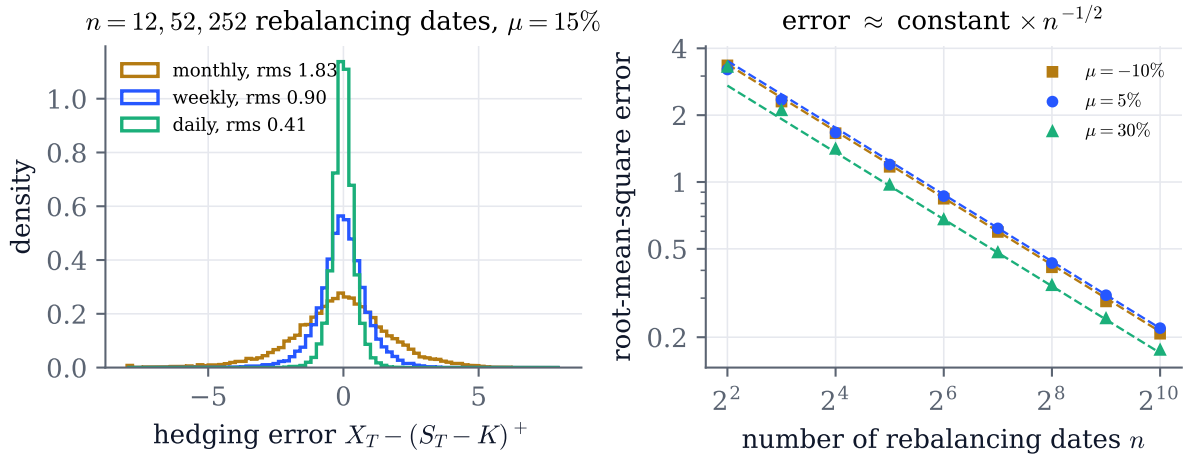

Figure 10.1: Discrete delta hedging of a call with $S_0 = K = 100$, $T = 1$, $r = 5\%$ and $\sigma = 20\%$; the call costs 10.45. Left: the terminal error $X_T^n - (S_T - K)^+$ for monthly, weekly and daily rebalancing, 20,000 paths with drift $\mu = 15\%$. Right: its root-mean-square against the number n of rebalancing dates, for three drifts (6,000 paths); the dashed lines are e^{rT} times the square root of the leading term of (10.13).

leaves the rates unchanged; Exercise 10.5). Geiss and Gobet (2011) survey these results; they state them for $\mu = r$ and note that they extend to $\mu \neq r$, the case of Figure 10.1, where $\mu = 15\%$.

The drift enters only through the law of the path. With $\mu = -10\%$, 5% and 30%, $\sqrt{n}$ times the root-mean-square error tends to 6.78, 7.02 and 5.44 (Figure 10.1, right): a large drift sends the stock away from the strike, where the gamma is small. The hedger does not earn the risk premium, and does not need it.

Two practical points complete the picture. First, rebalancing is not free: with proportional transaction costs, the cost of trading n times grows with n while the hedging error shrinks, and Leland (1985) shows how to compensate the costs, to first order, by hedging with an enlarged volatility. Second, the hedge uses σ, which must be estimated or implied from market prices (Chapter 15); hedging with the wrong volatility leaves an error that does not vanish as $n \to \infty$, and Exercise 10.7 computes it. Positions in other options can reduce the gamma of a book, and with it the error of (10.12) (Exercise 10.2).

▶ Interactive demo: rebalance a call hedge at n dates of your choice, for any drift, volatility and strike; watch the error distribution narrow like $n^{-1/2}$, and the one-period profit follow the gamma–theta parabola.

Further reading

Shreve (2004b), Section 4.5, derives the Black–Scholes–Merton equation by equating the evolutions of the portfolio and of the option value, as here, with the same notation for the hedge. The equation is due to Black and Scholes (1973), and Merton (1973) gave a more careful derivation and many extensions. Harrison and Pliska (1981) recast continuous-time hedging in terms of

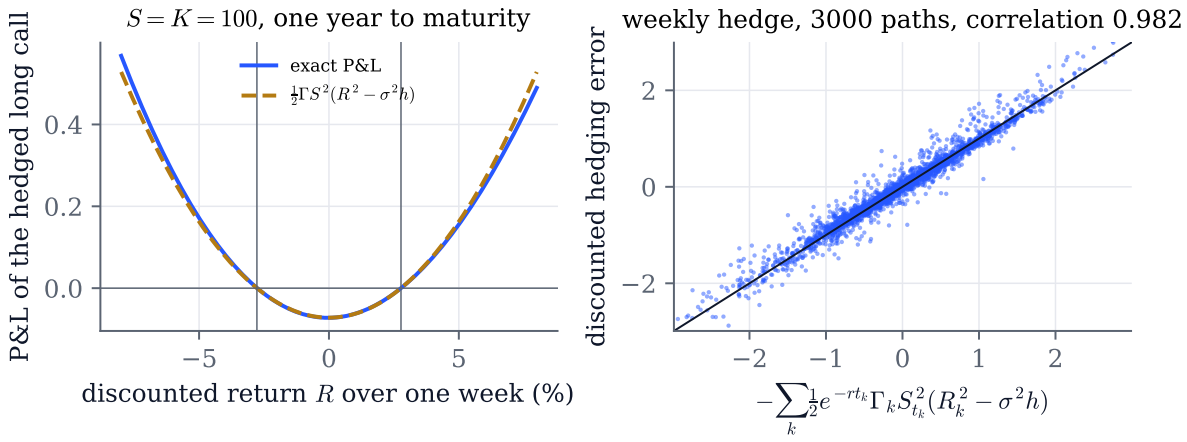

Figure 10.2: The gamma–theta decomposition for the call of Figure 10.1. Left: the discounted profit over one week of a long call hedged with $N(d_+)$ shares at $S = K$, one year before maturity, against the discounted return R of the stock, and the parabola $\frac{1}{2}\Gamma S^2(R^2 - \sigma^2 h)$; the vertical lines are at $\pm\sigma\sqrt{h}$. Right: the discounted error of a weekly hedge against the sum of the gamma–theta terms of (10.12), along 3,000 paths with $\mu = 15\%$.

martingales and stochastic integrals, which is the route of Chapter 11. On discrete hedging, Bertsimas et al. (2000) is a readable starting point and Geiss and Gobet (2011) a survey of the mathematics; El Karoui et al. (1998) study what happens when the volatility used to hedge is not the true one. Hull (2021) discusses delta and gamma hedging as practised by traders.

Chapter 11

Change of measure and risk-neutral pricing

Chapter 6 priced options as expectations under a postulated measure $\mathbb{Q}$, and Chapter 10 priced them by replication. This chapter connects the two. Girsanov's theorem builds $\mathbb{Q}$ from the real-world measure $\mathbb{P}$ by reweighting the paths, which changes the drift of the Brownian motion and nothing else. Under $\mathbb{Q}$ discounted wealth is a (local) martingale, which gives the risk-neutral pricing formula for every replicable claim, path-dependent or not, with random rates and volatilities. The martingale representation theorem shows that, when all the randomness comes from one Brownian motion, every claim is replicable: the fundamental theorems of Chapter 4 in continuous time. The chapter ends with numeraires, dividends and exchange rates.

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $T > 0$ is a horizon, and W is a Brownian motion with a filtration $(\mathcal{F}_t)_{t \geq 0}$ (Theorems 5.6 and 5.11). Stochastic integrals, Itô processes and quadratic variations are those of Chapters 7 and 8, as summarised in Theorem 8.2.

11.1 Densities on a filtration

Section 1.5 changed measures by a density $Z \geq 0$ with $\mathbb{E}[Z] = 1$: $\mathbb{Q}(A) = \mathbb{E}[Z\mathbf{1}_A]$. On a filtered space the density has a history.

Lemma 11.1 (Density process and Bayes' formula)**.** *Let Z_T be an $\mathcal{F}_T$-measurable random variable with $Z_T > 0$ almost surely and $\mathbb{E}[Z_T] = 1$, let $\mathbb{Q}(A) = \mathbb{E}[Z_T \mathbf{1}_A]$ for $A \in \mathcal{F}$, and let $Z_t = \mathbb{E}[Z_T \mid \mathcal{F}_t]$, $0 \leq t \leq T$.*

(i) $\mathbb{Q}$ is a probability measure equivalent to $\mathbb{P}$, (Z_t) is a $\mathbb{P}$-martingale with $Z_t > 0$ almost surely, and Z_t is the density of $\mathbb{Q}$ on $\mathcal{F}_t$: $\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}[Z_t X]$ for every $\mathcal{F}_t$-measurable X that is nonnegative or $\mathbb{Q}$-integrable.

(ii) (Bayes' formula) *For $0 \leq s \leq t \leq T$ and every $\mathcal{F}_t$-measurable Y with $\mathbb{E}^{\mathbb{Q}}|Y| < \infty$,*

$$\mathbb{E}^{\mathbb{Q}}[Y \mid \mathcal{F}_s] = \frac{\mathbb{E}[Z_t Y \mid \mathcal{F}_s]}{Z_s}.$$

(iii) An adapted process $(M_t)_{0 \leq t \leq T}$ with $\mathbb{E}^{\mathbb{Q}}|M_t| < \infty$ for every t is a $\mathbb{Q}$-martingale if and

only if $(Z_t M_t)_{0\le t\le T}$ is a $\mathbb{P}$-martingale.

Proof. (i) Equivalence is Theorem 1.23(b), and Z is a martingale by the tower property (Theorem 1.25). For $X \ge 0$ $\mathcal{F}_t$-measurable, $\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}[Z_T X] = \mathbb{E}[Z_t X]$ by (1.9) and (1.4); for $\mathbb{Q}$-integrable X, split $X = X^+ - X^-$. With $X = \mathbf{1}_{\{Z_t \le 0\}}$ this gives $\mathbb{Q}(Z_t \le 0) = \mathbb{E}[Z_t \mathbf{1}_{\{Z_t \le 0\}}] \le 0$, so $\mathbb{Q}(Z_t \le 0) = 0$, and $\mathbb{P}(Z_t \le 0) = 0$ by equivalence.

(ii) By (i), $\mathbb{E}[Z_t|Y|] = \mathbb{E}^{\mathbb{Q}}|Y| < \infty$. Let $R = \mathbb{E}[Z_t Y \mid \mathcal{F}_s]/Z_s$, which is $\mathcal{F}_s$-measurable, and $\mathbb{Q}$-integrable because $\mathbb{E}^{\mathbb{Q}}|R| = \mathbb{E}[Z_s|R|] \le \mathbb{E}[Z_t|Y|]$. For $A \in \mathcal{F}_s$, by (i) at time s, then the definition of conditional expectation, then (i) at time t,

$$\mathbb{E}^{\mathbb{Q}}[R\mathbf{1}_A] = \mathbb{E}\big[\mathbb{E}[Z_t Y \mid \mathcal{F}_s]\mathbf{1}_A\big] = \mathbb{E}[Z_t Y \mathbf{1}_A] = \mathbb{E}^{\mathbb{Q}}[Y\mathbf{1}_A].$$

So R satisfies the definition of $\mathbb{E}^{\mathbb{Q}}[Y \mid \mathcal{F}_s]$ (Theorem 1.14) in $(\Omega, \mathcal{F}, \mathbb{Q})$. Identities that hold $\mathbb{Q}$-almost surely hold $\mathbb{P}$-almost surely.

(iii) By (i), $\mathbb{E}|Z_t M_t| = \mathbb{E}^{\mathbb{Q}}|M_t| < \infty$, and by (ii), $\mathbb{E}^{\mathbb{Q}}[M_t \mid \mathcal{F}_s] = M_s$ if and only if $\mathbb{E}[Z_t M_t \mid \mathcal{F}_s] = Z_s M_s$. □

11.2 Girsanov's theorem

Let Θ be a progressively measurable process with $\int_0^T \Theta_u^2 \,\mathrm{d}u < \infty$ almost surely, that is $\Theta \in \mathbb{H}^2_{\mathrm{loc}}$ (Theorem 7.3), and set, for $0 \le t \le T$,

$$Z_t = \exp\Big(-\int_0^t \Theta_u \,\mathrm{d}W_u - \frac{1}{2}\int_0^t \Theta_u^2 \,\mathrm{d}u\Big), \qquad \widetilde{W}_t = W_t + \int_0^t \Theta_u \,\mathrm{d}u. \tag{11.1}$$

Theorem 11.2 (Girsanov's theorem)**.** *In this setting:*

(i) Z is a positive local martingale with $\mathrm{d}Z_t = -\Theta_t Z_t \,\mathrm{d}W_t$, and $\mathbb{E}[Z_T] \le 1$.

(ii) If $\mathbb{E}[Z_T] = 1$, then Z is a martingale on $[0,T]$, $\mathbb{Q}(A) = \mathbb{E}[Z_T \mathbf{1}_A]$ defines a probability measure equivalent to $\mathbb{P}$, and $(\widetilde{W}_t)_{0\le t\le T}$ is a Brownian motion under $\mathbb{Q}$, for which $(\mathcal{F}_t)_{t\le T}$ is a filtration.

(iii) (Novikov's condition) *If $\mathbb{E}\big[\exp\big(\frac{1}{2}\int_0^T \Theta_u^2 \,\mathrm{d}u\big)\big] < \infty$, then $\mathbb{E}[Z_T] = 1$. This holds in particular if Θ is bounded, for instance constant.*

The theorem is due to Girsanov (1960), after Cameron and Martin (1944) for deterministic Θ, and (iii) to Novikov (1972). We prove (i), and (ii)–(iii) when Θ is bounded, which covers the constant case and every model in this chapter. For general Θ see Karatzas and Shreve (1991), Sections 3.5.A and 3.5.D, or Revuz and Yor (1999), Chapter VIII, Section 1 (Theorem 1.4 and Novikov's criterion); Shreve (2004b), Section 5.2.1, proves (ii) when $\mathbb{E}\int_0^T \Theta_u^2 Z_u^2 \,\mathrm{d}u < \infty$. The proof rests on a lemma about exponentials of Itô integrals.

Lemma 11.3 (Exponential martingales)**.** *Let B be a Brownian motion with a filtration on some probability space, with expectation $\mathbb{E}$, let $\psi \in \mathbb{H}^2_{\mathrm{loc}}$, and let $\mathcal{E}_t = \exp\big(\int_0^t \psi \,\mathrm{d}B - \frac{1}{2}\int_0^t \psi^2 \,\mathrm{d}u\big)$, $t \le T$.*

(a) $\mathcal{E}_t = 1 + \int_0^t \psi_u \mathcal{E}_u \,\mathrm{d}B_u$ and $\mathbb{E}[\mathcal{E}_t] \le 1$.

(b) If $|\psi| \le b$ for a constant b, then $\mathbb{E}[\mathcal{E}_t^p] \le e^{p(p-1)b^2 t/2}$ for every $p \ge 1$, and $\mathcal{E}$ is a martingale

on $[0,T]$.

Proof. (a) The identity is Itô's formula for e^x (Theorem 8.13), computed in Theorem 8.15. The integrand $\psi\mathcal{E}$ is in $\mathbb{H}^2_{\mathrm{loc}}$, so with the stopping times τ_n of Theorem 7.16 the stopped process $\mathcal{E}_{t\wedge\tau_n}$ is a martingale and $\mathbb{E}[\mathcal{E}_{t\wedge\tau_n}]=1$. Almost surely $\tau_n = T$ for n large, so $\mathcal{E}_{t\wedge\tau_n}\to\mathcal{E}_t$, and Fatou's lemma[1] gives $\mathbb{E}[\mathcal{E}_t]\le 1$.

(b) Write $\mathcal{E}^{(p\psi)}$ for the exponential built from $p\psi$. Then $\mathcal{E}_t^p = \mathcal{E}_t^{(p\psi)}\exp\left(\frac{p^2-p}{2}\int_0^t\psi^2\,\mathrm{d}u\right) \le \mathcal{E}_t^{(p\psi)}e^{p(p-1)b^2t/2}$, and (a) for $p\psi$ gives the bound. With $p=2$, $\mathbb{E}\int_0^T\psi^2\mathcal{E}_u^2\,\mathrm{d}u \le b^2\int_0^T e^{b^2u}\,\mathrm{d}u<\infty$, so $\psi\mathcal{E}\in\mathbb{H}^2$ and $\mathcal{E}$ is a martingale by Theorem 7.7. □

Proof of Theorem 11.2. (i) is Theorem 11.3(a) with $B=W$ and $\psi=-\Theta$. Let now $|\Theta|\le c$. By Theorem 11.3(b), Z is a martingale and $\mathbb{E}[Z_T]=1$, which is (iii) in this case. As $Z_T>0$, Theorem 11.1 applies: $\mathbb{Q}$ is equivalent to $\mathbb{P}$ with density Z_t on $\mathcal{F}_t$. To use Lévy's characterisation, which is stated for all times, extend $\widetilde{W}$ past T: let $Z_t = Z_T$ for $t>T$, and let $\chi_t = \Theta_t\mathbf{1}_{\{t\le T\}}$ and $M_t = W_t + \int_0^t\chi_u\,\mathrm{d}u$ for $t\ge 0$, so that $M=\widetilde{W}$ on $[0,T]$ and $\mathrm{d}Z_t = -\chi_t Z_t\,\mathrm{d}W_t$. Fix $T'\ge T$; on $[0,T']$, $Z_t=\mathbb{E}[Z_T\mid\mathcal{F}_t]$ is the density process of $\mathbb{Q}$.

- *Quadratic variation.* M is an Itô process with diffusion coefficient 1, so $[M,M]_t = t$ (Theorem 8.11(i)): along partitions with mesh going to 0, the sums of squared increments converge to t in $\mathbb{P}$-probability. They converge in $\mathbb{Q}$-probability as well, because convergence in probability means that every subsequence has a further subsequence converging almost surely (Durrett (2019), Section 2.3), and $\mathbb{P}$ and $\mathbb{Q}$ have the same null sets.
- *Martingale.* By the product rule (Theorem 8.18),

$$\begin{aligned}\mathrm{d}(Z_tM_t) &= Z_t\,\mathrm{d}M_t + M_t\,\mathrm{d}Z_t + \mathrm{d}[Z,M]_t\\ &= Z_t\chi_t\,\mathrm{d}t + Z_t\,\mathrm{d}W_t - M_t\chi_tZ_t\,\mathrm{d}W_t - \chi_tZ_t\,\mathrm{d}t = Z_t(1-\chi_tM_t)\,\mathrm{d}W_t.\end{aligned}$$

  By Theorem 11.3(b), $\mathbb{E}[Z_t^2]\le e^{c^2T}$ and $\mathbb{E}[Z_t^4]\le e^{6c^2T}$, and $|M_t|\le|W_t|+cT$ gives $\mathbb{E}[M_t^4]\le 8(3t^2+c^4T^4)$. By Cauchy–Schwarz, $\mathbb{E}[Z_t^2(1-\chi_tM_t)^2]\le 2\mathbb{E}[Z_t^2]+2c^2\mathbb{E}[Z_t^4]^{1/2}\mathbb{E}[M_t^4]^{1/2}$ is bounded on $[0,T']$, so the integrand is in $\mathbb{H}^2[0,T']$ and ZM is a $\mathbb{P}$-martingale; and $\mathbb{E}^{\mathbb{Q}}|M_t| = \mathbb{E}[Z_t|M_t|]<\infty$. By Theorem 11.1(iii), M is a $\mathbb{Q}$-martingale on $[0,T']$.

As T' is arbitrary, M is a continuous $\mathbb{Q}$-martingale with $M_0=0$ and $[M,M]_t=t$ for all $t\ge 0$. By Lévy's characterisation (Theorem 8.20), applied under $\mathbb{Q}$, M is a $\mathbb{Q}$-Brownian motion and $(\mathcal{F}_t)$ is a filtration for it; on $[0,T]$, $M=\widetilde{W}$. Under $\mathbb{Q}$, M is not given as an Itô integral against a $\mathbb{Q}$-Brownian motion, so this is the general form of the theorem, cited in Chapter 8, not the case proved there. □

For a constant $\Theta_t=\theta$, $Z_t=\exp(-\theta W_t-\frac{\theta^2}{2}t)$ and $\widetilde{W}_t = W_t+\theta t$: under $\mathbb{Q}$, W is a Brownian motion with drift $-\theta$. At the single time T this is the Gaussian shift of Theorem 1.27 (with θ replaced by $-\theta$); Girsanov's theorem says that the whole path is shifted, with independent Gaussian increments under $\mathbb{Q}$.

Remark 11.4 (What a change of measure can and cannot do). (a) *Drift, not volatility.* The paths do not change; their weights do. Quadratic variation is a limit in probability along

[1] $\mathbb{E}[\liminf_n X_n]\le\liminf_n\mathbb{E}[X_n]$ for $X_n\ge 0$: apply monotone convergence to $\inf_{m\ge n}X_m$.

a single path, so it is the same under $\mathbb{P}$ and every equivalent $\mathbb{Q}$: $[\widetilde{W},\widetilde{W}]_t = [W,W]_t = t$. An equivalent change of measure can move the drift of an Itô process, never its diffusion coefficient: two measures under which $\log S$ has quadratic variations $\sigma_1^2 T \neq \sigma_2^2 T$ over $[0,T]$ are not equivalent.

(b) *A condition is needed.* The inequality $\mathbb{E}[Z_T] \leq 1$ can be strict, and then Z_T defines no probability measure. With $T=1$ and the processes X and B of Theorem 7.17, let $a>0$ and let τ be the first time $u<1$ at which $X_u + \frac{1}{2}\log(1-u) = -a$, that is, at which $B_v - \frac{v}{2}$ reaches $-a$ in the clock $v = -\log(1-u)$. Since $B_v/v \to 0$ (Exercise 5.5), $B_v - \frac{v}{2} \to -\infty$ and $\tau < 1$ almost surely. For $\Theta_s = -(1-s)^{-1/2}\mathbf{1}_{\{s\leq\tau\}}$ (set to 0 on the null event $\{\tau=1\}$, as in that example), $\int_0^1 \Theta_s^2\,\mathrm{d}s = -\log(1-\tau) < \infty$ and $-\int_0^1 \Theta\,\mathrm{d}W = X_\tau$, the stopped integral, so $Z_1 = \exp\left(X_\tau + \frac{1}{2}\log(1-\tau)\right) = e^{-a}$ almost surely.

(c) *Several Brownian motions.* Let $W = (W^1,\dots,W^d)$ be a d-dimensional Brownian motion, $\Theta = (\Theta^1,\dots,\Theta^d)$ with each $\Theta^i \in \mathbb{H}^2_{\mathrm{loc}}$, $Z_t = \exp\big(-\sum_i \int_0^t \Theta^i\,\mathrm{d}W^i - \frac{1}{2}\int_0^t |\Theta_u|^2\,\mathrm{d}u\big)$ and $\widetilde{W}^i_t = W^i_t + \int_0^t \Theta^i_u\,\mathrm{d}u$. If $\mathbb{E}[Z_T]=1$, for instance if Θ is bounded, then $\widetilde{W}$ is a d-dimensional $\mathbb{Q}$-Brownian motion on $[0,T]$. The proof for bounded Θ is the one above, with Theorem 8.17 and Lévy's characterisation in d dimensions; see Karatzas and Shreve (1991), Section 3.5.A, and Shreve (2004b), Section 5.4.1.

(d) *Reweighting is importance sampling* (Exercise 1.9). For constant θ, $\mathbb{E}[Z_T^2] = e^{\theta^2 T}$, and M paths simulated under $\mathbb{P}$ have an *effective sample size* $(\sum_i Z^i)^2/\sum_i (Z^i)^2$ of about $Me^{-\theta^2 T}$: a diagnostic of how unevenly the weights are spread, that is, of how few paths can dominate a weighted average. The variance of a weighted estimator depends on the payoff and can be larger or smaller than with paths simulated under $\mathbb{Q}$: for the call of Figure 11.1 it is about four times smaller.

Remark 11.5 (Stochastic integrals under $\mathbb{Q}$). In the setting of Theorem 11.2(ii), let $H \in \mathbb{H}^2_{\mathrm{loc}}$; the condition $\int_0^T H^2\,\mathrm{d}u < \infty$ is the same under $\mathbb{P}$ and $\mathbb{Q}$. Then, almost surely,

$$\int_0^t H_u\,\mathrm{d}\widetilde{W}_u = \int_0^t H_u\,\mathrm{d}W_u + \int_0^t H_u\Theta_u\,\mathrm{d}u \qquad \text{for all } t \leq T, \tag{11.2}$$

where the left side is the Itô integral built under $\mathbb{Q}$ from the $\mathbb{Q}$-Brownian motion $\widetilde{W}$, and the right side is built under $\mathbb{P}$. For simple H both sides are the same finite sum. In general, take simple H^n with $\int_0^T (H^n - H)^2\,\mathrm{d}u \to 0$ in probability; then the integrals of H^n converge to those of H uniformly on $[0,T]$ in probability, under $\mathbb{P}$ for W and under $\mathbb{Q}$ for $\widetilde{W}$ (Karatzas and Shreve (1991), Section 3.2.D), while $\int_0^t H^n\Theta\,\mathrm{d}u \to \int_0^t H\Theta\,\mathrm{d}u$ uniformly by Cauchy–Schwarz. As in the proof above, convergence in probability is the same under both measures, so the limits agree. In differential form, $H\,\mathrm{d}\widetilde{W} = H\,\mathrm{d}W + H\Theta\,\mathrm{d}t$: an Itô process under $\mathbb{P}$ is an Itô process under $\mathbb{Q}$, with the same diffusion coefficient and a new drift.

The market price of risk. In the Black–Scholes model, the stock is the geometric Brownian motion $S_t = S_0\exp\left(\sigma W_t + (\mu - \frac{\sigma^2}{2})t\right)$ under $\mathbb{P}$ (Theorem 6.1), and the money market grows at the rate r. The number

$$\theta = \frac{\mu - r}{\sigma}$$

is the *market price of risk*: the expected return in excess of the risk-free rate, per unit of volatility. With $\Theta \equiv \theta$, Theorem 11.2 gives a measure $\mathbb{Q}$ under which $\widetilde{W}_t = W_t + \theta t$ is a Brownian motion

on $[0,T]$, and since $\sigma W_t + \mu t = \sigma\widetilde{W}_t + rt$,

$$S_t = S_0 \exp\Big(\sigma\widetilde{W}_t + \Big(r - \frac{\sigma^2}{2}\Big)t\Big), \qquad \mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}\widetilde{W}_t.$$

This is the setting of Section 6.3: every price computed in Chapter 6 is an expectation under this $\mathbb{Q}$. The real-world drift μ enters only through θ, that is, through the weights Z_T, and it disappears from prices. Figure 11.1 shows the reweighting at work, and what happens if the weights are forgotten.

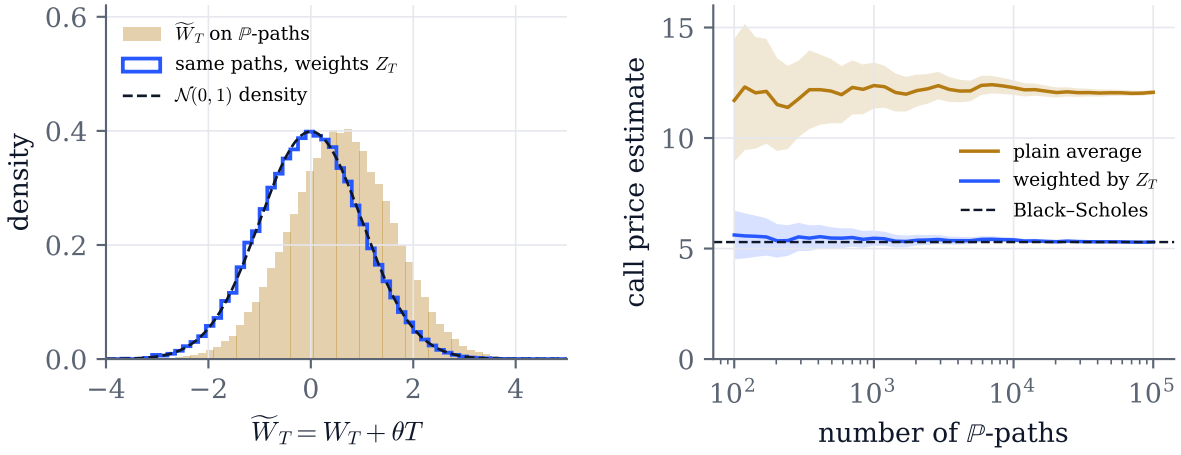

Figure 11.1: Girsanov's theorem on 10^5 paths simulated under $\mathbb{P}$ with $S_0 = 100$, $\mu = 15\%$, $r = 3\%$, $\sigma = 20\%$, $T = 1$, so $\theta = 0.6$. Left: $\widetilde{W}_T = W_T + \theta T$ is $\mathcal{N}(\theta, 1)$ on these paths; weighted by Z_T, the same values are $\mathcal{N}(0,1)$. Right: estimates of the call price with $K = 110$, with 95% confidence bands, as the number of paths grows. The weighted average converges to the Black–Scholes price 5.29; the plain average converges to $e^{-rT}\mathbb{E}[(S_T - K)^+] = 12.07$, which is not a price (`code/src/ieor222/riskneutral.py`).

▶ Interactive demo: reweight simulated paths by Z_T for any market price of risk, and watch the histogram of $\widetilde{W}_T$, the effective sample size and the price estimates change.

11.3 Risk-neutral pricing

The market. We now let the interest rate, the mean return and the volatility be random. Let r, α and σ be bounded progressively measurable processes, with $\sigma_t \geq \varepsilon$ for a constant $\varepsilon > 0$. The *money market* is worth $B_t = \exp\big(\int_0^t r_u\,\mathrm{d}u\big)$ at time t, and $D_t = 1/B_t$ is the *discount factor*: $\mathrm{d}D_t = -r_t D_t\,\mathrm{d}t$. The *stock* is

$$S_t = S_0 \exp\Big(\int_0^t \sigma_u\,\mathrm{d}W_u + \int_0^t \Big(\alpha_u - \frac{\sigma_u^2}{2}\Big)\mathrm{d}u\Big), \qquad \text{so that} \qquad \mathrm{d}S_t = \alpha_t S_t\,\mathrm{d}t + \sigma_t S_t\,\mathrm{d}W_t \quad (11.3)$$

by Theorem 8.13, with $S_0 > 0$. The Black–Scholes market is the case of constant r, $\alpha = \mu$ and σ. By the product rule, with $[D,S] = 0$ because D has no $\mathrm{d}W$-term,

$$\mathrm{d}(D_tS_t) = D_tS_t\big((\alpha_t - r_t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t\big) = \sigma_t D_t S_t(\Theta_t\,\mathrm{d}t + \mathrm{d}W_t), \qquad \Theta_t = \frac{\alpha_t - r_t}{\sigma_t}. \quad (11.4)$$

The market price of risk Θ is bounded, by $(\sup|\alpha| + \sup|r|)/\varepsilon$.

Definition 11.6 (Risk-neutral measure)**.** A *risk-neutral measure*, or *equivalent martingale measure*, is a probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$, equivalent to $\mathbb{P}$ on $\mathcal{F}_T$, under which the discounted price $(D_t S_t)_{0\le t\le T}$ of every traded asset that pays no dividends is a martingale. (Dividends are treated in Section 11.6.)

The money market itself trivially satisfies the condition: $D_t B_t = 1$. Prices are measured in units of the money market, and under $\mathbb{Q}$ every traded asset earns, on average, the same as the money market.

Proposition 11.7 (The risk-neutral measure of the model)**.** *Let $\mathbb{Q}$ and $\widetilde{W}$ be given by Theorem 11.2 with $\Theta = (\alpha - r)/\sigma$. Then $\widetilde{W}$ is a $\mathbb{Q}$-Brownian motion on $[0,T]$,*

$$\mathrm{d}S_t = r_t S_t\,\mathrm{d}t + \sigma_t S_t\,\mathrm{d}\widetilde{W}_t, \qquad D_t S_t = S_0 \exp\Big(\int_0^t \sigma_u\,\mathrm{d}\widetilde{W}_u - \frac{1}{2}\int_0^t \sigma_u^2\,\mathrm{d}u\Big),$$

and $(D_t S_t)$ is a $\mathbb{Q}$-martingale, so $\mathbb{Q}$ is a risk-neutral measure. In the Black–Scholes market, $\mathbb{Q}$ and $\widetilde{W}$ realise the setting of Section 6.3.

Proof. Θ is bounded, so Theorem 11.2(ii)–(iii) apply. By (11.2) with $H = \sigma$, $\int_0^t \sigma\,\mathrm{d}\widetilde{W} = \int_0^t \sigma\,\mathrm{d}W + \int_0^t (\alpha - r)\,\mathrm{d}u$, which turns (11.3) times D_t into the displayed formula; and $\sigma S\,\mathrm{d}W = \sigma S\,\mathrm{d}\widetilde{W} - (\alpha - r)S\,\mathrm{d}t$ gives the equation for S. Finally DS/S_0 is the exponential of Theorem 11.3 for the $\mathbb{Q}$-Brownian motion $\widetilde{W}$ and the bounded $\psi = \sigma$, hence a $\mathbb{Q}$-martingale. □

Strategies. A trader holds Δ_t shares at time t and keeps the rest of her wealth X_t in the money market. A *self-financing strategy* is a pair (X_0, Δ) with $X_0 \in \mathbb{R}$ and Δ progressively measurable with $\int_0^T \Delta_t^2 S_t^2\,\mathrm{d}t < \infty$ almost surely, whose wealth satisfies

$$\mathrm{d}X_t = \Delta_t\,\mathrm{d}S_t + r_t(X_t - \Delta_t S_t)\mathrm{d}t, \qquad \text{equivalently} \qquad \mathrm{d}(D_t X_t) = \Delta_t\,\mathrm{d}(D_t S_t). \tag{11.5}$$

The second form defines $D_t X_t = X_0 + \int_0^t \Delta_u\,\mathrm{d}(D_u S_u)$ explicitly, and the product rule $\mathrm{d}(DX) = D\,\mathrm{d}X - rDX\,\mathrm{d}t$ shows that the two forms agree. In the Black–Scholes market (constant r, μ, σ) this is Theorem 10.1. By (11.4) and (11.2), under the measure of Theorem 11.7,

$$D_t X_t = X_0 + \int_0^t \Delta_u \sigma_u D_u S_u\,\mathrm{d}\widetilde{W}_u, \tag{11.6}$$

an Itô integral with an integrand in $\mathbb{H}^2_{\mathrm{loc}}$, since σ and D are bounded. An *arbitrage* is a self-financing strategy with $X_0 = 0$, $X_T \ge 0$ almost surely and $\mathbb{P}(X_T > 0) > 0$. Without a restriction on Δ there are arbitrages: the doubling strategy of Theorem 7.17, with time rescaled from $[0,1]$ to $[0,T]$, run under $\mathbb{Q}$ with $\widetilde{W}$ in place of W and $\Delta_t \sigma_t D_t S_t$ in place of its integrand, turns zero into a sure discounted dollar, at the price of unbounded intermediate losses. A strategy is *tame* if its discounted wealth is bounded below, $D_t X_t \ge -c$ for a constant c: the trader has a credit line.

Theorem 11.8 (Risk-neutral pricing)**.** *Let $\mathbb{Q}$ be a risk-neutral measure and V_T an $\mathcal{F}_T$-measurable claim with $\mathbb{E}^{\mathbb{Q}}[D_T|V_T|] < \infty$.*

(i) If a self-financing strategy replicates the claim, $X_T = V_T$ almost surely, and its discounted

wealth (D_tX_t) is a $\mathbb{Q}$-martingale, then for $0 \le t \le T$

$$X_t = \mathbb{E}^{\mathbb{Q}}\Big[e^{-\int_t^T r_u\,\mathrm{d}u} V_T \,\Big|\, \mathcal{F}_t\Big], \qquad \textit{in particular} \qquad X_0 = \mathbb{E}^{\mathbb{Q}}[D_T V_T]. \tag{11.7}$$

All such strategies have the same wealth process.

(ii) For the measure of Theorem 11.7, the discounted wealth of every self-financing strategy is a $\mathbb{Q}$-local martingale. It is a $\mathbb{Q}$-martingale if $\mathbb{E}^{\mathbb{Q}}\int_0^T(\Delta_t\sigma_t D_t S_t)^2\,\mathrm{d}t < \infty$, and $\mathbb{E}^{\mathbb{Q}}[D_T X_T] \le X_0$ if the strategy is tame.

Proof. (i) $D_tX_t = \mathbb{E}^{\mathbb{Q}}[D_TX_T \mid \mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[D_TV_T \mid \mathcal{F}_t]$; divide by the $\mathcal{F}_t$-measurable $D_t > 0$ (taking out what is known) and use $D_T/D_t = e^{-\int_t^T r_u\,\mathrm{d}u}$.

(ii) The first two claims are (11.6) with Theorem 7.16 and Theorem 7.7, under $\mathbb{Q}$. Let $D_tX_t \ge -c$, and let τ_n be the stopping times of Theorem 7.16 for the integrand in (11.6). Then $\mathbb{E}^{\mathbb{Q}}[D_{T\wedge\tau_n}X_{T\wedge\tau_n}] = X_0$, and $D_{T\wedge\tau_n}X_{T\wedge\tau_n} \to D_TX_T$ almost surely because $\tau_n = T$ for large n. Fatou's lemma, applied to the nonnegative $D_{T\wedge\tau_n}X_{T\wedge\tau_n} + c$, gives $\mathbb{E}^{\mathbb{Q}}[D_TX_T] + c \le X_0 + c$. □

As in Chapter 4, the replication cost is the price. If the claim of (i) were offered at time 0 for $p > X_0$, one could sell it, run the replicating strategy from X_0, and put $p - X_0$ in the money market: at T the strategy settles the claim and $B_T(p - X_0) > 0$ is left. If $p < X_0$, buy the claim and run $(-X_0, -\Delta)$. Conversely, if the claim trades at X_t, its discounted price is a $\mathbb{Q}$-martingale, so $\mathbb{Q}$ remains risk-neutral for the market extended by the claim. The payoff may depend on the whole path, as for Asian and barrier options (Chapter 13); for $V_T = g(S_T)$ in the Black–Scholes market, (11.7) is the price of Theorem 6.9. With a random rate the discount factor stays inside the expectation: the claim $V_T = 1$, the zero-coupon bond, is worth $\mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u\,\mathrm{d}u} \mid \mathcal{F}_t]$ (Chapter 14). And the measure must be $\mathbb{Q}$: the same expectation under $\mathbb{P}$ depends on μ, and in Figure 11.1 it more than doubles the price of the call.

Example 11.9 (A forward contract, three ways)**.** In the Black–Scholes market, the long side of a forward contract receives $S_T - K$ at T. (a) Statically, hold one share and borrow Ke^{-rT} (Theorem 3.7): the value is $S_t - Ke^{-r(T-t)}$. (b) By (11.7), the value is $e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[S_T - K \mid \mathcal{F}_t] = S_t - Ke^{-r(T-t)}$, because $e^{-rt}S_t$ is a $\mathbb{Q}$-martingale. (c) The function $v(t,x) = x - Ke^{-r(T-t)}$ solves the Black–Scholes equation $\partial_t v + rx\partial_x v + \frac{1}{2}\sigma^2x^2\partial_{xx}v = rv$ with $v(T,x) = x - K$, so $v(t,S_t)$ is the replication value (Theorem 10.7), and its delta $\partial_x v = 1$ is the static hedge of (a). The three answers agree, and the forward price, the K that makes the contract worth zero, is $S_te^{r(T-t)}$. Here (a) is model-free, (b) uses only the martingale property of $e^{-rt}S_t$ under $\mathbb{Q}$, and (c) uses the full model.

11.4 Martingale representation and completeness

Theorem 11.8 prices claims that can be replicated. Which ones can? In the binomial model, all of them (Theorem 4.9); in the trinomial model, not (Theorem 4.14). The answer depends on the information: from now on, $(\mathcal{F}_t)$ is the filtration generated by W and augmented by the $\mathbb{P}$-null sets, $\mathcal{F}_t = \sigma(\{W_u : u \le t\} \cup \mathcal{N})$ with $\mathcal{N} = \{A \in \mathcal{F} : \mathbb{P}(A) = 0\}$. Adding null sets does not change independence, so this is a filtration for W, and $\mathcal{F}_0$ contains only events of probability 0 or 1.

Theorem 11.10 (Martingale representation). *Let $(\mathcal{F}_t)$ be the augmented filtration of W.*

(i) Every $\mathbb{P}$-martingale $(M_t)_{0\le t\le T}$ has a representation $M_t = M_0 + \int_0^t \Gamma_u \, \mathrm{d}W_u$, almost surely for each t, with M_0 constant and $\Gamma \in \mathbb{H}^2_{\mathrm{loc}}$; in particular M has a continuous version, which we use from now on. If $\mathbb{E}[M_T^2] < \infty$, then $\Gamma \in \mathbb{H}^2$, and Γ is unique up to $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-null sets.

(ii) Let $\mathbb{Q}$ and $\widetilde{W}$ be as in Theorem 11.2(ii). Every $\mathbb{Q}$-martingale $(\widetilde{M}_t)_{0\le t\le T}$ has a continuous version with $\widetilde{M}_t = \widetilde{M}_0 + \int_0^t \widetilde{\Gamma}_u \, \mathrm{d}\widetilde{W}_u$ for some $\widetilde{\Gamma} \in \mathbb{H}^2_{\mathrm{loc}}$.

Part (i) is cited: see Karatzas and Shreve (1991), Section 3.4.D, Revuz and Yor (1999), Chapter V, Section 3, Øksendal (2003), Theorem 4.3.4 (square-integrable case), or Shreve (2004b), Section 5.3.1. Uniqueness follows from the isometry: if $\int_0^T (\Gamma - \Gamma') \, \mathrm{d}W = 0$ then $\mathbb{E}\int_0^T (\Gamma - \Gamma')^2 \, \mathrm{d}u = 0$. The hypothesis on the filtration cannot be dropped: if $(\mathcal{F}_t)$ also carries a Brownian motion W' independent of W, then W' is a martingale that is not an integral against W (Exercise 11.7). Part (ii) needs care, because $(\mathcal{F}_t)$ is generated by W, not by $\widetilde{W}$, which may carry less information when Θ is random.

Proof of (ii). By Theorem 11.1(iii), $Y_t = Z_t \widetilde{M}_t$ is a $\mathbb{P}$-martingale, so by (i) $Y_t = Y_0 + \int_0^t \Gamma_u \, \mathrm{d}W_u$, continuous. Then Y/Z is a continuous version of $\widetilde{M}$. By Itô's formula for e^x, $1/Z_t = \exp\big(\int_0^t \Theta \, \mathrm{d}W + \tfrac{1}{2}\int_0^t \Theta^2 \, \mathrm{d}u\big)$ satisfies $\mathrm{d}(1/Z_t) = Z_t^{-1}(\Theta_t \, \mathrm{d}W_t + \Theta_t^2 \, \mathrm{d}t)$, and the product rule gives

$$\mathrm{d}\Big(\frac{Y_t}{Z_t}\Big) = \frac{Y_t}{Z_t}(\Theta_t \, \mathrm{d}W_t + \Theta_t^2 \, \mathrm{d}t) + \frac{\Gamma_t}{Z_t}\,\mathrm{d}W_t + \frac{\Gamma_t \Theta_t}{Z_t}\,\mathrm{d}t = \widetilde{\Gamma}_t(\mathrm{d}W_t + \Theta_t \, \mathrm{d}t), \qquad \widetilde{\Gamma}_t = \frac{\Gamma_t}{Z_t} + \widetilde{M}_t\Theta_t.$$

$\widetilde{\Gamma}$ is in $\mathbb{H}^2_{\mathrm{loc}}$ because $1/Z$ and $\widetilde{M}$ are continuous and $\Gamma, \Theta \in \mathbb{H}^2_{\mathrm{loc}}$, and $\widetilde{\Gamma}\,(\mathrm{d}W + \Theta \, \mathrm{d}t) = \widetilde{\Gamma}\,\mathrm{d}\widetilde{W}$ by (11.2). □

Theorem 11.11 (Fundamental theorems of asset pricing). *Consider the market of Section 11.3, in particular the Black–Scholes market, and let $\mathbb{Q}$ be the measure of Theorem 11.7.*

(i) (First theorem, one direction.) *$\mathbb{Q}$ is a risk-neutral measure, and no tame strategy is an arbitrage.*

If moreover $(\mathcal{F}_t)$ is the augmented filtration of W:

(ii) (Uniqueness.) *$\mathbb{Q}$ is the only risk-neutral measure.*

(iii) (Second theorem: completeness.) *Every $\mathcal{F}_T$-measurable claim V_T with $\mathbb{E}^{\mathbb{Q}}[D_T|V_T|] < \infty$ is replicated by a self-financing strategy whose discounted wealth is the $\mathbb{Q}$-martingale $D_t X_t = \mathbb{E}^{\mathbb{Q}}[D_T V_T \mid \mathcal{F}_t]$, so that its price is (11.7). The hedge is*

$$\Delta_t = \frac{\widetilde{\Gamma}_t}{\sigma_t D_t S_t}, \tag{11.8}$$

where $\widetilde{\Gamma}$ represents this martingale as in Theorem 11.10(ii). If $V_T \ge -c$ for a constant c, the strategy is tame.

Proof. (i) $\mathbb{Q}$ is risk-neutral by Theorem 11.7. Let X be tame, with $X_0 = 0$ and $X_T \ge 0$. By Theorem 11.8(ii), $\mathbb{E}^{\mathbb{Q}}[D_T X_T] \le 0$, so the nonnegative $D_T X_T$ vanishes $\mathbb{Q}$-almost surely, hence $\mathbb{P}$-almost surely, and $X_T = 0$.

(ii) Let $\mathbb{Q}'$ be risk-neutral, $Z'_T > 0$ its density with respect to $\mathbb{P}$ on $\mathcal{F}_T$ (Radon–Nikodym, Section 1.5) and $Z'_t = \mathbb{E}[Z'_T \mid \mathcal{F}_t]$. By Theorem 11.10(i), $Z'_t = 1 + \int_0^t \Gamma_u \, \mathrm{d}W_u$. By Theorem 11.1(iii), $Z'DS$ is a $\mathbb{P}$-martingale, so by Theorem 11.10(i) it is an Itô process without drift. The product rule and (11.4) give another decomposition,

$$\mathrm{d}(Z'_t D_t S_t) = \sigma_t D_t S_t (\Theta_t Z'_t + \Gamma_t)\mathrm{d}t + D_t S_t (\sigma_t Z'_t + \Gamma_t)\mathrm{d}W_t,$$

and by uniqueness of the decomposition (Theorem 8.11(ii)) its drift vanishes: $\Gamma = -\Theta Z'$, $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost everywhere, since $\sigma DS > 0$. So $\mathrm{d}Z'_t = -\Theta_t Z'_t \, \mathrm{d}W_t$, and with $\mathrm{d}(1/Z_t)$ as in the proof of Theorem 11.10(ii),

$$\mathrm{d}\Big(\frac{Z'_t}{Z_t}\Big) = \frac{Z'_t}{Z_t}(\Theta_t \, \mathrm{d}W_t + \Theta_t^2 \, \mathrm{d}t) - \frac{\Theta_t Z'_t}{Z_t} \, \mathrm{d}W_t - \frac{\Theta_t^2 Z'_t}{Z_t} \, \mathrm{d}t = 0.$$

Hence $Z'_T = Z_T$ and $\mathbb{Q}' = \mathbb{Q}$ on $\mathcal{F}_T$.

(iii) $\widetilde{M}_t = \mathbb{E}^{\mathbb{Q}}[D_T V_T \mid \mathcal{F}_t]$ is a $\mathbb{Q}$-martingale; write it as in Theorem 11.10(ii) and define Δ by (11.8). Since $\sigma \geq \varepsilon$ and D is bounded away from 0, $\int_0^T \Delta^2 S^2 \, \mathrm{d}t \leq \varepsilon^{-2} \sup_t D_t^{-2} \int_0^T \widetilde{\Gamma}^2 \, \mathrm{d}t < \infty$, so $(\widetilde{M}_0, \Delta)$ is a self-financing strategy, and by (11.6) its discounted wealth is $\widetilde{M}_0 + \int_0^t \widetilde{\Gamma} \, \mathrm{d}\widetilde{W} = \widetilde{M}_t$. At T, $D_T X_T = D_T V_T$. If $V_T \geq -c$, then $D_t X_t \geq -c \, \mathbb{E}^{\mathbb{Q}}[D_T \mid \mathcal{F}_t] \geq -c \, e^{T \sup |r|}$. □

Trading the claim at its price creates no arbitrage either (compare Theorem 4.10).

> **Corollary 11.12** (The market extended by a claim)**.** *In the setting of Theorem 11.11(ii)–(iii), let V_T be $\mathcal{F}_T$-measurable with $\mathbb{E}^{\mathbb{Q}}[D_T|V_T|] < \infty$, traded at $V_t = \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u \, \mathrm{d}u} V_T \mid \mathcal{F}_t]$, so that $D_t V_t = V_0 + \int_0^t \Gamma^V_u \, \mathrm{d}\widetilde{W}_u$ by Theorem 11.10(ii). Let a strategy hold Δ_t shares and Φ_t units of the claim, with Δ, Φ progressively measurable and $\int_0^T (\Delta_t^2 S_t^2 + \Phi_t^2 (\Gamma^V_t)^2) \, \mathrm{d}t < \infty$ almost surely, and let its discounted wealth be $D_t X_t = X_0 + \int_0^t \Delta_u \, \mathrm{d}(D_u S_u) + \int_0^t \Phi_u \, \mathrm{d}(D_u V_u)$. Then $D_t X_t = X_0 + \int_0^t (\Delta_u \sigma_u D_u S_u + \Phi_u \Gamma^V_u) \, \mathrm{d}\widetilde{W}_u$ is a $\mathbb{Q}$-local martingale, and no such strategy with discounted wealth bounded below is an arbitrage.*

Proof. The formula for $D_t X_t$ is (11.6) plus the representation of DV; its integrand is in $\mathbb{H}^2_{\mathrm{loc}}$. The Fatou argument in the proof of Theorem 11.8(ii) gives $\mathbb{E}^{\mathbb{Q}}[D_T X_T] \leq X_0$ when $DX \geq -c$, and the proof of Theorem 11.11(i) applies unchanged. □

The theorem rests on Girsanov's theorem, proved here for bounded Θ, and on the martingale representation theorem, cited. Two hypotheses carry it. If σ could vanish, Δ in (11.8) would be undefined, and where $\sigma = 0$ and $\alpha \neq r$ the stock would beat or trail the money market without risk, which is an arbitrage. If the filtration carries more randomness than W, for instance a second Brownian motion driving the volatility, there are many risk-neutral measures and some claims cannot be replicated (Exercise 11.7; the Heston model of Chapter 15). For locally bounded semimartingale prices (continuous prices are), no arbitrage in the strong sense of "no free lunch with vanishing risk" among strategies with wealth bounded below is equivalent to the existence of an equivalent measure under which discounted prices are local martingales (Delbaen and Schachermayer, 1994); and, under conditions met here, an arbitrage-free market is complete if and only if this measure is unique (Harrison and Pliska (1981); Harrison and Pliska (1983)), as in the discrete case of Theorem 4.11.

The hedge (11.8) exists, but the representation theorem does not say how to compute $\widetilde{\Gamma}$. For claims $g(S_T)$ in the Black–Scholes model the price is $v(t,S_t)$, and Itô's formula identifies the hedge as $\Delta_t = \partial_x v(t,S_t)$ (Theorem 10.7), for the call $\Delta_t = N(d_+)$ (Theorem 6.14); the same holds in local volatility models (Theorem 12.4). The two methods of the course are complementary: risk-neutral pricing handles any $\mathcal{F}_T$-measurable claim and random coefficients, but its hedge is not explicit; the PDE method of Chapters 10 and 12 needs a Markovian model and a payoff such as $g(S_T)$, and gives the hedge. Where both apply they agree, by the Feynman–Kac formula (Theorem 12.1).

11.5 Change of numeraire

Discounting by the money market is a choice. Any positive traded price N can serve as the unit of account, a *numeraire*, and each numeraire comes with its own martingale measure.

Proposition 11.13 (Change of numeraire)**.** *Let $\mathbb{Q}$ be a risk-neutral measure and N the positive price process of a traded asset that pays no dividends, so that (D_tN_t) is a $\mathbb{Q}$-martingale; let $\mathbb{Q}^N(A) = \mathbb{E}^{\mathbb{Q}}[D_T N_T \mathbf{1}_A]/N_0$ for $A \in \mathcal{F}_T$.*

(i) $\mathbb{Q}^N$ is a probability measure equivalent to $\mathbb{Q}$, with density D_tN_t/N_0 on $\mathcal{F}_t$.

(ii) If (D_tV_t) is a $\mathbb{Q}$-martingale, then (V_t/N_t) is a $\mathbb{Q}^N$-martingale. In particular, for every claim with $\mathbb{E}^{\mathbb{Q}}[D_T|V_T|] < \infty$,

$$\mathbb{E}^{\mathbb{Q}}\Big[e^{-\int_t^T r_u\,\mathrm{d}u} V_T \,\Big|\, \mathcal{F}_t\Big] = N_t\, \mathbb{E}^{\mathbb{Q}^N}\Big[\frac{V_T}{N_T} \,\Big|\, \mathcal{F}_t\Big].$$

(iii) (The stock as numeraire.) *In the Black–Scholes market with $N = S$, $\mathrm{d}\mathbb{Q}^S/\mathrm{d}\mathbb{Q} = e^{-rT}S_T/S_0 = \exp\big(\sigma\widetilde{W}_T - \tfrac{\sigma^2}{2}T\big)$, $W_t^S = \widetilde{W}_t - \sigma t$ is a $\mathbb{Q}^S$-Brownian motion on $[0,T]$,*

$$S_t = S_0 \exp\Big(\sigma W_t^S + \Big(r + \frac{\sigma^2}{2}\Big)t\Big),$$

the money market in units of stock, e^{rt}/S_t, is a $\mathbb{Q}^S$-martingale, and

$$e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+] = S_0\,\mathbb{Q}^S(S_T > K) - Ke^{-rT}\,\mathbb{Q}(S_T > K) = S_0 N(d_+) - Ke^{-rT}N(d_-). \tag{11.9}$$

Proof. (i) $D_TN_T/N_0 > 0$ has $\mathbb{Q}$-mean $D_0N_0/N_0 = 1$, so Theorem 11.1(i) applies with $(\mathbb{P},\mathbb{Q})$ replaced by $(\mathbb{Q},\mathbb{Q}^N)$; the density process is $\mathbb{E}^{\mathbb{Q}}[D_TN_T \mid \mathcal{F}_t]/N_0 = D_tN_t/N_0$. (ii) By Theorem 11.1(iii) with the same replacement, V/N is a $\mathbb{Q}^N$-martingale if and only if $(DN/N_0)(V/N) = DV/N_0$ is a $\mathbb{Q}$-martingale; and $\mathbb{E}^{\mathbb{Q}^N}|V_t/N_t| = \mathbb{E}^{\mathbb{Q}}[D_t|V_t|]/N_0 < \infty$. For the identity, apply this to $V_t = \mathbb{E}^{\mathbb{Q}}[(D_T/D_t)V_T \mid \mathcal{F}_t]$, for which $D_tV_t = \mathbb{E}^{\mathbb{Q}}[D_TV_T \mid \mathcal{F}_t]$ is a $\mathbb{Q}$-martingale, and use $V_t/N_t = \mathbb{E}^{\mathbb{Q}^N}[V_T/N_T \mid \mathcal{F}_t]$. (iii) $e^{-rT}S_T/S_0 = \exp(\sigma\widetilde{W}_T - \frac{\sigma^2}{2}T)$ is the density of Theorem 11.2 under $\mathbb{Q}$ with $\Theta = -\sigma$, so W^S is a $\mathbb{Q}^S$-Brownian motion, and $\sigma\widetilde{W}_t + (r - \frac{\sigma^2}{2})t = \sigma W_t^S + (r + \frac{\sigma^2}{2})t$. The money market e^{rt} has $D_te^{rt} = 1$, so e^{rt}/S_t is a $\mathbb{Q}^S$-martingale by (ii). Split the payoff as $S_T\mathbf{1}_{\{S_T>K\}} - K\mathbf{1}_{\{S_T>K\}}$. By the definition of $\mathbb{Q}^S$, $\mathbb{E}^{\mathbb{Q}}[e^{-rT}S_T\mathbf{1}_{\{S_T>K\}}] = S_0\mathbb{Q}^S(S_T > K)$; under $\mathbb{Q}^S$, $S_T > K$ if and only if $W_T^S/\sqrt{T} > -d_+$, an event of probability $N(d_+)$; and $\mathbb{Q}(S_T > K) = N(d_-)$ (Theorem 6.13). □

The two normal probabilities in the Black–Scholes formula are thus the probabilities of exercise under two measures, with the money market and with the stock as unit. Under $\mathbb{Q}^S$ the stock drifts at $r+\sigma^2$, so exercise is more likely (Figure 11.2). The same idea prices exchange options (Exercise 11.6) and currency options seen from either country (Exercise 11.5); with a zero-coupon bond as numeraire it gives the forward measure (Theorem 14.10).

A numeraire is also a Monte Carlo device. Since $e^{-rT}(S_T-K)^+ = (e^{-rT}S_T/S_0)\,S_0(1-K/S_T)^+$, the call price is $\mathbb{E}^{\mathbb{Q}^S}\big[S_0(1-K/S_T)^+\big]$: simulate S with drift $r+\sigma^2$ and average a payoff bounded by S_0. In Figure 11.2 this estimator has the smaller variance at every strike shown.

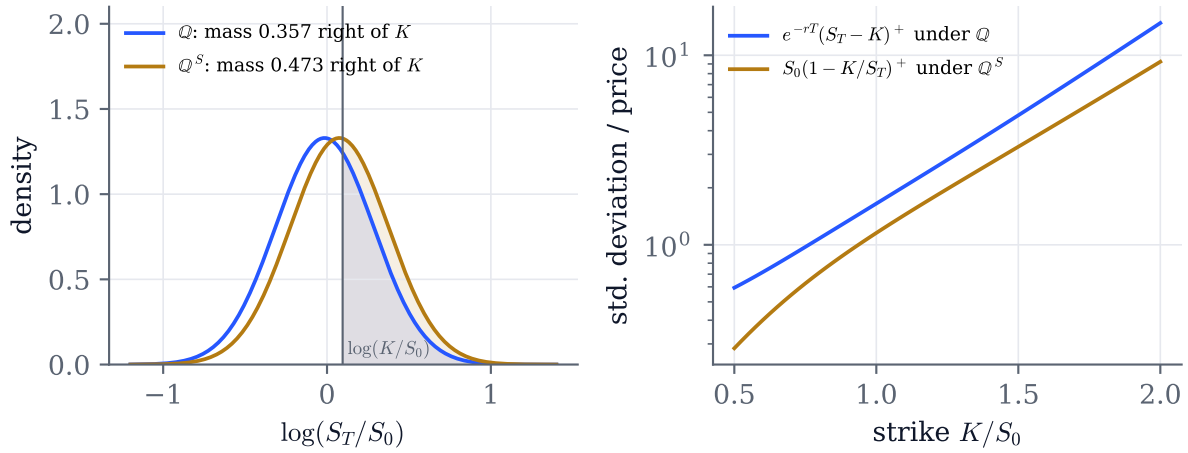

Figure 11.2: The stock as numeraire, with $S_0=100$, $K=110$, $r=3\%$, $\sigma=30\%$, $T=1$. Left: the law of $\log(S_T/S_0)$ under $\mathbb{Q}$ and under $\mathbb{Q}^S$; the shaded masses are $N(d_-)$ and $N(d_+)$ of (11.9). Right: standard deviation per path, relative to the price, of the two Monte Carlo estimators of the call, as the strike varies; computed exactly by quadrature.

▶ Interactive demo: the laws of S_T under $\mathbb{Q}$ and $\mathbb{Q}^S$ for any strike, rate and volatility, and the two Monte Carlo estimators of the call side by side.

11.6 Dividends and exchange rates

Suppose the stock pays dividends continuously: a holder of one share receives $qS_t\,\mathrm{d}t$ in cash over $[t,t+\mathrm{d}t]$, for a constant *dividend yield* $q\in\mathbb{R}$. The price S is then not the value of a self-financing position, and $e^{-rt}S_t$ should not be a martingale: part of the return is paid out. The wealth of a trader holding Δ_t shares satisfies

$$\mathrm{d}X_t = \Delta_t\,\mathrm{d}S_t + q\Delta_t S_t\,\mathrm{d}t + r(X_t-\Delta_t S_t)\mathrm{d}t. \tag{11.10}$$

Reinvesting every dividend in the stock turns one share at time 0 into e^{qt} shares at time t (the number n_t of shares solves $\mathrm{d}n_t = qn_t\,\mathrm{d}t$): a traded position without dividends, worth $e^{qt}S_t$.

Proposition 11.14 (Dividends and exchange rates)**.** *Let $r,q\in\mathbb{R}$ and $\sigma>0$ be constants, and let the stock, paying the dividend yield q, be $S_t=S_0\exp\big(\sigma W_t+(\mu-\frac{\sigma^2}{2})t\big)$ under $\mathbb{P}$. Let $\mathbb{Q}$ and $\widetilde{W}$ be given by Theorem 11.2 with $\Theta\equiv\theta_q=(\mu+q-r)/\sigma$.*

(i) *Under* $\mathbb{Q}$, $\mathrm{d}S_t = (r-q)S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}\widetilde{W}_t$; *the discounted value* $e^{-rt}e^{qt}S_t$ *of the reinvested position is a* $\mathbb{Q}$*-martingale; and the discounted wealth of every strategy* (11.10) *satisfies* $\mathrm{d}(e^{-rt}X_t) = \Delta_t\sigma e^{-rt}S_t\,\mathrm{d}\widetilde{W}_t$. *So Theorem 11.8 and Theorem 11.11 hold with* $\mathbb{Q}$, *and the forward price is* $S_t e^{(r-q)(T-t)}$.

(ii) *With* $\tau = T-t > 0$, $K > 0$ *and* $d_\pm = \big(\log(S_t/K) + (r - q \pm \frac{\sigma^2}{2})\tau\big)/(\sigma\sqrt{\tau})$, *the prices at time* t *of the call and the put are*

$$C_t = S_t e^{-q\tau}N(d_+) - Ke^{-r\tau}N(d_-), \qquad P_t = Ke^{-r\tau}N(-d_-) - S_t e^{-q\tau}N(-d_+),$$

and $C_t - P_t = S_t e^{-q\tau} - Ke^{-r\tau}$.

(iii) (Exchange rates.) *Let* X_t *be the price in domestic currency of one unit of foreign currency, a geometric Brownian motion under* $\mathbb{P}$, *and let* r *and* r_f *be the domestic and foreign interest rates. A unit of foreign currency held in the foreign money market is a traded asset worth* $X_t e^{r_f t}$ *in domestic currency. Hence (i)–(ii) hold with* $S = X$ *and* $q = r_f$*: the forward exchange rate is* $X_t e^{(r-r_f)\tau}$, *and the call on one unit of foreign currency is worth* $X_t e^{-r_f\tau}N(d_+) - Ke^{-r\tau}N(d_-)$ *(Garman and Kohlhagen, 1983).*

Proof. (i) $\sigma W_t + \mu t = \sigma\widetilde{W}_t + (r-q)t$, which gives the dynamics under $\mathbb{Q}$ (as in Section 11.2) and $e^{-(r-q)t}S_t = S_0\exp(\sigma\widetilde{W}_t - \frac{\sigma^2}{2}t)$, a $\mathbb{Q}$-martingale (Theorem 5.12). By the product rule and (11.10), $\mathrm{d}(e^{-rt}X_t) = e^{-rt}\Delta_t(\mathrm{d}S_t + (q-r)S_t\,\mathrm{d}t) = \Delta_t\sigma e^{-rt}S_t\,\mathrm{d}\widetilde{W}_t$, which is (11.6) for this market; with the reinvested position $e^{qt}S_t$ in place of S, the proofs of Theorems 11.8 and 11.11 apply unchanged. The forward contract pays $S_T - K$, worth $e^{-r\tau}\mathbb{E}^{\mathbb{Q}}[S_T \mid \mathcal{F}_t] - Ke^{-r\tau} = S_t e^{-q\tau} - Ke^{-r\tau}$, which vanishes for $K = S_t e^{(r-q)\tau}$.

(ii) Let $S'_u = e^{-q(T-u)}S_u$ for $u \in [0,T]$, the price of the *prepaid forward*, a claim to one share delivered at T. Then $S'_u = S'_0\exp\big(\sigma\widetilde{W}_u + (r - \frac{\sigma^2}{2})u\big)$ and $S'_T = S_T$: S' is the stock of (6.3) started from $S'_0 = e^{-qT}S_0$. By Theorem 6.11 applied to S', $C_t = c(t, S'_t)$ and $P_t = p(t, S'_t)$ with $S'_t = S_t e^{-q\tau}$, and $\log(S'_t/K) + (r \pm \frac{\sigma^2}{2})\tau = \log(S_t/K) + (r - q \pm \frac{\sigma^2}{2})\tau$.

(iii) The asset $X_t e^{r_f t}$ plays the role of the reinvested position $e^{qt}S_t$, and the definitions of (i)–(ii) only use this asset and the domestic money market. □

With $q > 0$, early exercise of an American call can be optimal (Exercise 11.4). In (iii), $\mathbb{Q}$ is the *domestic* risk-neutral measure; a foreign investor uses another one, obtained by a change of numeraire (Exercise 11.5).

Further reading

Shreve (2004b), Chapters 5 and 9, covers this material with the same notation; Björk (2019) develops the martingale approach and numeraires with many examples. Changes of numeraire were used by Margrabe (1978) and Jamshidian (1989), among others, and formalised by Geman et al. (1995). Karatzas and Shreve (1991), Sections 3.4–3.5, and Revuz and Yor (1999), Chapters V and VIII, prove the representation and Girsanov theorems in general. Arbitrage theory in continuous time started with Harrison and Kreps (1979) and Harrison and Pliska (1981); its general form is Delbaen and Schachermayer (1994).

Chapter 12

The Feynman–Kac formula

There are two ways to price a European claim in a diffusion model. The probabilistic way writes the price as an expected discounted payoff under a risk-neutral measure (Chapters 6 and 11). The analytic way characterizes the price as the solution of a partial differential equation, obtained by hedging (Chapter 10). The *Feynman–Kac formula* says that the two agree: for a diffusion X with generator $\mathcal{L}_t$, the function $v(t,x) = \mathbb{E}\big[e^{-\int_t^T k(u,X_u)\,\mathrm{d}u} h(X_T) \,\big|\, X_t = x\big]$ solves $\partial_t v + \mathcal{L}_t v - kv = 0$ with $v(T,\cdot) = h$, and, conversely, a solution of this equation with moderate growth is that expectation. The name joins Feynman's path integrals in quantum mechanics (Feynman, 1948) to Kac (1949), who proved the formula for Brownian motion with a potential k.

The formula is used in both directions: prices in models without closed forms, such as the local volatility models of Chapter 9, are computed by solving equations on a grid (Chapter 16), and solutions of equations by Monte Carlo. Along the way we reduce the Black–Scholes equation to the heat equation, find the equations satisfied by transition densities, which lead to Dupire's formula in Chapter 15, and study two contracts: the digital option, whose hedge ratio explodes, and the pay-later option.

Setting. Throughout, $T > 0$, W is a Brownian motion with a filtration $(\mathcal{F}_t)$ (Theorem 5.11), and, as in Chapter 9, every event of probability zero belongs to $\mathcal{F}_0$. The state space is an open interval $I = (\alpha, \beta)$, with $-\infty \le \alpha < \beta \le \infty$: $I = \mathbb{R}$ for most diffusions, $I = (0,\infty)$ for prices. Let $b, \sigma : [0,T] \times I \to \mathbb{R}$ be continuous, and consider

$$\mathrm{d}X_u = b(u, X_u)\,\mathrm{d}u + \sigma(u, X_u)\,\mathrm{d}W_u. \tag{12.1}$$

For $t \in [0,T)$, a *solution on* $[t,T]$ *with values in* I is defined as in Theorem 9.1, with $[t,T]$ in place of $[0,T]$ and an $\mathcal{F}_t$-measurable initial value X_t; under the hypotheses of Theorem 9.5 and with $\mathbb{E}[X_t^2] < \infty$ it exists and is unique, by the time shift $\widetilde{W}_s = W_{t+s} - W_t$, $\widetilde{\mathcal{F}}_s = \mathcal{F}_{t+s}$ described after Theorem 9.9. When $X_t = x$ is a constant we write $X^{t,x}$. The generator $\mathcal{L}_t$ is that of Theorem 9.10, acting on functions on I: $(\mathcal{L}_t f)(x) = \frac{1}{2}\sigma(t,x)^2 f''(x) + b(t,x) f'(x)$. For a function $v(t,x)$, $\mathcal{L}_t v$ acts on $v(t,\cdot)$, and $v \in C^{1,2}([0,T) \times I)$ means, as in Chapter 8, that $\partial_t v$, $\partial_x v$, $\partial_{xx} v$ exist and are continuous on $[0,T) \times I$.

12.1 From the Markov property to an equation

Here is why an equation should appear. Let h be a bounded payoff and $g(t,x) = \mathbb{E}[h(X_T^{t,x})]$. By the Markov property (Theorem 9.9), $\mathbb{E}[h(X_T) \mid \mathcal{F}_t] = g(t, X_t)$ for a solution X started at time 0, so, by the tower property, $M_t = g(t, X_t)$ is a martingale. If g is smooth, Itô's formula (Theorem 8.13) writes it as

$$\mathrm{d}M_t = (\partial_t g + \mathcal{L}_t g)(t, X_t)\,\mathrm{d}t + \partial_x g(t, X_t)\,\sigma(t, X_t)\,\mathrm{d}W_t.$$

A martingale has no drift: over a short interval $[t, t+\delta]$, $\mathbb{E}[M_{t+\delta} - M_t \mid \mathcal{F}_t] = 0$, while the $\mathrm{d}t$-term contributes about $(\partial_t g + \mathcal{L}_t g)(t, X_t)\,\delta$. So $\partial_t g + \mathcal{L}_t g$ must vanish at (t, X_t), and hence wherever the process can be: g solves the *backward equation* $\partial_t g + \mathcal{L}_t g = 0$, with $g(T, \cdot) = h$. With discounting, $e^{-\int_0^t k(u,X_u)\,\mathrm{d}u} g(t, X_t)$ is the martingale, and the product rule adds the term $-kg$. For $X = W$ the equation is the backward heat equation, and this is Chapter 8 read in reverse: there, a solution f of $\partial_t f + \frac{1}{2}\partial_{xx} f = 0$ made $f(t, W_t)$ a martingale.

The argument assumes that g is smooth, which is the hard part. The direction proved below, which is what pricing uses, starts from a smooth solution and identifies it with the expectation.

12.2 The Feynman–Kac formula

Theorem 12.1 (Feynman–Kac)**.** *Let $k : [0,T] \times I \to \mathbb{R}$ be continuous and bounded below, $k \geq -k_0$ for a constant $k_0 \geq 0$, and let $h : I \to \mathbb{R}$ be Borel. Let $v \in C^{1,2}([0,T) \times I)$ solve*

$$\partial_t v(t,x) + (\mathcal{L}_t v)(t,x) - k(t,x)\,v(t,x) = 0, \qquad (t,x) \in [0,T) \times I, \tag{12.2}$$

and let $E \subseteq I$ be a Borel set such that

$$\lim_{(s,y)\to(T,y_0),\ s<T} v(s,y) = h(y_0) \qquad \text{for every } y_0 \in I \setminus E. \tag{12.3}$$

Let $t \in [0,T)$ and let X be a solution of (12.1) on $[t,T]$ with values in I, such that $\mathbb{P}(X_T \in E) = 0$ and

$$\mathbb{E}\Big[\sup_{t \leq u < T} |v(u, X_u)|\Big] < \infty. \tag{12.4}$$

Then $e^{-\int_t^T k(u,X_u)\,\mathrm{d}u} h(X_T)$ is integrable and, almost surely,

$$v(t, X_t) = \mathbb{E}\Big[\exp\Big(-\int_t^T k(u, X_u)\,\mathrm{d}u\Big)\, h(X_T) \,\Big|\, \mathcal{F}_t\Big]. \tag{12.5}$$

In particular, if $X_t = x \in I$ is a constant, then $v(t,x) = \mathbb{E}\big[\exp\big(-\int_t^T k(u, X_u^{t,x})\,\mathrm{d}u\big)\, h(X_T^{t,x})\big]$.

If v is continuous on $[0,T] \times I$ with $v(T, \cdot) = h$, take $E = \emptyset$. The set E is there for payoffs such as $\mathbf{1}_{\{x>K\}}$, at whose jump no solution can be continuous; it is harmless when X_T has a density and E is countable.

Proof. By the time shift of the setting, replacing T by $T - t$ and v, k, b, σ by $(s,y) \mapsto v(t+s, y)$

and so on, we may and do assume $t = 0$. Let $G = \sup_{0\le u<T} |v(u, X_u)|$, a random variable since the supremum may be taken over rational u, and integrable by (12.4).

The discounted solution has no drift. Let $A_u = \int_0^u k(s, X_s)\,\mathrm{d}s$ and $Z_u = e^{-A_u}$. Then Z is continuous and adapted, $0 < Z_u \le e^{k_0 T}$, and $\mathrm{d}Z_u = -k(u, X_u)Z_u\,\mathrm{d}u$ has no $\mathrm{d}W$-term. Fix $T' < T$. Itô's formula on $[0, T']$ (Theorem 8.17 with one process and the open interval I; on $[0, T']$ it uses v only on $[0, T'] \times I$) gives $\mathrm{d}v(u, X_u) = (\partial_t v + \mathcal{L}_u v)(u, X_u)\,\mathrm{d}u + \partial_x v(u, X_u)\sigma(u, X_u)\,\mathrm{d}W_u$. By the product rule (Theorem 8.18), with $[Z, v(\cdot, X)] = 0$ (Theorem 8.11), and by (12.2), the process $Y_u = Z_u v(u, X_u)$ satisfies

$$\mathrm{d}Y_u = Z_u(\partial_t v + \mathcal{L}_u v - kv)(u, X_u)\,\mathrm{d}u + H_u\,\mathrm{d}W_u = H_u\,\mathrm{d}W_u, \qquad H_u = Z_u\,\partial_x v(u, X_u)\,\sigma(u, X_u).$$

Since T' is arbitrary, $Y_u = Y_0 + \int_0^u H_s\,\mathrm{d}W_s$ for all $u < T$, almost surely.

Localization. The integrand H is continuous and adapted, but it may be unbounded, so the stochastic integral need not be a martingale (Theorem 7.16). Choose $\alpha < \alpha_n < \beta_n < \beta$ with $\alpha_n \downarrow \alpha$ and $\beta_n \uparrow \beta$, let $T_n = T - 1/n$ (for $n > 1/T$), and let θ_n be the first time $u \in [0, T_n]$ with $X_u \notin (\alpha_n, \beta_n)$, or T_n if there is none. It is a stopping time: by continuity of the paths, for $u < T_n$ the event $\{\theta_n \le u\}$ is the event that the distance from X_s to $\mathbb{R} \setminus (\alpha_n, \beta_n)$ has infimum 0 over the rational $s \le u$ and $s = u$, which is in $\mathcal{F}_u$. If $0 < s \le \theta_n$, then $X_s \in [\alpha_n, \beta_n]$, so (s, X_s) lies in the compact set $[0, T_n] \times [\alpha_n, \beta_n]$, on which the continuous functions $\partial_x v$ and σ are bounded, by c_n say. Hence $\mathbf{1}_{\{0<s\le\theta_n\}}H_s$ is bounded by $e^{k_0 T}c_n$, so it is in $\mathbb{H}^2$ and its integral is a martingale (Theorem 7.7). Moreover $\int_0^{\theta_n} H_s\,\mathrm{d}W_s = \int_0^T \mathbf{1}_{\{0<s\le\theta_n\}}H_s\,\mathrm{d}W_s$ (the value of the integrand at the single time $s = 0$ does not matter): apply Theorem 7.15 on $[0, T_n]$ to $H\mathbf{1}_{[0,\tau_m]} \in \mathbb{H}^2$, with τ_m as in Theorem 7.16, and the stopping time θ_n, and let $m \to \infty$, using the isometry on the right. Taking conditional expectations,

$$v(0, X_0) = Y_0 = \mathbb{E}[Y_{\theta_n} \,|\, \mathcal{F}_0],$$

where Y_{θ_n} is integrable because $|Y_{\theta_n}| \le e^{k_0 T}G$.

Passage to the limit. Fix an outcome. The path X maps $[0, T]$ onto a compact subset of I, contained in (α_n, β_n) for n large, and then $\theta_n = T_n$. So $(\theta_n, X_{\theta_n}) \to (T, X_T)$ with $\theta_n < T$, and on the event $\{X_T \notin E\}$, which has probability one, (12.3) and the continuity of Z give $Y_{\theta_n} \to Z_T h(X_T)$. The limit satisfies $|Z_T h(X_T)| \le e^{k_0 T}G$, so it is integrable, and $Y_{\theta_n} \to Z_T h(X_T)$ in L^1 by dominated convergence. Conditional expectation is a contraction in L^1, since $\mathbb{E}|\mathbb{E}[U \mid \mathcal{F}_0]| \le \mathbb{E}|U|$ by Theorem 1.16(ii). Hence $\mathbb{E}[Y_{\theta_n} \mid \mathcal{F}_0] \to \mathbb{E}[Z_T h(X_T) \mid \mathcal{F}_0]$ in L^1, while the left side equals $v(0, X_0)$ for every n. This is (12.5). □

Remark 12.2 (On the hypotheses).

(a) *Integrability.* Condition (12.4) holds if v is bounded. It also holds if $|v(u, y)| \le C(1 + y^2)$ and $\mathbb{E}[\sup_{t\le u\le T} X_u^2] < \infty$; for $I = \mathbb{R}$ and coefficients satisfying the hypotheses of Theorem 9.5, the latter follows from $\mathbb{E}[X_t^2] < \infty$ by part (b) of that theorem, after the time shift. For geometric Brownian motion on $I = (0, \infty)$, $S_u = S_0 e^{\sigma W_u + (\mu - \sigma^2/2)u}$, growth like a power of y or of $1/y$ is allowed: for every $q \in \mathbb{R}$, $\mathbb{E}[\sup_{u\le T} S_u^q] < \infty$. Indeed $S_u^q \le S_0^q e^{c_q T} M_u$ with $c_q = |q(\mu - \frac{\sigma^2}{2})| + \frac{q^2\sigma^2}{2}$ and $M_u = e^{q\sigma W_u - q^2\sigma^2 u/2}$, a continuous martingale with $\mathbb{E}[M_T^2] = e^{q^2\sigma^2 T}$ (Theorem 5.12); Doob's maximal inequality (Theorem 7.6) gives $\mathbb{E}[\sup_{u\le T} M_u]^2 \le \mathbb{E}[\sup_{u\le T} M_u^2] \le 4e^{q^2\sigma^2 T}$.

(b) *Uniqueness.* If (12.4) and $\mathbb{P}(X_T^{t,x} \in E) = 0$ hold for the solutions from every starting point (t, x), two solutions of (12.2)–(12.3) are equal to the same expectation, hence to each other. For $I = \mathbb{R}$ and coefficients as in Theorem 9.5, there is at most one solution with $|v(t, x)| \le C(1 + x^2)$ (and $E = \emptyset$, say). Some growth condition is necessary: Tychonoff constructed a solution of the heat equation $\partial_t v + \frac{1}{2}\partial_{xx} v = 0$ with $v(T, \cdot) = 0$ that is not identically zero. Uniqueness holds among solutions bounded by Ce^{ax^2} (Karatzas and Shreve (1991), Section 4.3), so Tychonoff's solution grows faster than that, for all constants C and a.

(c) *The bound on k.* It makes the discount factor bounded, which is what the domination in the proof needs. Without it the right side of (12.5) can be infinite: for $k(x) = -\lambda x^2$ with $\lambda > 0$ and Brownian motion, it is infinite for large T (Exercise 12.3).

Remark 12.3 (The converse). That the expectation is a smooth solution needs analysis beyond these notes. Karatzas and Shreve (1991), Theorem 5.7.6, proves the representation (12.5), with a running payoff as well, for solutions of polynomial growth. For the converse, their Remark 5.7.8 recalls conditions under which the equation has a solution of polynomial growth: σ^2 bounded below by a positive constant, regular bounded coefficients, and a continuous payoff of polynomial growth; the solution is built by the parametrix method of Friedman (1964), Chapter 1, and by the theorem it is the expectation. Øksendal (2003), Theorem 8.2.1, proves the converse for coefficients that do not depend on t and payoffs $h \in C^2$ with compact support; Shreve (2004b), Theorems 6.4.1 and 6.4.3, argue as in Section 12.1. In the Black–Scholes model, Theorem 12.7 proves the converse for every payoff of polynomial growth.

12.3 Pricing equations for local volatility models

In this section, as in Theorem 9.13, $\sigma(t, x)$ is the volatility of returns: the diffusion coefficient of the stock is $x\sigma(t, x)$. We assume that σ is continuous and that, for constants M and ℓ and all $t \in [0, T]$, $x, x' > 0$,

$$0 \le \sigma(t, x) \le M, \qquad |\sigma(t, x) - \sigma(t, x')| \le \ell\,|\log x - \log x'|. \tag{12.6}$$

Then, for any probability measure, Brownian motion B and $\rho \in \mathbb{R}$, the equation $\mathrm{d}S_t = \rho S_t\,\mathrm{d}t + \sigma(t, S_t)S_t\,\mathrm{d}B_t$ with $S_0 > 0$ has exactly one positive solution (Theorem 9.14), and $\mathbb{E}[\sup_{t \le T} S_t^2] < \infty$: as shown in the proof of Theorem 9.14, $a(t, x) = x\sigma(t, x)$ for $x > 0$, $a(t, x) = 0$ for $x \le 0$, is Lipschitz on $\mathbb{R}$, so Theorem 9.5(b) applies to $\mathrm{d}S = \rho S\,\mathrm{d}t + a(t, S)\,\mathrm{d}B$, which S solves.

Corollary 12.4 (Pricing equations for local volatility models)**.** *Assume* (12.6) *and let* $r \in \mathbb{R}$. *Let* $\mathbb{Q}$ *be a probability measure and* $\widetilde{W}$ *a* $\mathbb{Q}$*-Brownian motion with filtration* $(\mathcal{F}_t)$*, and let* S *be the positive solution of*

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma(t, S_t)S_t\,\mathrm{d}\widetilde{W}_t, \qquad S_0 > 0.$$

Let $g : (0, \infty) \to \mathbb{R}$ *be Borel, and let* $v \in C^{1,2}([0, T) \times (0, \infty))$ *satisfy* $|v(t, x)| \le C(1 + x^2)$ *and solve the* pricing equation

$$\partial_t v + rx\,\partial_x v + \frac{\sigma(t, x)^2 x^2}{2}\,\partial_{xx} v = rv \qquad \text{on } [0, T) \times (0, \infty), \tag{12.7}$$

with the terminal condition (12.3) *for* $h = g$ *and a Borel set* E *such that* $\mathbb{Q}(S_T \in E) = 0$.

(a) Price. *For every $t \in [0,T)$, $v(t,S_t) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[g(S_T) \mid \mathcal{F}_t]$ almost surely; in particular $v(0,S_0) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[g(S_T)]$.*

(b) Hedge. *Let $\mathbb{P}$ be a probability measure, W a $\mathbb{P}$-Brownian motion for $(\mathcal{F}_t)$, and S a positive Itô process with $\mathrm{d}S_t = \mu_t S_t\,\mathrm{d}t + \sigma(t,S_t)S_t\,\mathrm{d}W_t$ and $S_0 > 0$, for some progressively measurable μ with $\int_0^T |\mu_t|\,\mathrm{d}t < \infty$ almost surely. Fix $T' < T$, and let X be the value of the self-financing strategy on $[0,T']$ (Theorem 10.1, extended in the obvious way to this S) that starts from $X_0 = v(0,S_0)$ and holds $\Delta_t = \partial_x v(t,S_t)$ shares: $e^{-rt}X_t = v(0,S_0) + \int_0^t \Delta_u\,\mathrm{d}(e^{-ru}S_u)$. Then, almost surely, $X_t = v(t,S_t)$ for every $t \le T'$; and $v(T',S_{T'}) \to g(S_T)$ as $T' \uparrow T$ if $S_T \notin E$.*

Proof. (a) Apply Theorem 12.1 under $\mathbb{Q}$, with $I = (0,\infty)$, drift $b(t,x) = rx$, diffusion coefficient $x\sigma(t,x)$, $k \equiv r$ and $h = g$: then (12.2) is (12.7). Condition (12.4) holds because $|v(u,S_u)| \le C(1 + \sup_{s\le T} S_s^2)$, which is integrable by the paragraph before the corollary.

(b) Fix $T' < T$. By Itô's formula on $[0,T']$ (Theorem 8.17 on $(0,\infty)$), the product rule and (12.7), which says that $\partial_t v + \frac{1}{2}\sigma^2 x^2 \partial_{xx} v - rv = -rx\,\partial_x v$,

$$\begin{aligned}\mathrm{d}\big(e^{-ru}v(u,S_u)\big) &= e^{-ru}\Big(\partial_t v + \mu_u S_u \partial_x v + \frac{\sigma^2 S_u^2}{2}\partial_{xx} v - rv\Big)\mathrm{d}u + e^{-ru}\partial_x v\,\sigma S_u\,\mathrm{d}W_u \\ &= \Delta_u\, e^{-ru}S_u\big((\mu_u - r)\,\mathrm{d}u + \sigma\,\mathrm{d}W_u\big),\end{aligned}$$

with $\sigma = \sigma(u,S_u)$ and the derivatives of v at (u,S_u). The right side is $\Delta_u\,\mathrm{d}(e^{-ru}S_u)$, by the product rule again. So $e^{-rt}v(t,S_t)$ and $e^{-rt}X_t$ have the same differential and the same initial value, so they are equal on $[0,T']$. As $T' \uparrow T$, $(T',S_{T'}) \to (T,S_T)$, and (12.3) gives $v(T',S_{T'}) \to g(S_T)$ when $S_T \notin E$. □

Part (b) is the local volatility version of the replication argument of Chapter 10: the real-world drift μ drops out, as in Theorem 10.6, and the price and the hedge depend on the model only through r and the function $\sigma(t,x)$. Part (a) identifies the replication cost with the expectation under the risk-neutral measure (Theorem 11.6). For constant σ, the call function c of Theorem 6.11 satisfies $0 \le c(t,x) \le x$ and solves (12.7) (Theorem 10.9), and the corollary recovers the formula from the equation.

Remark 12.5 (The hedge up to maturity). Part (b) stops at $T' < T$ because a strategy on $[0,T]$ needs $\int_0^T \Delta_t^2\,\mathrm{d}t < \infty$, which can fail: if the real-world drift turns $\log S$ into a bridge pinned at $\log K$, the delta of a digital satisfies $\Delta_t^2 \asymp 1/(T-t)$. It holds if $\sigma \ge \lambda > 0$ and $\mathbb{E}^{\mathbb{Q}}[g(S_T)^2] < \infty$. By (a), $M_t = e^{-rt}v(t,S_t) = \mathbb{E}^{\mathbb{Q}}[e^{-rT}g(S_T) \mid \mathcal{F}_t]$ is a square-integrable martingale, and by (b) under $\mathbb{Q}$, $M_t - M_0 = \int_0^t H_u\,\mathrm{d}\widetilde{W}_u$ for $t < T$, with $H_u = e^{-ru}\sigma(u,S_u)S_u\Delta_u$. Fix $T' < T$ and let τ_m be as in Theorem 7.16. The isometry gives $\mathbb{E}^{\mathbb{Q}}[(M_{T'\wedge\tau_m} - M_0)^2] = \mathbb{E}^{\mathbb{Q}}\int_0^{T'\wedge\tau_m} H_u^2\,\mathrm{d}u$; as $m \to \infty$, dominated convergence on the left (by Theorem 7.6, $\mathbb{E}^{\mathbb{Q}}[\sup_{t\le T'} M_t^2] < \infty$) and monotone convergence on the right give $\mathbb{E}^{\mathbb{Q}}\int_0^{T'} H_u^2\,\mathrm{d}u = \mathbb{E}^{\mathbb{Q}}[M_{T'}^2] - M_0^2 \le \mathbb{E}^{\mathbb{Q}}[M_T^2] - M_0^2$, by conditional Jensen. Letting $T' \uparrow T$, $\mathbb{E}^{\mathbb{Q}}\int_0^T H_u^2\,\mathrm{d}u \le \mathrm{Var}^{\mathbb{Q}}(e^{-rT}g(S_T)) < \infty$, as in Exercise 12.5(b). Since $\sigma \ge \lambda$ and S is continuous and positive, $\int_0^T \Delta_t^2\,\mathrm{d}t < \infty$ $\mathbb{Q}$-almost surely, hence almost surely under every measure equivalent to $\mathbb{Q}$: the hedge is then a self-financing strategy on $[0,T]$, with value $g(S_T)$ at T.

Remark 12.6 (Solving the equation). If σ is bounded below by a positive constant and regular enough, the equation, written in $y = \log x$, is uniformly parabolic with bounded coefficients, and

the theory recalled in Theorem 12.3, which allows continuous terminal data growing exponentially in y (Friedman (1964), Chapter 1), gives a classical solution; for discontinuous payoffs such as digitals, it does not apply as stated (but see Theorem 12.7 for Black–Scholes). Apart from a few models it has no closed form, and it is solved by finite differences on a grid (Chapter 16). One solve prices every initial state at once; Monte Carlo prices one, with a statistical error, but its cost does not grow exponentially with the number of state variables, as that of a grid does.

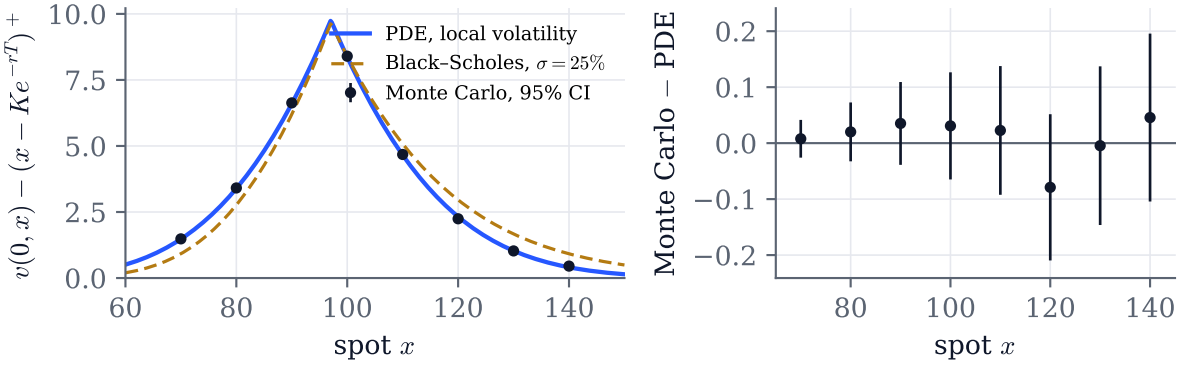

Figure 12.1: The price at time 0 of a call ($K = 100$, $T = 1$, $r = 3\%$) in the local volatility model with $\sigma(t,x) = 0.25 - 0.12\tanh(2\log(x/100))$, which satisfies (12.6). Left: the time value $v(0,x) - (x - Ke^{-rT})^+$, from one solution of (12.7) (Crank–Nicolson in $\log x$, 1601 points, 800 time steps), and from Monte Carlo at eight initial prices (10^5 paths each, Euler scheme for $\log S$ with 100 steps); dashed: Black–Scholes with the at-the-money volatility 25%. Right: Monte Carlo minus PDE, with 95% confidence intervals (`code/src/ieor222/feynmankac.py`).

Figures 12.1 and 12.2 use a bounded local volatility with a downward skew, as in the CEV model of Chapter 9: 25% at $x = 100$, between 13% and 37%. The two methods agree within the Monte Carlo error. Compared with Black–Scholes at 25%, the call is worth more when the stock is below the strike and less when it is above: what matters is the volatility between the current price and the strike, and the skew makes it higher below 100 and lower above.

▶ Interactive demo: the same price by Monte Carlo and by finite differences, and the local volatility price surface, for the skew, the strike and the maturity you choose.

12.4 The heat equation and the Black–Scholes equation

With constant coefficients, the Black–Scholes equation is the heat equation in disguise, and the heat equation can be solved explicitly. This gives the converse of the Feynman–Kac formula in the Black–Scholes model, for every payoff of polynomial growth, smooth or not. Let

$$G(s,z) = \frac{1}{\sqrt{2\pi s}}\, e^{-z^2/(2s)}, \qquad s > 0,\ z \in \mathbb{R},$$

be the *heat kernel*: the density of $\mathcal{N}(0,s)$, which is also the transition density $p(t,x;t+s,x+z)$ of Brownian motion. A direct computation gives $\partial_z G = -\frac{z}{s}G$, $\partial_{zz}G = (\frac{z^2}{s^2} - \frac{1}{s})G$ and $\partial_s G = (\frac{z^2}{2s^2} - \frac{1}{2s})G$, so $\partial_s G = \frac{1}{2}\partial_{zz}G$.

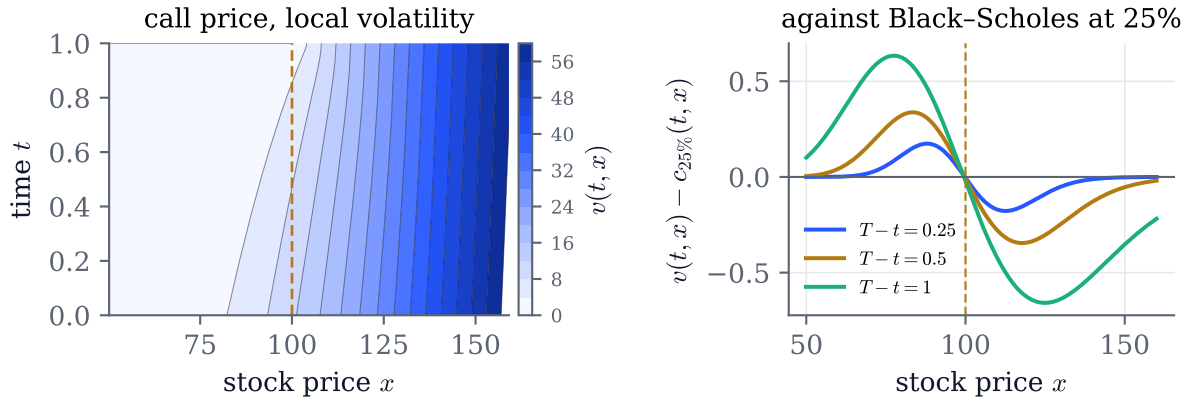

Figure 12.2: The same local volatility model. Left: the price $v(t,x)$ of the call with $K=100$, $T=1$, as a function of time and of the stock price, from one solution of (12.7) (contours every 4). Right: its difference from the Black–Scholes price with $\sigma = 25\%$, for three times to maturity.

Proposition 12.7 (The heat equation and the Black–Scholes equation)**.** *Let $r \in \mathbb{R}$ and $\sigma > 0$.*

(i) Let $v : [0,T) \times (0,\infty) \to \mathbb{R}$ and $u : (0, \sigma^2 T] \times \mathbb{R} \to \mathbb{R}$ be related by

$$v(t,x) = e^{-r(T-t)}\, u\Big(\sigma^2(T-t),\ \log x + \Big(r - \frac{\sigma^2}{2}\Big)(T-t)\Big). \tag{12.8}$$

Then $v \in C^{1,2}([0,T)\times(0,\infty))$ if and only if $u \in C^{1,2}((0,\sigma^2 T]\times\mathbb{R})$, and then, with the derivatives of u at the corresponding point (s,z),

$$\partial_t v + rx\,\partial_x v + \frac{\sigma^2 x^2}{2}\,\partial_{xx} v - rv = \sigma^2 e^{-r(T-t)}\Big(\frac{1}{2}\,\partial_{zz} u - \partial_s u\Big).$$

So v solves the Black–Scholes equation if and only if u solves the heat equation $\partial_s u = \frac{1}{2}\partial_{zz} u$.

(ii) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be Borel with $|\varphi(z)| \le a e^{p|z|}$ for constants $a, p \ge 0$, and for $s > 0$ let

$$u(s,z) = \int_{\mathbb{R}} \varphi(w)\, G(s, z-w)\, \mathrm{d}w = \mathbb{E}\big[\varphi(z + \sqrt{s}\, Y)\big], \qquad Y \sim \mathcal{N}(0,1).$$

Then u is infinitely differentiable on $(0,\infty)\times\mathbb{R}$, solves the heat equation, and $u(s,z) \to \varphi(z_0)$ as $(s,z) \to (0,z_0)$, at every point z_0 where φ is continuous.

(iii) Let $g : (0,\infty) \to \mathbb{R}$ be Borel with $|g(x)| \le a(x^p + x^{-p})$, and let

$$v(t,x) = e^{-r(T-t)}\, \mathbb{E}\Big[g\Big(x \exp\Big(\sigma\sqrt{T-t}\, Y + \Big(r - \frac{\sigma^2}{2}\Big)(T-t)\Big)\Big)\Big],$$

the function such that $v(t,S_t) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[g(S_T) \mid \mathcal{F}_t]$ in the Black–Scholes model (by the independence lemma, Theorem 1.19, as in Theorem 6.11). Then v is infinitely differentiable on $[0,T)\times(0,\infty)$, solves the Black–Scholes equation, and $v(t,x) \to g(x_0)$ as $(t,x) \to (T,x_0)$, at every point x_0 where g is continuous. If the points of discontinuity of g form a set of Lebesgue measure zero, v is the only solution with this terminal behaviour and $|v(t,x)| \le C(x^q + x^{-q})$ for some constants C, q.

Proof. (i) Write $\tau = T - t$, $s = \sigma^2\tau$, $z = \log x + (r - \frac{\sigma^2}{2})\tau$. The map $(t, x) \mapsto (s, z)$ is a smooth bijection from $[0, T) \times (0, \infty)$ onto $(0, \sigma^2 T] \times \mathbb{R}$, with smooth inverse $t = T - s/\sigma^2$, $x = \exp\left(z - (r - \frac{\sigma^2}{2})s/\sigma^2\right)$, which gives the equivalence of the regularity. Since $\partial_t \tau = -1$, the chain rule gives

$$\partial_t v = e^{-r\tau}\Big(ru - \sigma^2 \partial_s u - \Big(r - \frac{\sigma^2}{2}\Big)\partial_z u\Big), \quad x\,\partial_x v = e^{-r\tau}\partial_z u, \quad x^2 \partial_{xx} v = e^{-r\tau}(\partial_{zz} u - \partial_z u),$$

and the terms in u and $\partial_z u$ cancel in the combination.

(ii) Let $0 < s_1 < s_2$ and $R > 0$. Every partial derivative of $G(s, z)$ is a polynomial in z and $s^{-1/2}$ times $G(s, z)$, so for $s \in [s_1, s_2]$ it is bounded by $c\, e^{-z^2/(4s_2)}$ for a constant c. For $|z| \le R$, $|\varphi(w)| \le a e^{pR} e^{p|z-w|}$, so the derivatives of $\varphi(w) G(s, z - w)$ with respect to s and z are bounded, uniformly on $[s_1, s_2] \times [-R, R]$, by $c\, a e^{pR} e^{p|y| - y^2/(4s_2)}$ with $y = z - w$, which is integrable in w. By dominated convergence and the mean value theorem, we may differentiate under the integral sign as often as we wish; so u is smooth and $\partial_s u - \frac{1}{2}\partial_{zz} u = \int \varphi(w)\big(\partial_s G - \frac{1}{2}\partial_{zz} G\big)(s, z - w)\, \mathrm{d}w = 0$. If φ is continuous at z_0 and $(s_n, z_n) \to (0, z_0)$, then $\varphi(z_n + \sqrt{s_n}\, Y) \to \varphi(z_0)$ for every value of Y, and for $s_n \le 1$ and $|z_n - z_0| \le 1$ these variables are bounded by the integrable $a e^{p(|z_0|+1)} e^{p|Y|}$: dominated convergence again.

(iii) Let $\varphi(z) = g(e^z)$, so that $|\varphi(z)| \le a(e^{pz} + e^{-pz}) \le 2a e^{p|z|}$, and φ is continuous at $\log x_0$ when g is continuous at x_0. With the notation of (i), $x \exp(\sigma\sqrt{\tau}\, Y + (r - \frac{\sigma^2}{2})\tau) = \exp(z + \sqrt{s}\, Y)$, so v and the function u of (ii) are related by (12.8). Parts (i) and (ii) give the regularity and the equation, and $(t, x) \to (T, x_0)$ means $(s, z) \to (0, \log x_0)$. For uniqueness, let E be the set of discontinuity points of g. It is a Borel set: it is the union over $n \ge 1$ of the closed sets of points where the oscillation of g is at least $1/n$. By assumption it has Lebesgue measure zero. Under $\mathbb{Q}$, S_T given $S_t = x$ has a density, so $\mathbb{Q}(S_T^{t,x} \in E) = 0$, and Theorem 12.2(a) gives (12.4) for solutions with $|v(t, x)| \le C(x^q + x^{-q})$. Theorem 12.1, with $I = (0, \infty)$ and $k \equiv r$, identifies every such solution with $e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[g(S_T^{t,x})]$, which is $v(t, x)$; and v itself satisfies the growth bound, by the moments of S_T. □

The heat kernel smooths whatever it is given: before maturity, the price of a digital option, whose payoff jumps, is as smooth as that of a call. The price of the irregularity is paid near maturity, where derivatives blow up: $u(s, z) = N(z/\sqrt{s})$ solves the heat equation with $\varphi = \mathbf{1}_{(0,\infty)}$, and $\partial_z u(s, 0) = 1/\sqrt{2\pi s} \to \infty$ as $s \downarrow 0$ (Section 12.6). For the call, $\varphi(z) = (e^z - K)^+$, and the Gaussian integral of (ii) is the one computed in Theorem 6.11.

12.5 Transition densities and Kolmogorov's equations

In this section $I = \mathbb{R}$, and b and σ are continuous and satisfy the hypotheses of Theorem 9.5. A *transition density* of (12.1) is a Borel function $p(t, x; s, y)$, for $0 \le t < s \le T$ and $x, y \in \mathbb{R}$, such that

$$\mathbb{P}(X_s^{t,x} \in A) = \int_A p(t, x; s, y)\, \mathrm{d}y \qquad \text{for every Borel } A \subseteq \mathbb{R}.$$

By the Markov property, $\mathbb{E}[h(X_s) \mid \mathcal{F}_t] = \int h(y)\, p(t, X_t; s, y)\, \mathrm{d}y$ for bounded Borel h. The density satisfies an equation in the starting point (t, x) and one in the arrival point (s, y).

Theorem 12.8 (Kolmogorov's backward and forward equations)**.**

(i) Backward equation. *Suppose that $\sigma^2 \geq \lambda$ for a constant $\lambda > 0$, and that b and σ^2 are bounded and uniformly Hölder continuous on $[0,T]\times\mathbb{R}$. Then there is a transition density p which, for fixed (s,y), is in $C^{1,2}([0,s)\times\mathbb{R})$ as a function of (t,x) and satisfies*

$$\partial_t p(t,x;s,y) + \frac{\sigma(t,x)^2}{2}\,\partial_{xx}p(t,x;s,y) + b(t,x)\,\partial_x p(t,x;s,y) = 0. \tag{12.9}$$

(ii) Forward equation. *Let p be a transition density, and fix (t,x). Suppose that b has a continuous partial derivative $\partial_y b$ and $a = \sigma^2$ has continuous partial derivatives $\partial_y a$, $\partial_{yy} a$ on $[0,T]\times\mathbb{R}$, and that $(s,y) \mapsto p(t,x;s,y)$ is in $C^{1,2}((t,T]\times\mathbb{R})$. Then, for $(s,y) \in (t,T]\times\mathbb{R}$,*

$$\partial_s p(t,x;s,y) = \frac{1}{2}\,\partial_{yy}\big(a(s,y)\,p(t,x;s,y)\big) - \partial_y\big(b(s,y)\,p(t,x;s,y)\big). \tag{12.10}$$

In (i) the density solves, in the starting point, the backward equation of Section 12.1: formally, it is the Feynman–Kac solution with the payoff $h = \delta_y$. In (ii), it solves, in the arrival point, the equation of the formal adjoint $\mathcal{L}_s^* f = \frac{1}{2}(af)'' - (bf)'$, which moves probability mass forward in time. Equation (12.10) is also called the *Fokker–Planck equation*. The two equations are derived at the level of these notes, assuming the densities smooth, in Shreve (2004b), Exercises 6.8 and 6.9.

Proof. (i) Under these hypotheses the backward operator $\partial_t + \mathcal{L}_t$ has a fundamental solution $\Gamma(t,x;s,y)$: a function which, for fixed (s,y), is in $C^{1,2}([0,s)\times\mathbb{R})$ and solves (12.9), and such that for every bounded continuous h the function $w(t,x) = \int \Gamma(t,x;s,y)h(y)\,dy$ is bounded, is in $C^{1,2}([0,s)\times\mathbb{R})$, solves the backward equation, and tends to $h(x_0)$ as $(t,x)\to(s,x_0)$ (Friedman (1964), Chapter 1, which constructs Γ by the parametrix method). By Theorem 12.1, with s in place of T, $k=0$ and $E=\emptyset$, $w(t,x) = \mathbb{E}[h(X_s^{t,x})]$. So the probability measure $\mathbb{P}(X_s^{t,x}\in\cdot)$ and the measure with density $\Gamma(t,x;s,\cdot)$ integrate every bounded continuous function to the same value. Applying this to continuous functions increasing to the indicator of an open interval, they agree on open intervals, a π-system that generates the Borel sets, hence on all Borel sets by Dynkin's π–λ theorem. So $p = \Gamma$ is a transition density, and it satisfies (12.9).

(ii) Let $f \in C^2(\mathbb{R})$ vanish outside a bounded interval J. Then f' is bounded, and $\mathcal{L}_u f$ is bounded on $[0,T]\times\mathbb{R}$, since it vanishes off J and b, σ are continuous. By Dynkin's formula (Theorem 9.11, after the time shift), for $s \in (t,T]$,

$$\int f(y)\,p(t,x;s,y)\,dy = \mathbb{E}[f(X_s^{t,x})] = f(x) + \int_t^s \psi(u)\,du, \qquad \psi(u) = \int (\mathcal{L}_u f)(y)\,p(t,x;u,y)\,dy.$$

For u in a compact subset of $(t,T]$, the integrand of ψ is continuous and vanishes for $y \notin J$, so ψ is continuous on $(t,T]$, and the right side is differentiable in s with derivative $\psi(s)$. On the left we may differentiate under the integral sign, since $\partial_s p(t,x;\cdot,\cdot)$ is continuous, hence bounded on compact sets. Integrating by parts twice in y, with no boundary terms because f, f' vanish off J,

$$\int f\,\partial_s p\,dy = \int \Big(\frac{a}{2}f'' + bf'\Big)p\,dy = \int f\,\Big(\frac{1}{2}\partial_{yy}(ap) - \partial_y(bp)\Big)\,dy,$$

all functions being evaluated at (s,y). The difference of the two sides of (12.10) is therefore a continuous function of y whose integral against every such f vanishes. If it were positive (or

negative) at some y_0, it would be so on an interval around y_0, and a nonnegative $f \neq 0$ supported in that interval would give a nonzero integral. □

Example 12.9 (Brownian motion and geometric Brownian motion)**.** Write $\dot{G}$ for the derivative of the heat kernel G in its first argument and G', G'' for its derivatives in the second, so that $\dot{G} = \frac{1}{2}G''$ (Section 12.4).

(a) For $X = W$, $p(t,x;s,y) = G(s-t, y-x)$, so $\partial_t p = -\dot{G}$, $\partial_{xx} p = G''$, $\partial_s p = \dot{G}$ and $\partial_{yy} p = G''$: the identity $\dot{G} = \frac{1}{2}G''$ is both (12.9) and (12.10).

(b) For geometric Brownian motion $\mathrm{d}S = \mu S\,\mathrm{d}t + \sigma S\,\mathrm{d}W$, $\log S_s$ given $S_t = x$ is $\mathcal{N}(\log x + m\tau, \sigma^2\tau)$ with $m = \mu - \frac{\sigma^2}{2}$ and $\tau = s - t$ (Theorem 6.1), so

$$p(t,x;s,y) = \frac{1}{y}\,G(\sigma^2\tau,\ \log y - \log x - m\tau), \qquad x, y > 0.$$

The coefficients μx and σx are unbounded, so (i) does not apply, but both equations hold, as a computation in logarithmic variables shows. Let $\xi = \log x$ and $\eta = \log y$, and evaluate G and its derivatives at $(\sigma^2\tau, \eta - \xi - m\tau)$. In ξ the generator is $\mu x\partial_x + \frac{1}{2}\sigma^2 x^2 \partial_{xx} = m\,\partial_\xi + \frac{1}{2}\sigma^2\partial_{\xi\xi}$, and

$$\partial_t p + m\,\partial_\xi p + \frac{\sigma^2}{2}\,\partial_{\xi\xi} p = \frac{1}{y}\Big(-\sigma^2\dot{G} + mG' - mG' + \frac{\sigma^2}{2}G''\Big) = 0.$$

For the forward equation, $yp = G$ and $\partial_y = y^{-1}\partial_\eta$, so $\partial_y(\mu y\, p) = y^{-1}\mu G'$ and $\frac{1}{2}\partial_{yy}(\sigma^2 y^2 p) = \frac{\sigma^2}{2}\partial_y(G + G') = y^{-1}\frac{\sigma^2}{2}(G' + G'')$. Hence the right side of (12.10) is $y^{-1}(\frac{\sigma^2}{2}G'' - mG') = y^{-1}(\sigma^2\dot{G} - mG') = \partial_s p$.

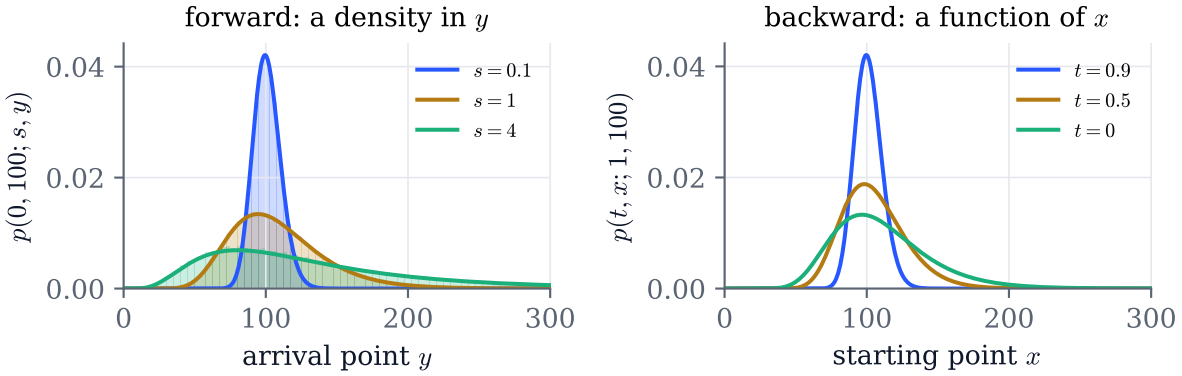

Figure 12.3: The transition density of geometric Brownian motion with $\mu = 8\%$, $\sigma = 30\%$. Left: as a function of the arrival point, $y \mapsto p(0, 100; s, y)$ for three times s, with histograms of 4×10^4 exact samples of S_s: a probability density, which spreads and drifts as s grows, as the forward equation says. Right: as a function of the starting point, $x \mapsto p(t, x; 1, 100)$ for three times t: not a density in x, it solves the backward equation and spreads as t decreases.

Remark 12.10 (Two equations, two uses). One solve of the backward equation prices one payoff from every initial state; one solve of the forward equation gives the law of X_s for all s, hence the prices of all payoffs at all maturities, from one initial state: the situation of a trader who sees calls of many strikes and maturities on one stock (Exercise 12.6). For a local volatility model under $\mathbb{Q}$,

(12.10) reads $\partial_s p = \frac{1}{2}\partial_{yy}(\sigma(s,y)^2 y^2 p) - \partial_y(ryp)$; integrated twice against $(y-K)^+$ it becomes an equation for the call prices as functions of (T,K), which is Dupire's formula (Theorem 15.7).

▶ Interactive demo: transition densities of Brownian motion and of geometric Brownian motion, with the residuals of both Kolmogorov equations computed by finite differences.

12.6 Digital and pay-later options

Digital options. In the Black–Scholes model, the digital call pays $\mathbf{1}_{\{S_T>K\}}$ and its price is $D(t,S_t)$ with $D(t,x) = e^{-r\tau}N(d_-)$, $\tau = T-t$ (Theorem 6.13). By Theorem 12.7(iii), D is smooth on $[0,T)\times(0,\infty)$, solves the Black–Scholes equation, and tends to the payoff at every $x_0 \neq K$; it is the only bounded solution with this behaviour (Theorem 12.1 with $E=\{K\}$). At (T,K) it cannot be continuous. The hedge ratio sees this discontinuity.

Proposition 12.11 (The delta of a digital is a density)**.** *In the Black–Scholes model, for $\tau = T-t>0$ and $x>0$,*

$$\partial_x D(t,x) = e^{-r\tau}\,\frac{K}{x}\,p(t,x;T,K) = \frac{e^{-r\tau}N'(d_-)}{\sigma x\sqrt{\tau}},$$

where p is the transition density of the stock under $\mathbb{Q}$, that of geometric Brownian motion with drift r (Theorem 12.9). Moreover $\partial_x D(t,x) \le 1/(K\sigma\sqrt{2\pi\tau})$, with equality where $d_- = -\sigma\sqrt{\tau}$, and $\partial_x D(t,K) \sim 1/(K\sigma\sqrt{2\pi\tau}) \to \infty$ as $\tau \downarrow 0$.

Proof. Let $R = \exp(\sigma\sqrt{\tau}\,Y + (r-\frac{\sigma^2}{2})\tau)$ with $Y\sim\mathcal{N}(0,1)$, and let f_R be its density. Then $D(t,x) = e^{-r\tau}\mathbb{Q}(xR>K) = e^{-r\tau}\int_{K/x}^{\infty} f_R(z)\,\mathrm{d}z$, so $\partial_x D = e^{-r\tau} f_R(K/x)\,K/x^2$. The law of S_T given $S_t=x$ is that of xR, whose density at y is $f_R(y/x)/x$; at $y=K$ this gives the first formula. The second is the expression of Theorem 12.9(b) with $\mu = r$: $\log K - \log x - (r-\frac{\sigma^2}{2})\tau = -\sigma\sqrt{\tau}\,d_-$, so $p(t,x;T,K) = N'(d_-)/(K\sigma\sqrt{\tau})$. By the identity $Ke^{-r\tau}N'(d_-) = xN'(d_+)$ of Theorem 6.14, $\partial_x D = N'(d_+)/(K\sigma\sqrt{\tau})$, which is largest where $d_+ = 0$, that is $d_- = -\sigma\sqrt{\tau}$. At $x=K$, $d_- = (r-\frac{\sigma^2}{2})\sqrt{\tau}/\sigma \to 0$. □

The delta is how fast the probability of finishing above K grows when the starting point moves up: the density of the terminal price at the strike. (Since D depends on x and K only through x/K, $x\,\partial_x D = -K\,\partial_K D$, and the identity is also Theorem 3.18.) Near maturity the density of S_T concentrates around S_t (Figure 12.3), and so does the delta (Figure 12.4). For the call the same singularity sits one derivative higher, in the gamma. A trader hedging a digital close to the money just before maturity must hold a very large position, which reverses whenever the stock crosses the strike, and a discretely rebalanced hedge converges more slowly than for a call (Chapter 10). In practice the digital is replaced by a call spread, which super-replicates it with a bounded hedge (Exercise 3.5).

Pay-later options. A *pay-later* (or *contingent premium*) call with strike K and maturity T costs nothing at time 0. At T, if $S_T > K$, the holder receives $S_T - K$ and pays a fee P fixed at time 0; otherwise nothing happens. The payoff to the holder is $(S_T - K - P)\mathbf{1}_{\{S_T>K\}}$: one call minus P digital calls, a static replication. The fee must make the contract worth zero.

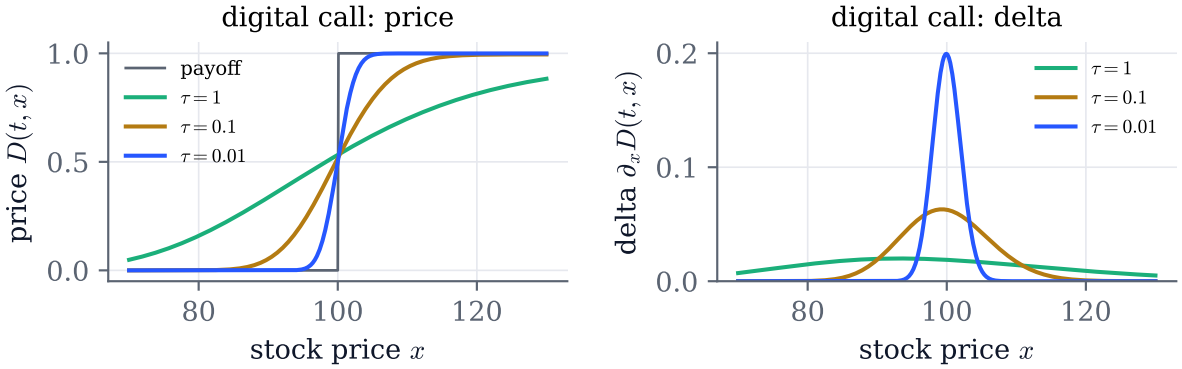

Figure 12.4: The digital call with $K = 100$, $r = 5\%$, $\sigma = 20\%$, for three times to maturity τ. Left: its price, which smooths the jump of the payoff over a range of width about $K\sigma\sqrt{\tau}$. Right: its delta, the discounted transition density of the stock at the strike (Theorem 12.11); at $\tau = 0.01$ its maximum is $1/(K\sigma\sqrt{2\pi\tau}) \approx 0.20$ shares per unit of payoff.

Proposition 12.12 (The pay-later fee)**.** *Let $\mathbb{Q}$ be a pricing measure (Theorem 3.5) for a market in which the bond, the stock, the call and the digital call with strike K and maturity T are traded, with discount factor $B = e^{-rT}$, and suppose $\mathbb{Q}(S_T > K) > 0$. Let C_0 and D_0 be the prices of the call and the digital. The fee that makes the pay-later call worth zero is*

$$P = \frac{C_0}{D_0} = \mathbb{E}^{\mathbb{Q}}\big[S_T - K \,\big|\, S_T > K\big] \;\geq\; e^{rT} C_0,$$

with equality if and only if $\mathbb{Q}(S_T > K) = 1$. In the Black–Scholes model, $P = S_0 e^{rT} N(d_+)/N(d_-) - K$, with $d_\pm = d_\pm(T, S_0)$.

Proof. The contract is worth $C_0 - PD_0$. Since $C_0 = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S_T - K)\mathbf{1}_{\{S_T > K\}}] > 0$ and $D_0 = e^{-rT}\mathbb{Q}(S_T > K) \leq e^{-rT}$, with equality if and only if $\mathbb{Q}(S_T > K) = 1$, the ratio is as stated. The Black–Scholes formula follows from Theorems 6.11 and 6.13. □

The fee is the average payoff of the call given that it pays. It exceeds the premium accrued at the riskless rate: the buyer is insured against paying for an option that expires worthless. It exceeds C_0 itself if $r \geq 0$, but not necessarily if $r < 0$. With $S_0 = K = 100$, $r = 5\%$, $\sigma = 20\%$ and $T = 1$: $C_0 = 10.45$, $D_0 = 0.532$, and $P = 19.63$, about 1.8 times $e^{rT}C_0 = 10.99$. The hedge of the holder's position $c - P\,D$, with delta $N(d_+) - P\,\partial_x D$, inherits the explosion of the digital's delta (Exercise 12.4).

Further reading

Shreve (2004b), Chapter 6, is at the level of these notes, with multidimensional versions. Karatzas and Shreve (1991), Sections 4.3–4.4 and 5.7, Øksendal (2003), Chapter 8, and Steele (2001), Chapters 11 and 15, give more; Friedman (1964) is the classical source on parabolic equations and their fundamental solutions.

Chapter 13

Exotic options

A call or a put pays a function of the price of the stock at maturity. Many contracts pay a function of the whole path instead: a *barrier option* is cancelled, or activated, when the price crosses a level; a *lookback option* pays the best price reached; an *Asian option* pays an average. Such contracts are called *exotic*, by contrast with the *vanilla* calls and puts. Risk-neutral pricing covers them without change (Theorem 11.8); what changes is the computation. Barriers and lookbacks need the joint law of a Brownian motion with drift and its running maximum, which the *reflection principle* gives in closed form. Asian options have no closed form, but their prices, if smooth, solve a partial differential equation with one more state variable, which a change of numeraire reduces to a single space dimension. The chapter ends with an American option whose price is explicit, the perpetual put.

13.1 Path-dependent claims

We work in the Black–Scholes setting of Section 6.3: under $\mathbb{Q}$, $\widetilde{W}$ is a Brownian motion with filtration $(\mathcal{F}_t)_{t\geq 0}$, $r \in \mathbb{R}$, $\sigma > 0$, and

$$S_t = S_0 \exp\Big(\sigma\widetilde{W}_t + \Big(r - \frac{\sigma^2}{2}\Big)t\Big) = S_0 e^{\sigma X_t}, \qquad X_t = \widetilde{W}_t + \nu t, \quad \nu = \frac{r}{\sigma} - \frac{\sigma}{2}. \tag{13.1}$$

So $\frac{1}{\sigma}\log(S_t/S_0)$ is a Brownian motion with drift ν. The running maximum and minimum are written $\overline{S}_t = \max_{u\leq t} S_u = S_0 e^{\sigma \overline{X}_t}$ and $\underline{S}_t = \min_{u\leq t} S_u = S_0 e^{\sigma \underline{X}_t}$, with $\overline{X}_t$, $\underline{X}_t$ the maximum and minimum of X on $[0,t]$. A *claim* with maturity T is an $\mathcal{F}_T$-measurable random variable V_T, which may depend on the whole path $(S_u)_{u\leq T}$. As in Chapter 6, its price at time $t \leq T$ is defined as

$$V_t = e^{-r(T-t)}\,\mathbb{E}^{\mathbb{Q}}\big[V_T \,\big|\, \mathcal{F}_t\big], \tag{13.2}$$

for $V_T \geq 0$ or $\mathbb{E}^{\mathbb{Q}}|V_T| < \infty$. This is the replication cost of the claim when it can be replicated (Theorem 11.8), and if $(\mathcal{F}_t)$ is the augmented filtration of the Brownian motion, every claim with $\mathbb{E}^{\mathbb{Q}}|V_T| < \infty$ can be (Theorem 11.11).

Conditioning on $\mathcal{F}_t$ uses one fact again and again: the Brownian motion starts afresh after t, independently of $\mathcal{F}_t$. Theorem 5.9(c) says so for the filtration of W only.

Lemma 13.1 (Restarting at a fixed time)**.** *Let $T \in (0,\infty]$, let W be a Brownian motion on $[0,T]$ (on $[0,\infty)$ if $T=\infty$) with a filtration $(\mathcal{F}_t)_{t\le T}$, and let $0 \le s < T$. Then $W'_u = W_{s+u} - W_s$, $0 \le u \le T-s$, is a Brownian motion independent of $\mathcal{F}_s$, and $(\mathcal{F}_{s+u})_u$ is a filtration for it.*

Proof. W' is a Brownian motion (Theorem 5.9(c)), and $(\mathcal{F}_{s+u})$ is a filtration for it by Theorem 5.11. Let $0 = u_0 < u_1 < \cdots < u_k \le T-s$, $D_j = W'_{u_j} - W'_{u_{j-1}}$, $A \in \mathcal{F}_s$ and $E_1, \dots, E_k$ Borel sets. The event $A \cap \bigcap_{j<k}\{D_j \in E_j\}$ is in $\mathcal{F}_{s+u_{k-1}}$, of which D_k is independent, so $\mathbb{P}(A \cap \bigcap_{j\le k}\{D_j \in E_j\}) = \mathbb{P}(A \cap \bigcap_{j<k}\{D_j \in E_j\})\mathbb{P}(D_k \in E_k)$, and by induction this is $\mathbb{P}(A)\prod_j \mathbb{P}(D_j \in E_j) = \mathbb{P}(A)\,\mathbb{P}(D \in E_1 \times \cdots \times E_k)$ for $D = (D_1, \dots, D_k)$ (take $A = \Omega$). Rectangles form a π-system that generates the Borel sets of $\mathbb{R}^k$, so by Dynkin's π–λ theorem $\mathbb{P}(A \cap \{D \in E\}) = \mathbb{P}(A)\mathbb{P}(D \in E)$ for every Borel E: D is independent of $\mathcal{F}_s$. Every vector $(W'_{t_1}, \dots, W'_{t_m})$ is a linear function of such a D, which proves the independence in the sense of Theorem 5.8. □

Remark 13.2 (Claims on finitely many dates). Payoffs $g(S_{t_1}, \dots, S_{t_n})$ are priced date by date with the independence lemma (Theorem 1.19), as in Theorem 4.7: power claims, forward-start calls and chooser options reduce in this way to the Black–Scholes formula (Exercise 6.2, Exercise 6.3, Exercise 6.4).

13.2 The reflection principle

Let W be a Brownian motion with filtration $(\mathcal{F}_t)$ under a probability $\mathbb{P}$, let $M_t = \max_{u\le t} W_u$, and for $m > 0$ let $\tau_m = \inf\{t \ge 0 : W_t = m\}$, with $\inf\emptyset = \infty$. A *stopping time* is a random time τ with values in $[0,\infty]$ such that $\{\tau \le t\} \in \mathcal{F}_t$ for every $t \ge 0$. By continuity and the intermediate value theorem, $\{\tau_m \le t\} = \{M_t \ge m\}$, and M_t is $\mathcal{F}_t$-measurable (it is a supremum over the rational times of $[0,t]$ and t), so τ_m is a stopping time.

After τ_m the path is as likely to go down as up, so flipping it after τ_m gives another Brownian path, and a path that ends at $w < m$ after touching m is matched with one that ends at $2m - w$ (Figure 13.1). Since τ_m is random, we approximate it by stopping times with finitely many values, at each of which the restart lemma applies.

Theorem 13.3 (Reflection principle)**.** *Let $T > 0$ and $m > 0$.*

(i) For every stopping time τ, the process $W^*_t = 2W_{t\wedge\tau} - W_t$, which equals W up to τ and is reflected about W_τ afterwards, is a Brownian motion.

(ii) For every Borel $h : \mathbb{R} \to [0,\infty)$ with $h = 0$ on (m,∞),

$$\mathbb{E}\big[h(W_T)\mathbf{1}_{\{M_T \ge m\}}\big] = \mathbb{E}\big[h(2m - W_T)\big]. \tag{13.3}$$

In particular $\mathbb{P}(M_T \ge m,\, W_T \le w) = \mathbb{P}(W_T \ge 2m - w) = N\big((w-2m)/\sqrt{T}\big)$ for $w \le m$; $\mathbb{P}(M_T \ge m) = 2\mathbb{P}(W_T \ge m)$, so that M_T has the law of $|W_T|$; and (M_T, W_T) has the density

$$f_T(a,w) = \frac{2(2a-w)}{T\sqrt{2\pi T}} \exp\Big(-\frac{(2a-w)^2}{2T}\Big), \qquad a > 0,\ w < a.$$

(iii) Let $\nu \in \mathbb{R}$, $X_t = W_t + \nu t$, and let $\overline{X}_T$, $\underline{X}_T$ be the maximum and minimum of X on $[0,T]$. For h as in (ii),

$$\mathbb{E}\big[h(X_T)\mathbf{1}_{\{\overline{X}_T \ge m\}}\big] = e^{2\nu m}\,\mathbb{E}\big[h(2m - W_T + \nu T)\big], \tag{13.4}$$

and for $a < 0$ and Borel $h \geq 0$ with $h = 0$ on $(-\infty, a)$, $\mathbb{E}[h(X_T)\mathbf{1}_{\{\overline{X}_T \leq a\}}] = e^{2\nu a}\,\mathbb{E}[h(2a - W_T + \nu T)]$. In particular, for $w \leq m$,

$$\mathbb{P}(\overline{X}_T \geq m,\, X_T \leq w) = e^{2\nu m} N\Big(\frac{w - 2m - \nu T}{\sqrt{T}}\Big),$$
$$\mathbb{P}(\overline{X}_T \geq m) = N\Big(\frac{\nu T - m}{\sqrt{T}}\Big) + e^{2\nu m} N\Big(\frac{-m - \nu T}{\sqrt{T}}\Big),$$

and $(\overline{X}_T, X_T)$ has the density $e^{\nu w - \nu^2 T/2} f_T(a, w)$.

Proof. (i) *Finitely many values.* Suppose first that τ takes values in a finite set $\{s_1, \dots, s_K\} \subset [0, \infty)$ and possibly ∞. On $\{\tau = \infty\}$, $W^* = W$. Fix $0 \leq u_1 < \cdots < u_p$ and a bounded Borel $f : \mathbb{R}^p \to \mathbb{R}$. On $\{\tau = s_k\}$ we have $W^*_{u_i} = \xi_i - \eta_i$ and $W_{u_i} = \xi_i + \eta_i$, with $\xi_i = W_{u_i \wedge s_k}$ and $\eta_i = W_{u_i \vee s_k} - W_{s_k}$. The pair $(\mathbf{1}_{\{\tau = s_k\}}, \xi)$ is $\mathcal{F}_{s_k}$-measurable, since $\{\tau = s_k\} = \{\tau \leq s_k\} \setminus \{\tau \leq s_{k-1}\}$; the vector η is a function of the restarted motion $W_{s_k + \cdot} - W_{s_k}$, hence independent of $\mathcal{F}_{s_k}$ (Theorem 13.1), and it is centred Gaussian, so $-\eta$ has the same law. The law of $(\mathbf{1}_{\{\tau = s_k\}}, \xi, \eta)$ is the product of the laws of its two independent parts, so it equals that of $(\mathbf{1}_{\{\tau = s_k\}}, \xi, -\eta)$, and $\mathbb{E}[\mathbf{1}_{\{\tau = s_k\}} f(\xi - \eta)] = \mathbb{E}[\mathbf{1}_{\{\tau = s_k\}} f(\xi + \eta)]$. Summing over k and adding $\{\tau = \infty\}$ gives $\mathbb{E}[f(W^*_{u_1}, \dots, W^*_{u_p})] = \mathbb{E}[f(W_{u_1}, \dots, W_{u_p})]$. So W^* has the finite-dimensional laws of a Brownian motion; it has continuous paths and starts at 0, so it is one (Theorem 5.6).

General τ. Let τ^n be the smallest point of $2^{-n}\mathbb{N}$ that is at least τ if $\tau \leq n$, and $\tau^n = \infty$ if $\tau > n$. For a grid point $s = k2^{-n} \leq n$, $\{\tau^n \leq s\} = \{\tau \leq s\} \in \mathcal{F}_s$, so the τ^n are stopping times with finitely many values, and $\tau^n \to \tau$ for every ω. By the first step, $W^{*,n}_t = 2W_{t \wedge \tau^n} - W_t$ is a Brownian motion, and $W^{*,n}_t \to W^*_t$ for every t and ω by continuity of the paths. For bounded continuous f, dominated convergence gives $\mathbb{E}[f(W^*_{u_1}, \dots, W^*_{u_p})] = \lim_n \mathbb{E}[f(W^{*,n}_{u_1}, \dots, W^{*,n}_{u_p})] = \mathbb{E}[f(W_{u_1}, \dots, W_{u_p})]$, and bounded continuous functions determine laws.

(ii) Take $\tau = \tau_m$ in (i). Before τ_m, $W^* = W < m$, and $W^*_{\tau_m} = m$; so τ_m is also the first passage of W^* at m, and $\{M^*_T \geq m\} = \{\tau_m \leq T\} = \{M_T \geq m\}$, where M^* is the running maximum of W^*. On this event $W^*_T = 2m - W_T$. Since W^* is a Brownian motion, (M^*_T, W^*_T) has the law of (M_T, W_T) (Theorem 5.10), and

$$\mathbb{E}[h(W_T)\mathbf{1}_{\{M_T \geq m\}}] = \mathbb{E}[h(W^*_T)\mathbf{1}_{\{M^*_T \geq m\}}] = \mathbb{E}[h(2m - W_T)\mathbf{1}_{\{M_T \geq m\}}] = \mathbb{E}[h(2m - W_T)];$$

in the last step, $h(2m - W_T) \neq 0$ forces $W_T \geq m$, hence $M_T \geq m$. With $h = \mathbf{1}_{(-\infty, w]}$ we get the first formula. With $w = m$ it gives $\mathbb{P}(M_T \geq m, W_T \leq m) = \mathbb{P}(W_T \geq m)$, and $\{W_T > m\} \subseteq \{M_T \geq m\}$, so $\mathbb{P}(M_T \geq m) = 2\mathbb{P}(W_T \geq m) = \mathbb{P}(|W_T| \geq m)$. Finally, $f_T(a, w) = -\partial_a \partial_w N\big((w - 2a)/\sqrt{T}\big)$, so the law with density f_T gives the same probability to every quadrant $\{a' \geq a,\, w' \leq w\}$ (for $w > a > 0$ add $\mathbb{P}(a < W_T \leq w)$; for $a \leq 0$ both are $\mathbb{P}(W_T \leq w)$), and quadrants determine laws (Dynkin's π–λ theorem).

(iii) The statements involve the paths on $[0, T]$ only, and the proofs of (i)–(ii) go through for a Brownian motion on $[0, T]$, with the grid $kT2^{-n}$ and $\tau^n = \infty$ if $\tau > T$ (at the grid point $s_k = T$ the vector η vanishes and the first step is trivial). By Girsanov's theorem (Theorem 11.2 with $\Theta \equiv \nu$), X is a Brownian motion on $[0, T]$ under $\widehat{\mathbb{P}}$, $\mathrm{d}\widehat{\mathbb{P}}/\mathrm{d}\mathbb{P} = Z_T = \exp(-\nu W_T - \frac{\nu^2}{2}T)$, with filtration $(\mathcal{F}_t)_{t \leq T}$. Since $1/Z_T = \exp(\nu X_T - \frac{\nu^2}{2}T)$, Theorem 1.23(b) and (ii) under $\widehat{\mathbb{P}}$, applied to $\varphi(x) = e^{\nu x - \nu^2 T/2} h(x)$, which vanishes on (m, ∞), give

$$\mathbb{E}[h(X_T)\mathbf{1}_{\{\overline{X}_T \geq m\}}] = \widehat{\mathbb{E}}[\varphi(X_T)\mathbf{1}_{\{\overline{X}_T \geq m\}}] = \widehat{\mathbb{E}}[\varphi(2m - X_T)] = e^{2\nu m}\,\widehat{\mathbb{E}}[e^{-\nu X_T - \nu^2 T/2} h(2m - X_T)].$$

Under $\widehat{\mathbb{P}}$, $X_T \sim \mathcal{N}(0,T)$, and by the Gaussian shift (Theorem 1.26 with $\theta = -\nu\sqrt{T}$, applied to $X_T/\sqrt{T}$) the last expectation is $\mathbb{E}[h(2m - G + \nu T)]$ for $G \sim \mathcal{N}(0,T)$, which is (13.4). The statement for the minimum is (13.4) for the Brownian motion $-W$, the drift $-\nu$, the level $-a$ and the function $x \mapsto h(-x)$. With $h = \mathbf{1}_{(-\infty,w]}$, $\mathbb{P}(2m - W_T + \nu T \le w) = N\big((w - 2m - \nu T)/\sqrt{T}\big)$; adding $\mathbb{P}(X_T > m)$ for $w = m$ gives the law of $\overline{X}_T$. For the density, $\mathbb{E}[g(\overline{X}_T, X_T)] = \widehat{\mathbb{E}}[e^{\nu X_T - \nu^2 T/2} g(\overline{X}_T, X_T)]$ for Borel $g \ge 0$, and under $\widehat{\mathbb{P}}$ the pair has density f_T by (ii). □

Two consequences are used below. First, $\overline{X}_T \le M_T + |\nu|T$ and M_T has the law of $|W_T|$, so $\mathbb{E}[e^{p\overline{X}_T}] < \infty$ for every p: in the model (13.1),

$$\mathbb{E}^{\mathbb{Q}}[\overline{S}_T^{\,p}] < \infty \qquad \text{for every } p \ge 0. \tag{13.5}$$

Second, letting $T \to \infty$ in the law of $\overline{X}_T$: if $\nu < 0$, the probability that X ever reaches m is $e^{2\nu m} < 1$ (Exercise 13.2). Part (i) is a consequence of the *strong Markov property*, of which the first step of its proof is a special case: after an almost surely finite stopping time τ, $(W_{\tau+t} - W_\tau)_{t \ge 0}$ is a Brownian motion independent of the information at τ (Mörters and Peres (2010), Theorem 2.16 and Section 2.2; Karatzas and Shreve (1991), Section 2.6); Shreve (2004b), Sections 3.7 and 7.2, derives (ii) and (iii) as here.

Remark 13.4 (Crossing between two dates). Given the values $a = X_s$ and $b = X_t$ at two dates $s < t$, both below m, the probability that X reached m in between is

$$p(a,b) = \exp\Big(-\frac{2(m-a)(m-b)}{t-s}\Big), \tag{13.6}$$

whatever the drift ν. Precisely,

$$\mathbb{E}\big[\mathbf{1}_{\{\max_{[s,t]} X \ge m\}} \varphi(X_s, X_t)\big] = \mathbb{E}\big[p(X_s, X_t)\, \varphi(X_s, X_t)\big]$$

for bounded Borel φ vanishing unless both arguments are below m. By Theorem 13.1 and the independence lemma it suffices to take $s = 0$, $a = 0$, and $h = t$; then $\int_m^\infty e^{\nu w - \nu^2 h/2} f_h(a', w)\, \mathrm{d}a'$, divided by the density $e^{\nu w - \nu^2 h/2} N'(w/\sqrt{h})/\sqrt{h}$ of X_h at w, is $e^{-2m(m-w)/h}$: the factor carrying ν cancels. Weighting each simulated path by its probability of not crossing between grid dates gives an unbiased estimator of a continuously monitored price, used in the demo (the Brownian-bridge technique of P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, 2004, Chapter 6).

▶ Interactive demo: reflect a path at any level, and compare the law of the maximum with drift with simulated maxima on coarse and fine grids.

13.3 Barrier options

Let $K > 0$ be a strike and $B > 0$ a barrier. The four barrier calls pay at T

$$\underbrace{(S_T - K)^+ \mathbf{1}_{\{\overline{S}_T < B\}}}_{\text{up-and-out}}, \quad \underbrace{(S_T - K)^+ \mathbf{1}_{\{\overline{S}_T \ge B\}}}_{\text{up-and-in}}, \quad \underbrace{(S_T - K)^+ \mathbf{1}_{\{\underline{S}_T > B\}}}_{\text{down-and-out}}, \quad \underbrace{(S_T - K)^+ \mathbf{1}_{\{\underline{S}_T \le B\}}}_{\text{down-and-in}}.$$

An *out* option is cancelled (knocked out) if the price reaches the barrier; an *in* option is worthless unless it does. Since $\overline{X}_T$ and $\underline{X}_T$ have densities, strict and weak inequalities give the same

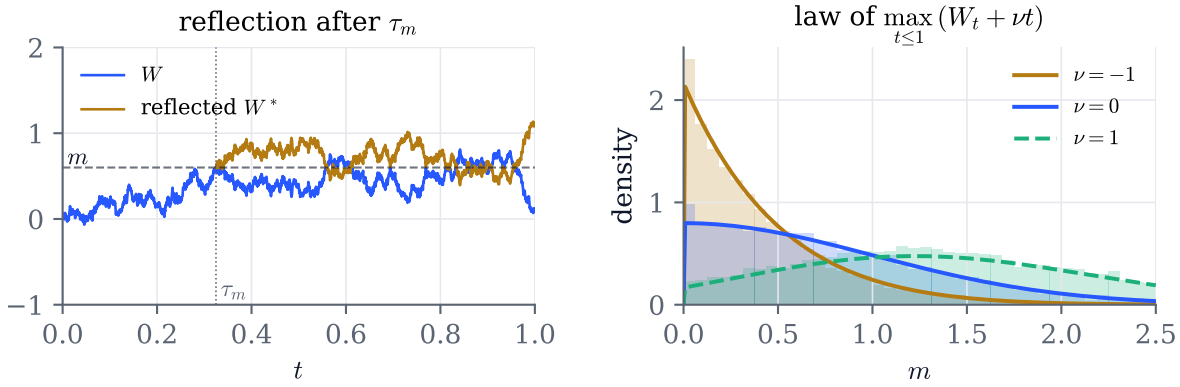

Figure 13.1: Left: a Brownian path and its reflection after the first passage at $m = 0.6$ (Theorem 13.3(i)). Right: the density of $\overline{X}_1$ for three drifts, from Theorem 13.3(iii), against histograms of 10^4 simulated maxima on a grid of 2000 steps; the grid misses the top of some excursions, which shifts the histograms slightly to the left.

prices. Some cases are trivial: if $S_0 \ge B$ the up-and-out call is knocked out at once, and so is the down-and-out call if $S_0 \le B$; if $K \ge B$ the up-and-out call pays nothing, because $S_T \le \overline{S}_T < B \le K$ on the event where it is alive.

Write $c(x;k)$ for the Black–Scholes price at time 0 of a call with strike k and maturity T when the stock is at x, and $\delta(x;k) = e^{-rT}N(d_-(T,x))$, with d_- computed with the strike k, for the digital call of Theorem 6.13.

Proposition 13.5 (Barrier options)**.** *Let $\theta = 2r/\sigma^2 - 1$.*

(i) (Up-and-out.) *If $S_0 < B$ and $K < B$, the up-and-out call costs*

$$UO_0 = u(S_0) - \Big(\frac{B}{S_0}\Big)^{\theta} u\Big(\frac{B^2}{S_0}\Big), \qquad u(x) = c(x;K) - c(x;B) - (B-K)\,\delta(x;B).$$

(ii) (Down-and-out.) *If $S_0 > B$, the down-and-out call costs*

$$DO_0 = \ell(S_0) - \Big(\frac{B}{S_0}\Big)^{\theta} \ell\Big(\frac{B^2}{S_0}\Big), \qquad \ell(x) = \begin{cases} c(x;K) & \text{if } B \le K, \\ c(x;B) + (B-K)\,\delta(x;B) & \text{if } B > K. \end{cases}$$

(iii) (In–out parity.) $UI_0 + UO_0 = DI_0 + DO_0 = c(S_0;K)$.

(iv) (Later dates.) *For $t < T$, $UO_t = \mathbf{1}_{\{\overline{S}_t < B\}}\, UO(T-t, S_t)$, where $UO(T',x)$ is the price in (i) with maturity T' and initial price x; similarly for the down-and-out call, with $\mathbf{1}_{\{\underline{S}_t > B\}}$.*

The function $u(x)$ is the price of the European claim $(S_T - K)^+\mathbf{1}_{\{S_T < B\}}$, a call whose payoff is cut off at B, and $\ell(x)$ that of $(S_T - K)^+\mathbf{1}_{\{S_T > B\}}$. Each knock-out price is the European price minus an *image* term: the same European claim on a stock started from the mirror point B^2/S_0, weighted by $(B/S_0)^{\theta}$. For $B \le K$, (ii) is the formula of Merton (1973) for the down-and-out call.

Proof. (i) Let $b = \frac{1}{\sigma}\log(B/S_0) > 0$, so that $\overline{S}_T < B$ if and only if $\overline{X}_T < b$. Let $h(x) =$

$(S_0e^{\sigma x}-K)^+\mathbf{1}_{\{x<b\}}$, which vanishes on $[b,\infty)$. As $X_T\le \overline{X}_T$, the payoff is $h(X_T)\mathbf{1}_{\{\overline{X}_T<b\}} = h(X_T)-h(X_T)\mathbf{1}_{\{\overline{X}_T\ge b\}}$, and Theorem 13.3(iii), under $\mathbb{Q}$ with $W=\widetilde{W}$, gives

$$\mathbb{E}^{\mathbb{Q}}\big[h(X_T)\mathbf{1}_{\{\overline{X}_T\ge b\}}\big] = e^{2\nu b}\,\mathbb{E}^{\mathbb{Q}}\big[h(2b-\widetilde{W}_T+\nu T)\big] = e^{2\nu b}\,\mathbb{E}^{\mathbb{Q}}\big[h(2b+X_T)\big],$$

since $-\widetilde{W}_T$ and $\widetilde{W}_T$ have the same law. Now $S_0e^{\sigma(2b+X_T)} = (B^2/S_0)\,e^{\sigma X_T}$ is the price at T of the stock started from B^2/S_0, and $2b+X_T<b$ means that this price is below B. So $e^{-rT}\mathbb{E}^{\mathbb{Q}}[h(2b+X_T)] = u(B^2/S_0)$ and $e^{-rT}\mathbb{E}^{\mathbb{Q}}[h(X_T)] = u(S_0)$, where $u(x)$ is the price of $(S_T-K)^+\mathbf{1}_{\{S_T<B\}}$ for a stock started from x. For $K<B$, $(y-K)^+\mathbf{1}_{\{y<B\}} = (y-K)^+-(y-B)^+-(B-K)\mathbf{1}_{\{y\ge B\}}$, whose price is the formula for u (Theorems 6.11 and 6.13). Finally $e^{2\nu b} = (B/S_0)^{2\nu/\sigma}$ and $2\nu/\sigma=\theta$.

(ii) The same computation with $a=\frac{1}{\sigma}\log(B/S_0)<0$, the statement for the minimum in Theorem 13.3(iii), and $h(x)=(S_0e^{\sigma x}-K)^+\mathbf{1}_{\{x>a\}}$, which vanishes on $(-\infty,a]$: the price is $\ell(S_0)-e^{2\nu a}\ell(B^2/S_0)$, where $\ell(x)$ prices $(S_T-K)^+\mathbf{1}_{\{S_T>B\}}$. If $B\le K$ this is the call payoff; if $B>K$ it is $(y-B)^+ + (B-K)\mathbf{1}_{\{y>B\}}$.

(iii) The in and out payoffs add up to $(S_T-K)^+$.

(iv) For $u\ge t$, $S_u = S_tR_{u-t}$ with $R_v=\exp\big(\sigma W'_v+(r-\frac{\sigma^2}{2})v\big)$, where $W'_v=\widetilde{W}_{t+v}-\widetilde{W}_t$ is a Brownian motion independent of $\mathcal{F}_t$ (Theorem 13.1). The payoff is $\mathbf{1}_{\{\overline{S}_t<B\}}(S_tR_{T-t}-K)^+\mathbf{1}_{\{S_t\max_{v\le T-t}R_v<B\}}$; apply the independence lemma (Theorem 1.19) with $X=S_t$ and $Y=(R_{T-t},\max_{v\le T-t}R_v)$. □

Remark 13.6 (Reading the formulas). (a) As $S_0\uparrow B$ in (i), $UO_0\to u(B)-u(B)=0$: a call about to be knocked out is worth nothing, although it is in the money. With $S_0=K=100$, $B=120$, $r=5\%$, $\sigma=20\%$ and $T=1$, the up-and-out call costs 1.18 against 10.45 for the call (Figure 13.2).

(b) The up-and-out price is not monotone in S_0: its delta is negative near the barrier, and at the barrier it is of order $-(B-K)/(\sigma B\sqrt{T-t})$ close to maturity, so it blows up (Exercise 13.3). The hedge is large and changes sign near the barrier, the same problem as for the digital option of Exercise 6.5, and for the same reason: the payoff jumps from $B-K$ to 0 at the barrier. By (iv), the price is $v(t,S_t)$ while the option is alive, and v solves the Black–Scholes equation on $\{x<B\}$ with $v(t,B)=0$; Shreve (2004b), Section 7.3, derives the hedge this way.

Remark 13.7 (Discrete monitoring). In practice the barrier is checked at finitely many dates, say daily. A path can cross the barrier and come back between two checks unnoticed, so a discretely monitored up-and-out call is worth more than the continuous one, by an amount that shrinks only like $1/\sqrt{m}$ for m equally spaced dates (Figure 13.2). Broadie et al. (1997) showed that the discretely monitored price equals the continuous one with the barrier moved away from S_0 by the factor $e^{\beta\sigma\sqrt{T/m}}$, up to an error $o(1/\sqrt{m})$, where $\beta=-\zeta(\frac{1}{2})/\sqrt{2\pi}\approx 0.5826$ and ζ is the Riemann zeta function. The constant is the mean overshoot of a Gaussian random walk over a distant level, met in Exercise 5.8. In Figure 13.2 the shift removes most of the error from about $m=26$ dates on, and is poor for $m=4$.

▶ Interactive demo: barrier prices against the spot for any barrier, strike and volatility; Monte Carlo with discrete monitoring, with the Brownian-bridge weights of Theorem 13.4, and the shifted barrier.

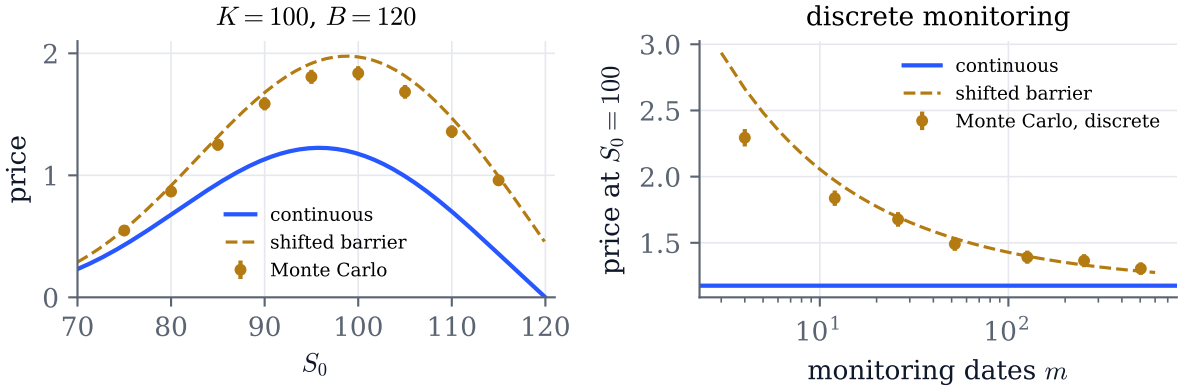

Figure 13.2: Up-and-out call with $K = 100$, $B = 120$, $r = 5\%$, $\sigma = 20\%$, $T = 1$. Left: price against S_0, from Theorem 13.5(i) (continuous monitoring), and with $m = 12$ monitoring dates: Monte Carlo with 2×10^4 paths (bars: two standard errors) and the closed form with the shifted barrier of Theorem 13.7. Right: price at $S_0 = 100$ against the number m of monitoring dates.

13.4 Lookback options

A *floating-strike lookback put* pays $\overline{S}_T - S_T$ at T: it is the put struck, with hindsight, at the highest price of the period, and it is never out of the money.

Proposition 13.8 (Lookback put)**.** *Let $r \neq 0$ and $d_\pm = (r \pm \frac{\sigma^2}{2})\sqrt{T}/\sigma$. The floating-strike lookback put costs*

$$LP_0 = S_0\big[e^{-rT}N(-d_-) - N(-d_+)\big] + S_0\,\frac{\sigma^2}{2r}\big[N(d_+) - e^{-rT}N(-d_-)\big].$$

For $r = 0$ the price is the limit as $r \to 0$: $LP_0 = S_0\big[N(\eta) - N(-\eta) + 2\eta N'(\eta) + 2\eta^2 N(\eta)\big]$ with $\eta = \sigma\sqrt{T}/2$. At a later date $t < T$, $LP_t = S_t\,\lambda(T-t, \overline{S}_t/S_t)$, where $\lambda(u,y) = e^{-ru}\mathbb{E}^{\mathbb{Q}}\big[\max\big(y, e^{\sigma \overline{X}_u}\big)\big] - 1$ for $y \geq 1$.

The first bracket is the price of the at-the-money European put divided by S_0; the second term, nonnegative because $d_+ + d_- = 2r\sqrt{T}/\sigma$ has the sign of r, is what looking back adds. With $S_0 = 100$, $r = 5\%$, $\sigma = 20\%$, $T = 1$, the lookback put costs 14.29 against 5.57 for the at-the-money put.

Proof. By the martingale property, $LP_0 = e^{-rT}\mathbb{E}^{\mathbb{Q}}[\overline{S}_T] - S_0$, and $\overline{S}_T = S_0 e^{\sigma \overline{X}_T}$. Writing $e^{\sigma \overline{X}_T} - 1 = \int_0^\infty \sigma e^{\sigma m}\mathbf{1}_{\{\overline{X}_T \geq m\}}\,\mathrm{d}m$ and using Tonelli's theorem and Theorem 13.3(iii),

$$\mathbb{E}^{\mathbb{Q}}\big[e^{\sigma \overline{X}_T}\big] = 1 + \underbrace{\int_0^\infty \sigma e^{\sigma m} N\Big(\frac{\nu T - m}{\sqrt{T}}\Big)\mathrm{d}m}_{I_1} + \underbrace{\int_0^\infty \sigma e^{(\sigma+2\nu)m} N\Big(\frac{-m - \nu T}{\sqrt{T}}\Big)\mathrm{d}m}_{I_2}.$$

Note that $\nu\sqrt{T} = d_-$, $(\nu+\sigma)\sqrt{T} = d_+$, $\sigma\nu + \frac{\sigma^2}{2} = r$ and $\sigma + 2\nu = 2r/\sigma =: \kappa$. Integrate by parts (the boundary terms at infinity vanish, as the normal tail beats the exponential), substitute

$m = \nu T + \sqrt{T}z$ in I_1 and $m = -\nu T + \sqrt{T}z$ in I_2, and use $\int_{-c}^{\infty} e^{\alpha z} N'(z)\,\mathrm{d}z = e^{\alpha^2/2} N(c+\alpha)$:

$$I_1 = -N(d_-) + e^{\sigma\nu T}\int_{-d_-}^{\infty} e^{\sigma\sqrt{T}z} N'(z)\,\mathrm{d}z = e^{rT}N(d_+) - N(d_-),$$
$$I_2 = \frac{\sigma}{\kappa}\Big(- N(-d_-) + e^{-\kappa\nu T}\int_{d_-}^{\infty} e^{\kappa\sqrt{T}z} N'(z)\,\mathrm{d}z\Big) = \frac{\sigma^2}{2r}\Big(e^{rT}N(d_+) - N(-d_-)\Big),$$

because $-\kappa\nu + \frac{\kappa^2}{2} = r$ and $\kappa\sqrt{T} - d_- = d_+$. Collecting terms, $e^{-rT}(1 + I_1 + I_2) - 1$ is the bracketed formula divided by S_0. The price is continuous in r by dominated convergence, since $\overline{X}_T$ depends continuously on ν and $e^{\sigma\overline{X}_T} \le e^{\sigma(M_T + |\nu|T)}$; the case $r = 0$ follows by letting $r \to 0$ (l'Hôpital in the second term). For $t < T$, write $\overline{S}_T = \max\big(\overline{S}_t,\, S_t \max_{v \le T-t} R_v\big)$ with R as in the proof of Theorem 13.5(iv), and apply Theorem 13.1 and the independence lemma with $X = (S_t, \overline{S}_t)$; the result is homogeneous of degree one in $(S_t, \overline{S}_t)$. □

The price depends on the path through $(S_t, \overline{S}_t)$, a Markov pair; Shreve (2004b), Section 7.4, computes λ and the hedge. Lookbacks were first priced by Goldman et al. (1979); the fixed-strike lookback call is Exercise 13.4.

13.5 Asian options

Let $Y_t = \int_0^t S_u\,\mathrm{d}u$, so that Y_T/T is the average price over $[0,T]$. The *(arithmetic, fixed-strike) Asian call* pays $(Y_T/T - K)^+$. Asian options are common on commodities and currencies: an average is harder to manipulate than a closing price. An average of lognormal variables is not lognormal, and there is no closed formula; but the *geometric* average is lognormal.

Proposition 13.9 (Geometric Asian options)**.** *Let $K > 0$, $G_T = \exp\big(\frac{1}{T}\int_0^T \log S_u\,\mathrm{d}u\big)$ and, for $n \ge 1$, $G^{(n)} = \big(\prod_{i=1}^n S_{iT/n}\big)^{1/n}$. Then $\log(G_T/S_0)$ and $\log(G^{(n)}/S_0)$ are normal under $\mathbb{Q}$, with means and variances*

$$\mu = \Big(r - \frac{\sigma^2}{2}\Big)\frac{T}{2}, \quad s^2 = \frac{\sigma^2 T}{3}; \qquad \mu_n = \Big(r - \frac{\sigma^2}{2}\Big)\frac{T(n+1)}{2n}, \quad s_n^2 = \frac{\sigma^2 T(n+1)(2n+1)}{6n^2}.$$

The geometric Asian calls cost $e^{-rT}\mathbb{E}^{\mathbb{Q}}[(G_T - K)^+] = e^{-rT}\big(F\,N(\delta_+) - K\,N(\delta_-)\big)$ with $F = S_0 e^{\mu + s^2/2}$ and $\delta_\pm = \big(\log(F/K) \pm \frac{s^2}{2}\big)/s$, and the same with (μ_n, s_n) for $G^{(n)}$.

Proof. $\log(S_u/S_0) = (r - \frac{\sigma^2}{2})u + \sigma\widetilde{W}_u$. Averaging over $u \in [0,T]$ gives the mean $(r - \frac{\sigma^2}{2})\frac{T}{2}$ plus $\frac{\sigma}{T}\int_0^T \widetilde{W}_u\,\mathrm{d}u$, which is $\mathcal{N}(0, \sigma^2 T/3)$ by Theorem 8.19. For the discrete average, the mean of $\frac{1}{n}\sum_i t_i$ with $t_i = iT/n$ is $\frac{T(n+1)}{2n}$, and $\frac{1}{n}\sum_i \widetilde{W}_{t_i}$ is centred Gaussian (Theorem 5.7) with variance $\frac{T}{n^3}\sum_{i,j=1}^n \min(i,j) = \frac{T}{n^3}\sum_{k=1}^n (n-k+1)^2 = \frac{T(n+1)(2n+1)}{6n^2}$. For $G = S_0 e^{\mu + s\xi}$ with $\xi \sim \mathcal{N}(0,1)$ and $s > 0$, $G > K$ if and only if $\xi > -\delta_-$, and Theorem 6.10 with $a = s$ gives $\mathbb{E}[(G-K)^+] = FN(\delta_- + s) - KN(\delta_-)$. □

For $n = 1$ this is the Black–Scholes call, and $(\mu_n, s_n) \to (\mu, s)$ as $n \to \infty$. Averaging cuts the volatility to $\sigma/\sqrt{3}$, and if $r \ge 0$ Asian calls are cheaper than calls (Exercise 13.5). By the inequality between arithmetic and geometric means, $G_T \le Y_T/T$, so the geometric call is a

lower bound for the arithmetic one. The two are also strongly correlated, which makes the geometric payoff, whose price is known, an efficient *control variate* for Monte Carlo estimates of the arithmetic price (Kemna and Vorst, 1990): in Figure 13.3 it divides the standard deviation by a factor between 14 and 27.

The arithmetic price is a function of three variables. Given $\mathcal{F}_t$, the average still to come depends on the past only through S_t, but the payoff also depends on the accrued Y_t: the price depends on (S_t, Y_t), not on S_t alone (the pair (S, Y) is Markov, Y alone is not).

Proposition 13.10 (The Asian PDE)**.** *Let $K \in \mathbb{R}$ and let V_t be the price of $(Y_T/T - K)^+$.*

(i) $V_t = v(t, S_t, Y_t)$, where, for a Brownian motion W' and $R_u = \exp\left(\sigma W'_u + (r - \frac{\sigma^2}{2})u\right)$,

$$v(t,x,y) = e^{-r(T-t)}\,\mathbb{E}\Big[\Big(\frac{y + x\int_0^{T-t} R_u\,\mathrm{d}u}{T} - K\Big)^+\Big].$$

(ii) Let w be continuous on $[0,T]\times(0,\infty)\times\mathbb{R}$, with continuous $\partial_t w$ and continuous first and second derivatives in (x,y) on $[0,T)\times(0,\infty)\times\mathbb{R}$, such that

$$\partial_t w + rx\,\partial_x w + x\,\partial_y w + \frac{\sigma^2 x^2}{2}\,\partial_{xx} w = rw, \qquad w(T,x,y) = \Big(\frac{y}{T} - K\Big)^+, \tag{13.7}$$

with $|\partial_x w| \le c$ and $|w(t,x,y)| \le c(1 + x + |y|)$ for a constant c. Then $V_t = w(t, S_t, Y_t)$ for $t \le T$, and holding $\Delta_t = \partial_x w(t, S_t, Y_t)$ shares replicates the claim: $e^{-rt}V_t = V_0 + \int_0^t \Delta_u\,\mathrm{d}(e^{-ru}S_u)$.

Proof. (i) $Y_T = Y_t + S_t\int_0^{T-t} R_u\,\mathrm{d}u$ with $R_u = S_{t+u}/S_t$, a function of the restarted Brownian motion of Theorem 13.1; apply the independence lemma with $X = (S_t, Y_t)$.

(ii) Y is an Itô process with $\mathrm{d}Y_t = S_t\,\mathrm{d}t$, so $[Y,Y] = [S,Y] = 0$, while $\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}\widetilde{W}_t$ (Theorem 8.8 under $\mathbb{Q}$). Fix $s < T$. The multidimensional Itô formula (Theorem 8.17, on $[0,s]$ and the convex open set $(0,\infty)\times\mathbb{R}$) applied to $e^{-rt}w(t,x,y)$ gives, by (13.7),

$$\begin{aligned}\mathrm{d}(e^{-rt}w(t,S_t,Y_t)) &= e^{-rt}(\partial_t w + rS_t\partial_x w + S_t\partial_y w + \tfrac{1}{2}\sigma^2 S_t^2\partial_{xx} w - rw)\mathrm{d}t + e^{-rt}\sigma S_t\,\partial_x w\,\mathrm{d}\widetilde{W}_t\\ &= \Delta_t\,\mathrm{d}(e^{-rt}S_t),\end{aligned}$$

with the derivatives of w evaluated at (t, S_t, Y_t). The integrand $e^{-rt}\sigma S_t\Delta_t$ is bounded by $c\sigma e^{|r|T}S_t$ in absolute value and $\mathbb{E}^{\mathbb{Q}}[S_t^2]$ is bounded on $[0,T]$, so it is in $\mathbb{H}^2$ and $e^{-rt}w(t,S_t,Y_t)$ is a martingale on $[0,s]$: $e^{-rt}w(t,S_t,Y_t) = \mathbb{E}^{\mathbb{Q}}[e^{-rs}w(s,S_s,Y_s) \mid \mathcal{F}_t]$ for $t \le s < T$. As $s \uparrow T$, $w(s,S_s,Y_s) \to (Y_T/T - K)^+$ by continuity, and $|w(s,S_s,Y_s)| \le c(1 + \overline{S}_T + T\overline{S}_T)$, which is integrable by (13.5). Conditional dominated convergence gives $e^{-rt}w(t,S_t,Y_t) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(Y_T/T - K)^+ \mid \mathcal{F}_t] = e^{-rt}V_t$. □

Equation (13.7) has two space variables and no diffusion in y, which makes it awkward for finite differences. Večeř (2001) reduced it to one: $Y_T/T - K$ is the terminal wealth of a simple self-financing portfolio, the option is a call on it with strike zero, and in units of the stock this portfolio is a one-dimensional Markov process.

Theorem 13.11 (Večeř's equation)**.** *Let $\gamma(t) = \frac{1 - e^{-r(T-t)}}{rT}$ (and $\gamma(t) = \frac{T-t}{T}$ if $r = 0$), $K \in \mathbb{R}$, and V_t the price of $(Y_T/T - K)^+$.*

(i) $\Pi_t = e^{-r(T-t)}(Y_t/T - K) + \gamma(t)S_t$ *satisfies* $\mathrm{d}(e^{-rt}\Pi_t) = \gamma(t)\,\mathrm{d}(e^{-rt}S_t)$*: it is the wealth of the self-financing strategy that holds* $\gamma(t)$ *shares (Theorem 10.1), started from* $\Pi_0 = \gamma(0)S_0 - e^{-rT}K$*. Its terminal value is* $\Pi_T = Y_T/T - K$*, so the Asian call pays* Π_T^+*.*

(ii) $Z_t = \Pi_t/S_t$ *satisfies* $\mathrm{d}Z_t = \sigma\big(\gamma(t) - Z_t\big)\big(\mathrm{d}\widetilde{W}_t - \sigma\,\mathrm{d}t\big)$*, and* $V_t = S_t\,g(t, Z_t)$*, where, for a Brownian motion* W' *and* $\Lambda_u = \exp\big(-\sigma(W'_T - W'_u) - (r + \tfrac{\sigma^2}{2})(T-u)\big)$*,*

$$g(t,z) = \mathbb{E}\Big[\Big(e^{r(T-t)}\Lambda_t\big(z - \gamma(t)\big) + \frac{1}{T}\int_t^T \Lambda_u\,\mathrm{d}u\Big)^+\Big].$$

(iii) Let $\tilde g$ *be continuous on* $[0,T]\times\mathbb{R}$*, with continuous* $\partial_t\tilde g$*,* $\partial_z\tilde g$*,* $\partial_{zz}\tilde g$ *on* $[0,T)\times\mathbb{R}$*, such that*

$$\partial_t\tilde g + \frac{\sigma^2}{2}\big(\gamma(t) - z\big)^2\partial_{zz}\tilde g = 0, \qquad \tilde g(T,z) = z^+, \tag{13.8}$$

with $|\partial_z\tilde g| \le c$*. Then* $V_t = S_t\,\tilde g(t, Z_t)$ *for* $t \le T$*; in particular* $V_0 = S_0\,\tilde g\big(0, \gamma(0) - e^{-rT}K/S_0\big)$*. Holding* $\Delta_t = \tilde g(t,Z_t) + \big(\gamma(t) - Z_t\big)\partial_z\tilde g(t,Z_t)$ *shares replicates the claim.*

Proof. (i) $e^{-rt}\Pi_t = e^{-rT}(Y_t/T - K) + \gamma(t)e^{-rt}S_t$ and $\gamma'(t) = -e^{-r(T-t)}/T$. By the product rule with the deterministic C^1 function γ, $\mathrm{d}(e^{-rt}\Pi_t) = e^{-rT}\frac{S_t}{T}\mathrm{d}t + \gamma'(t)e^{-rt}S_t\,\mathrm{d}t + \gamma(t)\,\mathrm{d}(e^{-rt}S_t)$, and the first two terms cancel. Since $\gamma(T) = 0$, $\Pi_T = Y_T/T - K$.

(ii) Let $U_t = e^{-rt}S_t$, so $\mathrm{d}U = \sigma U\,\mathrm{d}\widetilde{W}$. Itô's formula for $1/x$ on $(0,\infty)$ (Theorem 8.14) gives $\mathrm{d}(1/U) = (1/U)(-\sigma\,\mathrm{d}\widetilde{W} + \sigma^2\mathrm{d}t)$, and the product rule (Theorem 8.18) applied to $Z = (e^{-rt}\Pi)\cdot(1/U)$, with $\mathrm{d}(e^{-rt}\Pi) = \gamma\sigma U\,\mathrm{d}\widetilde{W}$, gives $\mathrm{d}Z = Z(-\sigma\,\mathrm{d}\widetilde{W} + \sigma^2\mathrm{d}t) + \gamma\sigma\,\mathrm{d}\widetilde{W} - \gamma\sigma^2\mathrm{d}t$, the formula. For the price, use the stock as numeraire: by Theorem 11.13(ii)–(iii), $V_t = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[\Pi_T^+ \mid \mathcal{F}_t] = S_t\,\mathbb{E}^{\mathbb{Q}^S}[\Pi_T^+/S_T \mid \mathcal{F}_t] = S_t\,\mathbb{E}^{\mathbb{Q}^S}[Z_T^+ \mid \mathcal{F}_t]$, where $W_t^S = \widetilde{W}_t - \sigma t$ is a $\mathbb{Q}^S$-Brownian motion on $[0,T]$ and $S_u = S_0\exp\big(\sigma W_u^S + (r + \frac{\sigma^2}{2})u\big)$. From the definition of Π, $Y_t/T - K = e^{r(T-t)}(\Pi_t - \gamma(t)S_t)$, hence

$$Z_T = \frac{\Pi_T}{S_T} = e^{r(T-t)}\frac{S_t}{S_T}\big(Z_t - \gamma(t)\big) + \frac{1}{T}\int_t^T \frac{S_u}{S_T}\,\mathrm{d}u,$$

and $S_u/S_T = \exp\big(-\sigma(W_T^S - W_u^S) - (r + \frac{\sigma^2}{2})(T-u)\big)$ for $u \ge t$ depends only on the increments of W^S after t, which under $\mathbb{Q}^S$ form a Brownian motion independent of $\mathcal{F}_t$ (Theorem 13.1). The independence lemma under $\mathbb{Q}^S$ gives $\mathbb{E}^{\mathbb{Q}^S}[Z_T^+ \mid \mathcal{F}_t] = g(t,Z_t)$.

(iii) By Itô's formula on $[0,s]$, $s < T$, and (13.8), $\mathrm{d}\tilde g(t,Z_t) = \sigma(\gamma - Z_t)\partial_z\tilde g\,(\mathrm{d}\widetilde{W}_t - \sigma\,\mathrm{d}t)$, with $\tilde g$ and its derivatives evaluated at (t,Z_t). The product rule with $\mathrm{d}U = \sigma U\,\mathrm{d}\widetilde{W}$ gives

$$\mathrm{d}\big(U_t\tilde g(t,Z_t)\big) = \sigma U_t\tilde g\,\mathrm{d}\widetilde{W}_t + \sigma U_t(\gamma - Z_t)\partial_z\tilde g\,(\mathrm{d}\widetilde{W}_t - \sigma\,\mathrm{d}t) + \sigma^2 U_t(\gamma - Z_t)\partial_z\tilde g\,\mathrm{d}t = \Delta_t\,\mathrm{d}U_t;$$

the $\mathrm{d}t$-terms cancel. Since $|\partial_z\tilde g| \le c$ and $\tilde g(\cdot,0)$ is bounded on $[0,T]$, $|\tilde g(t,z)| + |(\gamma(t) - z)\partial_z\tilde g(t,z)| \le c'(1 + |z|)$, so $|U_t\Delta_t| \le c'(S_t + |\Pi_t|) \le c''(1 + \overline{S}_T)$ by the definition of Π. The rest is the end of the proof of Theorem 13.10(ii): the integrand is in $\mathbb{H}^2$, $U_t\tilde g(t,Z_t)$ is a martingale on $[0,s]$, and as $s \uparrow T$, $S_s\tilde g(s,Z_s) \to S_T Z_T^+ = \Pi_T^+$ with an integrable bound. □

Under $\mathbb{Q}^S$, $\mathrm{d}Z_t = \sigma(\gamma(t) - Z_t)\,\mathrm{d}W_t^S$ has no drift, as the ratio of two traded prices must under the measure of its denominator; hence no first-order term in (13.8). For $z \ge \gamma(t)$, that is $Y_t/T \ge K$, the option will surely be exercised, and the function g of (ii) is $g(t,z) = z$, since $\mathbb{E}[\Lambda_u] = e^{-r(T-u)}$; with $g \to 0$ as $z \to -\infty$, this gives the boundary conditions of a finite-difference

grid. Solved by Crank–Nicolson, (13.8) reproduces the benchmark prices of continuously averaged Asian calls of Linetsky (2004) to five decimals (`code/tests/test_exotics.py`). The function of Theorem 13.10 is $w(t,x,y) = x\,\tilde g(t,\pi(t,x,y)/x)$ with $\pi(t,x,y) = e^{-r(T-t)}(y/T-K)+\gamma(t)x$ (Exercise 13.6).

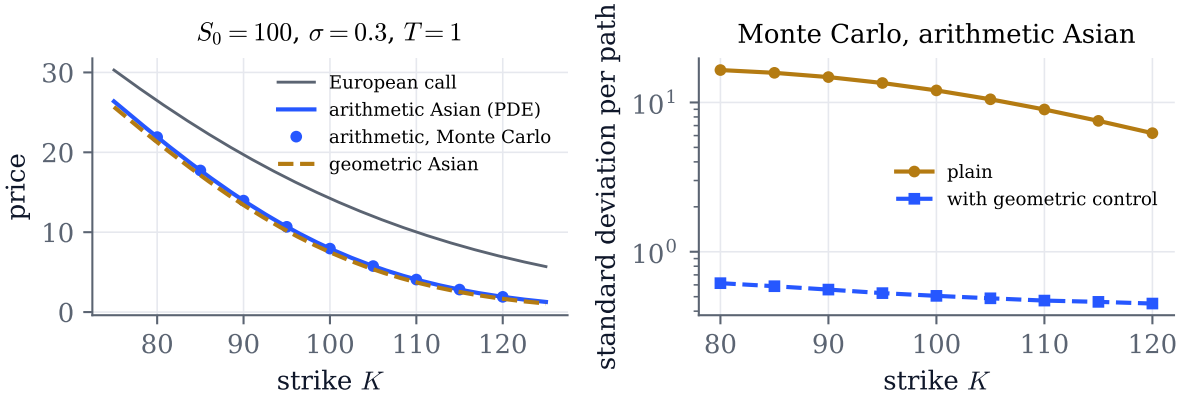

Figure 13.3: Asian calls with $S_0 = 100$, $r = 5\%$, $\sigma = 30\%$, $T = 1$. Left: prices against the strike: the European call, the arithmetic Asian call from (13.8) (line) and by Monte Carlo (dots), and the geometric Asian call of Theorem 13.9. Right: standard deviation per path of the Monte Carlo estimator of the arithmetic price, plain and with the geometric payoff as a control variate (10^4 paths, 100 time steps).

▶ Interactive demo: arithmetic and geometric Asian prices for any volatility and maturity, Večeř's function $\tilde g(t,\cdot)$, and the variance reduction of the control variate.

13.6 The perpetual American put

An American put can be exercised at any time before maturity. A *perpetual* American put has no maturity: its holder chooses when, if ever, to receive $K - S_\tau$. By analogy with Theorem 4.13(ii), its price is the largest expected discounted payoff over all exercise rules. Let $\mathcal{T}$ be the set of stopping times with values in $[0,\infty]$, and for an initial price $S_0 = x > 0$ let

$$v(x) = \sup_{\tau\in\mathcal{T}} \mathbb{E}^{\mathbb{Q}}\big[e^{-r\tau}(K-S_\tau)^+ \mathbf{1}_{\{\tau<\infty\}}\big].$$

That v is the smallest initial capital from which the seller can hedge the put whatever the buyer does is shown in Shreve (2004b), Section 8.3; here we compute it. With no maturity, time disappears from the problem, and the optimal rule is to exercise when the price first falls to a fixed level.

Proposition 13.12 (Perpetual American put)**.** *Let $r > 0$, $K > 0$, $\beta = 2r/\sigma^2$ and $L^* = \frac{\beta K}{\beta+1} = \frac{2rK}{2r+\sigma^2}$. Then*

$$v(x) = \begin{cases} K - x, & 0 < x \le L^*, \\ (K-L^*)\,(x/L^*)^{-\beta}, & x \ge L^*, \end{cases}$$

and the supremum is attained by $\tau^* = \inf\{t \ge 0 : S_t \le L^*\}$. The function v is C^1, with

$v'(L^*) = -1$, it satisfies $rv = rxv' + \frac{\sigma^2 x^2}{2} v''$ on (L^*, ∞), and $v > (K - x)^+$ there.

Proof. A martingale. Let $N_t = e^{-rt} S_t^{-\beta}$. Since $-\beta(r - \frac{\sigma^2}{2}) - r = -\frac{1}{2}\beta^2\sigma^2$, $N_t = x^{-\beta} \exp\big(-\beta\sigma\widetilde{W}_t - \frac{1}{2}\beta^2\sigma^2 t\big)$ is the exponential martingale of Theorem 5.12(iii), and $N_t = N_0 - \int_0^t \beta\sigma N_u \, \mathrm{d}\widetilde{W}_u$ (Theorem 8.5) with an integrand in $\mathbb{H}^2$ on every $[0, T']$. For a stopping time τ and $t \le T'$, Theorem 7.15 applied to $\tau \wedge T'$ shows that $N_{t\wedge\tau}$ is a martingale, so $\mathbb{E}^{\mathbb{Q}}[N_{t\wedge\tau}] = x^{-\beta}$ for all $t \ge 0$.

Lower bound. For $0 < L \le x$ let $\tau_L = \inf\{t : S_t \le L\}$, a stopping time. Before τ_L, $S \ge L$, so $0 \le N_{t\wedge\tau_L} \le L^{-\beta}$; as $t \to \infty$, $N_{t\wedge\tau_L} \to L^{-\beta} e^{-r\tau_L}$ on $\{\tau_L < \infty\}$, and $N_t \le e^{-rt} L^{-\beta} \to 0$ on $\{\tau_L = \infty\}$. Dominated convergence gives $\mathbb{E}^{\mathbb{Q}}[e^{-r\tau_L} \mathbf{1}_{\{\tau_L < \infty\}}] = (x/L)^{-\beta}$. Since $S_{\tau_L} = L$, exercising at τ_L yields $(K - L)(x/L)^{-\beta}$ for $L \le K$. For $x \ge L^*$ take $L = L^*$; for $x < L^*$ exercise at once, $\tau = 0$. So v is at least the formula.

Upper bound. For $x_0 \in (0, L^*]$, let $f(y) = a + cy^{-\beta}$ with $c = x_0^{\beta+1}/\beta$ and $a = K - x_0 \frac{\beta+1}{\beta}$, which is nonnegative because $x_0 \le L^*$. Then f is convex, $f(x_0) = K - x_0$ and $f'(x_0) = -1$: the line $K - y$ is tangent to f at x_0, so $f(y) \ge K - y$ for all $y > 0$, and $f \ge a \ge 0$; hence $f(y) \ge (K - y)^+$. For every $\tau \in \mathcal{T}$ and $t \ge 0$,

$$\mathbb{E}^{\mathbb{Q}}\big[e^{-r(t\wedge\tau)} f(S_{t\wedge\tau})\big] = a\,\mathbb{E}^{\mathbb{Q}}\big[e^{-r(t\wedge\tau)}\big] + c\,\mathbb{E}^{\mathbb{Q}}[N_{t\wedge\tau}] \le a + cx^{-\beta} = f(x),$$

using $a \ge 0$ and $r > 0$. As $t \to \infty$, the nonnegative variable inside the expectation tends to $e^{-r\tau} f(S_\tau)$ on $\{\tau < \infty\}$, so by Fatou's lemma (a consequence of monotone convergence) $\mathbb{E}^{\mathbb{Q}}[e^{-r\tau}(K - S_\tau)^+ \mathbf{1}_{\{\tau<\infty\}}] \le f(x)$. Taking $x_0 = x$ if $x \le L^*$ gives $v(x) \le K - x$; taking $x_0 = L^*$, for which $a = 0$ and $c = L^{*\beta+1}/\beta = (K - L^*)L^{*\beta}$, gives $v(x) \le (K - L^*)(x/L^*)^{-\beta}$ for every x.

Properties. At L^* both expressions equal $K - L^*$, and the right derivative is $-\beta(K - L^*)/L^* = -\beta K/((\beta + 1)L^*) = -1$. The equation holds because $y^{-\beta}$ solves it, as in Exercise 6.2(c). For $x > L^*$, $v(x) = f_{L^*}(x) > K - x$, because the convex f_{L^*} touches its tangent line only at L^* (it is strictly convex), and $v > 0$. □

The optimal level L^* maximizes $(K - L)L^\beta$, the value of waiting for L, and there the exercise value and the value of waiting meet with the same slope, $v'(L^*) = -1$: the *smooth-fit* condition. If $r = 0$, $v \equiv K$: the levels $L \downarrow 0$ are reached almost surely, as $\nu < 0$, and $K - S_\tau < K$, so no rule attains it. An American put with maturity T is worth at most v, since it allows fewer exercise rules, and at least the value of the rule $\tau^* \wedge T$, which tends to v as $T \to \infty$; its exercise boundary depends on time and has no closed form (Chapter 16). The perpetual put was priced by Merton (1973); the perpetual call on a stock paying dividends is Exercise 13.7.

▶ Interactive demo: the perpetual put and its exercise level for any rate and volatility, with binomial American puts of growing maturity approaching it.

Further reading

Shreve (2004b), Chapters 7 and 8, uses the same tools; Hull (2021) describes the markets. E. Reiner and M. Rubinstein, *Breaking down the barriers* (*Risk*, 1991), collect the closed forms of all barrier calls and puts, with rebates.

Chapter 14

Short-rate models

Until now the interest rate was a constant, or a bounded process given in advance (Section 11.3). A zero-coupon bond pays 1 at its maturity T; by Theorem 11.8, its price at time t is $\mathbb{E}^{\mathbb{Q}}\big[e^{-\int_t^T r_u\,\mathrm{d}u}\,\big|\,\mathcal{F}_t\big]$, so a model for the short rate r under a risk-neutral measure $\mathbb{Q}$ prices every bond, and with it the yield curve. The affine models give bond prices in closed form through Riccati equations (Theorem 14.5): the Ho–Lee and Vasicek models, with Gaussian rates that can be negative, and the Cox–Ingersoll–Ross model, whose rate is nonnegative. Bond options are priced with the forward measure, and a last remark describes the approach of Heath, Jarrow and Morton.

Setting. Throughout, $(\Omega, \mathcal{F}, \mathbb{Q})$ is a probability space, $\widetilde{W}$ is a $\mathbb{Q}$-Brownian motion with a filtration $(\mathcal{F}_t)$ (Theorem 5.11, under $\mathbb{Q}$), every $\mathbb{Q}$-null event belongs to $\mathcal{F}_0$, and $T^* > 0$ is the longest maturity considered. Stochastic integrals, Itô processes and Itô's formula are those of Chapters 7 and 8 (Theorem 8.2), applied under $\mathbb{Q}$, and solutions of stochastic differential equations are those of Theorem 9.1, with $\widetilde{W}$ in place of W. Why we work under $\mathbb{Q}$ from the start is explained in Theorem 14.4.

14.1 Bonds, yields and forward rates

Definition 14.1 (Bonds, yields, forward rates)**.** A *zero-coupon bond* with maturity $T \le T^*$ pays 1 at time T; its price at time $t \le T$ is $P(t,T) > 0$, with $P(T,T) = 1$. For $t < T$, the *yield* is $Y(t,T) = -\log P(t,T)/(T-t)$: the constant continuously compounded rate at which $P(t,T)$ grows to 1 by time T. For $t \le T_1 < T_2 \le T^*$, the *forward rate* for $[T_1, T_2]$ seen at time t is

$$F(t;T_1,T_2) = \frac{\log P(t,T_1) - \log P(t,T_2)}{T_2 - T_1}. \tag{14.1}$$

If $T \mapsto P(t,T)$ is continuously differentiable, the *instantaneous forward rate* is $f(t,T) = -\partial_T \log P(t,T)$, so that $P(t,T) = \exp\big(-\int_t^T f(t,u)\,\mathrm{d}u\big)$. The *short rate* is $r_t = \lim_{T \downarrow t} Y(t,T)$, when the limit exists. The maps $T \mapsto Y(t,T)$ and $T \mapsto f(t,T)$ are the *yield curve* and the *forward curve* at time t.

The yield is the average of the forward rates, $Y(t,T) = \frac{1}{T-t}\int_t^T f(t,u)\,\mathrm{d}u$, and the short rate is $f(t,t)$ when f is continuous. The forward rate is the rate of a loan that can be locked in today by trading bonds. At time t, buy one bond maturing at T_1 and sell $P(t,T_1)/P(t,T_2)$ bonds maturing at T_2. This costs nothing, pays 1 at T_1 and costs $P(t,T_1)/P(t,T_2) = e^{F(t;T_1,T_2)(T_2-T_1)}$ at T_2: a loan from T_1 to T_2 at the continuously compounded rate $F = F(t;T_1,T_2)$, agreed at t. A costless contract to receive 1 at T_1 against $e^{k(T_2-T_1)}$ at T_2 with $k < F$ would be an arbitrage (Theorem 3.1): enter it, sell one T_1-bond and buy $e^{k(T_2-T_1)}$ bonds maturing at T_2, which brings $P(t,T_2)(e^{F(T_2-T_1)} - e^{k(T_2-T_1)}) > 0$ today and nothing later; for $k > F$, reverse it. A bond paying coupons $c_1,\dots,c_n$ at times $T_1 < \cdots < T_n$ is, by the same static argument, worth $\sum_i c_i P(t,T_i)$, which is how zero-coupon prices are extracted from coupon bonds (Exercise 14.1).

14.2 Short-rate models

Definition 14.2 (Short-rate model)**.** Let $b, \sigma : [0,T^*]\times\mathbb{R}\to\mathbb{R}$ be Borel functions and $I\subseteq\mathbb{R}$ an interval. A *short-rate model* is a solution r on $[0,T^*]$, with values in I, of

$$\mathrm{d}r_t = b(t,r_t)\,\mathrm{d}t + \sigma(t,r_t)\,\mathrm{d}\widetilde{W}_t, \tag{14.2}$$

with $r_0\in I$ a constant, such that $\mathbb{E}^{\mathbb{Q}}[D_T]<\infty$ for every $T\le T^*$, where $D_t = \exp\big(-\int_0^t r_u\,\mathrm{d}u\big)$ is the discount factor; the money market is worth $1/D_t$ at time t. The *bond prices* of the model are, for $0\le t\le T\le T^*$,

$$P(t,T) = \mathbb{E}^{\mathbb{Q}}\Big[e^{-\int_t^T r_u\,\mathrm{d}u}\,\Big|\,\mathcal{F}_t\Big] = \frac{\mathbb{E}^{\mathbb{Q}}[D_T\mid\mathcal{F}_t]}{D_t}. \tag{14.3}$$

The condition $\mathbb{E}^{\mathbb{Q}}[D_T]<\infty$ holds if r is bounded below. The conditional expectation in (14.3) is that of a positive variable (Theorem 1.17), and taking out what is known gives its second form, so $D_tP(t,T) = \mathbb{E}^{\mathbb{Q}}[D_T\mid\mathcal{F}_t]$: for every maturity, the discounted bond price is a positive $\mathbb{Q}$-martingale on $[0,T]$, and the discounted money market is constant. So $\mathbb{Q}$ is a risk-neutral measure (Theorem 11.6) for the market made of the money market and the bonds, with respect to any real-world measure equivalent to $\mathbb{Q}$; prices of other claims are given by Theorem 11.8. When the discounted bond prices are Itô processes, $\mathrm{d}\big(D_tP(t,T)\big) = D_tP(t,T)\,\sigma_t^P\,\mathrm{d}\widetilde{W}_t$, as in all the models below, the discounted wealth of a self-financing strategy in the money market and finitely many bonds is a stochastic integral against $\widetilde{W}$, a $\mathbb{Q}$-local martingale, and the argument of Theorem 11.8(ii) and Theorem 11.11(i) shows that no tame strategy is an arbitrage.

In a diffusion model the bond price should be a function of the current rate, and this function solves a partial differential equation, the analogue of the Black–Scholes equation.

Proposition 14.3 (The bond-pricing equation)**.** *Let r be a short-rate model, $T\le T^*$, $F\in C^{1,2}([0,T]\times\mathbb{R})$, and*

$$G(t,x) = \partial_t F(t,x) + b(t,x)\,\partial_x F(t,x) + \tfrac{1}{2}\sigma(t,x)^2\,\partial_{xx}F(t,x) - x\,F(t,x).$$

(i) If $G = 0$ on $[0,T)\times I$ and $F(T,\cdot) = 1$ on I, then $D_tF(t,r_t) = F(0,r_0) + \int_0^t D_u\sigma(u,r_u)\partial_x F(u,r_u)\,\mathrm{d}\widetilde{W}_u$ is a $\mathbb{Q}$-local martingale on $[0,T]$. If it is a $\mathbb{Q}$-martingale, then $P(t,T) = F(t,r_t)$ for $0\le t\le T$.

(ii) Conversely, let $(\mathcal{F}_t)$ be the filtration generated by $\widetilde{W}$ and augmented by the $\mathbb{Q}$-null sets,

and suppose that $P(t,T) = F(t,r_t)$ *for* $t \in [0,T]$. *Then* $G(t,r_t) = 0$ *for* $\mathrm{d}t \otimes \mathrm{d}\mathbb{Q}$*-almost every* $(t,\omega) \in [0,T] \times \Omega$, *and*

$$\mathrm{d}P(t,T) = r_t P(t,T)\,\mathrm{d}t + \sigma(t,r_t)\,\partial_x F(t,r_t)\,\mathrm{d}\widetilde{W}_t. \tag{14.4}$$

Proof. By Itô's formula (Theorem 8.13) and the product rule (Theorem 8.18), with $\mathrm{d}D_t = -r_t D_t\,\mathrm{d}t$ and $[D, F(\cdot,r)] = 0$ because D has no $\mathrm{d}\widetilde{W}$-part,

$$\mathrm{d}\big(D_t F(t,r_t)\big) = D_t\,G(t,r_t)\,\mathrm{d}t + D_t\,\sigma(t,r_t)\,\partial_x F(t,r_t)\,\mathrm{d}\widetilde{W}_t.$$

The integrand of the stochastic integral is in $\mathbb{H}^2_{\mathrm{loc}}$, since $\sigma(\cdot,r)$ is (Theorem 9.1) and D and $\partial_x F(\cdot,r)$ are continuous. (i) If $G = 0$ on $[0,T) \times I$, the $\mathrm{d}t$-term vanishes and the integral is a local martingale (Theorem 7.16). If it is a martingale, then $D_t F(t,r_t) = \mathbb{E}^{\mathbb{Q}}[D_T F(T,r_T) \mid \mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[D_T \mid \mathcal{F}_t] = D_t P(t,T)$. (ii) Now $D_t F(t,r_t) = D_t P(t,T)$ is a $\mathbb{Q}$-martingale, so by Theorem 11.10(i), applied under $\mathbb{Q}$ to $\widetilde{W}$, it is an Itô process with zero drift. By uniqueness of the decomposition (Theorem 8.11(ii)), $D_t G(t,r_t) = 0$ for $\mathrm{d}t \otimes \mathrm{d}\mathbb{Q}$-almost every (t,ω), and $D > 0$. Then $P(t,T) = D_t P(t,T) \cdot D_t^{-1}$ with $\mathrm{d}(D_t^{-1}) = r_t D_t^{-1}\,\mathrm{d}t$, and the product rule gives (14.4). □

If r_t has a density $p_t(x) > 0$, continuous in $(t,x) \in (0,T] \times I$ (Vasicek), then by (ii) and Fubini $\mathbb{E}^{\mathbb{Q}} \int_0^T |G(t,r_t)|\,\mathrm{d}t = \int_0^T \int_I |G| p_t\,\mathrm{d}x\,\mathrm{d}t = 0$, so $G = 0$ on $(0,T) \times I^\circ$ if b, σ are continuous. Part (i) is the Feynman–Kac formula (Theorem 12.1) with the discount rate $k(t,x) = x$; that theorem applies directly when the rate is bounded below and stays in an open interval, as in the CIR model under Feller's condition, with $I = (0,\infty)$. For Gaussian rates k is unbounded below, and the martingale property is checked by hand in Theorem 14.5. Formula (14.4) says that every bond earns the short rate on average under $\mathbb{Q}$, with a volatility $\sigma\,\partial_x \log F$ that depends on its maturity.

Remark 14.4 (Where is the real-world measure?). The short rate is not the price of a traded asset, so, unlike the drift of the stock in Section 11.2, its drift under $\mathbb{Q}$ is not determined by its drift under the real-world measure. Let λ be a bounded progressively measurable process and $\mathbb{P}$ the measure of Theorem 11.2, applied under $\mathbb{Q}$ with $\Theta = -\lambda$, under which $W_t = \widetilde{W}_t - \int_0^t \lambda_u\,\mathrm{d}u$ is a Brownian motion on $[0,T^*]$. Under $\mathbb{P}$, $\mathrm{d}r_t = \big(b(t,r_t) + \sigma(t,r_t)\lambda_t\big)\mathrm{d}t + \sigma(t,r_t)\,\mathrm{d}W_t$, and a bond with $\mathrm{d}P(t,T) = r_t P(t,T)\,\mathrm{d}t + \sigma_t^P P(t,T)\,\mathrm{d}\widetilde{W}_t$ satisfies

$$\mathrm{d}P(t,T) = \big(r_t + \lambda_t \sigma_t^P\big) P(t,T)\,\mathrm{d}t + \sigma_t^P P(t,T)\,\mathrm{d}W_t.$$

Every bond earns, in excess of the short rate, λ_t times its volatility: λ is the *market price of interest-rate risk*, the same for all maturities (as in Exercise 11.3). The short rate reveals $b + \sigma\lambda$, bond prices reveal b: so short-rate models are specified under $\mathbb{Q}$ and fitted to bond prices, and the risk premium is modelled separately (Exercise 14.2).

14.3 Affine term structures

Guess a bond price of the form $F(t,x) = e^{A - Bx}$, with A and B functions of time. Then $\partial_x F = -BF$ and $\partial_{xx} F = B^2 F$, and the bond-pricing equation, divided by F, is a polynomial in x of degree one if b and σ^2 are affine in x. Each coefficient must vanish, which gives two ordinary differential equations.

Theorem 14.5 (Affine term structures)**.** *Let $\beta_0 : [0, T^*] \to \mathbb{R}$ be continuous, and let $\beta_1 \in \mathbb{R}$ and $\alpha_0, \alpha_1 \geq 0$ be constants. Let $I = \mathbb{R}$ if $\alpha_1 = 0$, and $I = [-\alpha_0/\alpha_1, \infty)$ if $\alpha_1 > 0$, so that $\alpha_0 + \alpha_1 x \geq 0$ on I. Let r be a solution with values in I of*

$$\mathrm{d}r_t = (\beta_0(t) + \beta_1 r_t)\mathrm{d}t + \sigma_t \,\mathrm{d}\widetilde{W}_t, \qquad \sigma_t = \sqrt{\alpha_0 + \alpha_1 r_t}, \tag{14.5}$$

that is, of (14.2) with $\sigma(t,x) = \sqrt{(\alpha_0 + \alpha_1 x)^+}$, with $r_0 \in I$.

(i) Let $\gamma = \sqrt{\beta_1^2 + 2\alpha_1}$. The function

$$B(\tau) = \frac{2(e^{\gamma\tau} - 1)}{(\gamma - \beta_1)(e^{\gamma\tau} - 1) + 2\gamma} \quad (\gamma > 0), \qquad B(\tau) = \tau \quad (\gamma = 0), \tag{14.6}$$

is the unique solution on $[0, \infty)$ of the Riccati equation *$B'(\tau) = 1 + \beta_1 B(\tau) - \frac{1}{2}\alpha_1 B(\tau)^2$, $B(0) = 0$, and $B(\tau) > 0$ for $\tau > 0$.*

(ii) For $0 \leq t \leq T \leq T^*$ let $A(t,T) = \int_t^T \big(-\beta_0(s)B(T-s) + \frac{1}{2}\alpha_0 B(T-s)^2\big)\,\mathrm{d}s$. Then $\mathbb{E}^{\mathbb{Q}}[D_T] < \infty$, so r is a short-rate model, and

$$P(t,T) = \exp\big(A(t,T) - B(T-t)\, r_t\big), \tag{14.7}$$

$$D_t P(t,T) = P(0,T) \exp\Big(-\int_0^t B(T-u)\sigma_u \,\mathrm{d}\widetilde{W}_u - \frac{1}{2}\int_0^t B(T-u)^2 \sigma_u^2 \,\mathrm{d}u\Big). \tag{14.8}$$

In particular $\mathrm{d}P(t,T) = r_t P(t,T)\,\mathrm{d}t - B(T-t)\sigma_t P(t,T)\,\mathrm{d}\widetilde{W}_t$, and the yield $Y(t,T) = \big(B(T-t)r_t - A(t,T)\big)/(T-t)$ and the forward rate $f(t,T) = -\partial_T A(t,T) + B'(T-t)\, r_t$ are affine functions of r_t.

(iii) If β_0 is constant, then $A(t,T) = A(T-t)$, where $A(\tau) = \int_0^\tau \big(-\beta_0 B(u) + \frac{1}{2}\alpha_0 B(u)^2\big)\mathrm{d}u$ solves $A' = -\beta_0 B + \frac{1}{2}\alpha_0 B^2$, $A(0) = 0$.

Proof. (i) If $\gamma = 0$ then $\beta_1 = \alpha_1 = 0$ and $B(\tau) = \tau$ solves $B' = 1$. Let $\gamma > 0$, write $u = e^{\gamma\tau} - 1$ and $d = (\gamma - \beta_1)u + 2\gamma$. Since $\gamma \geq |\beta_1|$, $d \geq 2\gamma > 0$, and $B = 2u/d > 0$ for $\tau > 0$. Differentiating, $B' = 2\gamma e^{\gamma\tau}\big(d - (\gamma - \beta_1)u\big)/d^2 = 4\gamma^2(u+1)/d^2$. On the other side, using $2\alpha_1 = \gamma^2 - \beta_1^2$,

$$d^2\Big(1 + \beta_1 B - \frac{\alpha_1}{2}B^2\Big) = d^2 + 2\beta_1 u d - (\gamma^2 - \beta_1^2)u^2 = 4\gamma^2 u + 4\gamma^2,$$

because the coefficient of u^2 is $(\gamma-\beta_1)^2 + 2\beta_1(\gamma-\beta_1) - \gamma^2 + \beta_1^2 = 0$ and that of u is $4\gamma(\gamma-\beta_1) + 4\beta_1\gamma$. So B solves the equation. If $\tilde{B}$ is another solution on some $[0, \tau_0]$, both are bounded there by some L, the right-hand side $q(y) = 1 + \beta_1 y - \frac{1}{2}\alpha_1 y^2$ is Lipschitz on $[-L, L]$ with some constant c, and $|B - \tilde{B}|(\tau) \leq c\int_0^\tau |B - \tilde{B}|$; Gronwall's lemma (Theorem 9.3) gives $\tilde{B} = B$.

(ii) Fix T and let $M_t = D_t \exp\big(A(t,T) - B(T-t)r_t\big)$ for $t \leq T$. Its logarithm $A(t,T) - B(T-t)r_t - \int_0^t r_u\,\mathrm{d}u$ is affine in r_t with C^1 coefficients, so no second derivative appears in its differential. With $\partial_t A(t,T) = \beta_0(t)B(T-t) - \frac{1}{2}\alpha_0 B(T-t)^2$ and $\mathrm{d}\big(B(T-t)r_t\big) = -B'(T-t)r_t\,\mathrm{d}t + B(T-t)\,\mathrm{d}r_t$, and writing B for $B(T-t)$,

$$\begin{aligned}\mathrm{d}\log M_t &= \Big(\beta_0 B - \frac{\alpha_0}{2}B^2 + B' r_t - B(\beta_0 + \beta_1 r_t) - r_t\Big)\mathrm{d}t - B\sigma_t\,\mathrm{d}\widetilde{W}_t \\ &= \Big(-\frac{\alpha_0}{2}B^2 + r_t(B' - \beta_1 B - 1)\Big)\mathrm{d}t - B\sigma_t\,\mathrm{d}\widetilde{W}_t = -\frac{1}{2}B^2\sigma_t^2\,\mathrm{d}t - B\sigma_t\,\mathrm{d}\widetilde{W}_t,\end{aligned}$$

by the Riccati equation and $\sigma_t^2 = \alpha_0 + \alpha_1 r_t$ on I. So $M_t = M_0 Z_t$, where Z is the exponential of Theorem 11.2 (under $\mathbb{Q}$, for $\widetilde{W}$) with $\Theta_u = B(T-u)\sigma_u$, a continuous process. We show that M is a $\mathbb{Q}$-martingale on $[0,T]$. If $\alpha_1 = 0$, $\Theta_u = \sqrt{\alpha_0}\, B(T-u)$ is deterministic and bounded on $[0,T]$, and Theorem 11.2(ii)–(iii) apply. If $\alpha_1 > 0$, then $r_t \geq -c$ with $c = \alpha_0/\alpha_1$, so $D_t \leq e^{cT}$ and $e^{-B(T-t)r_t} \leq e^{cB(T-t)}$; as $A(\cdot,T)$ and B are continuous on $[0,T]$, M is bounded by a constant. By Theorem 11.2(i), M is a local martingale: with the stopping times τ_n of Theorem 7.16, for $s \leq t \leq T$ and $E \in \mathcal{F}_s$, $\mathbb{E}^{\mathbb{Q}}[M_{t\wedge\tau_n}\mathbf{1}_E] = \mathbb{E}^{\mathbb{Q}}[M_{s\wedge\tau_n}\mathbf{1}_E]$, and since $\tau_n = T$ for n large, almost surely, dominated convergence gives $\mathbb{E}^{\mathbb{Q}}[M_t\mathbf{1}_E] = \mathbb{E}^{\mathbb{Q}}[M_s\mathbf{1}_E]$.

Since $A(T,T) = 0$ and $B(0) = 0$, $M_T = D_T$. Hence $\mathbb{E}^{\mathbb{Q}}[D_T] = M_0 < \infty$ and $D_t P(t,T) = \mathbb{E}^{\mathbb{Q}}[D_T \mid \mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[M_T \mid \mathcal{F}_t] = M_t$, which is (14.7); and $M_t = M_0 Z_t$ with $M_0 = P(0,T)$ is (14.8). The product rule applied to $P(t,T) = M_t/D_t$, with $\mathrm{d}M_t = -B(T-t)\sigma_t M_t\, \mathrm{d}\widetilde{W}_t$ (Theorem 11.2(i)), gives the dynamics. The yield and the forward rate follow from the definitions; $A(t,T)$ is differentiable in T, with $\partial_T A(t,T) = \int_t^T \big(-\beta_0(s) + \alpha_0 B(T-s)\big)B'(T-s)\,\mathrm{d}s$, because the integrand of A vanishes at $s = T$.

(iii) Substitute $u = T - s$. □

Remark 14.6 (Affine drift and variance are also necessary). Suppose that b and σ do not depend on t and that, for every maturity, $F(t,x) = \exp\big(A(T-t) - B(T-t)x\big)$ solves the bond-pricing equation of Theorem 14.3 on $[0,T)\times I$. Dividing by F, $-A'(\tau) + B'(\tau)x - b(x)B(\tau) + \frac{1}{2}\sigma(x)^2 B(\tau)^2 - x = 0$ for all $\tau > 0$ and $x \in I$. Here B is continuous, $B(0) = 0$, and $B \not\equiv 0$ (otherwise $-A' = x$ on I), so it takes two distinct nonzero values $B(\tau_1)$, $B(\tau_2)$. The two equations are then a linear system for $\big(b(x), \sigma(x)^2\big)$ with nonzero determinant $\frac{1}{2}B(\tau_1)B(\tau_2)\big(B(\tau_1) - B(\tau_2)\big)$ and right-hand side affine in x: so b and σ^2 are affine. Duffie and Kan (1996) develop the multi-factor version, in which r is an affine function of a diffusion in $\mathbb{R}^d$ with affine drift and covariance, and the Riccati equations become a system; see also Shreve (2004b), Section 10.2.

The simplest affine model has $\beta_1 = \alpha_1 = 0$.

Proposition 14.7 (The Ho–Lee model). *Let $\sigma > 0$, $r_0 \in \mathbb{R}$ and $\beta : [0,T^*] \to \mathbb{R}$ continuous, and let $r_t = r_0 + \int_0^t \beta(s)\,\mathrm{d}s + \sigma\widetilde{W}_t$, the solution of $\mathrm{d}r_t = \beta(t)\,\mathrm{d}t + \sigma\,\mathrm{d}\widetilde{W}_t$.*

(i) $P(t,T) = \exp\big(-\int_t^T (T-s)\beta(s)\,\mathrm{d}s + \frac{\sigma^2}{6}(T-t)^3 - (T-t)\,r_t\big)$, and $f(0,T) = r_0 + \int_0^T \beta(s)\,\mathrm{d}s - \frac{\sigma^2}{2}T^2$.

(ii) Let $f^M : [0,T^*] \to \mathbb{R}$ be continuously differentiable (a forward curve observed in the market) and $P^M(0,T) = \exp\big(-\int_0^T f^M(u)\,\mathrm{d}u\big)$. Then $P(0,T) = P^M(0,T)$ for every $T \in [0,T^*]$ if and only if $r_0 = f^M(0)$ and $\beta(t) = (f^M)'(t) + \sigma^2 t$ for every $t \in [0,T^*]$.

Proof. (i) Apply Theorem 14.5 with $\beta_0 = \beta$, $\beta_1 = 0$, $\alpha_0 = \sigma^2$, $\alpha_1 = 0$: then $\gamma = 0$, $B(\tau) = \tau$ and $A(t,T) = \int_t^T \big(-(T-s)\beta(s) + \frac{\sigma^2}{2}(T-s)^2\big)\mathrm{d}s$. Differentiating $-\log P(0,T) = \int_0^T (T-s)\beta(s)\,\mathrm{d}s - \frac{\sigma^2}{6}T^3 + Tr_0$ in T gives $f(0,T)$. (ii) The functions $\log P(0,\cdot)$ and $\log P^M(0,\cdot)$ are continuously differentiable and vanish at 0, so they agree on $[0,T^*]$ if and only if their derivatives do: $f(0,T) = f^M(T)$ for all T. By (i), at $T = 0$ this says $r_0 = f^M(0)$; given this, the two sides agree for all T if and only if their derivatives do, $\beta(T) - \sigma^2 T = (f^M)'(T)$. □

Alternatively, as in Theorem 8.19, $\int_t^T r_u\,\mathrm{d}u$ is, given $\mathcal{F}_t$, normal with variance $\sigma^2(T-t)^3/3$, and $\mathbb{E}^{\mathbb{Q}}[e^{-X}] = e^{-m+v/2}$ for $X \sim \mathcal{N}(m,v)$. The model of Ho and Lee (1986), originally a binomial

tree, is the prototype of the models built to reproduce any initial curve exactly. Its weakness is visible in (i): the rate does not revert to a mean, its variance $\sigma^2 t$ grows without bound, and for constant β the term $\frac{\sigma^2}{6}T^3$ eventually wins, so that $P(0,T) \to \infty$ as $T \to \infty$. Hull and White (1990) add mean reversion, $\mathrm{d}r_t = (\beta(t) - \kappa r_t)\,\mathrm{d}t + \sigma\,\mathrm{d}\widetilde{W}_t$, and keep the exact fit (Exercise 14.5).

14.4 The Vasicek model

Proposition 14.8 (The Vasicek model)**.** *Let $\kappa > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$, $r_0 \in \mathbb{R}$, and let r be the Ornstein–Uhlenbeck process $\mathrm{d}r_t = \kappa(\theta - r_t)\,\mathrm{d}t + \sigma\,\mathrm{d}\widetilde{W}_t$ of Theorem 9.8. Let $\tau = T - t$,*

$$B(\tau) = \frac{1 - e^{-\kappa\tau}}{\kappa}, \qquad R_\infty = \theta - \frac{\sigma^2}{2\kappa^2}, \qquad A(\tau) = R_\infty(B(\tau) - \tau) - \frac{\sigma^2}{4\kappa}B(\tau)^2.$$

(i) $P(t,T) = \exp\big(A(\tau) - B(\tau)r_t\big)$ *for* $0 \le t \le T$.

(ii) For $t < T$, the yield and the forward rate are

$$Y(t,T) = R_\infty + (r_t - R_\infty)\frac{B(\tau)}{\tau} + \frac{\sigma^2 B(\tau)^2}{4\kappa\tau}, \qquad f(t,T) = \mathbb{E}^{\mathbb{Q}}[r_T \mid \mathcal{F}_t] - \frac{\sigma^2}{2}B(\tau)^2;$$

$Y(t,T) \to r_t$ *as* $\tau \downarrow 0$ *and* $Y(t,T) \to R_\infty$ *as* $\tau \to \infty$.

(iii) $\mathrm{d}P(t,T) = r_t P(t,T)\,\mathrm{d}t - \sigma B(T-t)P(t,T)\,\mathrm{d}\widetilde{W}_t$, *and* $D_t P(t,T) = P(0,T)\exp\big(-\sigma\int_0^t B(T-u)\,\mathrm{d}\widetilde{W}_u - \frac{\sigma^2}{2}\int_0^t B(T-u)^2\,\mathrm{d}u\big)$.

(iv) For $t > 0$, r_t and all yields $Y(t,T)$ are normal, and $\mathbb{Q}(r_t < 0) = N\big(-m(t)/\sqrt{v(t)}\big) > 0$ with $m(t) = \theta + (r_0 - \theta)e^{-\kappa t}$ and $v(t) = \sigma^2(1 - e^{-2\kappa t})/(2\kappa)$; as $t \to \infty$ it tends to $N\big(-\theta\sqrt{2\kappa}/\sigma\big)$.

Proof. (i) and (iii) are Theorem 14.5 with $\beta_0 = \kappa\theta$, $\beta_1 = -\kappa$, $\alpha_0 = \sigma^2$, $\alpha_1 = 0$, in the time-homogeneous form (iii) of that theorem. Here $\gamma = \kappa$ and (14.6) gives $B(\tau) = (1 - e^{-\kappa\tau})/\kappa$, the solution of $B' = 1 - \kappa B$. Next, $A(\tau) = -\kappa\theta\int_0^\tau B + \frac{\sigma^2}{2}\int_0^\tau B^2$, with $\int_0^\tau B = (\tau - B(\tau))/\kappa$ and, since $1 - e^{-2\kappa u} = 2\kappa B(u) - \kappa^2 B(u)^2$,

$$\int_0^\tau B(u)^2\,\mathrm{d}u = \frac{1}{\kappa^2}\int_0^\tau (1 - 2e^{-\kappa u} + e^{-2\kappa u})\mathrm{d}u = \frac{\tau - B(\tau) - \frac{\kappa}{2}B(\tau)^2}{\kappa^2};$$

collecting terms gives the formula for A. (ii) $Y = (Br_t - A)/\tau$ is the displayed expression. As $\tau \downarrow 0$, $B(\tau)/\tau \to 1$ and $B(\tau)^2/\tau \to 0$; as $\tau \to \infty$, B tends to $1/\kappa$, so the last two terms vanish. By (i), $f(t,T) = -A'(\tau) + B'(\tau)r_t = \kappa\theta B - \frac{\sigma^2}{2}B^2 + e^{-\kappa\tau}r_t$, and $\kappa\theta B + e^{-\kappa\tau}r_t = \theta + (r_t - \theta)e^{-\kappa\tau} = \mathbb{E}^{\mathbb{Q}}[r_T \mid \mathcal{F}_t]$ by Theorem 9.8(i). (iv) r_t is normal with mean $m(t)$ and variance $v(t)$ by Theorem 9.8(ii), and the yields are affine functions of r_t. □

Reading the formulas. The forward rate is the expected future short rate minus a *convexity correction* $\frac{\sigma^2}{2}B(\tau)^2$, and the long yield R_∞ lies below the long-run mean θ by $\sigma^2/(2\kappa^2)$: a bond price is the expectation of the convex function e^{-x} of the integrated rate, and Jensen's inequality pushes prices up and yields down (Exercise 14.3). The yield curve has three possible shapes (Vasicek, 1977): increasing if $r_t \le R_\infty - \sigma^2/(4\kappa^2)$, decreasing if $r_t \ge R_\infty + \sigma^2/(2\kappa^2) = \theta$, and

humped in between (Figure 14.1; checked numerically in `code/tests/test_rates.py`). With one source of randomness, yields of all maturities are affine in r_t, hence perfectly correlated, with slope $B(\tau)/\tau$ decreasing in τ.

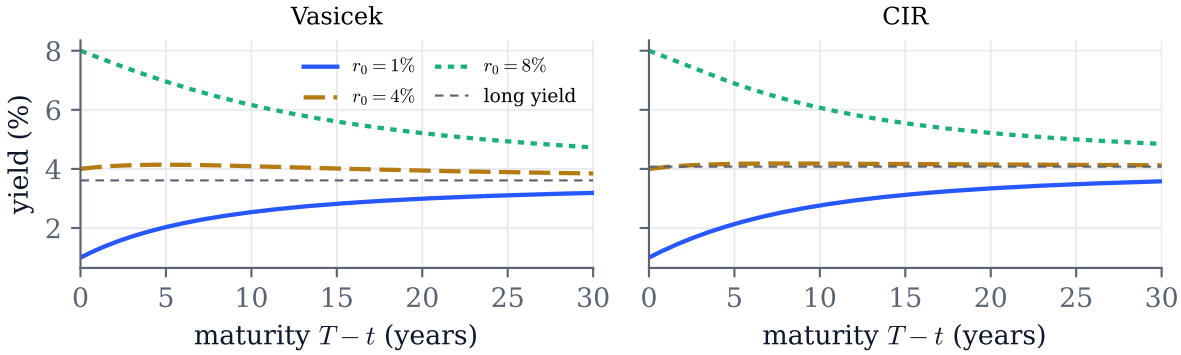

Figure 14.1: Yield curves $\tau \mapsto Y(t, t+\tau)$ from three values of the short rate. Left: Vasicek, $\kappa = 0.15$, $\theta = 5\%$, $\sigma = 2.5\%$, so $R_\infty = 3.61\%$; the curve is increasing for $r_t \le 2.92\%$, decreasing for $r_t \ge 5\%$, and humped in between. Right: CIR with the same κ, θ and the same volatility of rates at $r = \theta$, $\sigma\sqrt{\theta} = 2.5\%$ ($\sigma = 11.2\%$, $\nu = 1.2$), whose long yield is $2\kappa\theta/(\gamma+\kappa) = 4.08\%$.

Negative rates. By Theorem 14.8(iv), the Vasicek rate is negative with positive probability at every time, and the probability is not small when θ is low relative to the stationary standard deviation $\sigma/\sqrt{2\kappa}$ (Figure 14.2). In the 2010s, central bank deposit rates were negative in the euro area, Switzerland, Denmark and Japan, so this is not in itself a defect; but the Gaussian tails put no bound on how negative rates can be. A more serious consequence concerns long bonds: if $\sigma^2 > 2\kappa^2\theta$, then $R_\infty < 0$ and $P(0,T) = e^{-TY(0,T)} \to \infty$ as $T \to \infty$: the price of a promise to pay one dollar at T grows without bound as T increases.

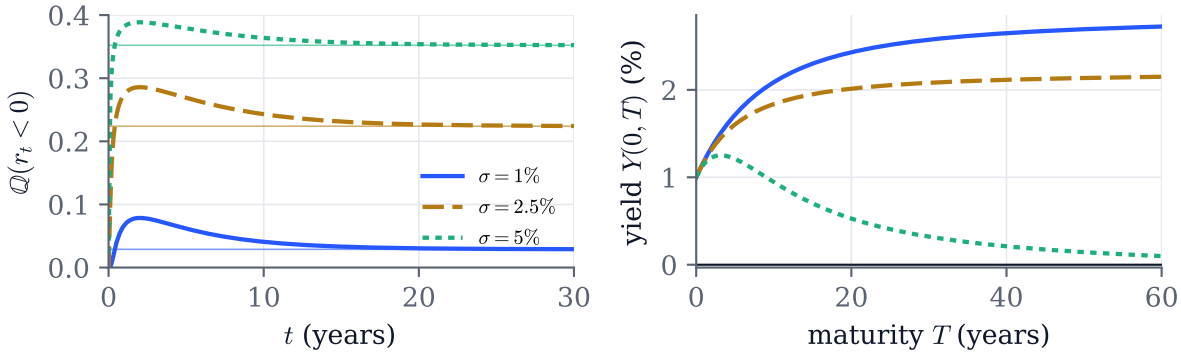

Figure 14.2: Vasicek with $\kappa = 0.2$, $\theta = 3\%$, $r_0 = 1\%$ and three volatilities. Left: $\mathbb{Q}(r_t < 0)$, with its stationary limits 2.9%, 22.4% and 35.2% (thin lines). Right: the yield curve at time 0; the long yields are $R_\infty = 2.875\%$, 2.219% and -0.125%.

Simulation. Monte Carlo bond prices need D_T along each path; in the Vasicek model $(r_{t+h}, \int_t^{t+h} r_u \, du)$ is Gaussian given r_t and can be simulated exactly (Exercise 14.3, Figure 14.3).

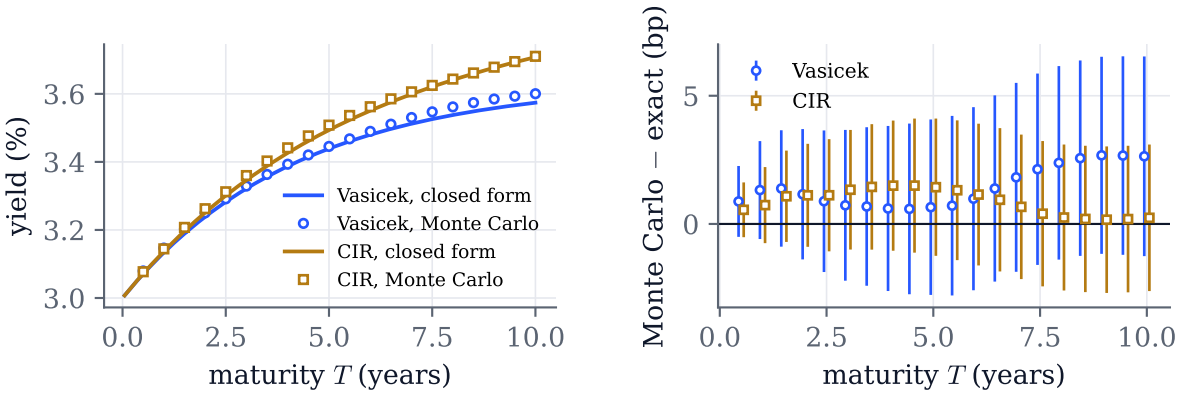

Figure 14.3: Monte Carlo against the closed forms, with $r_0 = 3\%$ and the parameters of Figure 14.1, from 2×10^4 paths on a grid of step 0.05. Vasicek: exact simulation of $(r, \int r)$; CIR: exact simulation of r (Section 14.5) and the trapezoidal rule for $\int r$. Left: yields. Right: the errors of the simulated yields in basis points, with 95% confidence intervals.

▶ Interactive demo: Vasicek and CIR yield curves for any parameters, Monte Carlo bond prices against the closed forms, and the probability of negative rates.

14.5 The Cox–Ingersoll–Ross model

Cox et al. (1985) replace the constant volatility of the Vasicek model by one proportional to $\sqrt{r_t}$: rates are less volatile when they are low, and cannot become negative. In Chapter 9 this *square-root diffusion* appeared as the CEV model with $\gamma = \frac{1}{2}$ (Theorem 9.15); here it has a mean-reverting drift.

Proposition 14.9 (The CIR model)**.** *Let $\kappa, \theta, \sigma > 0$, $r_0 \geq 0$, and consider*

$$\mathrm{d}r_t = \kappa(\theta - r_t)\,\mathrm{d}t + \sigma\sqrt{r_t^+}\,\mathrm{d}\widetilde{W}_t. \tag{14.9}$$

(i) On every interval $[0, T]$, (14.9) has a unique solution, and almost surely $r_t \geq 0$ for all t.

(ii) (Feller condition.) *If $2\kappa\theta \geq \sigma^2$ and $r_0 > 0$, then almost surely $r_t > 0$ for all $t \geq 0$.*

(iii) Let $\gamma = \sqrt{\kappa^2 + 2\sigma^2}$. For $0 \leq t \leq T$, $P(t,T) = \exp\big(A(T-t) - B(T-t)r_t\big)$, where

$$B(\tau) = \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \kappa)(e^{\gamma\tau} - 1) + 2\gamma}, \qquad A(\tau) = \frac{2\kappa\theta}{\sigma^2} \log \frac{2\gamma e^{(\gamma+\kappa)\tau/2}}{(\gamma + \kappa)(e^{\gamma\tau} - 1) + 2\gamma}.$$

(iv) $Y(t,T) \to r_t$ as $T \downarrow t$ and $Y(t,T) \to 2\kappa\theta/(\gamma + \kappa)$ as $T \to \infty$; and $\mathrm{d}P(t,T) = r_t P(t,T)\,\mathrm{d}t - \sigma B(T-t)\sqrt{r_t}\,P(t,T)\,\mathrm{d}\widetilde{W}_t$.

Proof. (i) The drift is Lipschitz, and $|\sqrt{x^+} - \sqrt{y^+}| \leq \sqrt{|x - y|}$ (for $x, y \geq 0$ because $(\sqrt{x} - \sqrt{y})^2 \leq$

$|\sqrt{x}-\sqrt{y}|(\sqrt{x}+\sqrt{y})$; if $y<0\le x$, because $\sqrt{x}\le\sqrt{x-y}$). So pathwise uniqueness holds by the Yamada–Watanabe criterion (Theorem 9.6(b)). For existence, truncate: for $n\ge 1$, $b_n(x)=\kappa(\theta-(x\vee -n)\wedge n)$ and $\sigma_n(x)=\sigma\sqrt{x^+\wedge n}$ are bounded and continuous, equal the coefficients on $[-n,n]$, and satisfy the same Lipschitz and square-root bounds. Their equation has a weak solution (Karatzas and Shreve (1991), Theorem 5.4.22) and pathwise uniqueness (their Proposition 5.2.13), hence a strong solution r^n (their Corollary 5.3.23). Let $\rho_n=\inf\{t:|r^n_t|\ge n\}$. For $m>n$, r^m and r^n solve the n-th equation until one of them leaves $(-n,n)$, and the Yamada–Watanabe argument, applied to the processes stopped at that time, shows that they agree up to it; so $r^m=r^n$ on $[0,\rho_n]$ and $\rho_m\ge\rho_n$. Set $r=r^n$ on $[0,\rho_n]$. By Itô's formula for x^2 stopped at ρ_n, since $2xb(x)+\sigma^2x^+\le C(1+x^2)$, and Gronwall's lemma, $\mathbb{E}^{\mathbb{Q}}[(r^n_{t\wedge\rho_n})^2]\le(r_0^2+Ct)e^{Ct}$ for all n, so $\mathbb{Q}(\rho_n\le t)\le(r_0^2+Ct)e^{Ct}/n^2\to 0$: $\rho_n\to\infty$, and r solves (14.9) on every $[0,T]$. For nonnegativity, let $g(x)=(x^-)^3$, a C^2 function with $g'(x)=-3(x^-)^2$ and $g''(x)=6x^-$. Then $g'(x)\sqrt{x^+}=0$ and $g''(x)x^+=0$ for every x, so by Itô's formula (Theorem 8.13) the stochastic integral and the second-order term vanish, and

$$g(r_t)=g(r_0)-3\kappa\int_0^t(r_s^-)^2(\theta-r_s)\,\mathrm{d}s.$$

Here $g(r_0)=0$, and the integrand is zero when $r_s\ge 0$ and positive when $r_s<0$. So $0\le g(r_t)\le 0$, that is $r_t\ge 0$, for all t, almost surely.

(ii) Fix $t>0$, $0<\varepsilon<\min(r_0,1)$ and $N>r_0$, and let $\tau_\varepsilon=\inf\{s:r_s\le\varepsilon\}$, $\sigma_N=\inf\{s:r_s\ge N\}$ (with $\inf\emptyset=\infty$) and $\rho=\tau_\varepsilon\wedge\sigma_N\wedge t$. On $[0,\rho]$, r takes values in $[\varepsilon,N]$. Apply Itô's formula to a C^2 function that equals $\log$ on $[\varepsilon,\infty)$, stopped at ρ; the stopped stochastic integral is $\int_0^t\mathbf{1}_{\{s\le\rho\}}\sigma r_s^{-1/2}\,\mathrm{d}\widetilde{W}_s$ (Theorem 7.15, as in Theorem 9.15(b)), whose integrand is bounded by $\sigma/\sqrt{\varepsilon}$, so it has mean zero. Hence, as $2\kappa\theta\ge\sigma^2$,

$$\mathbb{E}^{\mathbb{Q}}[\log r_\rho]=\log r_0+\mathbb{E}^{\mathbb{Q}}\int_0^\rho\Big(\frac{\kappa\theta-\sigma^2/2}{r_s}-\kappa\Big)\mathrm{d}s\ge\log r_0-\kappa t.$$

In the same way, with an integrand bounded by $\sigma\sqrt{N}$, $\mathbb{E}^{\mathbb{Q}}[r_\rho]=r_0+\kappa\mathbb{E}^{\mathbb{Q}}\int_0^\rho(\theta-r_s)\,\mathrm{d}s\le r_0+\kappa\theta t$. On the event $E=\{\tau_\varepsilon\le t,\ \tau_\varepsilon<\sigma_N\}$, $r_\rho=\varepsilon$; elsewhere $\log r_\rho\le r_\rho$. Therefore $\log r_0-\kappa t\le\mathbb{Q}(E)\log\varepsilon+\mathbb{E}^{\mathbb{Q}}[r_\rho]$, and

$$\mathbb{Q}(E)\le\frac{r_0+\kappa\theta t+\kappa t-\log r_0}{\log(1/\varepsilon)}.$$

As $N\to\infty$, $\sigma_N>t$ eventually on every path (continuous paths are bounded on $[0,t]$), so the bound holds for $\mathbb{Q}(\tau_\varepsilon\le t)$. If r vanishes before t, then $\tau_\varepsilon\le t$ for every ε; letting $\varepsilon\to 0$, r does not vanish on $[0,t]$, almost surely. Take $t=1,2,\dots$.

(iii) Apply Theorem 14.5 with $\beta_0=\kappa\theta$, $\beta_1=-\kappa$, $\alpha_0=0$, $\alpha_1=\sigma^2$ and $I=[0,\infty)$: by (i), r takes values in I, where $\sigma\sqrt{r^+}=\sqrt{\alpha_0+\alpha_1 r}$. Then $\gamma=\sqrt{\kappa^2+2\sigma^2}$, $\gamma-\beta_1=\gamma+\kappa$, and (14.6) is the formula for B. For A, it suffices that $A(0)=0$ and $A'=-\kappa\theta B$. With $d=(\gamma+\kappa)(e^{\gamma\tau}-1)+2\gamma$,

$$A'(\tau)=\frac{2\kappa\theta}{\sigma^2}\Big(\frac{\gamma+\kappa}{2}-\frac{(\gamma+\kappa)\gamma e^{\gamma\tau}}{d}\Big)=\frac{\kappa\theta(\gamma+\kappa)(\kappa-\gamma)(e^{\gamma\tau}-1)}{\sigma^2 d}=-\kappa\theta B(\tau),$$

since $d-2\gamma e^{\gamma\tau}=(\kappa-\gamma)(e^{\gamma\tau}-1)$ and $\kappa^2-\gamma^2=-2\sigma^2$.

(iv) As $\tau\downarrow 0$, $B(\tau)/\tau\to B'(0)=1$ and $A(\tau)/\tau\to A'(0)=0$. As $\tau\to\infty$, $B(\tau)\to 2/(\gamma+\kappa)$ and $\log d/\tau\to\gamma$, so $A(\tau)/\tau\to\frac{2\kappa\theta}{\sigma^2}\cdot\frac{\kappa-\gamma}{2}$, and $-\frac{\kappa\theta(\kappa-\gamma)}{\sigma^2}=\frac{\kappa\theta(\gamma^2-\kappa^2)}{\sigma^2(\gamma+\kappa)}=\frac{2\kappa\theta}{\gamma+\kappa}$. The dynamics are those of Theorem 14.5(ii). □

The number $\nu = 2\kappa\theta/\sigma^2$ measures the strength of the pull away from zero against the noise. If $\nu < 1$ and $r_0 > 0$, zero is reached with positive probability, by Feller's test (Karatzas and Shreve (1991), Section 5.5.C), and in fact with probability one: Exercise 14.8 derives this from the scale function, granting that the rate leaves every bounded interval in finite time; see Göing-Jaeschke and Yor (2003), who write the CIR process as a time-changed squared Bessel process of dimension 2ν, and Feller (1951) for the classification of the boundary. At zero the drift $\kappa\theta > 0$ pushes the rate back up, and by (i) it never goes below. The stationary law is the gamma law with shape ν and scale $\sigma^2/(2\kappa)$ (Cox et al., 1985); for $\nu \geq 1$ this is Exercise 9.5(d). Its density, proportional to $x^{\nu-1}e^{-2\kappa x/\sigma^2}$, is unbounded near 0 when $\nu < 1$ (Figure 14.4). The mean of r_t is that of the Vasicek model with the same κ and θ (Exercise 14.4).

The Euler scheme of Section 9.7 can make the rate negative, and is usually run with r^+ in both coefficients. But given r_t, $r_{t+h} = cX$ with $c = \sigma^2(1 - e^{-\kappa h})/(4\kappa)$ and X noncentral chi-square with $4\kappa\theta/\sigma^2$ degrees of freedom and noncentrality $r_t e^{-\kappa h}/c$ (Cox et al., 1985): Figures 14.3 and 14.4 use this exact transition.

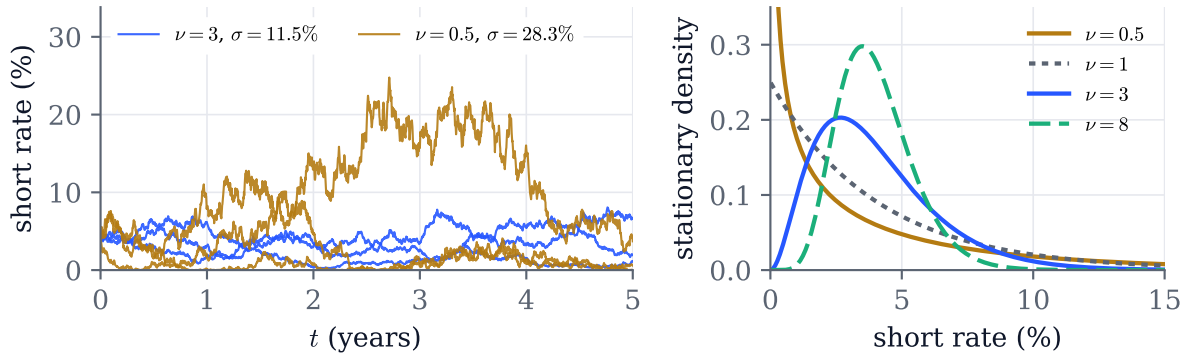

Figure 14.4: The Feller condition, with $\kappa = 0.5$, $\theta = 4\%$, exact simulation. Left: three paths with $\nu = 2\kappa\theta/\sigma^2 = 3$ and three with $\nu = 0.5$, from $r_0 = 4\%$; the latter keep returning to 0. Right: the stationary gamma densities for four values of ν, all with mean θ; for $\nu < 1$ the density is unbounded at 0.

▶ Interactive demo: CIR paths and densities near zero as the Feller ratio $\nu = 2\kappa\theta/\sigma^2$ crosses 1, and a comparison of the full-truncation Euler scheme with exact simulation.

14.6 The forward measure and bond options

The price of a claim V_T in a short-rate model is $\mathbb{E}^{\mathbb{Q}}[D_T V_T]$: the payoff and the discount factor are both random, and correlated. Changing the unit of account from the money market to the bond maturing at T removes the discount factor (Theorem 11.13).

Proposition 14.10 (The forward measure)**.** *Let r be a short-rate model, $T \leq T^*$, and let $\mathbb{Q}^T(E) = \mathbb{E}^{\mathbb{Q}}[D_T \mathbf{1}_E]/P(0,T)$ for $E \in \mathcal{F}_T$: the* T-forward measure.

(i) $\mathbb{Q}^T$ is a probability measure equivalent to $\mathbb{Q}$ on $\mathcal{F}_T$, with density $D_t P(t,T)/P(0,T)$ on

$\mathcal{F}_t$. *For every $\mathcal{F}_T$-measurable V_T with $\mathbb{E}^{\mathbb{Q}}[D_T|V_T|] < \infty$, its price at time $t \le T$ is*

$$\frac{\mathbb{E}^{\mathbb{Q}}[D_T V_T \mid \mathcal{F}_t]}{D_t} = P(t,T)\,\mathbb{E}^{\mathbb{Q}^T}[V_T \mid \mathcal{F}_t],$$

and for $T \le S \le T^*$ the forward bond price $P(t,S)/P(t,T)$ *is a $\mathbb{Q}^T$-martingale on $[0,T]$.*

(ii) In the Vasicek model of Theorem 14.8, let $T > 0$. Then $W_t^T = \widetilde{W}_t + \sigma\int_0^t B(T-u)\,\mathrm{d}u$ is a $\mathbb{Q}^T$-Brownian motion on $[0,T]$; under $\mathbb{Q}^T$, r_T is normal with mean $f(0,T)$ and variance $v(T) = \sigma^2(1-e^{-2\kappa T})/(2\kappa)$, and for $T < S \le T^*$, $\log P(T,S)$ is normal with variance Σ^2 and mean $\log\big(P(0,S)/P(0,T)\big) - \frac{1}{2}\Sigma^2$, where

$$\Sigma = \sigma B(S-T)\sqrt{\frac{1-e^{-2\kappa T}}{2\kappa}}.$$

(iii) (Bond options in the Vasicek model.) *Let $0 < T < S \le T^*$, $K > 0$ and $d_\pm = \big(\log\frac{P(0,S)}{KP(0,T)} \pm \frac{1}{2}\Sigma^2\big)/\Sigma$. The prices at time 0 of the call and the put with strike K and expiry T on the bond maturing at S, with payoffs $(P(T,S)-K)^+$ and $(K-P(T,S))^+$ at T, are*

$$\begin{aligned}\text{ZBC} &= P(0,S)N(d_+) - KP(0,T)N(d_-),\\ \text{ZBP} &= KP(0,T)N(-d_-) - P(0,S)N(-d_+).\end{aligned}$$

Proof. (i) $Z_T = D_T/P(0,T) > 0$ has $\mathbb{E}^{\mathbb{Q}}[Z_T] = 1$, so by Theorem 11.1, with $(\mathbb{Q},\mathbb{Q}^T)$ in place of $(\mathbb{P},\mathbb{Q})$, $\mathbb{Q}^T$ is equivalent to $\mathbb{Q}$ with density $\mathbb{E}^{\mathbb{Q}}[Z_T \mid \mathcal{F}_t] = D_tP(t,T)/P(0,T)$ on $\mathcal{F}_t$, and, as $\mathbb{E}^{\mathbb{Q}^T}|V_T| = \mathbb{E}^{\mathbb{Q}}[D_T|V_T|]/P(0,T) < \infty$, Bayes' formula gives $\mathbb{E}^{\mathbb{Q}^T}[V_T \mid \mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[D_TV_T \mid \mathcal{F}_t]/(D_tP(t,T))$. For $V_T = P(T,S)$, the tower property gives $\mathbb{E}^{\mathbb{Q}}[D_TP(T,S) \mid \mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[D_S \mid \mathcal{F}_t] = D_tP(t,S)$, so $\mathbb{E}^{\mathbb{Q}^T}[P(T,S) \mid \mathcal{F}_t] = P(t,S)/P(t,T)$. This is Theorem 11.13(i)–(ii) for the numeraire $P(\cdot,T)$, without a bound on r.

(ii) By Theorem 14.8(iii), the density process $D_tP(t,T)/P(0,T)$ is the exponential of Theorem 11.2 for $\widetilde{W}$ under $\mathbb{Q}$ with $\Theta_u = \sigma B(T-u)$, deterministic and bounded. By that theorem, applied with $(\mathbb{Q},\mathbb{Q}^T)$ in place of $(\mathbb{P},\mathbb{Q})$, W^T is a $\mathbb{Q}^T$-Brownian motion on $[0,T]$, for which $(\mathcal{F}_t)$ is a filtration. Substituting $\sigma\,\mathrm{d}\widetilde{W}_t = \sigma\,\mathrm{d}W_t^T - \sigma^2B(T-t)\,\mathrm{d}t$,

$$\mathrm{d}r_t = (\kappa\theta - \sigma^2B(T-t) - \kappa r_t)\mathrm{d}t + \sigma\,\mathrm{d}W_t^T,$$

a linear equation with additive noise. By Theorem 9.7(b) under $\mathbb{Q}^T$, r_T is normal with variance $v(T)$ and mean $\theta + (r_0-\theta)e^{-\kappa T} - \sigma^2\int_0^T e^{-\kappa u}B(u)\,\mathrm{d}u = \mathbb{E}^{\mathbb{Q}}[r_T] - \frac{\sigma^2}{2}B(T)^2 = f(0,T)$, because $e^{-\kappa u} = B'(u)$, and by Theorem 14.8(ii). Then $\log P(T,S) = A(S-T) - B(S-T)r_T$ is normal with variance $B(S-T)^2v(T) = \Sigma^2$. By (i), at times 0 and T, $\mathbb{E}^{\mathbb{Q}^T}[P(T,S)] = P(0,S)/P(0,T)$; for a normal variable X with variance Σ^2, $\mathbb{E}[e^X] = e^{\mathbb{E}[X]+\Sigma^2/2}$, which gives the mean.

(iii) The payoffs are bounded by $P(T,S)+K$, and $\mathbb{E}^{\mathbb{Q}}[D_T(P(T,S)+K)] = P(0,S)+KP(0,T) < \infty$. By (i) at $t=0$, the call is worth $P(0,T)\,\mathbb{E}^{\mathbb{Q}^T}[(P(T,S)-K)^+]$. By (ii), $P(T,S) = F_0e^{\Sigma Y - \Sigma^2/2}$ with $F_0 = P(0,S)/P(0,T)$ and Y standard normal under $\mathbb{Q}^T$, and $P(T,S) > K$ if and only if $Y > -d_-$. By Theorem 6.10 with $a = \Sigma$, $\mathbb{E}^{\mathbb{Q}^T}[(P(T,S)-K)^+] = F_0N(d_-+\Sigma) - KN(d_-)$, and $d_-+\Sigma = d_+$. For the put, use $(x-K)^+ - (K-x)^+ = x-K$: ZBC − ZBP $= P(0,T)(F_0-K) = P(0,S) - KP(0,T)$, and $1-N(d) = N(-d)$. □

The formula is due to Jamshidian (1989). The forward measure was used by him and by Geman (1989), and changes of numeraire were formalised by Geman et al. (1995) (Chapter 11). It is Black's formula (Exercise 11.4) for the forward bond price $P(0,S)/P(0,T)$, whose volatility $\sigma e^{-\kappa(T-t)}B(S-T)$ is deterministic, with Σ^2 the integral of its square over $[0,T]$. Caplets are puts on zero-coupon bonds (Exercise 14.6), and options on coupon bonds are portfolios of options on zero-coupon bonds (Exercise 14.7). By (ii), $f(0,T)$ is the expected short rate under $\mathbb{Q}^T$, not under $\mathbb{Q}$; and under $\mathbb{Q}^T$ a Monte Carlo price needs r_T only, not the discount factor.

▶ Interactive demo: bond options by Jamshidian's formula and by Monte Carlo under $\mathbb{Q}$ and $\mathbb{Q}^T$, and $\mathbb{E}^{\mathbb{Q}}[r_T]$ against $f(0,T)=\mathbb{E}^{\mathbb{Q}^T}[r_T]$.

14.7 Beyond short-rate models

A one-factor model fits today's curve only through a time-dependent drift (Theorem 14.7) and moves all yields together. Heath et al. (1992) model the whole forward curve instead.

Remark 14.11 (The HJM drift condition). Suppose that, for each $T \le T^*$, the forward rate evolves under $\mathbb{Q}$ as $f(t,T) = f(0,T) + \int_0^t \alpha(u,T)\,\mathrm{d}u + \int_0^t \sigma_f(u,T)\,\mathrm{d}\widetilde{W}_u$, $t \le T$, with $r_t = f(t,t)$ and α, σ_f regular enough to exchange the integrals in u and in T (a stochastic Fubini theorem). Writing $\log P(t,T) = -\int_t^T f(t,u)\,\mathrm{d}u$ and $\Sigma(t,T) = \int_t^T \sigma_f(t,u)\,\mathrm{d}u$, one finds that $\mathrm{d}(D_tP(t,T))/(D_tP(t,T))$ has drift $\frac{1}{2}\Sigma(t,T)^2 - \int_t^T \alpha(t,u)\,\mathrm{d}u$ and diffusion coefficient $-\Sigma(t,T)$. If the drift vanishes for every T, the discounted bonds of all maturities are local martingales; and, differentiating in T, it vanishes for every T if and only if the *HJM drift condition* holds:

$$\alpha(t,T) = \sigma_f(t,T)\int_t^T \sigma_f(t,u)\,\mathrm{d}u.$$

Under $\mathbb{Q}$ the drift of the forward curve is thus determined by its volatilities; only these are modelled. For example, in the Vasicek model, $f(t,T) = \theta + e^{-\kappa(T-t)}(r_t-\theta) - \frac{\sigma^2}{2}B(T-t)^2$ by Theorem 14.8(ii), so $\sigma_f(t,T) = \sigma e^{-\kappa(T-t)}$ and the condition gives $\alpha(t,T) = \sigma^2 e^{-\kappa(T-t)}B(T-t)$, which agrees with the drift computed from the dynamics of r; for Ho–Lee, $\sigma_f = \sigma$ and $\alpha(t,T) = \sigma^2(T-t)$. A deterministic σ_f gives a Gaussian model; in general the short rate of an HJM model need not be a Markov process. See Shreve (2004b), Sections 10.3.3–10.3.4, and Brigo and Mercurio (2006), Chapter 5.

Further reading

Shreve (2004b), Sections 6.5, 9.4 and 10.2–10.4, and Brigo and Mercurio (2006), Chapters 3 and 5, cover this chapter and much more. Cox et al. (1985) derive their model from a general equilibrium.

Chapter 15

Implied volatility and the smile

The Black–Scholes model prices all options on a stock with one volatility; inverting it on market prices gives a different volatility for each strike and maturity, the *implied volatility surface*. We start from what is wrong with the model, characterize the surfaces free of static arbitrage, fit them with local volatility (Dupire's formula) and stochastic volatility (the Heston model), explain why exotics must be priced in one model consistent with all the vanillas, and price variance swaps and the VIX from the smile without a model.

Setting. Prices are at time 0. The interest rate r and the dividend yield q are constants (Theorem 11.14), $C(K,T)$ and $P(K,T)$ are the prices of the call and the put with strike $K>0$ and maturity $T>0$, and $F_T = S_0 e^{(r-q)T}$ is the forward price. The prices *come from a pricing measure* $\mathbb{Q}$ if $S_T \ge 0$, $\mathbb{E}^{\mathbb{Q}}[S_T] = F_T$, $C(K,T) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S_T-K)^+]$ and $P(K,T) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(K-S_T)^+]$, as in Theorems 3.5 and 11.6. The Black–Scholes call price with dividend yield q is (Theorem 11.14)

$$C^{\mathrm{BS}}(K,T,\sigma) = S_0 e^{-qT} N(d_+) - K e^{-rT} N(d_-), \qquad d_\pm = \frac{\log(F_T/K)}{\sigma\sqrt{T}} \pm \frac{\sigma\sqrt{T}}{2}. \tag{15.1}$$

Since $S_0 e^{-qT} e^k = K e^{-rT}$ for the *log-moneyness* $k = \log(K/F_T)$, it depends only on k and the *total volatility* $s = \sigma\sqrt{T}$: $C^{\mathrm{BS}} = S_0 e^{-qT}\, b(k,s)$ with

$$b(k,s) = N\Big(-\frac{k}{s} + \frac{s}{2}\Big) - e^k N\Big(-\frac{k}{s} - \frac{s}{2}\Big), \qquad s>0. \tag{15.2}$$

We call $c = C/(S_0 e^{-qT})$ the *normalized* price. With $d_\pm = -k/s \pm s/2$, we have $d_-^2 - d_+^2 = 2k$, hence $e^k N'(d_-) = N'(d_+)$ (Theorem 6.14), and $\partial_k b = -e^k N(d_-)$.

15.1 The limits of the Black–Scholes model

In the Black–Scholes model, daily log returns are independent and normal. Returns of stocks and indices share three *stylised facts* that contradict this (Cont, 2001). *Heavy tails*: the kurtosis $\mathbb{E}[(y-\mathbb{E}y)^4]/\operatorname{Var}(y)^2$ of daily log returns y is well above the normal value 3. *Volatility clustering*: returns are nearly uncorrelated, but their absolute values are positively correlated over weeks and months. *Leverage effect*: returns are negatively correlated with future volatility. Figure 15.1

shows the three on a series simulated from the Heston model of Section 15.5 under the real-world measure. The first two come from a random variance alone; the third also needs a correlation between returns and variance.

Proposition 15.1 (Mixtures of normals). *Let $m \in \mathbb{R}$, let $Z_1, \dots, Z_n$ be independent $\mathcal{N}(0,1)$ variables, independent of a vector $(V_1, \dots, V_n)$ of nonnegative variables with $\mathbb{E}[V_i^2] < \infty$ and $\mathbb{E}[V_i] > 0$, and let $y_i = m + \sqrt{V_i}\, Z_i$.*

(i) The kurtosis of y_i is $3\,\mathbb{E}[V_i^2]/\mathbb{E}[V_i]^2 \geq 3$, with equality if and only if V_i is almost surely constant.

(ii) For $i \neq j$, $\mathrm{Cov}(y_i, y_j) = 0$, while $\mathrm{Cov}\left((y_i - m)^2, (y_j - m)^2\right) = \mathrm{Cov}(V_i, V_j)$.

(iii) For all i, j and every bounded Borel f, $\mathrm{Cov}\left(y_i, f(V_j)\right) = 0$.

Proof. By independence, $\mathbb{E}[f(V)h(Z)] = \mathbb{E}[f(V)]\,\mathbb{E}[h(Z)]$ when both factors are integrable. Hence $\mathbb{E}[(y_i - m)^2] = \mathbb{E}[V_i]$, $\mathbb{E}[(y_i - m)^4] = 3\mathbb{E}[V_i^2]$, and $\mathbb{E}[V_i^2] \geq \mathbb{E}[V_i]^2$ with equality if and only if $\mathrm{Var}(V_i) = 0$. For $i \neq j$, $\mathbb{E}[(y_i - m)(y_j - m)] = \mathbb{E}[\sqrt{V_i V_j}]\,\mathbb{E}[Z_i Z_j] = 0$ and $\mathbb{E}[(y_i - m)^2 (y_j - m)^2] = \mathbb{E}[V_i V_j]$. Finally $\mathbb{E}[(y_i - m) f(V_j)] = \mathbb{E}[\sqrt{V_i} f(V_j)]\,\mathbb{E}[Z_i] = 0$. □

Returns can be uncorrelated without being independent: a persistent variance produces clustering, a random one fat tails. By (iii), shocks independent of the variances give no leverage effect; in the Heston model it comes from a correlation $\rho < 0$ between the two shocks (Exercise 15.5).

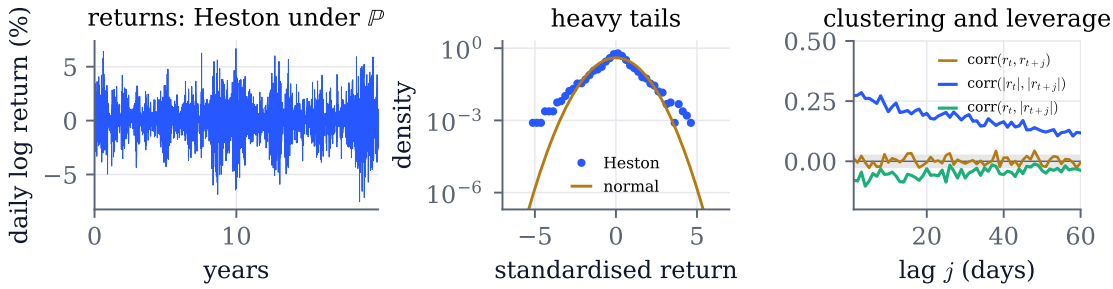

Figure 15.1: Twenty years of daily log returns simulated from the Heston model under $\mathbb{P}$ ($\mu = 7\%$, $v_0 = \theta = 0.04$, $\kappa = 4$, $\xi = 0.6$, $\rho = -0.7$). Middle: the density against the normal one; the kurtosis is 5.3 (2.9 for a normal sample of the same size). Right: autocorrelations of returns and of absolute returns (0.27 at one day, 0.12 at sixty days), and the leverage correlation between returns and later absolute returns (about -0.08); the band is $\pm 2/\sqrt{n}$.

▶ Interactive demo: tails, clustering and leverage as functions of ξ and ρ.

15.2 Implied volatility

Proposition 15.2 (Existence, uniqueness and bounds). *Let $S_0, K, T > 0$ and $r, q \in \mathbb{R}$.*

(i) The map $\sigma \mapsto C^{\mathrm{BS}}(K, T, \sigma)$ is continuous and strictly increasing on $(0, \infty)$, with derivative $S_0 e^{-qT} N'(d_+)\sqrt{T} > 0$, and tends to $(S_0 e^{-qT} - K e^{-rT})^+$ as $\sigma \downarrow 0$ and to $S_0 e^{-qT}$ as

$\sigma \to \infty$. Hence $C^{\mathrm{BS}}(K,T,\sigma) = C$ has a solution $\sigma > 0$ if and only if

$$\left(S_0 e^{-qT} - K e^{-rT}\right)^+ < C < S_0 e^{-qT}, \tag{15.3}$$

and the solution is then unique.

(ii) If $C = C(K,T)$ comes from a pricing measure $\mathbb{Q}$, then C lies in the closed interval of (15.3), and in the open interval as soon as $\mathbb{Q}(S_T > K) > 0$ and $\mathbb{Q}(S_T < K) > 0$.

(iii) Call and put prices that satisfy parity, $C - P = S_0 e^{-qT} - K e^{-rT}$, have the same solution: $P = C^{\mathrm{BS}}(K,T,\sigma) - S_0 e^{-qT} + K e^{-rT}$ if and only if $C = C^{\mathrm{BS}}(K,T,\sigma)$.

(iv) With $c = C/(S_0 e^{-qT})$ and $k = \log(K/F_T)$, the solution satisfies

$$\sigma\sqrt{T} \le -2N^{-1}\Big(\frac{1-c}{1+e^k}\Big), \qquad \text{with equality when } k = 0. \tag{15.4}$$

Proof. (i) Work with $b(k,s)$, $s = \sigma\sqrt{T}$. Differentiating (15.2), with $\partial_s d_\pm = k/s^2 \pm \frac{1}{2}$ and $e^k N'(d_-) = N'(d_+)$, gives $\partial_s b = N'(d_+) > 0$ (the vega of Theorem 6.14, with $S_0 e^{-qT}$ in place of S_0). As $s \downarrow 0$, $d_\pm \to +\infty$ if $k < 0$, $-\infty$ if $k > 0$, and 0 if $k = 0$, so $b \to (1 - e^k)^+$; as $s \to \infty$, $d_\pm \to \pm\infty$ and $b \to 1$. Conclude by the intermediate value theorem.

(ii) By Jensen's inequality, $\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+] \ge (F_T - K)^+$, and $(S_T - K)^+ \le S_T$; multiply by e^{-rT}, with $e^{-rT} F_T = S_0 e^{-qT}$. If $F_T \ge K$, the lower gap $\mathbb{E}^{\mathbb{Q}}[(K - S_T)^+]$ is positive when $\mathbb{Q}(S_T < K) > 0$; if $F_T < K$, it is $\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+]$, positive when $\mathbb{Q}(S_T > K) > 0$. The upper gap is $\mathbb{E}^{\mathbb{Q}}[\min(S_T, K)] \ge K\,\mathbb{Q}(S_T > K)$. (iii) The Black–Scholes prices satisfy parity (Theorem 11.14).

(iv) Fix $s > 0$ and let $\beta(k) = 1 - b(k,s) - (1 + e^k)N(-s/2)$. Then $\beta(0) = 1 - (2N(s/2) - 1) - 2N(-s/2) = 0$ and $\beta'(k) = e^k(N(-k/s - s/2) - N(-s/2))$, positive for $k < 0$ and negative for $k > 0$. So $b(k,s) \ge 1 - (1 + e^k)N(-s/2)$, with equality only at $k = 0$; at the solution $c = b(k,s)$, which rearranges to (15.4). □

Definition 15.3 (Implied volatility)**.** Let C be the price of the call with strike K and maturity T, and suppose that it satisfies (15.3). Its *implied volatility* $\sigma_{\mathrm{imp}}(K,T)$ is the unique $\sigma > 0$ with $C^{\mathrm{BS}}(K,T,\sigma) = C$ (Theorem 15.2). The *total implied variance* is $w(k,T) = \sigma_{\mathrm{imp}}(K,T)^2 T$, as a function of $k = \log(K/F_T)$ and T. The map $(K,T) \mapsto \sigma_{\mathrm{imp}}(K,T)$ is the *implied volatility surface*, and $K \mapsto \sigma_{\mathrm{imp}}(K,T)$ the *smile* at maturity T.

By (ii) and (iii), prices from a pricing measure under which S_T has a positive density have an implied volatility, the same for calls and puts. In Black and Scholes (1973)'s model the surface is flat; for equity indices the smile is a *skew*, decreasing in the strike, more steeply at short maturities (Gatheral (2006), Chapter 1). Out-of-the-money puts insure against the large falls of Section 15.1, and are priced accordingly.

Remark 15.4 (Computing implied volatility). At $k = 0$, $\sigma_{\mathrm{imp}} = \frac{2}{\sqrt{T}} N^{-1}\big(\frac{1+c}{2}\big)$ by (15.4), the exact form of the rule of Brenner and Subrahmanyam (1988) (Theorem 6.16): an at-the-money 3-month option priced 36 on a spot of 1000 ($r = q = 0$) has $\sigma_{\mathrm{imp}} = 18.05\%$ (18% by the rule). Otherwise the bound (15.4) (Tehranchi (2016), Proposition 4.3) brackets the root, and Newton's method for $\log b(k,s) = \log c$, safeguarded by bisection, converges for every admissible price (`code/src/ieor222/volatility.py`). Since $\delta\sigma \approx \delta C / \partial_\sigma C^{\mathrm{BS}}$, price errors are amplified where the vega is small: in the wings and at short maturities.

Remark 15.5 (Bachelier versus Black–Scholes). In Bachelier (1900)'s model $S_T = F_T + \sigma_{\mathrm{N}}\sqrt{T}\,Z$, $Z \sim \mathcal{N}(0,1)$, and the call costs $e^{-rT}\big[(F_T-K)N(d) + \sigma_{\mathrm{N}}\sqrt{T}N'(d)\big]$, $d = (F_T-K)/(\sigma_{\mathrm{N}}\sqrt{T})$; every price in (15.3) has a *normal implied volatility* σ_{N}. Implied volatility belongs to a price and a model: at the forward $\sigma_{\mathrm{N}} \approx F_T\sigma_{\mathrm{imp}}$, and a flat normal smile is a lognormal skew (Exercise 15.2).

15.3 Static arbitrage in implied volatility

Call prices must be convex in the strike (Theorem 3.12) and, by Jensen's inequality, nondecreasing in the maturity when the strike moves with the forward. For the total implied variance these conditions become an inequality in k and a monotonicity in T. For a function w of k, let

$$g(k) = \Big(1 - \frac{kw'(k)}{2w(k)}\Big)^2 - \frac{w'(k)^2}{4}\Big(\frac{1}{w(k)} + \frac{1}{4}\Big) + \frac{w''(k)}{2}. \tag{15.5}$$

Proposition 15.6 (Static arbitrage in total implied variance)**.** *Let $w : \mathbb{R}\times(0,T^*] \to (0,\infty)$ and $C(K,T) = S_0e^{-qT}\,b(k,\sqrt{w(k,T)})$, $k = \log(K/F_T)$: the prices whose total implied variance is w.*

(i) (Butterflies.) *Fix T, and let $w = w(\cdot,T)$ be twice continuously differentiable. Then $C(\cdot,T)$ is twice continuously differentiable on $(0,\infty)$ and, with $d_\pm(k) = -k/\sqrt{w(k)} \pm \sqrt{w(k)}/2$,*

$$\partial_{KK}C(K,T) = e^{-rT}\,\frac{N'(d_-(k))}{K\sqrt{w(k)}}\,g(k). \tag{15.6}$$

There is a probability measure μ on $[0,\infty)$ with $\int x\,\mu(\mathrm{d}x) = F_T$ and $C(K,T) = e^{-rT}\int(x-K)^+\mu(\mathrm{d}x)$ for all $K>0$ if and only if $g \ge 0$ on $\mathbb{R}$ and $d_+(k) \to -\infty$ as $k\to\infty$.

(ii) (Calendar spreads.) *If the prices come from a pricing measure $\mathbb{Q}$ under which $(e^{-(r-q)t}S_t)_{t\le T^*}$ is a martingale, then $T \mapsto w(k,T)$ is nondecreasing for every k.*

Proof. (i) Let $B(k,v) = b(k,\sqrt{v})$ and $\psi = \partial_v B = N'(d_+)/(2\sqrt{v})$. From $d_\pm = -k/\sqrt{v} \pm \sqrt{v}/2$, $\partial_k b = -e^k N(d_-)$ and $e^k N'(d_-) = N'(d_+)$, a direct computation gives

$$\partial_{kk}B - \partial_k B = 2\psi, \qquad 2\partial_{kv}B - \partial_v B = -\frac{2k}{v}\,\psi, \qquad \partial_{vv}B = \Big(\frac{k^2}{2v^2} - \frac{1}{8} - \frac{1}{2v}\Big)\psi;$$

for the last, $\log\psi = -\frac12 d_+^2 - \frac12\log v + \text{const}$ and $d_+\partial_v d_+ = \frac18 - \frac{k^2}{2v^2}$. Let $\gamma(k) = B(k,w(k))$. As $K^2\partial_{KK} = \partial_{kk} - \partial_k$,

$$K^2\partial_{KK}\gamma = \gamma'' - \gamma' = (\partial_{kk}B - \partial_k B) + (2\partial_{kv}B - \partial_v B)w' + \partial_{vv}B\,w'^2 + \partial_v B\,w'' = 2\psi\,g,$$

and $S_0e^{-qT}N'(d_+) = Ke^{-rT}N'(d_-)$ gives (15.6), a continuous function of K.

If μ exists, $C(\cdot,T)$ is convex, so $g\ge 0$, and $C(K,T)\to 0$ as $K\to\infty$ by dominated convergence. If $d_+(k_n) \ge -M$ along some $k_n\to\infty$, then $\sqrt{w(k_n)}\to\infty$ (otherwise $d_+(k_n) \le -k_n/A + A/2 \to -\infty$ along a subsequence with $\sqrt{w(k_n)} \le A$) and $d_-(k_n) \le -\sqrt{w(k_n)}/2 \to -\infty$, so, with $N(x) \le N'(x)/|x|$ for $x<0$,

$$b(k_n,\sqrt{w(k_n)}) \ge N(-M) - e^{k_n}N(d_-) \ge N(-M) - \frac{N'(d_+)}{|d_-|} \longrightarrow N(-M) > 0,$$

against $C \to 0$. Conversely, let $g \geq 0$, $d_+ \to -\infty$ and $\gamma_T(K) = e^{rT}C(K,T)$. Then γ_T is convex, $(F_T - K)^+ < \gamma_T(K) < F_T$ by Theorem 15.2(i), and $\gamma_T(K) \leq F_T N(d_+) \to 0$ as $K \to \infty$. A convex function with a finite limit at infinity is nonincreasing, and $\gamma_T' \geq -1$: if $\gamma_T'(K_1) < -1$, then $\gamma_T' < -1$ on $(0, K_1)$ and $F_T - \gamma_T(K_1) = \int_0^{K_1}(-\gamma_T') > K_1$, against $\gamma_T(K_1) > F_T - K_1$. Let μ be the probability measure on $[0,\infty)$ with $\mu((x,\infty)) = -\gamma_T'(x)$ for $x > 0$ (nonincreasing, continuous, with limit 0 at infinity since $|\gamma_T'(x)| \leq 2\gamma_T(x/2)/x$), and atom $1 + \gamma_T'(0+)$ at 0. By Tonelli, $\int (y-K)^+ \mu(\mathrm{d}y) = \int_K^\infty \mu((x,\infty))\,\mathrm{d}x = \gamma_T(K)$ for $K > 0$, and $\int y\,\mu(\mathrm{d}y) = \gamma_T(0+) = F_T$.

(ii) $X_t = e^{-(r-q)t}S_t/S_0 = S_t/F_t$ is a martingale. For $T_1 < T_2$, conditional Jensen gives $\mathbb{E}^{\mathbb{Q}}[(X_{T_2} - e^k)^+ \mid \mathcal{F}_{T_1}] \geq (X_{T_1} - e^k)^+$, so $\mathbb{E}^{\mathbb{Q}}[(X_T - e^k)^+] = e^{rT}C(F_T e^k, T)/F_T = b(k, \sqrt{w(k,T)})$ is nondecreasing in T, and b is strictly increasing in its second argument. □

The comparison in (ii) is at fixed log-moneyness. The condition $g \geq 0$ is *Durrleman's condition*, and the conditions of (i) and (ii) are also sufficient: if every slice satisfies (i), $w(k,\cdot)$ is nondecreasing and $w(k,T) \to 0$ as $T \downarrow 0$, the prices come from a martingale (Roper (2010), Theorem 2.9, for $r = q = 0$; apply it to S_t/F_t in general). The limit in (i) bounds the wings: $w(k) < 2k$ for large k, and $w(k) < 2|k|$ for $k \ll 0$ if $\mathbb{Q}(S_T = 0) = 0$ (Exercise 15.3); in general $\limsup_{|k|\to\infty} w(k)/|k| \leq 2$, part of Lee (2004)'s moment formula.

Quotes exist for finitely many strikes and maturities, and a surface is fitted to them. The *raw SVI* slice $w(k) = a + b(\rho(k-m) + \sqrt{(k-m)^2 + \varsigma^2})$ has linear wings, as the moment formula suggests, but can fit quotes well and still violate (i): in an example of A. Vogt reported by Gatheral and Jacquier (2014), (15.6) is negative for $k \in [0.64, 1.26]$, so butterflies centred there have negative prices (Figure 15.2). The same authors give the surface parameterization SSVI, $w(k,T) = \frac{\theta_T}{2}(1 + \rho\varphi(\theta_T)k + \sqrt{(\varphi(\theta_T)k + \rho)^2 + 1 - \rho^2})$, with explicit conditions on θ and φ that exclude both arbitrages (their Theorems 4.1 and 4.2). Chapter 17 builds arbitrage-free surfaces directly from quotes.

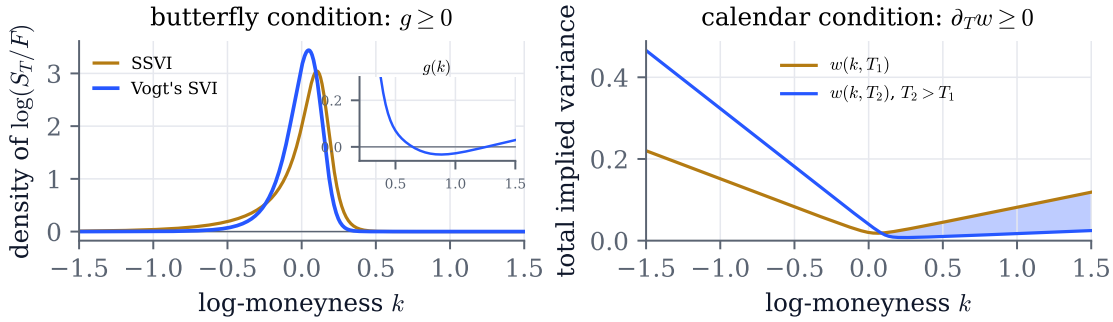

Figure 15.2: Left: the density $N'(d_-)g/\sqrt{w}$ of $\log(S_T/F_T)$ (see (15.6)) for an SSVI slice ($\theta = 0.04$, $\rho = -0.7$, $\varphi(\theta) = \theta^{-1/2}$) and for raw SVI with Vogt's parameters $(a,b,\rho,m,\varsigma) = (-0.041, 0.1331, 0.306, 0.3586, 0.4153)$, $T = 1$, for which $g < 0$ on $[0.64, 1.26]$ (inset); there the density is negative, by less than 2×10^{-4}. Right: SSVI slices with $\varphi(\theta) = 1.5\,\theta^{-1/2}$, and $(\theta,\rho) = (0.02, -0.3)$ at T_1, $(0.04, -0.9)$ at $T_2 > T_1$; where they cross, there is a calendar arbitrage.

▶ Interactive demo: butterfly and calendar checks on a parametric surface.

15.4 Local volatility and Dupire's formula

In the local volatility models of Theorem 9.13 the volatility is a function $\sigma(t, S_t)$. Call prices for all strikes and maturities determine it: Dupire (1994) observed that the forward equation of Theorem 12.8, integrated twice in the strike, is an equation for the call prices, which can be solved for σ.

Theorem 15.7 (Dupire's formula)**.** *Let $r, q \in \mathbb{R}$, $T^* > 0$, and let $\sigma : [0, T^*] \times (0, \infty) \to [0, \infty)$ be continuous, with $\sigma \le M$ and $|\sigma(t,x) - \sigma(t,x')| \le \ell\,|\log x - \log x'|$ for constants M, ℓ, and such that $a = \sigma^2$ has continuous partial derivatives $\partial_x a$, $\partial_{xx} a$. Under $\mathbb{Q}$, let S be the positive solution of $\mathrm{d}S_t = (r-q)S_t\,\mathrm{d}t + \sigma(t,S_t)S_t\,\mathrm{d}\widetilde{W}_t$, $S_0 > 0$, and suppose that $Y = \log S$ has a transition density p_Y with $(T,\eta) \mapsto p_Y(0, \log S_0; T, \eta)$ in $C^{1,2}((0,T^*]\times\mathbb{R})$. Then $C(K,T) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+]$ is in $C^{1,2}((0,T^*]\times(0,\infty))$ and*

$$\partial_T C(K,T) = \frac{\sigma(T,K)^2 K^2}{2}\,\partial_{KK} C(K,T) - (r-q)K\,\partial_K C(K,T) - q\,C(K,T). \tag{15.7}$$

In particular, wherever $\partial_{KK} C(K,T) > 0$,

$$\sigma(T,K)^2 = \frac{2(\partial_T C + (r-q)K\,\partial_K C + qC)}{K^2\,\partial_{KK} C}. \tag{15.8}$$

Proof. The forward equation for the price. Y solves $\mathrm{d}Y_t = \tilde{b}(t,Y_t)\,\mathrm{d}t + \tilde{\sigma}(t,Y_t)\,\mathrm{d}\widetilde{W}_t$ with $\tilde{\sigma}(t,y) = \sigma(t,e^y)$ and $\tilde{b} = r - q - \tilde{\sigma}^2/2$, which are bounded and Lipschitz in y, so it has a unique solution (Theorem 9.5), and $S = e^Y$ by Itô's formula. As in Chapter 9, S also solves an equation on $\mathbb{R}$ with Lipschitz coefficients (extend $x\sigma(t,x)$ by 0 for $x \le 0$), so $\mathbb{E}^{\mathbb{Q}}[\sup_{t\le T^*} S_t^2] < \infty$. By Theorem 12.8(ii), $\varrho(T,\eta) = p_Y(0,\log S_0; T, \eta)$ satisfies $\partial_T \varrho = \frac{1}{2}\partial_{\eta\eta}(\tilde{a}\varrho) - \partial_\eta(\tilde{b}\varrho)$, $\tilde{a} = \tilde{\sigma}^2$. The density of S_T is $p(T,x) = \varrho(T, \log x)/x$; with $\nu = r - q$ and $\partial_\eta = x\partial_x$, $x\,\partial_T p = \frac{1}{2}x\partial_x(x\partial_x(axp)) - x\partial_x(\nu x p) + \frac{1}{2}x\partial_x(axp)$, and $x\partial_x(axp) + axp = \partial_x(ax^2 p)$, so

$$\partial_T p = \tfrac{1}{2}\partial_{xx}(ax^2 p) - \partial_x(\nu x p) \qquad \text{on } (0,T^*]\times(0,\infty). \tag{15.9}$$

Integrating twice. Let $\phi(T,K) = \mathbb{E}^{\mathbb{Q}}[(S_T - K)^+]$. For $0 < T_1 < T_2 \le T^*$ and $0 < K < R$, integrate (15.9) against $x - K$ over $[T_1, T_2]\times[K, R]$, where all functions are continuous, exchange the integrals and integrate by parts in x:

$$\int_K^R (x-K)(p(T_2,x) - p(T_1,x))\mathrm{d}x = \int_{T_1}^{T_2}\Big(\frac{a(t,K)K^2 p(t,K)}{2} + \nu\int_K^R x p(t,x)\,\mathrm{d}x\Big)\mathrm{d}t + \alpha_R - \beta_R K,$$

with $\beta_R = \int_{T_1}^{T_2}(\frac{1}{2}\partial_x(ax^2p) - \nu x p)(t,R)\,\mathrm{d}t$ and $\alpha_R = R\beta_R - \frac{1}{2}\int_{T_1}^{T_2}(ax^2p)(t,R)\,\mathrm{d}t$. As $R\to\infty$, by monotone convergence the left side tends to $\phi(T_2,K) - \phi(T_1,K)$ and the inner integral to $\mathbb{E}^{\mathbb{Q}}[S_t \mathbf{1}_{\{S_t > K\}}] \le F_t$. So $\alpha_R - \beta_R K$ converges for every K; with two values of K, $\beta_R \to \beta$ and $\alpha_R \to \alpha$:

$$\phi(T_2,K) - \phi(T_1,K) = \int_{T_1}^{T_2}\Big(\frac{a(t,K)K^2 p(t,K)}{2} + \nu\,\mathbb{E}^{\mathbb{Q}}\big[S_t\mathbf{1}_{\{S_t>K\}}\big]\Big)\mathrm{d}t + \alpha - \beta K. \tag{15.10}$$

As $K\to\infty$, the left side and the term in ν tend to 0 (dominated convergence, with $S_t \le \sup_{u\le T^*} S_u$), so $h(K) = \frac{1}{2}\int_{T_1}^{T_2} a(t,K)K^2 p(t,K)\,\mathrm{d}t = \beta K - \alpha + o(1)$. But $h \ge 0$ and, by Tonelli, $\int_0^\infty h\,\mathrm{d}K \le \frac{M^2}{2}\int_{T_1}^{T_2}\mathbb{E}^{\mathbb{Q}}[S_t^2]\,\mathrm{d}t < \infty$; this forces $\beta = 0$, then $\alpha = 0$.

Differentiating. The integrand of (15.10) is continuous in t, since p is and since S has continuous paths and $\mathbb{Q}(S_t = K) = 0$ (dominated convergence again). So $\partial_T \phi = \frac{1}{2} a K^2 p + \nu\, \mathbb{E}^{\mathbb{Q}}[S_T \mathbf{1}_{\{S_T > K\}}]$. By Theorem 3.18 (with $B = 1$), $\partial_K \phi = -\mathbb{Q}(S_T > K)$ and $\partial_{KK} \phi = p(T, K)$, so $\mathbb{E}^{\mathbb{Q}}[S_T \mathbf{1}_{\{S_T > K\}}] = \phi - K \partial_K \phi$ and $\partial_T \phi = \frac{1}{2} \sigma(T,K)^2 K^2 \partial_{KK} \phi + \nu(\phi - K \partial_K \phi)$; with $C = e^{-rT} \phi$ this is (15.7). These derivatives are jointly continuous by the same arguments. □

The numerator of (15.8) prices a calendar spread per unit of time (adjusted for carry), the denominator a butterfly: the local variance is positive and finite when the conditions of Section 15.3 hold strictly. In total implied variance,

$$\sigma(T, F_T e^k)^2 = \frac{\partial_T w(k,T)}{g(k,T)}, \tag{15.11}$$

with g from (15.5) on the slice $w(\cdot, T)$ (Exercise 15.6; Gatheral (2006), Chapter 1). The hypothesis on the density is that of Theorem 12.8(ii); for σ bounded away from 0 and with enough bounded Hölder derivatives, it follows from the adjoint equation of Friedman (1964), Chapter 1, Section 8.

Remark 15.8 (Using the formula). Read backwards, (15.8) defines σ from call prices. If the model with this σ satisfies the hypotheses of the theorem, its prices and the given ones solve (15.7) with the limit $(S_0 - K)^+$ as $T \downarrow 0$, and it reproduces the calls when that solution is unique. As it differentiates twice, the formula is applied to a smooth arbitrage-free fit, not to raw quotes. Finally, vanilla prices do not identify the dynamics: if $\mathrm{d}S_t = (r-q) S_t \,\mathrm{d}t + \sqrt{v_t} S_t \,\mathrm{d}\widetilde{W}_t$, the local volatility equation with $\sigma(t,x)^2 = \mathbb{E}^{\mathbb{Q}}[v_t \mid S_t = x]$ has a weak solution with the same one-dimensional marginals, hence the same call prices (Gyöngy (1986), for bounded, uniformly non-degenerate coefficients; G. Brunick and S. Shreve, *Ann. Appl. Probab.*, 2013, under an integrability condition, without uniqueness).

15.5 Stochastic volatility and the Heston model

Random volatility alone produces a smile. If the variance v is independent of the shocks of the price, then given its path $\log(S_T/F_T)$ is Gaussian with variance $V = \int_0^T v_t \,\mathrm{d}t$ and mean $-V/2$ (Hull and White, 1987): $S_T = F_T \exp(\sqrt{V} Z - V/2)$ with $Z \sim \mathcal{N}(0,1)$ independent of V.

Proposition 15.9 (Random variance makes a smile). *Let $V \geq 0$, not almost surely 0, with $\mathbb{E}[V] < \infty$, be independent of $Z \sim \mathcal{N}(0,1)$, let the prices come from the pricing measure under which $S_T = F_T \exp(\sqrt{V} Z - V/2)$, and let $w(k)$ be the total implied variance at maturity T.*

(i) (Mixing.) *$C(K,T) = \mathbb{E}^{\mathbb{Q}}\big[S_0 e^{-qT} b(k, \sqrt{V})\big]$, an average of Black–Scholes prices.*

(ii) (Symmetry.) *w is defined on $\mathbb{R}$ and $w(k) = w(-k)$ for every k.*

(iii) (Smile.) *If V is not almost surely constant, then $w(0) < \mathbb{E}^{\mathbb{Q}}[V]$; if moreover $V \leq \bar{v}$, then $w(k) > \mathbb{E}^{\mathbb{Q}}[V]$ for all $|k| > \sqrt{\bar{v} + \bar{v}^2/4}$.*

Proof. (i) Given $V = v$, the computation of Theorem 6.11 with total variance v gives $\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+] = F_T b(k, \sqrt{v})$; apply the independence lemma (Theorem 1.19). (ii) S_T has a positive density on $\{V > 0\}$, so $\mathbb{Q}(S_T > K) > 0$ and $\mathbb{Q}(S_T < K) > 0$, and $w(k)$ exists by Theorem 15.2. By (15.2), $b(k,s) - (1 - e^k) = e^k b(-k, s)$ for $s \geq 0$: both sides are the normalized put (put–call symmetry). The prices $\bar{c}(k) = \mathbb{E}^{\mathbb{Q}}[b(k, \sqrt{V})]$ satisfy the same identity, so $b\big(-k, \sqrt{w(k)}\big) = e^{-k}\big(\bar{c}(k) - 1 + e^k\big) =$

$\bar c(-k) = b(-k, \sqrt{w(-k)})$, and $w(k) = w(-k)$. (iii) Let $h_k(v) = b(k, \sqrt v)$: continuous on $[0,\infty)$, increasing, and by the proof of Theorem 15.6, $h_k'' = h_k'(\frac{k^2}{2v^2} - \frac18 - \frac{1}{2v})$ on $(0,\infty)$. So h_0 is strictly concave, and Jensen gives $h_0(w(0)) = \mathbb{E}^{\mathbb{Q}}[h_0(V)] < h_0(\mathbb{E}^{\mathbb{Q}}[V])$. If $|k| > \sqrt{\bar v + \bar v^2/4}$, then $k^2 > v + v^2/4$ on $(0,\bar v]$, h_k is strictly convex on $[0,\bar v]$, and $h_k(w(k)) = \mathbb{E}^{\mathbb{Q}}[h_k(V)] > h_k(\mathbb{E}^{\mathbb{Q}}[V])$. □

Uncorrelated random volatility thus produces a symmetric smile, below the average volatility at the money and above it in the wings; Renault and Touzi (1996) show that it is U-shaped, with its minimum at the forward. A skew needs a correlation between the price and its volatility, as in the model of Heston (1993), which also keeps prices computable.

Lemma 15.10 (A Fourier formula for calls)**.** *Let X be a random variable with $\mathbb{E}[e^X] = 1$, and $\varphi(z) = \mathbb{E}[e^{izX}]$ for complex z with $-1 \le \operatorname{Im} z \le 0$. Then for every $k \in \mathbb{R}$,*

$$\mathbb{E}[(e^X - e^k)^+] = 1 - \frac{e^{k/2}}{\pi}\int_0^\infty \operatorname{Re}\Big[e^{-iuk}\varphi(u - \tfrac{i}{2})\Big]\frac{\mathrm{d}u}{u^2 + \frac14}.$$

Proof. Write $(e^X - e^k)^+ = e^X - \min(e^X, e^k)$ and $\min(e^x, e^k) = e^{k/2}e^{x/2}\psi(x)$ with $\psi(x) = e^{-|x-k|/2}$. Then $\int e^{iux}\psi(x)\,\mathrm{d}x = e^{iuk}/(u^2 + \frac14)$ is integrable, so by Fourier inversion (Durrett (2019), Section 3.3, for the Laplace density $\psi/4$), $\psi(x) = \frac{1}{2\pi}\int e^{-iux}e^{iuk}(u^2 + \frac14)^{-1}\mathrm{d}u$ for every x. As $\mathbb{E}[e^{X/2}] \le \mathbb{E}[e^X]^{1/2} = 1$, Fubini's theorem gives

$$\mathbb{E}[\min(e^X, e^k)] = \frac{e^{k/2}}{2\pi}\int_{\mathbb{R}} e^{iuk}\,\mathbb{E}[e^{(1/2-iu)X}]\frac{\mathrm{d}u}{u^2 + \frac14},$$

with $\mathbb{E}[e^{(1/2-iu)X}] = \overline{\varphi(u - \frac{i}{2})}$. The integrand at $-u$ is the conjugate of that at u, so the integral is twice the integral over $(0,\infty)$ of the real part, which is the claim. □

Proposition 15.11 (The Heston model)**.** *Let $\kappa, \theta, \xi, v_0 > 0$ and $\rho \in [-1,1]$. Under $\mathbb{Q}$, let $\widetilde W$ and B be Brownian motions with $[\widetilde W, B]_t = \rho t$, let v be the square-root (CIR) process $\mathrm{d}v_t = \kappa(\theta - v_t)\,\mathrm{d}t + \xi\sqrt{v_t}\,\mathrm{d}B_t$, which is nonnegative, and positive if $2\kappa\theta \ge \xi^2$ (Theorem 14.9), and let*

$$S_t = S_0 \exp\Big(\int_0^t \sqrt{v_u}\,\mathrm{d}\widetilde W_u + \int_0^t \Big(r - q - \frac{v_u}{2}\Big)\mathrm{d}u\Big), \qquad \mathrm{d}S_t = (r-q)S_t\,\mathrm{d}t + \sqrt{v_t}\,S_t\,\mathrm{d}\widetilde W_t.$$

(i) $\mathbb{E}^{\mathbb{Q}}[v_t] = \theta + (v_0 - \theta)e^{-\kappa t}$, so that $\frac1T\mathbb{E}^{\mathbb{Q}}[\int_0^T v_t\,\mathrm{d}t] = \theta + (v_0 - \theta)\frac{1 - e^{-\kappa T}}{\kappa T}$.

(ii) $(e^{-(r-q)t}S_t)$ is a martingale, and for real u, $\mathbb{E}^{\mathbb{Q}}[e^{iuX_T}] = \exp\big(A(u,T) + D(u,T)v_0\big)$ for $X_T = \log(S_T/F_T)$, where, with $\beta = \kappa - \rho\xi iu$, $d = \sqrt{\beta^2 + \xi^2(iu + u^2)}$ (principal root) and $\gamma = (\beta - d)/(\beta + d)$,

$$D = \frac{\beta - d}{\xi^2}\,\frac{1 - e^{-dT}}{1 - \gamma e^{-dT}}, \qquad A = \frac{\kappa\theta}{\xi^2}\Big((\beta - d)T - 2\log\frac{1 - \gamma e^{-dT}}{1 - \gamma}\Big). \tag{15.12}$$

(iii) If $\rho = 0$, the smile is symmetric in k, and below $\big(\frac1T\mathbb{E}^{\mathbb{Q}}\int_0^T v_t\,\mathrm{d}t\big)^{1/2}$ at $k = 0$, at every maturity.

Proof. (i) As $v \geq 0$, $v_t \leq v_0 + \kappa\theta t + \xi M_t$ with $M = \int_0^{\cdot} \sqrt{v}\,\mathrm{d}B$ a local martingale; localizing (Theorem 7.16) and applying Fatou's lemma, $\mathbb{E}^{\mathbb{Q}}[v_t] \leq v_0 + \kappa\theta t$. So M is a martingale (Theorem 7.7), and $m(t) = \mathbb{E}^{\mathbb{Q}}[v_t]$ solves $m(t) = v_0 + \int_0^t \kappa(\theta - m(u))\,\mathrm{d}u$, as for the Ornstein–Uhlenbeck mean in Chapter 9. (ii) The martingale property is Proposition 2.5 of Andersen and Piterbarg (2007). The characteristic function is equation (17) of Heston (1993), without the drift $(r-q)T$, in the equivalent form of Albrecher et al. (2007), which, unlike Heston's, does not cross the branch cut of the complex logarithm as u or T grows. (iii) If $\rho = 0$, $(\widetilde{W}, B)$ is a two-dimensional Brownian motion (Theorem 8.20). By pathwise uniqueness (Theorem 14.9(i)) and the Yamada–Watanabe theorem (Karatzas and Shreve (1991), Section 5.3.D), v is a functional of B, hence independent of $\widetilde{W}$, and given v, $\log(S_T/F_T)$ is $\mathcal{N}(-V/2, V)$ (Romano and Touzi (1997), who make Hull and White (1987)'s argument rigorous); V is not constant. Apply Theorem 15.9. □

Behind (ii) is the affine structure of Theorem 14.5: A and D solve Riccati equations (Exercise 15.7). Prices use Theorem 15.10 (a formula of A. Lewis, 2001). As $\mathbb{E}^{\mathbb{Q}}[e^{pX_T}] \leq 1$ for $p \in [0,1]$, φ is analytic on $-1 < \operatorname{Im} z < 0$ and continuous on the closure; at $z = u - \frac{i}{2}$, (15.12) with $iz + z^2 = u^2 + \frac{1}{4}$, $\beta = \kappa - \frac{\rho\xi}{2} - i\rho\xi u$ is this continuation if its root and logarithm cross no branch cut (Albrecher et al. (2007), Theorem 3; R. Lord and C. Kahl, *Math. Finance*, 2010). We checked this for every parameter set used here against the Riccati equations solved numerically (to 10^{-11}). v_0 and θ set the short and long levels, ρ tilts the smile, ξ curves it. The variance is not traded, so risk-neutral measures are many (Exercise 11.7); *calibration* to quoted vanillas selects one.

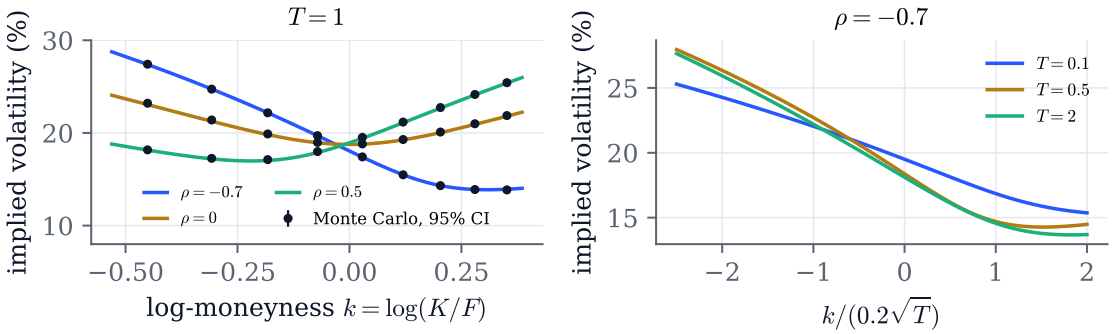

Figure 15.3: Heston smiles, $S_0 = 100$, $r = 2\%$, $q = 0$, $v_0 = \theta = 0.04$, $\kappa = 2$, $\xi = 0.5$. Left, $T = 1$: by Theorem 15.10 and (15.12) (lines), and by Monte Carlo (10^5 paths of the full-truncation Euler scheme of Lord et al. (2010), 250 steps). Right, $\rho = -0.7$: three maturities, against $k/(0.2\sqrt{T})$.

▶ Interactive demo: Heston smiles by Fourier and Monte Carlo; random-variance smiles.

15.6 One model for all prices

Pricing an exotic by Black–Scholes "at the right volatility" fails: the price depends on the shape of the surface, not only on its level. The digital call is the simplest example.

Proposition 15.12 (A digital under the smile)**.** *Let the prices $C(\cdot, T)$ come from a pricing measure $\mathbb{Q}$ with $\mathbb{Q}(S_T = K) = 0$, and let $\sigma_{\mathrm{imp}}(\cdot, T)$ be differentiable at K. The digital call*

paying $\mathbf{1}_{\{S_T>K\}}$ *costs*

$$e^{-rT}\mathbb{Q}(S_T > K) = -\partial_K C(K,T) = e^{-rT}N(d_-) - \mathcal{V}\,\partial_K \sigma_{\text{imp}}(K,T),$$

where $d_\pm$ *and* $\mathcal{V} = S_0 e^{-qT} N'(d_+)\sqrt{T}$ *are those of* (15.1) *at* $\sigma = \sigma_{\text{imp}}(K,T)$.

Proof. By Theorem 3.18(i), the one-sided derivatives of $C(\cdot,T)$ at K are $-e^{-rT}\mathbb{Q}(S_T > K)$ and $-e^{-rT}\mathbb{Q}(S_T \geq K)$, which coincide. Differentiate $C(K,T) = C^{\text{BS}}(K,T,\sigma_{\text{imp}}(K,T))$, with $\partial_K C^{\text{BS}} = -e^{-rT}N(d_-)$ and $\partial_\sigma C^{\text{BS}} = \mathcal{V}$ (Theorem 6.14, with $S_0 e^{-qT}$ for S_0). □

In the Heston market of Figure 15.3 ($\rho = -0.7$, $T = 1$), at $K = F_T = 102.02$ the implied volatility is 18.05%, falling by 2.26 points per 10 of strike, and $\mathcal{V} = 39.7$: the Black–Scholes digital at 18.05% costs 0.455, the correct price 0.545. Quoting digitals that way can be arbitraged: $1/h$ call spreads between K and $K+h$ pay at most $\mathbf{1}_{\{S_T>K\}}$ (Exercise 3.5) and cost 0.534 for $h = 1$; buying the digital and selling them collects 0.079.

Path-dependent claims depend on more than the vanillas. The local volatility of the Heston surface (Theorem 15.8) has the same vanilla prices, as Figure 15.4 confirms, but an up-and-out call with barrier 120 costs 3.32 in the Heston model and 2.78 with local volatility (Monte Carlo, ±0.03): a barrier option depends on the law of the whole path. Hedges differ too, since a smile-model delta includes the move of the implied volatility with the spot, which Hagan et al. (2002) argue local volatility gets wrong.

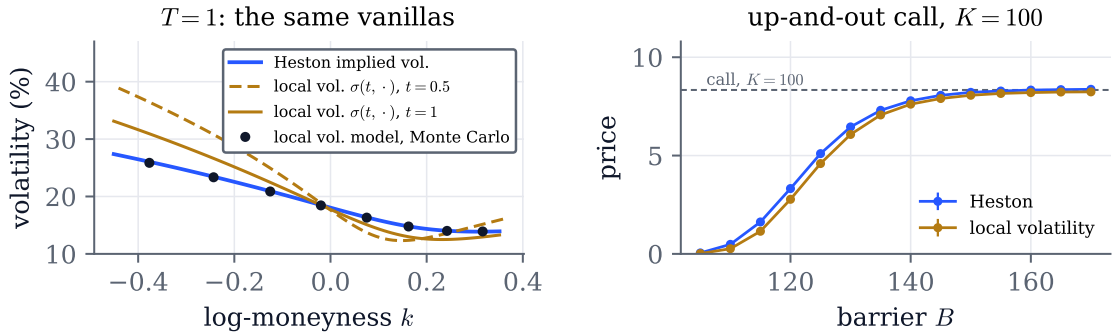

Figure 15.4: The Heston model of Figure 15.3 and its local volatility, from (15.8) in forward variables with Fourier derivatives of the prices. Left: local volatility at $t = 0.5$ and $t = 1$ against $\log(x/F_t)$, steeper than the implied volatility at $T = 1$, which the local volatility model reproduces (10^5 paths, 200 Euler steps). Right: up-and-out calls with $K = 100$, $T = 1$, the barrier monitored at the 200 dates, with 95% confidence intervals.

▶ Interactive demo: Dupire on a smooth surface, repricing check, barriers against Heston.

The lesson is a consistency principle. (a) Exotics and hedges must be priced in one model, under one pricing measure calibrated to all the vanillas; prices with a volatility per product are not expectations under one measure, and Theorem 3.6 no longer protects them. (b) The calibration surface must be free of static arbitrage, or (15.8) fails. (c) Models fitting the same vanillas still disagree on exotics; their spread, or model-free bounds such as those of Brown et al. (2001) for barrier options, measure this model risk. Chapter 17 builds arbitrage-free surfaces from bid and ask quotes, the first step of this programme.

15.7 Variance swaps and the VIX

A *variance swap* pays at T the realized variance $\frac{1}{T}\sum_i(\log S_{t_i}/S_{t_{i-1}})^2$ over a grid of $[0,T]$, usually daily, minus a strike K_{var}. If $\mathrm{d}S_t = S_t(\alpha_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t)$, it converges in probability to $\frac{1}{T}[\log S, \log S]_T = \frac{1}{T}\int_0^T \sigma_t^2\,\mathrm{d}t$ as the grid is refined (Theorems 7.12 and 8.11). We price this limit, without a model for σ (Carr and Madan, 1998; Demeterfi et al., 1999).

Proposition 15.13 (Variance swaps)**.** *Under $\mathbb{Q}$, let $S_t = S_0\exp\big(\int_0^t \sigma_u\,\mathrm{d}\widetilde{W}_u + \int_0^t(r-q-\frac{1}{2}\sigma_u^2)\mathrm{d}u\big)$, with σ progressively measurable and bounded, so that $(e^{-(r-q)t}S_t)$ is a martingale (Novikov's condition, as in Theorem 11.2) and $\mathbb{E}^{\mathbb{Q}}[S_T] = F_T$, and let the option prices come from $\mathbb{Q}$.*

(i) Along every path, $\int_0^T \sigma_t^2\,\mathrm{d}t = 2\int_0^T \sigma_t\,\mathrm{d}\widetilde{W}_t - 2\log(S_T/F_T)$.

(ii) The fair strike, which makes the swap worth 0, is

$$K_{\text{var}} = \frac{\mathbb{E}^{\mathbb{Q}}\int_0^T \sigma_t^2\,\mathrm{d}t}{T} = -\frac{2}{T}\mathbb{E}^{\mathbb{Q}}\Big[\log\frac{S_T}{F_T}\Big] = \frac{2e^{rT}}{T}\Big(\int_0^{F_T}\frac{P(K,T)}{K^2}\,\mathrm{d}K + \int_{F_T}^{\infty}\frac{C(K,T)}{K^2}\,\mathrm{d}K\Big). \tag{15.13}$$

(iii) The swap is replicated at zero cost by $\frac{2\,\mathrm{d}K}{TK^2}$ puts at strikes $K < F_T$ and calls at $K > F_T$, $\frac{2}{TF_T}$ short forwards with delivery price F_T (together worth $e^{-rT}K_{\text{var}}$), a loan of $e^{-rT}K_{\text{var}}$, and the self-financing strategy started from 0 that holds $\Delta_t = 2e^{-r(T-t)}/(TS_t)$ shares, dividends reinvested in the money market.

Proof. (i) $\log S_T = \log F_T + \int_0^T \sigma\,\mathrm{d}\widetilde{W} - \frac{1}{2}\int_0^T\sigma^2\,\mathrm{d}t$. (ii) The stochastic integral has mean 0 (Theorem 7.7), and $\log(S_T/F_T)$ is integrable by (i). Apply Theorem 3.17 with $B = e^{-rT}$, $f(x) = -\log(x/F_T)$ and $\kappa = F_T$: $f(\kappa) = 0$, $f''(K) = K^{-2}$, and the forward term vanishes since the claim paying S_T costs $e^{-rT}F_T$. Condition (3.8) holds: by Tonelli and Theorem 3.16, the sum of the two integrals is $e^{-rT}\mathbb{E}^{\mathbb{Q}}[f(S_T) + (S_T - F_T)/F_T] = e^{-rT}\mathbb{E}^{\mathbb{Q}}[f(S_T)] < \infty$. (iii) With dividends reinvested, $\mathrm{d}(e^{-rt}X_t) = \Delta_t e^{-rt}S_t\sigma_t\,\mathrm{d}\widetilde{W}_t$ (the computation of Theorem 11.14(i), with σ_t for σ), here $\frac{2}{T}e^{-rT}\sigma_t\,\mathrm{d}\widetilde{W}_t$, so $X_T = \frac{2}{T}\int_0^T\sigma\,\mathrm{d}\widetilde{W}$. By Theorem 3.16 the options pay $\frac{2}{T}\big(f(S_T) + (S_T - F_T)/F_T\big)$ and the forwards $-\frac{2}{T}(S_T - F_T)/F_T$; by (i) the total is $\frac{1}{T}\int_0^T\sigma^2\,\mathrm{d}t - K_{\text{var}}$. □

The strike averages the smile with weights $1/K^2$; a skew lifts it above the at-the-money implied variance (Figure 15.5). If S jumps, (i) fails: a jump J of $\log S$ adds J^2/T to the realized variance but $2(e^J - 1 - J)/T = (J^2 + J^3/3 + \dots)/T$ to the hedge.

Definition 15.14 (The VIX)**.** For a maturity T, let $K_1 < \cdots < K_n$ be the listed strikes, F the forward price, R the interest rate, K_0 the largest strike $\le F$, and $Q(K_i)$ the mid-quote of the put if $K_i < K_0$, of the call if $K_i > K_0$, and their average at K_0; let $\Delta K_i = \frac{1}{2}(K_{i+1} - K_{i-1})$ (one-sided at the ends). The CBOE variance estimate is

$$\sigma_T^2 = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2}\,e^{RT}Q(K_i) - \frac{1}{T}\Big(\frac{F}{K_0} - 1\Big)^2. \tag{15.14}$$

The *VIX* is $100\sqrt{\frac{365}{30}V_{30}}$, where V_{30} is the linear interpolation, in the time to maturity measured in minutes, of the total variances $T_1\sigma_{T_1}^2$ and $T_2\sigma_{T_2}^2$ of S&P 500 options with maturities $T_1 < T_2$ around 30 days, at 30 days (Chicago Board Options Exchange, 2009); further rules select the

strikes and maturities.

The VIX is a discretized 30-day variance-swap strike. Indeed, Theorem 3.17 with $\kappa = K_0$ gives, with $x = F/K_0 - 1$,

$$-\frac{2}{T}\mathbb{E}^{\mathbb{Q}}\Big[\log\frac{S_T}{F}\Big] = \frac{2e^{RT}}{T}\Big(\int_0^{K_0}\frac{P(K)}{K^2}\,\mathrm{d}K + \int_{K_0}^{\infty}\frac{C(K)}{K^2}\,\mathrm{d}K\Big) - \frac{2}{T}(x - \log(1+x)),$$

and $x - \log(1+x) = \frac{x^2}{2} + O(x^3)$. The sum in (15.14) is the trapezoidal rule for the integral, averaging put and call at K_0, where the integrand jumps. Up to terms of order x, its derivative jumps there by $-e^{-RT}/K_0^2$ (put and call slopes differ by e^{-RT}), so the error formula of the trapezoidal rule gives an upward bias of about $h^2/(6TK_0^2)$ for strikes spaced by h (Figure 15.5). With jumps, VIX^2 prices the log contract.

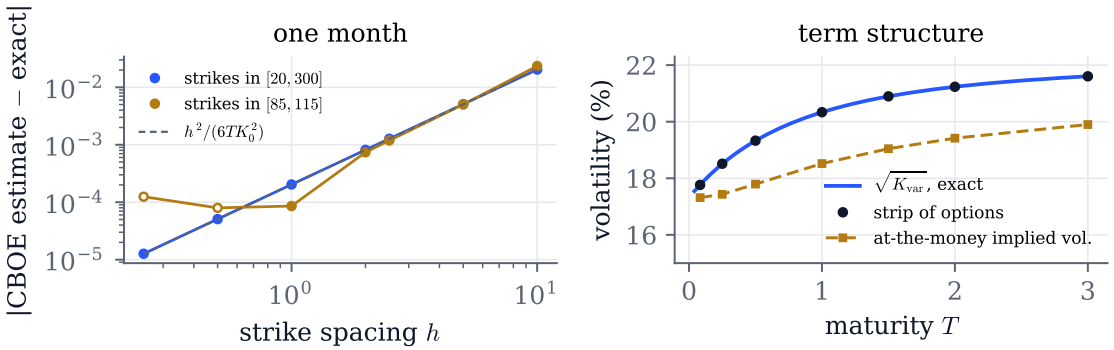

Figure 15.5: Heston model with $S_0 = 100$, $r = 2\%$, $v_0 = 0.03$, $\theta = 0.05$, $\kappa = 2$, $\xi = 0.5$, $\rho = -0.7$. Left: error of (15.14) at $T = 30/365$ against Theorem 15.11(i), for strike spacings h on a wide and a narrow range (open markers: negative errors, from the missing wings). Right: $\sqrt{K_{\text{var}}}$, exact and from a fine strip, above the at-the-money implied volatility.

▶ Interactive demo: variance swaps from a strip of options, and the VIX formula.

Further reading

Gatheral (2006) is the practitioner's account of the surface and its models; Lee (2004) treats the wings and Demeterfi et al. (1999) variance swaps.

Chapter 16

Numerical methods

Closed-form prices are the exception: a local volatility, an arithmetic average or an early-exercise right usually destroys them. Two characterizations of the price survive: it is an expected discounted payoff under a risk-neutral measure (Theorem 11.8), and it solves a partial differential equation (Theorem 12.4). Each leads to a numerical method. *Monte Carlo* averages simulated payoffs; *finite differences* replace the derivatives in the equation by difference quotients on a grid. Trees, the third classical method, were the subject of Chapter 4 and Theorem 6.4; the simplest finite-difference scheme is in fact a trinomial tree. Two questions organize the chapter: how fast does a method converge, and what makes it fail? For Monte Carlo the answers come from the central limit theorem and from the discretization of the dynamics; for finite differences, from consistency and stability, whose combination is the subject of the Lax equivalence theorem. We end with the American put, which requires a complementarity problem at every time step, and with physics-informed neural networks.

Setting. We work in the Black–Scholes model under $\mathbb{Q}$: $S_t = S_0 \exp\left(\sigma \widetilde{W}_t + (r - \frac{\sigma^2}{2})t\right)$ with $S_0 > 0$, $\sigma > 0$, $r \in \mathbb{R}$ (Theorem 6.11). Its closed forms make it a test bench: every error below is measured against them.

16.1 Monte Carlo

Let Y be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{E}|Y| < \infty$ and $p = \mathbb{E}[Y]$; for a European claim, $\mathbb{P}$ is the risk-neutral measure $\mathbb{Q}$, $Y = e^{-rT} g(S_T)$, and p is the price at time 0. Let $Y_1, Y_2, \dots$ be independent copies of Y, and define the *Monte Carlo estimator* and the *sample variance*

$$\hat{p}_M = \frac{1}{M} \sum_{i=1}^{M} Y_i, \qquad s_M^2 = \frac{1}{M-1} \sum_{i=1}^{M} \left(Y_i - \hat{p}_M\right)^2 \quad (M \geq 2).$$

In the Black–Scholes model the copies are produced by *exact sampling*: $S_T^{(i)} = S_0 \exp\left(\sigma \sqrt{T}\, Z_i + (r - \frac{\sigma^2}{2})T\right)$ with $Z_1, Z_2, \dots$ independent standard normal variables.

Theorem 16.1 (Monte Carlo)**.**

(i) Unbiased and consistent. $\mathbb{E}[\hat{p}_M] = p$ *for every* $M \geq 1$, *and* $\hat{p}_M \to p$ *almost surely as* $M \to \infty$.

(ii) Mean square error. *If* $\mathbb{E}[Y^2] < \infty$ *and* $s^2 = \mathrm{Var}(Y)$, *then* $\mathbb{E}\big[(\hat{p}_M - p)^2\big] = s^2/M$, *and* $\mathbb{P}\big(|\hat{p}_M - p| \geq \varepsilon\big) \leq s^2/(M\varepsilon^2)$ *for every* $\varepsilon > 0$ *and every* M.

(iii) Confidence intervals. *If moreover* $s > 0$, *then* $\sqrt{M}(\hat{p}_M - p)/s$ *converges in distribution to* $\mathcal{N}(0,1)$, $s_M \to s$ *almost surely, and for every* $\alpha \in (0,1)$, *with* $z_\alpha = N^{-1}(1-\alpha/2)$,

$$\mathbb{P}\Big(\hat{p}_M - z_\alpha \frac{s_M}{\sqrt{M}} \leq p \leq \hat{p}_M + z_\alpha \frac{s_M}{\sqrt{M}}\Big) \longrightarrow 1 - \alpha \qquad (M \to \infty). \tag{16.1}$$

Proof. (i) Linearity of expectation, and the strong law of large numbers (Durrett (2019), Theorem 2.4.1). (ii) By independence, $\mathrm{Var}(\hat{p}_M) = Ms^2/M^2$; the bound is Chebyshev's inequality.

(iii) The convergence in distribution is the central limit theorem (Durrett (2019), Theorem 3.4.1). Next, $s_M^2 = \frac{M}{M-1}\big(\frac{1}{M}\sum_i Y_i^2 - \hat{p}_M^2\big) \to \mathbb{E}[Y^2] - p^2 = s^2$ almost surely, by the strong law for Y_i^2 and Y_i. Let $Z_M = \sqrt{M}(\hat{p}_M - p)/s$. Since N is continuous, $\mathbb{P}(Z_M \leq w) \to N(w)$ for every w, and also $\mathbb{P}(Z_M < w) \to N(w)$, because $\mathbb{P}(Z_M \leq w - \eta) \leq \mathbb{P}(Z_M < w) \leq \mathbb{P}(Z_M \leq w)$ for $\eta > 0$; hence $\mathbb{P}(|Z_M| \leq w) \to 2N(w) - 1$. The event in (16.1) is $\{|Z_M| \leq z_\alpha s_M/s\}$. Fix $\delta \in (0,1)$; the event $A_M = \{|s_M/s - 1| \leq \delta\}$ has $\mathbb{P}(A_M) \to 1$, and on A_M, $\{|Z_M| \leq z_\alpha(1-\delta)\} \subseteq \{|Z_M| \leq z_\alpha s_M/s\} \subseteq \{|Z_M| \leq z_\alpha(1+\delta)\}$. So the limits inferior and superior of the probability in (16.1) lie between $2N(z_\alpha(1-\delta)) - 1$ and $2N(z_\alpha(1+\delta)) - 1$; let $\delta \downarrow 0$. □

The error is of order $s/\sqrt{M}$: each additional digit costs a hundred times more paths, but the rate does not depend on the dimension. Y may be a function of a whole path or of a hundred assets, which is why Monte Carlo is often the method of choice for path-dependent claims and models with many factors (Theorem 12.6). The interval (16.1), with $z_{0.05} \approx 1.96$, is asymptotic, and for small M its coverage can be far from nominal (Exercise 16.1); part (ii) holds for every M but needs s, and gives the cruder half-width $s/\sqrt{0.05M} \approx 4.47\, s/\sqrt{M}$. For the call with $S_0 = K = 100$, $r = 5\%$, $\sigma = 20\%$, $T = 1$, $p = 10.45$ and the discounted payoff has the standard deviation $s = 14.72$ (a closed form, in `code/src/ieor222/numerics.py`): with $M = 10^4$ paths the half-width is 0.29, almost 3% of the price (Figure 16.1).

Remark 16.2 (Variance reduction). Besides M, the other lever is s: replace Y by a variable with the same mean and a smaller variance. The main methods are antithetic variates (use Z_i and $-Z_i$), control variates (subtract a correlated variable of known mean, such as the geometric Asian option of Theorem 13.9 for the arithmetic one), importance sampling (a change of drift by Theorem 11.2; see Exercise 1.9), stratification and quasi-Monte Carlo. They are the subject of the companion notes for MFE 230D, *Derivatives: Quantitative Methods* (`github.com/GennaroAlberto/mfe230d-derivatives`).

16.1.1 Discretized dynamics

In a local volatility model (Theorem 9.13) one simulates a discretization, typically the Euler–Maruyama scheme (9.9) with n steps of size $h = T/n$, and averages M copies of $Y_h = e^{-rT} g(S_T^h)$. The estimator $\hat{p}_{M,h}$ then has a *bias* $\beta(h) = \mathbb{E}[Y_h] - p$, the weak error of Theorem 9.16, and since

$\mathbb{E}[(X-p)^2] = (\mathbb{E}X - p)^2 + \mathrm{Var}(X)$,

$$\mathbb{E}\big[(\hat{p}_{M,h} - p)^2\big] = \beta(h)^2 + \frac{\mathrm{Var}(Y_h)}{M}. \tag{16.2}$$

A path costs n normal draws, the estimator nM, and the budget must be shared between steps and paths.

Proposition 16.3 (Cost of a biased estimator)**.** *Suppose that there are $h_0 \in (0, T]$ and constants $0 < c_1 \le c_2$ and $0 < v_1 \le v_2$ such that $c_1 h \le |\beta(h)| \le c_2 h$ and $v_1 \le \mathrm{Var}(Y_h) \le v_2$ for every $h = T/n \le h_0$, and let $0 < \varepsilon \le \min(\sqrt{2}\, c_2 h_0, \sqrt{2v_2})$.*

(i) With $n = \lceil \sqrt{2}\, c_2 T/\varepsilon \rceil$ steps and $M = \lceil 2v_2/\varepsilon^2 \rceil$ paths, the root mean square error is at most ε, and $nM \le 12\, c_2 v_2 T\, \varepsilon^{-3}$.

(ii) Conversely, if $T/n \le h_0$ and the root mean square error is at most ε, then $n \ge c_1 T/\varepsilon$ and $M \ge v_1/\varepsilon^2$, so that $nM \ge c_1 v_1 T\, \varepsilon^{-3}$.

Proof. (i) Here $h = T/n \le \varepsilon/(\sqrt{2}\, c_2) \le h_0$, so $\beta(h)^2 \le \varepsilon^2/2$, and $\mathrm{Var}(Y_h)/M \le \varepsilon^2/2$; apply (16.2). Since $\sqrt{2}\, c_2 T/\varepsilon \ge 1$ and $2v_2/\varepsilon^2 \ge 1$, rounding up at most doubles each factor, so $nM \le (2\sqrt{2}\, c_2 T/\varepsilon)(4v_2/\varepsilon^2) \le 12\, c_2 v_2 T \varepsilon^{-3}$. (ii) By (16.2), $\varepsilon^2 \ge \beta(h)^2 \ge c_1^2 T^2/n^2$ and $\varepsilon^2 \ge \mathrm{Var}(Y_h)/M \ge v_1/M$. □

Without bias, accuracy ε costs of the order of ε^{-2}; a bias of order h raises the cost to the order of ε^{-3}, whatever the allocation with $T/n \le h_0$. With a budget C, the error decreases like $C^{-1/3}$ instead of $C^{-1/2}$, the step shrinking like $C^{-1/3}$ (Duffie and Glynn (1995); Glasserman (2004), Section 1.1.3). Schemes of higher weak order, Richardson extrapolation (Exercise 16.5) and multilevel Monte Carlo (Giles (2008)), which brings the cost of the Euler scheme down to the order of $\varepsilon^{-2}(\log \varepsilon)^2$, recover most of the loss; see also the MFE 230D notes.

Example 16.4 (A power claim)**.** Let $Y = e^{-rT} S_T^2$, with price $p = S_0^2 e^{(r+\sigma^2)T}$, and simulate S itself by the Euler scheme. As in Theorem 9.17, $S_T^h = S_0 \prod_{k<n}(1 + rh + \sigma\Delta_k)$ with independent $\Delta_k \sim \mathcal{N}(0, h)$, so $\mathbb{E}[Y_h] = e^{-rT} S_0^2 m_2^n$ and $\mathrm{Var}(Y_h) = e^{-2rT} S_0^4 (m_4^n - m_2^{2n})$, where $m_2 = (1+rh)^2 + \sigma^2 h$ and $m_4 = (1+rh)^4 + 6(1+rh)^2\sigma^2 h + 3\sigma^4 h^2$ are the second and fourth moments of one factor. Expanding $n \log m_2$ in $h = T/n$ gives $\beta(h) = pT(r^2 - \frac{1}{2}(2r+\sigma^2)^2)h + O(h^2)$. For $S_0 = 1$, $r = 5\%$, $\sigma = 40\%$, $T = 1$: $p = 1.2337$ and $\beta(h) \approx -0.0386\, h$; numerically, $|\beta(h)|/h$ increases with n from 0.033 ($n = 1$) to this limit, and $\mathrm{Var}(Y_h)$ from 0.68 to 1.36. So Theorem 16.3 applies with $h_0 = T$, and Figure 16.1 (right) shows the two regimes.

▶ Interactive demo: Monte Carlo prices with confidence intervals, their coverage over many repetitions, and the error against the cost for any number of time steps.

16.2 Finite differences

The equation. By Theorem 12.4, the price of $g(S_T)$ is $v(t, S_t)$, where $\partial_t v + rx\,\partial_x v + \frac{1}{2}\sigma^2 x^2 \partial_{xx} v = rv$ and $v(T, \cdot) = g$. In the time to maturity $\tau = T - t$ and the log-price $y = \log x$, let $u(\tau, y) = v(T - \tau, e^y)$. As in the proof of Theorem 12.7, $\partial_t v = -\partial_\tau u$, $x\,\partial_x v = \partial_y u$ and

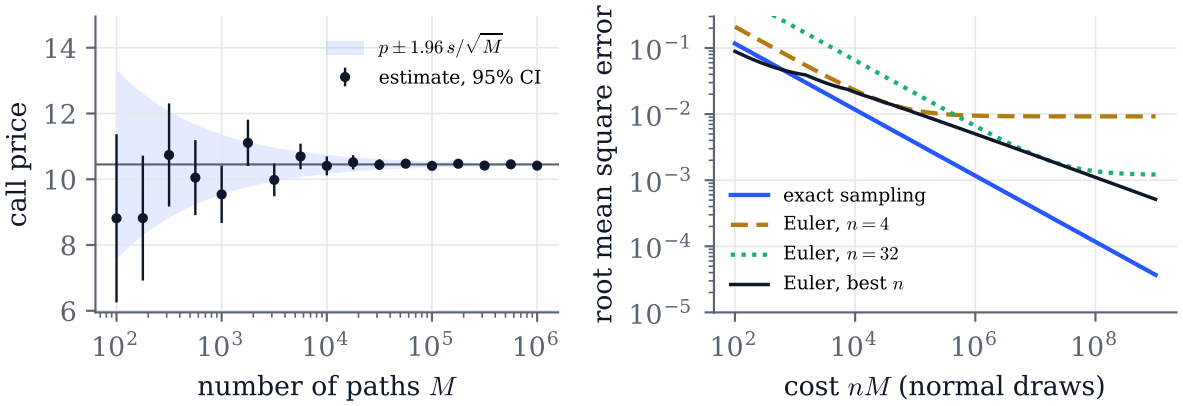

Figure 16.1: Left: Monte Carlo estimates of the call ($S_0 = K = 100$, $r = 5\%$, $\sigma = 20\%$, $T = 1$) with 95% confidence intervals (16.1), from independent samples; the band is $p \pm 1.96\, s/\sqrt{M}$. Right: root mean square error of the price of S_T^2 (Theorem 16.4) against the cost, computed exactly from (16.2): exact sampling (slope $-\frac{1}{2}$); the Euler scheme with $n = 4$ and $n = 32$ steps, which stalls at its bias; and with the best n for each budget (slope $-\frac{1}{3}$).

$x^2 \partial_{xx} v = \partial_{yy} u - \partial_y u$, so

$$\partial_\tau u = a\, \partial_{yy} u + b\, \partial_y u - c\, u, \qquad u(0, y) = g(e^y), \qquad a = \frac{\sigma^2}{2}, \quad b = r - \frac{\sigma^2}{2}, \quad c = r. \tag{16.3}$$

With $a = \frac{1}{2}$, $b = c = 0$ this is the heat equation. For local volatility a and b depend on (τ, y), and the schemes below apply unchanged. In x, by contrast, the diffusion coefficient $\frac{1}{2}\sigma^2 x^2$ grows with x, which matters for stability (Theorem 16.12).

The grid. Truncate to $[y_{\min}, y_{\max}]$, of length L, with J intervals of length $h = L/J$ and nodes $y_j = y_{\min} + jh$; take N time steps of size $k = T/N$, $\tau_n = nk$. The unknowns are $U_j^n \approx u(\tau_n, y_j)$, with prescribed end values $U_0^n = \ell_n$ and $U_J^n = \rho_n$ (Section 16.3).

Lemma 16.5 (Difference quotients)**.** *If f is four times continuously differentiable on $[y - h, y + h]$, there are $\xi, \eta \in (y - h, y + h)$ with*

$$\frac{f(y+h) - 2f(y) + f(y-h)}{h^2} = f''(y) + \frac{h^2}{12} f''''(\xi), \qquad \frac{f(y+h) - f(y-h)}{2h} = f'(y) + \frac{h^2}{6} f'''(\eta).$$

Proof. Taylor's formula gives $f(y \pm h) = f(y) \pm h f'(y) + \frac{h^2}{2} f''(y) \pm \frac{h^3}{6} f'''(y) + \frac{h^4}{24} f''''(\xi_\pm)$; add the two expansions, and use the intermediate value theorem to write $\frac{1}{2}(f''''(\xi_+) + f''''(\xi_-)) = f''''(\xi)$. For f', expand to third order and subtract. □

With these quotients, the operator of (16.3) becomes, for $1 \le j \le J - 1$,

$$(L_h V)_j = p_- V_{j-1} + p_0 V_j + p_+ V_{j+1}, \qquad p_\pm = \frac{a}{h^2} \pm \frac{b}{2h}, \qquad p_0 = -\frac{2a}{h^2} - c. \tag{16.4}$$

Definition 16.6 (The θ-scheme)**.** Let $\theta \in [0,1]$. The θ-scheme for (16.3) is

$$\frac{U_j^{n+1} - U_j^n}{k} = \theta\,(L_h U^{n+1})_j + (1-\theta)\,(L_h U^n)_j, \qquad 1 \le j \le J-1,\ 0 \le n < N, \tag{16.5}$$

with $U_j^0 = g(e^{y_j})$ and the boundary values ℓ_n, ρ_n. It is *explicit* for $\theta = 0$, *implicit* for $\theta = 1$, and the *Crank–Nicolson* scheme (Crank and Nicolson (1947)), the trapezoidal rule in time, for $\theta = \frac{1}{2}$.

For $\theta > 0$ each step solves a tridiagonal system, in $O(J)$ operations (the Thomas algorithm). With $V^n = (U_1^n, \dots, U_{J-1}^n)$ and A the tridiagonal matrix with rows (p_-, p_0, p_+),

$$BV^{n+1} = CV^n + kF^n, \qquad B = I - \theta k A, \qquad C = I + (1-\theta)kA, \tag{16.6}$$

where F^n vanishes except for its first entry, $p_-(\theta\ell_{n+1} + (1-\theta)\ell_n)$, and its last, $p_+(\theta\rho_{n+1} + (1-\theta)\rho_n)$. The *iteration matrix* is $G = B^{-1}C$.

Definition 16.7 (Consistency and stability)**.** Let u solve (16.3) on $[0,T] \times [y_{\min}, y_{\max}]$ and $u_j^n = u(\tau_n, y_j)$. The *truncation error* is $T_j^n = \frac{1}{k}(u_j^{n+1} - u_j^n) - \theta\,(L_h u^{n+1})_j - (1-\theta)\,(L_h u^n)_j$, and the scheme is *consistent of order* (p,q) if $|T_j^n| \le C(k^p + h^q)$ for all n, j. Given a norm on $\mathbb{R}^{J-1}$ and a family of grids, the scheme is *stable* if there is a constant C_s, the same for all grids, with $\|G^m\| \le C_s$ and $\|G^m B^{-1}\| \le C_s$ for all $m \le N$, in the operator norm.

Stability means that the steps do not amplify errors: an error δ in V^n becomes $G^m\delta$ after m further steps, and an error $k\delta$ in the right side of (16.6) becomes $kG^mB^{-1}\delta$. We use the norms $\|V\|_\infty = \max_j |V_j|$ and $\|V\|_{2,h} = \left(h\sum_j V_j^2\right)^{1/2}$.

Proposition 16.8 (Stability of the explicit scheme)**.** *Consider the heat equation $\partial_\tau u = a\,\partial_{yy} u$ ($b = c = 0$), the explicit scheme, and $\lambda = ak/h^2$.*

(i) If $\lambda \le \frac{1}{2}$, then $U_j^{n+1} = \lambda U_{j-1}^n + (1-2\lambda)U_j^n + \lambda U_{j+1}^n$ is a convex combination. Hence the discrete maximum principle *holds: every U_j^n lies between the smallest and the largest of the values U_i^0, ℓ_m, ρ_m, $m \le n$; and $\|G^m\|_\infty \le 1$ for every m.*

(ii) If $\lambda > \frac{1}{2}$, the scheme is unstable. G has the eigenvalue $g = 1 - 4\lambda\cos^2\left(\frac{\pi}{2J}\right)$, which is smaller than -1 as soon as $\cos^2(\frac{\pi}{2J}) > \frac{1}{2\lambda}$, hence for all large J, and then $\|G^m\|_\infty \ge |g|^m$; along grids with $h \to 0$ and $N = aT/(\lambda h^2)$, $\|G^N\|_\infty \to \infty$. On the infinite grid $h\mathbb{Z}$, the initial values $U_j^0 = \varepsilon(-1)^j$ give $U_j^n = \varepsilon(1-4\lambda)^n(-1)^j$.

Proof. (i) The weights λ, $1-2\lambda$, λ are nonnegative with sum 1, and a convex combination lies between the smallest and the largest of the combined values; induct on n. With zero boundary values, the same remark gives $\|GV\|_\infty \le \|V\|_\infty$.

(ii) For $1 \le m \le J-1$, $w_j = \sin(m\pi j/J)$ vanishes at $j = 0, J$ and $w_{j+1} + w_{j-1} = 2\cos(m\pi/J)w_j$, so $(Aw)_j = -\frac{4a}{h^2}\sin^2\left(\frac{m\pi}{2J}\right)w_j$ and $Gw = \left(1 - 4\lambda\sin^2\frac{m\pi}{2J}\right)w$; for $m = J-1$ this is g. An eigenvalue of G^m is g^m, and the modulus of an eigenvalue is at most any operator norm. On $h\mathbb{Z}$, $(-1)^{j-1} + (-1)^{j+1} = -2(-1)^j$, so one step multiplies $\varepsilon(-1)^j$ by $1 - 4\lambda$. □

The sawtooth $(-1)^j$ is the highest frequency the grid carries. For $\lambda > \frac{1}{2}$ a rounding error of size 10^{-16} reaches size 1 after $16/\log_{10}|1-4\lambda|$ steps, about 200 for $\lambda = 0.55$; for a call the growth starts from the kink of the payoff, which contains all frequencies, and is visible much sooner (Figure 16.2, left).

Remark 16.9 (The explicit scheme is a trinomial tree). For (16.3) the explicit scheme reads $U_j^{n+1} = kp_- U_{j-1}^n + (1+kp_0)U_j^n + kp_+ U_{j+1}^n$, with weights adding up to $1-ck$. If they are nonnegative ($h|b| \le 2a$ and $k(2a/h^2 + c) \le 1$) and $ck < 1$, this is backward induction in a trinomial tree for the log-price, with moves $\pm h$ or 0, probabilities equal to the weights divided by $1-ck$, and the discount factor $1-rk \approx e^{-rk}$ per period: trees like those of Theorem 4.15, with $h = \delta x$, whose first two moments agree with the Black–Scholes ones up to $O(k^2)$ instead of exactly. For the heat equation, the stability condition $\lambda \le \frac{1}{2}$ of Theorem 16.8 is exactly the nonnegativity of the middle probability (Exercise 16.2).

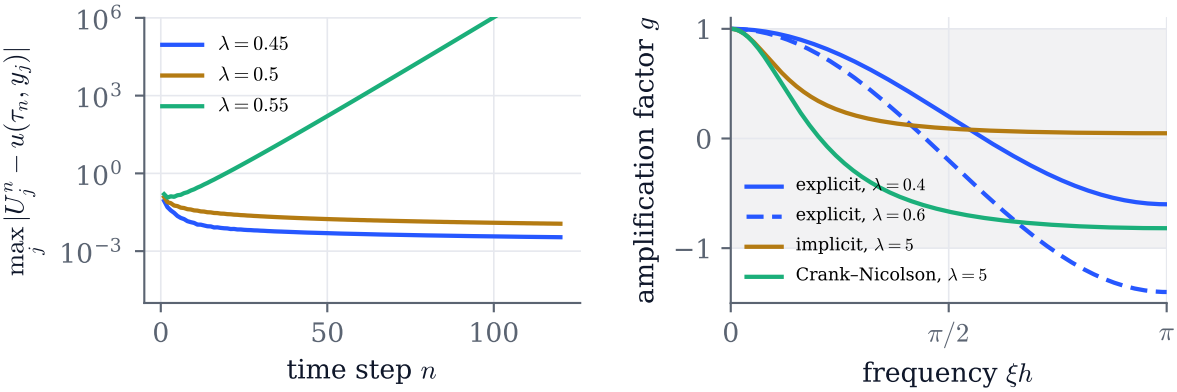

Figure 16.2: Left: the explicit scheme for the call of Figure 16.1 in log-price (201 nodes over six standard deviations of $\log S_T$ on each side of $\log K$), $k = \lambda h^2/a$: the largest error over the grid against the step; for $\lambda = 0.55$ it grows like 1.2^n. Right: the amplification factors of the three schemes for the heat equation against the frequency, $1-4\lambda s$, $1/(1+4\lambda s)$ and $(1-2\lambda s)/(1+2\lambda s)$ with $s = \sin^2(\xi h/2)$; $|g| \le 1$ is shaded.

For the full equation and all three schemes, consistency and stability give convergence.

Theorem 16.10 (Consistency, stability and convergence). *Let $a > 0$, $b \in \mathbb{R}$, $c \ge 0$ and $\theta \in [0,1]$. Consider grids with $h|b| \le a$ and, if $\theta < \frac{1}{2}$, $(1-2\theta)k(4a/h^2 + c) \le 2$. Let u solve (16.3) on $[0,T] \times [y_{\min}, y_{\max}]$, with $\partial_\tau^i u$, $i \le 3$, and $\partial_y^j u$, $j \le 4$, continuous on this closed rectangle and bounded by M_u. Let U be the θ-scheme with initial values $u(0, y_j)$ and boundary values $\ell_n = u(\tau_n, y_{\min})$, $\rho_n = u(\tau_n, y_{\max})$.*

(i) Consistency. *$|T_j^n| \le C_1(k + h^2)$ for every θ, and $|T_j^n| \le C_1(k^2 + h^2)$ for $\theta = \frac{1}{2}$, with $C_1 = M_u(\frac{3}{2} + \frac{a}{12} + \frac{|b|}{6})$.*

(ii) Stability. *$\|G^m\|_{2,h} \le e^{|b|L/a}$ and $\|G^m B^{-1}\|_{2,h} \le e^{|b|L/a}$ for all $m \ge 0$.*

(iii) Convergence. *With $u^n = (u_1^n, \dots, u_{J-1}^n)$, for every $n \le N$, $\|u^n - V^n\|_{2,h} \le e^{|b|L/a}\, T\sqrt{L}\, C_1(k + h^2)$, and $k^2 + h^2$ for Crank–Nicolson.*

(iv) Maximum norm. *If moreover $(1-\theta)k(2a/h^2 + c) \le 1$, then $\|G^m\|_\infty \le 1$, $\|G^m B^{-1}\|_\infty \le 1$, and $|u_j^n - U_j^n| \le T\, C_1(k + h^2)$ (resp. $k^2 + h^2$) for all n, j. Two solutions of the scheme with the same initial values and boundary values (ℓ, ρ) and (ℓ', ρ') differ at every node by at most $\max_n \max(|\ell_n - \ell'_n|, |\rho_n - \rho'_n|)$.*

Proof. (i) By Theorem 16.5, $(L_h u^m)_j$ differs from $(a\partial_{yy}u + b\partial_y u - cu)(\tau_m, y_j)$, which is $\partial_\tau u(\tau_m, y_j)$

(also at $\tau = 0$, by continuity), by at most $h^2 M_u(\frac{a}{12} + \frac{|b|}{6})$. It remains to compare $\frac{1}{k}(u_j^{n+1} - u_j^n)$ with $\theta\,\partial_\tau u_j^{n+1} + (1-\theta)\,\partial_\tau u_j^n$. By Taylor's formula in τ, $u_j^{n+1} - u_j^n = k\,\partial_\tau u_j^n + \frac{k^2}{2}\partial_{\tau\tau} u(\zeta, y_j)$ and $\partial_\tau u_j^{n+1} = \partial_\tau u_j^n + k\,\partial_{\tau\tau} u(\zeta', y_j)$, so the difference is at most $(\frac{1}{2} + \theta) k M_u$. For $\theta = \frac{1}{2}$, expand around $\bar\tau = \tau_n + \frac{k}{2}$ as in Theorem 16.5: $\frac{1}{k}(u_j^{n+1} - u_j^n)$ and $\frac{1}{2}(\partial_\tau u_j^{n+1} + \partial_\tau u_j^n)$ differ from $\partial_\tau u(\bar\tau, y_j)$ by $\frac{k^2}{24}\partial_\tau^3 u$ and $\frac{k^2}{8}\partial_\tau^3 u$ at intermediate points, so by at most $\frac{k^2}{6} M_u$ in all.

(ii) Since $h|b| \le a$, $p_\pm > 0$. Let D be diagonal with entries $d_j = (p_-/p_+)^{j/2}$. Then $S = D^{-1}AD$ is symmetric tridiagonal, with diagonal p_0 and off-diagonal $q = \sqrt{p_- p_+}$, since $d_j^{-1} p_+ d_{j+1} = d_{j+1}^{-1} p_- d_j = q$. As in the proof of Theorem 16.8, its eigenvalues are $\mu_m = p_0 + 2q\cos\frac{m\pi}{J}$, and $0 < q \le \frac{1}{2}(p_- + p_+) = a/h^2$ puts them in $[-4a/h^2 - c, -c]$. Now $G = DG_S D^{-1}$ and $B^{-1} = D(I - \theta k S)^{-1} D^{-1}$, where $G_S = (I - \theta k S)^{-1}(I + (1-\theta)kS)$ is symmetric with eigenvalues $g(\mu) = \frac{1 + (1-\theta)k\mu}{1 - \theta k \mu}$. For $\mu \le 0$, $g(\mu) \le 1$, and $g(\mu) \ge -1$ if and only if $(1 - 2\theta)k|\mu| \le 2$, which holds by assumption when $\theta < \frac{1}{2}$. The spectral norm of a symmetric matrix is the largest modulus of its eigenvalues, so $\|G_S^m\|_2 \le 1$, $\|(I - \theta k S)^{-1}\|_2 \le 1$, and $\|G^m B^{-1}\|_2 = \|D G_S^m (I - \theta k S)^{-1} D^{-1}\|_2 \le \|D\|_2 \|D^{-1}\|_2$; the same holds for G^m, and $\|\cdot\|_{2,h}$ has the same operator norms as the Euclidean norm. Finally $\|D\|_2\|D^{-1}\|_2 = \max_j d_j / \min_j d_j \le \exp\left(\frac{J}{2}|\log\frac{p_-}{p_+}|\right)$, and with $x = \frac{|b|h}{2a} \le \frac{1}{2}$, $|\log\frac{p_-}{p_+}| = \log\frac{1+x}{1-x} = 2\sum_{i\ge 0}\frac{x^{2i+1}}{2i+1} \le \frac{2x}{1-x^2} \le \frac{8}{3}x$, so the exponent is at most $\frac{2}{3}|b|L/a$.

(iii) The exact solution satisfies $Bu^{n+1} = Cu^n + kF^n + kT^n$, with the same F^n as the scheme. Subtracting (16.6), $e^n = u^n - V^n$ satisfies $e^{n+1} = Ge^n + kB^{-1}T^n$ and $e^0 = 0$, so $e^n = k\sum_{m<n} G^{n-1-m} B^{-1} T^m$ and, by (ii), $\|e^n\|_{2,h} \le nk\, e^{|b|L/a} \max_m \|T^m\|_{2,h}$, with $nk \le T$ and $\|T^m\|_{2,h} \le \sqrt{L}\max_j |T_j^m|$.

(iv) One step maps $W = (\ell_n, V^n, \rho_n) \in \mathbb{R}^{J+1}$ to $Z = (\ell_{n+1}, V^{n+1}, \rho_{n+1})$ with $\max_j |Z_j| \le \max(|\ell_{n+1}|, |\rho_{n+1}|, \max_j |W_j|)$ (discrete maximum principle). Indeed, $R_j = W_j + (1-\theta)k(L_h W)_j$ has the weights $(1-\theta)kp_\pm$ and $1 - (1-\theta)k(2a/h^2 + c)$, nonnegative with sum $1 - (1-\theta)kc \le 1$, so $|R_j| \le \max_i |W_i|$. Next, $Z_j - \theta k(L_h Z)_j = R_j$ for $1 \le j \le J-1$; if $|Z_j|$ is maximal at an interior node j^*, then $|R_{j^*}| \ge (1 + \theta k(\frac{2a}{h^2} + c))|Z_{j^*}| - \theta k(p_- + p_+)|Z_{j^*}| = (1 + \theta k c)|Z_{j^*}| \ge |Z_{j^*}|$. With zero boundary values this gives $\|G\|_\infty \le 1$ and $\|B^{-1}\|_\infty \le 1$, and the error bound follows as in (iii). The difference of two solutions solves the scheme with zero initial values and boundary values $\ell - \ell'$, $\rho - \rho'$; induct on n. □

For $r < 0$ the same proofs hold, with the bounds multiplied by a constant depending only on $|r|T$, provided $k|r| \le \frac{1}{2}$. The condition $h|b| \le a$ is mild: $h \le \frac{2}{3}$ for $\sigma = 20\%$, $r = 5\%$.

Remark 16.11 (The Lax equivalence theorem). For a consistent linear scheme approximating a well-posed linear initial-value problem, stability is necessary and sufficient for convergence (Lax and Richtmyer (1956); Strikwerda (2004), Theorem 1.5.1, proved in Chapter 10). We proved the sufficiency, with a rate; Theorem 16.8(ii) shows the necessity at work. Theorem 16.10 also prices the schemes: for an error ε, the explicit scheme ($k \le h^2/(2a)$) and the implicit one (error $k + h^2$) need work JN of the order of $\varepsilon^{-3/2}$, Crank–Nicolson with $k \approx h \approx \sqrt{\varepsilon}$ only ε^{-1}, if its hypotheses hold; the next section is about when they do not.

Remark 16.12 (Grids in the price itself). On the grid $x_j = j\,\delta x$, $0 \le j \le M$, the explicit scheme for the equation in x reads $U_j \leftarrow a_j U_{j-1} + b_j U_j + c_j U_{j+1}$ with $a_j = \frac{1}{2}\delta t(\sigma^2 j^2 - rj)$, $b_j = 1 - \delta t(\sigma^2 j^2 + r)$ and $c_j = \frac{1}{2}\delta t(\sigma^2 j^2 + rj)$. The local ratio $\frac{1}{2}\sigma^2 j^2 \delta t$ grows with j, and the middle weights are nonnegative if and only if $\delta t(\sigma^2(M-1)^2 + r) \le 1$: the top of the grid sets the time step, and refining in x at fixed δt eventually breaks the scheme. With $K = 100$, $\sigma = 20\%$,

$r = 5\%$, $T = 1$, $x_M = 400$ and 10^4 time steps, the middle weights are nonnegative for $M \le 500$; the spectral radius of the iteration matrix stays below 1 up to $M = 509$ and equals 1.0055 for $M = 510$, which multiplies errors by about $1.0055^{10^4} \approx 10^{24}$: the prices explode near x_M. In the log-price, $\lambda = ak/h^2$ is the same at every node.

▶ Interactive demo: the explicit scheme just below and above its stability limit, in the log-price and on a grid in the price itself, where refining the grid makes it explode.

16.3 The Black–Scholes equation on a grid

Domain and boundary values. For a call or a put we take $[y_{\min}, y_{\max}] = [\log S_0 - m\sigma\sqrt{T}, \log S_0 + m\sigma\sqrt{T}]$ and impose at the ends the discounted payoff of zero volatility, the limit of the price far from the strike: for the call $\ell_n = 0$, $\rho_n = e^{y_{\max}} - Ke^{-r\tau_n}$; for the put $\ell_n = Ke^{-r\tau_n} - e^{y_{\min}}$, $\rho_n = 0$. By Theorem 16.10(iv), for the implicit scheme and for the others under its condition, this changes the discrete solution by at most the error in the boundary values: for the call, by put–call parity, the larger of the call price at $e^{y_{\min}}$ and the put price at $e^{y_{\max}}$. With the parameters of Figure 16.1 and $m = 6$, both are below 10^{-8}. The influence of the boundary in fact decays fast away from it (Kangro and Nicolaides (2000)).

Non-smooth payoffs. Theorem 16.10 needs bounded derivatives up to maturity, and a call does not have them: $\partial_{yy}u = x^2\Gamma + x\Delta$, and $x^2\Gamma = xN'(d_+)/(\sigma\sqrt{\tau})$ is of order $K/(\sigma\sqrt{2\pi\tau})$ near $(0, \log K)$ (Theorem 6.14). The kink of the payoff feeds every frequency, and the amplification factors (Figure 16.2, right) tell what happens to them. The implicit scheme damps the high ones at each step, and the explicit scheme under $\lambda \le \frac{1}{2}$ is a positive average. The Crank–Nicolson factor $(1 - 2\lambda s)/(1 + 2\lambda s)$, however, tends to -1 as $\lambda \to \infty$: high frequencies survive, changing sign at every step. With k proportional to h, the natural choice for a second-order scheme, λ grows like $1/h$: oscillations appear in the delta and the gamma near the strike, and the price loses its second order. With a fixed ratio k/h, the Crank–Nicolson delta does not converge and its gamma diverges, in the maximum norm (Giles and Carter (2006)).

Rannacher's start. The remedy of Rannacher (1984) replaces the first Crank–Nicolson steps by implicit ones. Giles and Carter (2006) prove that replacing the first two steps by four implicit half-steps restores second-order convergence in the maximum norm for the value, delta and gamma of calls and puts, and of digitals when the payoff is averaged over each cell. The half-steps multiply a high frequency by $(1 + 2\lambda s)^{-4}$, after which Crank–Nicolson proceeds much as if the data were smooth. Treatments of the payoff complement this start (they do not remove the oscillation: in Figure 16.3 the strike is already a node): place the strike on a node or midway between two, or average the payoff over each cell, which digitals need (Pooley et al. (2003)); compare the smoothing of a tree (Theorem 6.7, Exercise 6.8).

On successive grids (Figure 16.3) the error of the explicit scheme decreases by factors 4.2, 3.9, 4.1, 4.0, that of the implicit scheme by factors close to 2 (with $k = h/\sigma$ the time error dominates), and that of Crank–Nicolson with Rannacher's start by factors close to 4. Plain Crank–Nicolson is erratic, and its gamma at the strike is off by 0.30 on the finest grid, where the true gamma is 0.019.

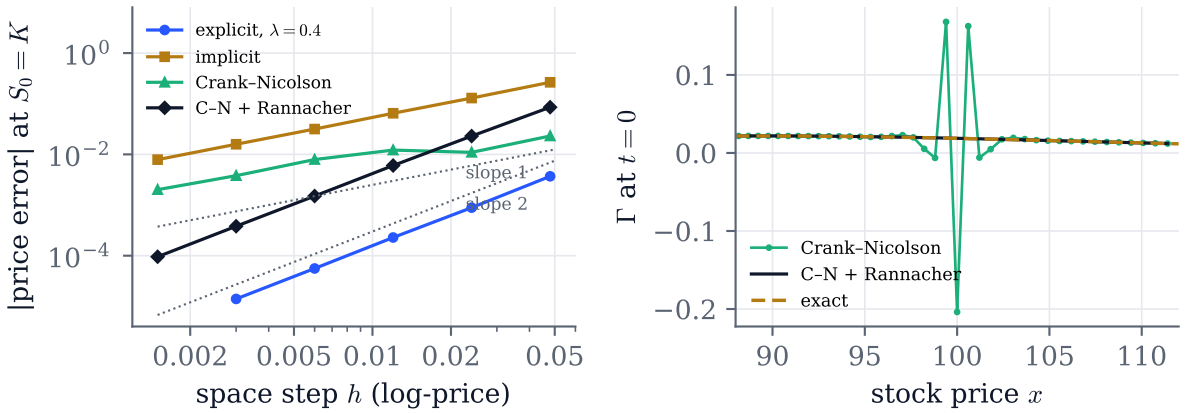

Figure 16.3: The call of Figure 16.1 on grids in log-price centered at $S_0 = K$, $m = 6$. Left: error of the price at S_0 against h, for $J = 50, \dots, 1600$; explicit scheme with $k = 0.4\,h^2/a$, the others with $k = h/\sigma$. Right: the gamma at $t = 0$ from Crank–Nicolson with $J = 400$ and $N = 25$ ($\lambda = ak/h^2 \approx 22$), with and without Rannacher's start, and the Black–Scholes gamma.

▶ Interactive demo: convergence orders for any strike and maturity, and the Crank–Nicolson gamma with and without Rannacher's start as the ratio k/h grows.

16.4 The American put on a grid

As in Theorem 4.13(ii) and Theorem 13.12, the price of the American put is the largest expected discounted payoff over exercise rules: with $S_u^{t,x} = x \exp\left(\sigma(\widetilde{W}_u - \widetilde{W}_t) + (r - \frac{\sigma^2}{2})(u-t)\right)$,

$$v(t,x) = \sup_{\tau} \mathbb{E}^{\mathbb{Q}}\left[e^{-r(\tau - t)}\left(K - S_\tau^{t,x}\right)^+\right], \qquad t \in [0,T],\ x > 0, \tag{16.7}$$

over the stopping times τ of the filtration of $(\widetilde{W}_u - \widetilde{W}_t)_{u \ge t}$ with values in $[t,T]$. With $g(x) = (K-x)^+$ and $\mathcal{A}v = \partial_t v + rx\,\partial_x v + \frac{1}{2}\sigma^2 x^2 \partial_{xx} v - rv$, v solves the *linear complementarity problem*

$$v \ge g, \qquad \mathcal{A}v \le 0, \qquad (v-g)\,\mathcal{A}v = 0, \qquad v(T,\cdot) = g : \tag{16.8}$$

at each point either exercise is optimal and $v = g$, or v solves the Black–Scholes equation (stated and discussed in Shreve (2004b), Section 8.4; Jaillet et al. (1990) give the rigorous form, a variational inequality of which v is the unique solution). If $r > 0$, the *exercise region* $\{x : v(t,x) = g(x)\}$ at a time $t < T$ is an interval $(0, b(t)]$, with b nondecreasing and $L^* \le b(t) < K$ for the perpetual level L^* of Theorem 13.12; if $r = 0$ it is empty (Exercise 16.7). The price is continuously differentiable in x across $b(t)$ (smooth fit; Shreve (2004b), Section 8.4) but in general not twice, as for the perpetual put, whose second derivative jumps at L^*. Near maturity $K - b(t) \sim \sigma K \sqrt{(T-t)\,|\log(T-t)|}$ when $r > 0$ (Barles et al. (1995)). No closed form is known.

The projected scheme. In log-price, each step of the θ-scheme asks for the complementarity conditions instead of the equation: given V^n, find V^{n+1} with

$$V^{n+1} \ge \mathbf{g}, \qquad BV^{n+1} - q^n \ge 0, \qquad (V^{n+1} - \mathbf{g}) \cdot (BV^{n+1} - q^n) = 0, \tag{16.9}$$

where $q^n = CV^n + kF^n$, $\mathbf{g}_j = (K - e^{y_j})^+$, inequalities hold entry by entry, and the boundary values are $\ell_n = K - e^{y_{\min}}$ and $\rho_n = 0$. The *projected SOR* method (PSOR) of Cryer (1971) sweeps over the nodes $j = 1, \dots, J-1$ with a relaxation parameter ω:

$$z_j^{(i+1)} = \max\Big(\mathbf{g}_j,\ z_j^{(i)} + \omega\Big(\frac{q_j - B_{j,j-1}z_{j-1}^{(i+1)} - B_{j,j+1}z_{j+1}^{(i)}}{B_{jj}} - z_j^{(i)}\Big)\Big), \tag{16.10}$$

a Gauss–Seidel step for $Bz = q$ when $\omega = 1$, overrelaxed when $\omega > 1$, projected on $\{z \ge \mathbf{g}\}$.

Proposition 16.13 (American put on a grid)**.** *Let $c \ge 0$, $\theta \in (0,1]$, and $h|b| \le a$.*

(i) For all $q, \mathbf{g} \in \mathbb{R}^{J-1}$, the problem $z \ge \mathbf{g}$, $Bz - q \ge 0$, $(z - \mathbf{g}) \cdot (Bz - q) = 0$ has exactly one solution.

(ii) For every $\omega \in (0,2)$ and every starting point, the iterates (16.10) converge to it.

(iii) If moreover $(1-\theta)k(2a/h^2 + c) \le 1$, and V_E solves (16.6) with $V_E^0 \le V^0$ and boundary values not larger than ℓ_n, ρ_n (a European put), then $V_E^n \le V^n$ for every n.

Proof. Let D, S be as in the proof of Theorem 16.10(ii), and $P = I - \theta k S$, symmetric with eigenvalues $1 - \theta k\mu_m \ge 1$, so positive definite; then $B = DPD^{-1}$.

(i) Write $z = Dw$. Then $Bz - q = D(Pw - D^{-1}q)$, and since D is diagonal with positive entries, z solves the problem if and only if w solves it with P, $D^{-1}q$, $D^{-1}\mathbf{g}$ in place of B, q, $\mathbf{g}$. The function $F(w) = \frac{1}{2}w \cdot Pw - w \cdot D^{-1}q$ is strictly convex and $F(w) \to \infty$ as $|w| \to \infty$, since $w \cdot Pw \ge |w|^2$; so it has a minimizer w^* on the closed convex set $\{w \ge D^{-1}\mathbf{g}\}$. For each j, $t \mapsto F(w^* + te_j)$ is minimal at $t = 0$ over $t \ge (D^{-1}\mathbf{g})_j - w_j^*$, so its derivative $(Pw^* - D^{-1}q)_j$ at 0 is nonnegative, and zero if $w_j^* > (D^{-1}\mathbf{g})_j$: w^* solves the problem. If w and w' both solve it, then $(w - w') \cdot (Pw - D^{-1}q) = -(w' - D^{-1}\mathbf{g}) \cdot (Pw - D^{-1}q) \le 0$, and symmetrically; adding, $(w - w') \cdot P(w - w') \le 0$, so $w = w'$.

(ii) If $z^{(i)} = Dw^{(i)}$, dividing (16.10) by $d_j > 0$ gives the same iteration for w with P, $D^{-1}q$, $D^{-1}\mathbf{g}$, because $P_{jl} = d_j^{-1}B_{jl}d_l$. For a symmetric positive definite matrix, projected SOR converges for every $\omega \in (0,2)$ (Cryer (1971)).

(iii) First, $Bx \ge 0$ implies $x \ge 0$: if the smallest entry x_{j^*} were negative, then, as $-\theta k p_\pm \le 0$ and $x_{j^*\pm1} \ge x_{j^*}$, $(Bx)_{j^*} \le (1 + \theta kc)x_{j^*} < 0$ (a missing neighbor at $j^* = 1$ or $J-1$ only helps). Next, C has nonnegative entries and F^n is nondecreasing in the boundary values. If $V_E^n \le V^n$, then $BV_E^{n+1} = CV_E^n + kF_E^n \le q^n \le BV^{n+1}$, so $V_E^{n+1} \le V^{n+1}$. □

The grid's exercise boundary at τ_n is the largest node where $V_j^n = \mathbf{g}_j > 0$. Our code sweeps the nodes in red–black order (odd, then even), which is Gauss–Seidel for a reordering of the unknowns, so (ii) still applies. Because the exercise region of the put lies at the bottom of the grid, (16.9) can also be solved directly, by one elimination from the top and a projected back-substitution from the bottom: the algorithm of M. Brennan and E. Schwartz (*Journal of Finance*, 1977), justified by Jaillet et al. (1990) for steps whose matrices have the sign structure of the implicit scheme. With Crank–Nicolson, Rannacher's start, $J = 800$ and $N = 400$, outside that setting, it agrees with PSOR (6.089962 against 6.089961): the put of Figure 16.4 costs 6.0900 at $S_0 = K$, against 6.0902 from a 5000-step tree and 5.5735 for the European put; $b(0) \approx 80.8$, above $L^* = 71.4$. Here $\lambda = ak/h^2 \approx 5.6$, so (iii) does not apply; the computed American prices still dominate the European ones, which is observed, not proved (for the implicit scheme (iii)

applies, and the tests check it). PSOR with $\omega = 1.2$ needs about 11 sweeps per step for $J = 400$, $N = 200$, and about 80 for $J = 3200$, $N = 1600$.

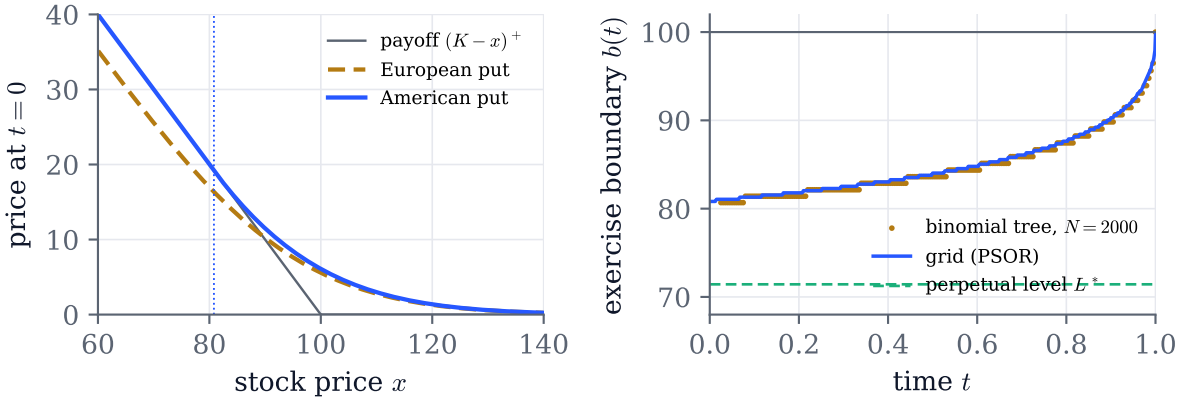

Figure 16.4: The American put with $K = 100$, $r = 5\%$, $\sigma = 20\%$, $T = 1$, by projected Crank–Nicolson with PSOR. Left: American and European prices at $t = 0$; the dotted line is $b(0)$. Right: the exercise boundary $b(t)$ from the grid and from a binomial tree with 2000 steps, with K and the perpetual level $L^* = 2rK/(2r + \sigma^2)$.

▶ Interactive demo: the American put and its exercise boundary for any rate, volatility and maturity, from the grid and from a tree, and the number of PSOR sweeps against ω.

16.5 Physics-informed neural networks

A *physics-informed neural network* (PINN) represents the solution by a network $u_w(\tau, y)$ with parameters w, and asks it to satisfy the equation at sampled points (Raissi et al. (2019); the closely related deep Galerkin method is due to Sirignano and Spiliopoulos (2018)). For (16.3), with points (τ_i, y_i) in the domain, y_i' at $\tau = 0$ and (τ_i'', y_i'') on the lateral boundary, where the values are β_i, the parameters minimize

$$\frac{1}{N_r}\sum_i \left|\partial_\tau u_w - a\,\partial_{yy}u_w - b\,\partial_y u_w + c\,u_w\right|^2(\tau_i, y_i) + \frac{1}{N_0}\sum_i \left|u_w(0, y_i') - g(e^{y_i'})\right|^2 + \frac{1}{N_b}\sum_i \left|u_w(\tau_i'', y_i'') - \beta_i\right|^2,$$

by a stochastic gradient method such as Adam, the derivatives in (τ, y) and in w being computed by automatic differentiation.

Remark 16.14 (What PINNs solve, and how they fail).

(a) *What they solve.* The same equation as a grid, over a class of functions instead of a set of nodes. There is no mesh, so the number of state variables enters only through the network and the sampling, which is the attraction where grids are out of reach (Theorem 12.6): Sirignano and Spiliopoulos (2018) solve free-boundary problems of the kind American options on many assets lead to, in up to 200 dimensions. The network is a smooth function, whose derivatives give the Greeks.

(b) *No a priori control.* A small loss does not by itself mean a small error: turning the residuals at finitely many points into an error bound needs a stability estimate for the equation and a bound on the quadrature error (Mishra and Molinaro (2023)). Nor is a small loss guaranteed: the problem is not convex, and its result depends on the initialization.

(c) *Optimization.* The loss terms have different scales and badly unbalanced gradients (Wang et al. (2021)); training fails for strong convection or reaction terms on problems a grid solves easily (A. S. Krishnapriyan, A. Gholami, S. Zhe, R. M. Kirby and M. W. Mahoney, *NeurIPS*, 2021); and networks fit smooth components first, so kinks are learned last.

(d) *Non-smooth payoffs.* The residual involves $\partial_{yy}u$, of order $\tau^{-1/2}$ near the strike at maturity; the network smooths the kink and its error concentrates there, the difficulty that Rannacher's start resolves for Crank–Nicolson.

(e) *Cost.* In one or two dimensions a grid is faster and more accurate by orders of magnitude. In the notebook a network with one hidden layer of 24 units, trained with hand-written gradients for 4000 Adam steps (about three seconds), reaches a loss of 8×10^{-5} on the call with $K = 1$, $r = 5\%$, $\sigma = 20\%$; its largest error is 1.6×10^{-2} near the strike at $\tau = 0.05$ and about 2×10^{-3} at $\tau = 0.5$ and 1, while Crank–Nicolson with Rannacher's start on 200 nodes is accurate to about 10^{-4}, in about a millisecond.

▶ Interactive demo: a physics-informed network for the Black–Scholes equation in NumPy, with hand-written gradients: its loss, its error near the strike and its cost, against a grid.

Further reading

Glasserman (2004) is the reference on Monte Carlo in finance (Chapters 1, 4 and 6 for the principles, variance reduction and discretization). Wilmott et al. (1995), Chapters 8 and 9, present finite differences and projected methods for American options at the level of these notes; Strikwerda (2004), Chapters 1, 2 and 10, develops consistency, stability, von Neumann analysis and the Lax equivalence theorem. Giles and Carter (2006) analyze Rannacher's start for option payoffs, and Jaillet et al. (1990) the variational inequality of the American put and its approximation.

Part V

Extras

Chapter 17

Arbitrage-free option surfaces

Chapter 15 ended with a programme: price exotics and hedges in one model calibrated to all the vanillas, starting from a surface of call prices that is free of static arbitrage. Markets do not provide such a surface. They provide bid and ask quotes at a few maturities and strikes, and their mid prices need not be free of arbitrage. This chapter presents two recent papers that turn quotes into arbitrage-free surfaces, and surfaces into arbitrage-free scenarios. *SANOS* (Buehler et al., 2026a) builds smooth surfaces that are free of static arbitrage by construction and calibrates them by linear programming inside the bid–ask spreads. *DYSANOS* (Buehler et al., 2026b) lets the parameters of these surfaces evolve in time, which generates a surface free of static arbitrage on every day, and tests what is left: dynamic arbitrage. The tools are those of Chapter 3 (convexity in the strike, discrete pricing measures) and Chapter 15 (implied volatility, Dupire's formula), plus one new notion: convex order. Both papers are preprints; results are cited with their numbering (arXiv versions 4), and we say which statements we prove here.

Setting. As in Chapter 15, r and q are constants, $F_T = S_0 e^{(r-q)T}$, and prices come from a pricing measure $\mathbb{Q}$ under which $(e^{-(r-q)t}S_t)$ is a martingale. Then $Z_T = S_T/F_T$ is a nonnegative martingale with $Z_0 = 1$, and in units of the forward, with *normalized strike* $x = K/F_T$,

$$c(T,x) = \frac{e^{rT}C(xF_T,T)}{F_T} = \mathbb{E}^{\mathbb{Q}}\big[(Z_T - x)^+\big], \qquad x \geq 0. \tag{17.1}$$

The stock is the call with strike 0, $c(T,0) = 1$, and parity reads $p(T,x) = c(T,x) - 1 + x$. Both papers work in these units (SANOS Section 2; DYSANOS Remark 2.1). From now on all prices are normalized; the results of Chapter 3 apply with $B = 1$ and $S_0 = 1$, since in these units the bond and the stock both cost 1.

17.1 Why an arbitrage-free surface

Three reasons. (a) Consistency (Section 15.6): exotics must be priced under one pricing measure calibrated to the vanillas, and quotes with a static arbitrage come from no pricing measure (Theorem 17.6 below), so no model reproduces them. (b) Dupire's formula needs a smooth surface with positive butterflies and calendar spreads (Theorem 15.8). (c) Quotes are finite and noisy, and interpolating them carelessly creates arbitrage that was not in the data.

Our running example is the Heston market of Figure 15.3 ($v_0 = \theta = 0.04$, $\kappa = 2$, $\xi = 0.5$, $\rho = -0.7$), quoted at the expiries $T_j = 1, 2, 3, 6, 9, 12$ months and the 21 strikes $x = \exp(0.2z\sqrt{T_j})$, $z = -3, -2.75, \dots, 2$. The half-spread is half a volatility point times the vega, and at least 10^{-4}; the mid is the Heston price plus a uniform error of at most 0.8 half-spreads, so the Heston price lies inside every spread (`code/src/ieor222/surfaces.py`). Already the mids are not free of arbitrage: at four of the six expiries some butterfly of adjacent mids has a negative price.

The common practice is to interpolate implied volatilities. Linear interpolation creates kinks, and a kink of the wrong sign is an arbitrage.

Proposition 17.1 (Concave kinks are butterfly arbitrages)**.** *Let $w : \mathbb{R} \to (0, \infty)$ be continuous, and continuously differentiable on $(-\infty, k_0]$ and on $[k_0, \infty)$ with one-sided derivatives $w'(k_0\pm)$, and let $c(x) = b(\log x, \sqrt{w(\log x)})$ be the normalized prices with total implied variance w (see (15.2)). Then c has one-sided derivatives at $x_0 = e^{k_0}$, and*

$$\partial_x c(x_0+) - \partial_x c(x_0-) = \frac{N'(d_+)}{2x_0\sqrt{w(k_0)}}(w'(k_0+) - w'(k_0-)), \qquad d_+ = -\frac{k_0}{\sqrt{w(k_0)}} + \frac{\sqrt{w(k_0)}}{2}.$$

If $w'(k_0+) < w'(k_0-)$, the butterfly $c(x_0 - h) - 2c(x_0) + c(x_0 + h)$ is negative for all small $h > 0$, and the prices come from no pricing measure.

Proof. On each side of x_0, c is continuously differentiable with $\partial_x c = (\partial_k b + \partial_s b\, w'/(2\sqrt{w}))/x$, where $\partial_s b = N'(d_+)$ (proof of Theorem 15.2) and $\partial_k b$ is continuous across x_0; subtract the two one-sided limits. By Taylor's formula on each side, $c(x_0 \pm h) = c(x_0) \pm h\, \partial_x c(x_0\pm) + o(h)$, so the butterfly equals $h(\partial_x c(x_0+) - \partial_x c(x_0-)) + o(h)$. Under a pricing measure c would be convex (Theorem 3.12), with nonnegative butterflies. □

The jump of $\partial_x c$ is the mass that the implied law puts at x_0 (Theorem 3.18): a concave kink is a negative atom. Linear interpolation of the total variance of noisy mids (or of their implied volatility, which has the same kinks) has a kink at every quote, concave about half of the time: in our data at 7 to 11 of the 19 inner quotes of each expiry, with negative atoms down to -0.065. A cubic spline in implied volatility has no kinks, but it oscillates: its density is negative on whole intervals (Figure 17.1), and at the one-month expiry, where it meets its flat extrapolation, it breaks the slope bound $\partial_x c \geq -1$ of Theorem 3.12 (a put spread with a negative price).

Interpolating the *prices* linearly is safe: if quotes at the strikes $0 = x^0 < x^1 < \cdots < x^n$ satisfy the conditions of Theorem 3.14, their linear interpolation is the call function of the discrete law built there (SANOS Definition 2.11 and Propositions 2.13–2.14). It is also, in a precise sense, too high.

Proposition 17.2 (Linear interpolation is the most expensive)**.** *Let $c : [0, \infty) \to [0, \infty)$ be convex and nonincreasing, let $0 = x^0 < \cdots < x^n$ with $c(x^n) = 0$, and let $\hat{c}$ be the linear interpolation of $(x^i, c(x^i))$, with $\hat{c} = 0$ beyond x^n. Then $c \leq \hat{c}$, and $c < \hat{c}$ inside every interval (x^{i-1}, x^i) on which c is strictly convex.*

Proof. On $[x^{i-1}, x^i]$ a convex function lies below its chord, strictly inside the interval if it is strictly convex there. Beyond x^n, $0 \leq c \leq c(x^n) = 0$. □

So every surface free of static arbitrage that matches the quotes at the strikes lies below their

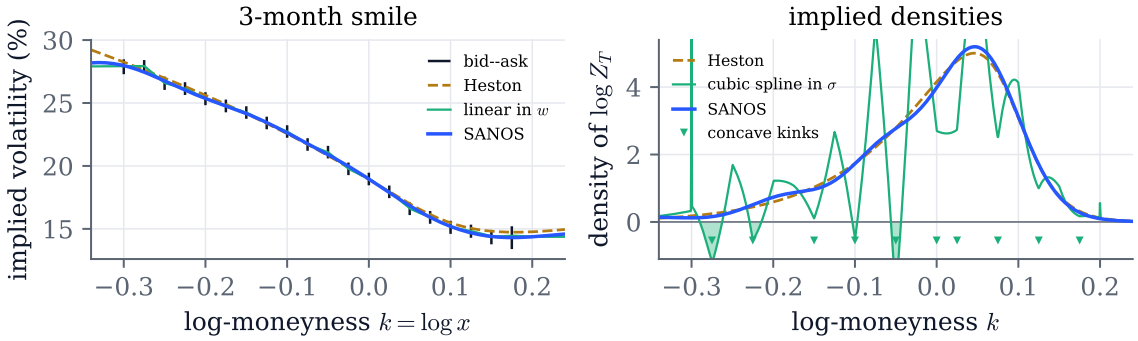

Figure 17.1: The 3-month quotes of the running example. Left: bid–ask intervals in implied volatility, the Heston smile, linear interpolation of the mid total variance, and the SANOS fit of Section 17.3.1. Right: implied densities of $\log Z_T$; the cubic spline through the mid volatilities oscillates beyond the frame and is negative on the shaded intervals; triangles mark the concave kinks of the linear interpolation, each a negative atom.

linear interpolation, strictly between strikes if its density is positive there; SANOS attributes this observation to Buehler (2006). Conversely, smoothing the discrete law of the quotes raises prices (Theorem 17.8(iii) below). Smoothness must be paid for by moving prices, and the natural place to move them is inside the bid–ask spreads. This is the idea of SANOS.

17.2 Static arbitrage on a grid

Definition 17.3 (Static arbitrage)**.** Let $0 < T_1 < \cdots < T_M$ and, for each j, let $X_j \subset [0,\infty)$ be a set of normalized strikes with call prices $c(T_j, x)$, $x \in X_j$. These prices are *free of static arbitrage* if there are a probability space and a nonnegative martingale $(Z_j)_{j=0,\dots,M}$ with $Z_0 = 1$ such that $c(T_j, x) = \mathbb{E}[(Z_j - x)^+]$ for all j and $x \in X_j$. A surface $c : [0,\infty)^2 \to \mathbb{R}$ is free of static arbitrage if $c(T, x) = \mathbb{E}[(Z_T - x)^+]$ for all $T, x \geq 0$, for a nonnegative martingale $(Z_T)_{T\geq 0}$ with $Z_0 = 1$; it is *strictly* free of static arbitrage if Z can be chosen positive.

This is SANOS Definition 2.1 (which asks for a positive Z) and DYSANOS Definition 2.3. The name refers to *semi-static* portfolios: fixed positions in the calls and the bond, and a position in the stock that may be changed, within bounds, at the dates T_j. If the prices come from a martingale, the normalized cost of such a portfolio is the expectation of its payoff (the gains of the stock position form a martingale transform with mean zero, Theorem 2.8), so, as in Theorem 3.6, none is an arbitrage. Conversely, Davis and Hobson (2007), Theorem 4.2, show that finitely many quotes at several maturities that come from no martingale admit a model-independent arbitrage or a weak one (like that of Exercise 3.8). A calendar violation $c(T_2, x) < c(T_1, x)$, $T_1 < T_2$, is exploited by buying the T_2 call, selling the T_1 call and, if the latter ends in the money, selling the stock forward from T_1 to T_2 (Exercise 17.1).

Across maturities the key notion is convex order. For laws μ, ν with finite means, $\mu \leq_{\mathrm{cx}} \nu$ if $\int f \,\mathrm{d}\mu \leq \int f \,\mathrm{d}\nu$ for every convex f. By conditional Jensen, the marginals of a martingale increase in convex order; the converse is a theorem.

Lemma 17.4 (Convex order on a grid)**.** *Let μ, ν be probability measures on a finite set $G = \{g^0 < g^1 < \cdots < g^m\} \subset [0, \infty)$ with the same mean. Then $\mu \leq_{\text{cx}} \nu$ if and only if $\int (z - g^i)^+ \mu(\mathrm{d}z) \leq \int (z - g^i)^+ \nu(\mathrm{d}z)$ for $i = 1, \ldots, m-1$.*

Proof. The condition is necessary because $z \mapsto (z - g^i)^+$ is convex. Conversely, let f be convex and let $\hat{f}$ be continuous, piecewise linear, equal to f on G, with kinks only at $g^1, \ldots, g^{m-1}$ (affine before g^1 and after g^{m-1}). The slopes of the chords of a convex function increase, so the changes of slope w_i of $\hat{f}$ are nonnegative, and Theorem 3.15 gives $\hat{f}(z) = \hat{f}(0) + \hat{f}'(0+)z + \sum_{i=1}^{m-1} w_i (z - g^i)^+$. Both laws live on G, where $f = \hat{f}$, and have the same mass and mean, so $\int f \, \mathrm{d}\nu - \int f \, \mathrm{d}\mu = \sum_i w_i \big(\int (z - g^i)^+ \mathrm{d}\nu - \int (z - g^i)^+ \mathrm{d}\mu \big) \geq 0$. □

Lemma 17.5 (Strassen)**.** *Let μ, ν be probability measures on $\mathbb{R}$ with finite means. There are random variables $Z_1 \sim \mu$ and $Z_2 \sim \nu$ with $\mathbb{E}[Z_2 \mid Z_1] = Z_1$ if and only if $\mu \leq_{\text{cx}} \nu$ (Strassen, 1965, Theorem 8).*

For laws on finite sets the coupling solves a linear feasibility problem: the code finds one with `linprog`. Chaining couplings gives a Markov chain, hence a discrete-time martingale.

Theorem 17.6 (Quotes on a common grid)**.** *Let $0 = x^0 < x^1 < \cdots < x^n$ be strikes common to the maturities $T_1 < \cdots < T_M$, and let c_j^i be the call prices, with $c_j^0 = 1$ (the stock) and $c_j^n = 0$ (a strike beyond which no call has value). Let $s_j^i = (c_j^i - c_j^{i-1})/(x^i - x^{i-1})$, $i = 1, \ldots, n$. The prices are free of static arbitrage if and only if*

(a) $-1 \leq s_j^1 \leq s_j^2 \leq \cdots \leq s_j^n \leq 0$ for every j (bounds, monotonicity, butterflies), and

(b) $c_j^i \leq c_{j+1}^i$ for every i and $j < M$ (calendar spreads).

In that case Z can be chosen to be a Markov chain on the strikes with $\mathbb{Q}(Z_j = x^i) = p_j^i := s_j^{i+1} - s_j^i$ ($s_j^0 := -1$, $s_j^{n+1} := 0$); this chain is positive if and only if $s_j^1 = -1$ for every j.

Proof. Necessity. If $c_j^i = \mathbb{E}[(Z_j - x^i)^+]$, then $x \mapsto \mathbb{E}[(Z_j - x)^+]$ equals 1 at 0 and is convex with difference quotients in $[-1, 0]$ (proof of Theorem 3.12), which gives (a); conditional Jensen, $\mathbb{E}[(Z_{j+1} - x)^+ \mid Z_j] \geq (Z_j - x)^+$, gives (b).

Sufficiency. For each j, the construction in the proof of Theorem 3.14 (with $B = S_0 = 1$, and the last atom at x^n because $c_j^n = 0$) gives a law μ_j with weights p_j^i on the strikes, mean 1, and call function the linear interpolation $\hat{c}_j$ of the c_j^i, with $\hat{c}_j = 0$ beyond x^n. By (b), $\hat{c}_j \leq \hat{c}_{j+1}$ at the strikes, hence everywhere, both being affine between strikes; so $\mu_j \leq_{\text{cx}} \mu_{j+1}$ by Theorem 17.4, and $\delta_1 \leq_{\text{cx}} \mu_1$ by Jensen's inequality. Let $\pi_j(z, \cdot)$ be the conditional laws of a coupling of μ_{j-1} and μ_j given by Theorem 17.5 ($\mu_0 = \delta_1$), and let Z be the Markov chain with $Z_0 = 1$ and transitions π_j. Then $Z_j \sim \mu_j$ and $\mathbb{E}[Z_j \mid Z_0, \ldots, Z_{j-1}] = \mathbb{E}[Z_j \mid Z_{j-1}] = Z_{j-1}$. Finally $\mathbb{Q}(Z_j = 0) = p_j^0 = 1 + s_j^1$. □

In normalized strikes a common grid means strikes proportional to the forwards; for such strikes Davis and Hobson (2007), Corollary 4.1, reduce their conditions to those of single maturities and calendar spreads, as in Theorem 17.6. SANOS Theorem 2.6 treats different grids: it replaces (b) by the requirement that the linear interpolation of slice j lie below c_{j+1} at the strikes of $j + 1$, which is sufficient but not necessary (Exercise 17.2); exact conditions are in Davis and Hobson (2007), Theorem 4.2, and Cousot (2007). For a continuum of strikes and maturities, SANOS Theorem 2.2 characterizes the surfaces strictly free of static arbitrage by $c(T, 0) = 1$,

$\partial_x c(T,0) = -1$, $c(T,x) \to 0$ as $x \to \infty$, convexity in x and monotonicity in T; its one-maturity part goes back to Buehler (2006), and the martingale comes from Kellerer's theorem (Kellerer, 1972), which extends Theorem 17.5 to a continuum of times. In total implied variance these conditions correspond to those of Theorem 15.6.

17.3 SANOS

The linear interpolation of Theorem 17.2 is the call function of a discrete law, a sum of atoms. SANOS replaces each atom by a small lognormal law. Write

$$\mathrm{Call}(s,x,v) = sN(d_+) - xN(d_-), \qquad d_\pm = \frac{\log(s/x) \pm v/2}{\sqrt{v}}, \qquad \mathrm{Call}(s,x,0) = (s-x)^+, \tag{17.2}$$

for $s > 0$, $x \geq 0$ and a total variance $v \geq 0$ (SANOS eq. (3)): $\mathrm{Call}(s,x,v) = \mathbb{E}[(sY - x)^+]$ when $\log Y \sim \mathcal{N}(-v/2, v)$, so $\mathbb{E}[Y] = 1$ (Theorem 6.11), and $\mathrm{Call}(s,x,v) = s\, b(\log(x/s), \sqrt{v})$.

Definition 17.7 (SANOS surface)**.** Fix *model strikes* $0 < K^1 < \cdots < K^N$ and *model expiries* $0 = T_0 < T_1 < \cdots < T_M$. A *martingale density* is a family $q_1, \ldots, q_M$ of probability vectors on the model strikes with $\sum_i q_j^i K^i = 1$ and $\sum_i q_j^i (K^i - x)^+$ nondecreasing in j for every x; let q_0 be the unit mass at 1. Let $v : [0,\infty) \to [0,\infty)$ be continuous and nondecreasing with $v(0) = 0$ (the *smoothing variance*), and $\alpha_j : [T_{j-1}, T_j] \to [0,1]$ continuous and nondecreasing with $\alpha_j(T_{j-1}) = 0$, $\alpha_j(T_j) = 1$. For $T \in (T_{j-1}, T_j]$ let $\mu_T = \alpha_j(T) q_j + (1 - \alpha_j(T)) q_{j-1}$, and $\mu_T = q_M$ for $T > T_M$. The *SANOS surface* is

$$c(T,x) = \sum_i \mu_T^i \, \mathrm{Call}(K^i, x, v(T)), \qquad T > 0,\ x \geq 0,$$

where the sum runs over the atoms of μ_T ($K^1, \ldots, K^N$, and 1 for $T \leq T_1$).

This is SANOS eq. (13)–(15) with the lognormal driver of their production model (Section 3.1), in which $v(T_j) = \eta V_j$ with V_j the at-the-money total implied variance of the market and a smoothing parameter $\eta \in [0,1)$, $\eta = 0.25$ being their recommended default. DYSANOS (Definition 2.6) interpolates the total volatility $\sqrt{v(T)}$ linearly in T.

Theorem 17.8 (SANOS surfaces are strictly free of static arbitrage)**.** *In Theorem 17.7:*

(i) there is a positive martingale $(Z_T)_{T \geq 0}$ with $Z_0 = 1$ and $c(T,x) = \mathbb{E}[(Z_T - x)^+]$ for all $T, x \geq 0$: c is strictly free of static arbitrage;

(ii) if $v(T) > 0$, then $c(T,\cdot)$ is infinitely differentiable on $(0,\infty)$ with $\partial_{xx} c(T,x) = \sum_i \mu_T^i N'(d_-^i)/(x\sqrt{v(T)}) > 0$, $d_-^i = \big(\log(K^i/x) - v(T)/2\big)/\sqrt{v(T)}$: Z_T has a smooth positive density, every butterfly has a positive price, and every strike has an implied volatility;

(iii) $c(T_j,x) \geq \sum_i q_j^i (K^i - x)^+$, with equality for all x if $v(T_j) = 0$.

Proof. (i) We build two independent martingales. First, let $\xi_0 = 1, \xi_1, \ldots, \xi_M$ be a Markov chain whose transitions are couplings of q_{j-1} and q_j given by Theorem 17.5 (the q_j increase in convex order), so that $\xi_j \sim q_j$ and $\mathbb{E}[\xi_j \mid \xi_0, \ldots, \xi_{j-1}] = \xi_{j-1}$, and let $U_1, \ldots, U_M$ be uniform on $(0,1)$, independent of each other and of the chain. Put $X_0 = \xi_0 = 1$; for $T \in (T_{j-1}, T_j]$ put $X_T = \xi_j$ if

$U_j \le \alpha_j(T)$ and $X_T = \xi_{j-1}$ otherwise, and $X_T = \xi_M$ for $T > T_M$; then $X_T \sim \mu_T$, and $X_{T_j} = \xi_j$. Let $\mathcal{G}_T = \sigma(U_1, \dots, U_M, X_t : t \le T)$ and $\mathcal{H} = \sigma(U_1, \dots, U_M, \xi_0, \dots, \xi_{j-1})$. For $T_{j-1} \le s < t \le T_j$, $X_t - X_s = (\xi_j - \xi_{j-1})\mathbf{1}_E$ with $E = \{\alpha_j(s) < U_j \le \alpha_j(t)\}$. On E every X_u, $u \le s$, is a function of $\xi_0, \dots, \xi_{j-1}$ and the U's, so each $A \in \mathcal{G}_s$ satisfies $A \cap E = A' \cap E$ for some $A' \in \mathcal{H}$ (the sets A with this property form a σ-field containing the U's and the events $\{X_u \in B\}$, $u \le s$), and $\mathbb{E}[(X_t - X_s)\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[\xi_j - \xi_{j-1} \mid \mathcal{H}]\,\mathbf{1}_{E\cap A'}] = 0$, because ξ_j is independent of the U's given $\xi_0, \dots, \xi_{j-1}$. With the tower property, X is a $(\mathcal{G}_T)$-martingale. Second, let W be a Brownian motion independent of all this, and $Y_T = \exp\left(W_{v(T)} - v(T)/2\right)$, a martingale in its own filtration because v is nondecreasing. Then $Z = XY$ is a positive martingale in the filtration generated by both (by independence, the martingale identity holds on the π-system of sets $A \cap B$, $A \in \mathcal{G}_s$, $B \in \sigma(Y_u : u \le s)$, hence on the σ-field it generates by Dynkin's π–λ theorem; X is bounded). Conditioning on X_T, $\mathbb{E}[(Z_T - x)^+] = \sum_i \mu_T^i \mathbb{E}[(K^i Y_T - x)^+] = c(T, x)$.

(ii) Differentiate the finite sum: $\partial_x \mathrm{Call}(s, x, v) = -N(d_-)$ and $\partial_{xx}\mathrm{Call} = N'(d_-)/(x\sqrt{v})$. Since Z_T has a positive density, $\mathbb{Q}(Z_T > x) > 0$ and $\mathbb{Q}(Z_T < x) > 0$, and Theorem 15.2(ii) gives implied volatilities.

(iii) Jensen: $\mathbb{E}[(K^i Y - x)^+] \ge (K^i \mathbb{E}[Y] - x)^+$. □

The paper's proof of its Theorem 3.1 (for any positive martingale Y with full support, independent of X, such as a Heston price, SANOS Remark 3.2) checks the conditions of its Theorem 2.2: the strike conditions directly, and the calendar condition at the expiries by Jensen's inequality and the change of measure $\mathrm{d}\mathbb{P}^Y = Y_j\,\mathrm{d}\mathbb{P}$; the martingale then comes, through its Theorem 2.2, from Kellerer's theorem, and $Z = XY$ is its eq. (16). The construction above makes the martingale explicit, and shows how SANOS moves between expiries: X jumps from ξ_{j-1} to ξ_j at a random time. By (iii), smoothing the discrete law of the quotes raises prices; SANOS Proposition 3.8 shows that for quotes free of arbitrage and strictly inside their spreads, some positive amount of smoothing of the linear interpolation still prices inside them. Per-strike variances (Remark 3.3) keep (ii) but can break the calendar condition; the iterative model of Theorem 3.9, with per-strike variance increments, is again free of static arbitrage.

17.3.1 Calibration by linear programming

The prices of Theorem 17.7 are linear in q once v is fixed, and the constraints on q are linear.

Proposition 17.9 (SANOS calibration is a linear program)**.** *Fix model strikes with $K^1 < 1 < K^N$, expiries and smoothing variances $0 \le v_1 \le \cdots \le v_M$. Let the quotes be $(T_{j(\ell)}, x_\ell)$ with bid b_ℓ, ask a_ℓ, mid m_ℓ and weight $\omega_\ell > 0$, $\ell = 1, \dots, L$, and $\varepsilon \ge 0$. Let $\mathcal{Q}$ be the set of $q = (q_1, \dots, q_M)$ with*

$$\begin{aligned} q_j \ge 0, \qquad \sum_i q_j^i = 1, \qquad \sum_i q_j^i K^i = 1 \qquad (j = 1, \dots, M), \\ \sum_i q_j^i (K^i - K^l)^+ \ge \sum_i q_{j-1}^i (K^i - K^l)^+ \qquad (j = 2, \dots, M,\ l = 1, \dots, N), \end{aligned} \tag{17.3}$$

and $c_\ell(q) = \sum_i q_{j(\ell)}^i \mathrm{Call}(K^i, x_\ell, v_{j(\ell)})$. Then:

(i) $\mathcal{Q}$ is a nonempty compact polytope, and every $q \in \mathcal{Q}$ is a martingale density, so it defines a surface strictly free of static arbitrage by Theorem 17.8;

(ii) $J(q) = \sum_\ell \omega_\ell[(c_\ell - a_\ell)^+ + (b_\ell - c_\ell)^+ + \varepsilon|c_\ell - m_\ell|]$ *attains its minimum on* $\mathcal{Q}$*, and the minimum equals that of the linear program: minimize* $\sum_\ell \omega_\ell(u_\ell + d_\ell + \varepsilon e_\ell)$ *over* $q \in \mathcal{Q}$ *and* $u, d, e \geq 0$ *subject to* $c_\ell(q) - u_\ell \leq a_\ell$*,* $c_\ell(q) + d_\ell \geq b_\ell$ *and* $|c_\ell(q) - m_\ell| \leq e_\ell$*;*

(iii) with $\varepsilon = 0$*, the minimum is* 0 *if and only if some SANOS surface with these* v_j *prices every quote inside its spread.*

Proof. (i) $\mathcal{Q}$ is defined by finitely many linear constraints and lies in a product of simplices, so it is a compact polytope; it contains $q_1 = \cdots = q_M =$ the law with mean 1 on the two model strikes around 1. The calendar constraints at the model strikes give convex order by Theorem 17.4 (with G the model strikes; $(K^i - K^1)^+$ and $(K^i - K^N)^+$ are fixed by the mass and the mean), and v may be interpolated as in Theorem 17.7. (ii) J is continuous, hence attains its minimum on the compact $\mathcal{Q}$. For fixed q the constraints force $u_\ell \geq (c_\ell - a_\ell)^+$, $d_\ell \geq (b_\ell - c_\ell)^+$, $e_\ell \geq |c_\ell - m_\ell|$, with equality optimal since $\omega_\ell > 0$; so minimizing over the slacks gives $J(q)$. (iii) $J \geq 0$, with $J(q) = 0$ exactly when every $c_\ell(q) \in [b_\ell, a_\ell]$. □

This is the linear program of SANOS Section 4.2 for model expiries equal to market expiries (the paper also interpolates market expiries between model expiries). Their eq. (33) has a switch: in the position used in (17.3), the calendar constraint is on the weights and gives strict absence of arbitrage; in the other it is imposed on the smoothed prices, which, the paper notes, no longer guarantees strict absence of arbitrage. Their Remark 4.1 lists three treatments of the spread and recommends the penalty of (ii) with $\varepsilon = 10^{-8}$ and $\omega_\ell = 1/(a_\ell - b_\ell)$ (eq. (34)), which we use. On S&P 500 options on 6 May 2025 (48 expiries, 1000 options) the fit puts 91.4% of the options inside their spreads, the median error of the others is 21% of the half-spread, and it runs in under a second (Section 5).

Replication. On the running example, with 80 model strikes and $\eta = 0.25$, SciPy's `linprog` (with the HiGHS solver) solves the program in a fraction of a second. Every fitted price is inside its spread, up to the solver's tolerance (10^{-7} half-spreads), and the fitted implied volatilities are within about 0.4 points of Heston's except in the one-month wings (about 1 point). The solver returns a vertex of the feasible polytope, and the optimal q_j have only about ten nonzero weights out of 80 (8 to 13 in our run; the optimum need not be unique, and the value $\varepsilon = 10^{-8}$ is at the solver's tolerance, so another solver may return another vertex and slightly different numbers); the smoothing turns them into a density close to Heston's (Figure 17.2). On a fine grid of strikes and of maturities, including maturities between expiries, butterflies, calendar spreads and bounds hold up to rounding (10^{-8}); the two naive interpolations of Section 17.1 fail the butterfly test at every expiry.

▶ Interactive demo: naive interpolation against SANOS; the linear program, with sliders for the spreads, the noise and the model grid; the martingale behind a SANOS surface; a real DAX surface, downloaded at run time.

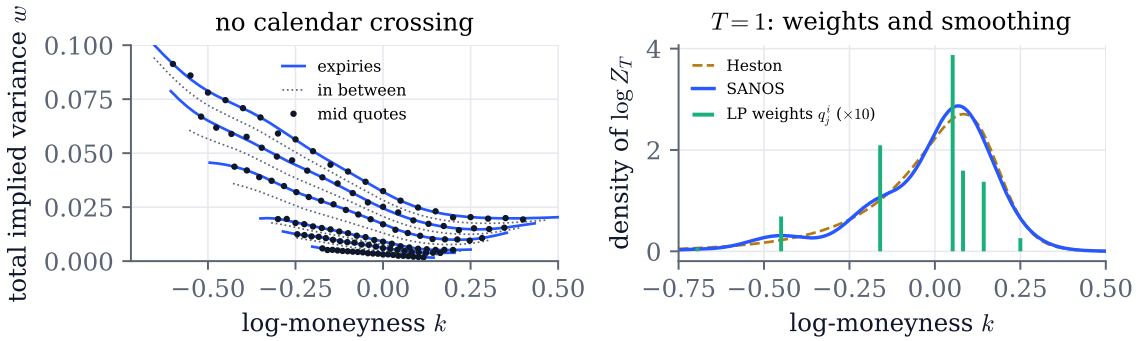

Figure 17.2: The SANOS fit of the running example ($\eta = 0.25$). Left: total implied variance at the six expiries and halfway between them, with the mid quotes; the curves are ordered in T (no calendar arbitrage). Right: at $T = 1$, the nonzero weights q_j^i of the linear program (scaled by 10), the smoothed SANOS density of $\log Z_T$, and Heston's.

17.4 Discrete local volatility

The constraints (17.3) can be traded for positivity constraints. Fix a grid $K^1 < \cdots < K^N$ containing 1, and for $\Delta > 0$ and *discrete local volatilities* $\Sigma^2, \dots, \Sigma^{N-1} \geq 0$ let (SANOS eq. (19))

$$w^{i+} = \frac{(\Sigma^i K^i)^2 \Delta}{(K^{i+1} - K^{i-1})(K^{i+1} - K^i)}, \qquad w^{i-} = \frac{(\Sigma^i K^i)^2 \Delta}{(K^{i+1} - K^{i-1})(K^i - K^{i-1})}, \tag{17.4}$$

$w^{1\pm} = w^{N\pm} = 0$, and let $\mathbf{M}$ be the tri-band matrix with $(\mathbf{M}p)^i = (1 + w^{i-} + w^{i+})p^i - w^{(i-1)+}p^{i-1} - w^{(i+1)-}p^{i+1}$ (SANOS eq. (20), where $\mathbf{M} = Q^{-1}_{j|j-1}$).

Proposition 17.10 (Discrete local volatility)**.** *Let* $\mathbf{M}$ *be as above.*

(i) $\mathbf{M}$ *is invertible and* $\mathbf{Q} = \mathbf{M}^{-1}$ *is a martingale transition matrix:* $\mathbf{Q} \geq 0$ *entrywise, its columns sum to* 1*, and* $\sum_i K^i \mathbf{Q}_{il} = K^l$*. Hence for any* $\Sigma_1, \dots, \Sigma_M \geq 0$ *and expiries* T_j*, the vectors* $q_j = \mathbf{Q}_j q_{j-1}$ *(*$\Delta = T_j - T_{j-1}$*,* q_0 *the unit mass at* 1*) form a martingale density.*

(ii) Conversely, let $q_1, \dots, q_M$ *be a martingale density on the grid with* $q_j^i > 0$ *for* $1 < i < N$*, let* $c_j^i = \sum_l q_j^l (K^l - K^i)^+$*, and let*

$$(\Sigma_j^i)^2 = \frac{2\,\Theta_j^i}{(K^i)^2 \Gamma_j^i}, \qquad \Theta_j^i = \frac{c_j^i - c_{j-1}^i}{T_j - T_{j-1}}, \qquad \Gamma_j^i = \frac{q_j^i}{\frac{1}{2}(K^{i+1} - K^{i-1})}. \tag{17.5}$$

Then $\mathbf{M}_j q_j = q_{j-1}$*: the discrete local volatilities reproduce the density.*

Proof. (i) Column l of $\mathbf{M}$ has $1 + w^{l-} + w^{l+}$ on the diagonal and $-w^{l-}$, $-w^{l+}$ in rows $l-1$, $l+1$: it sums to 1, and $\sum_i K^i \mathbf{M}_{il} = K^l + w^{l-}(K^l - K^{l-1}) - w^{l+}(K^{l+1} - K^l) = K^l$ by (17.4). So $\mathbf{M}^\top$ has positive diagonal, nonpositive off-diagonal entries and rows summing to 1. If $\mathbf{M}^\top y = g \geq 0$ and y_i is the smallest entry of y, then $g_i = y_i + \sum_{l \neq i} |\mathbf{M}^\top_{il}|(y_i - y_l) \leq y_i$, so $y \geq 0$ (a maximum principle). Applied to $\pm y$ with $g = 0$, this shows that $\mathbf{M}^\top$ is invertible, and applied to $g = e_k$ that $(\mathbf{M}^\top)^{-1} \geq 0$; so $\mathbf{Q} = ((\mathbf{M}^\top)^{-1})^\top \geq 0$. Multiplying $\mathbf{1}^\top \mathbf{M} = \mathbf{1}^\top$ and $K^\top \mathbf{M} = K^\top$ by $\mathbf{Q}$

on the right gives the column sums and the means. Column l of $\mathbf{Q}_j$ is a law with mean K^l, so (q_{j-1}, q_j) is the law of a martingale pair, and conditional Jensen gives convex order. (ii) Let $D^i = c^i_j - c^i_{j-1} \geq 0$ (convex order). Then $w^{i\pm}q^i_j = D^i/|K^{i\pm1} - K^i|$, and $D^1 = D^N = 0$ since all mass lies in $[K^1, K^N]$. For $1 < i < N$,

$$(\mathbf{M}_j q_j)^i = q^i_j - \Big(\frac{D^{i+1} - D^i}{K^{i+1} - K^i} - \frac{D^i - D^{i-1}}{K^i - K^{i-1}}\Big) = q^i_j - (q^i_j - q^i_{j-1}),$$

because the change of slope of c_j at K^i is q^i_j (Theorem 17.6). Rows 1 and N are the same computation with the slopes -1 before K^1 and 0 after K^N. □

This is SANOS Theorem 3.4, with its Theorem A.4 for the inverse (a fact about M-matrices that the paper takes from the literature), and Theorem 3.7; discrete local volatilities come from H. Buehler and E. Ryskin, *Discrete local volatility for large time steps* (working paper, 2015). The paper's Theorem 3.7 uses the convention $0/0 := 0$ in (17.5); when $q^i_j = 0 < \Theta^i_j$ no finite Σ^i_j exists, which is why (ii) assumes $q^i_j > 0$. It suffices that $\Theta^i_j = 0$ whenever $q^i_j = 0$, with $\Sigma^i_j := 0$ there: the proof only uses $w^{i\pm}q^i_j = D^i/|K^{i\pm1} - K^i|$, which then holds with both sides zero. Theorem 3.6 maps a density between two different grids.

Dupire on a grid. By (17.5), $c^i_j - c^i_{j-1} = \frac{1}{2}(\Sigma^i_j K^i)^2(T_j - T_{j-1})\Gamma^i_j$, with Γ^i_j the second difference of c_j at K^i: one fully implicit finite-difference step of Dupire's equation $\partial_T c = \frac{1}{2}\sigma^2 x^2 \partial_{xx} c$ (Equation (15.7) in normalized units), the scheme with which J. Andreasen and B. Huge (*Risk* 24, 2011) interpolate volatility; written for the weights, $\mathbf{M}q_j = q_{j-1}$ is the implicit Euler step of the forward equation (15.9) (Exercise 17.5). So (17.5) is a finite-difference Dupire formula, and as the grid is refined the implied discrete local volatilities of Heston prices approach the local volatility of Figure 15.4: at $T = 0.5$, the largest error on $\log x \in [-0.4, 0.3]$ is 2.7, 0.6 and 0.13 volatility points for $(\Delta, h) = (1/12, 0.05)$, $(1/52, 0.02)$, $(1/252, 0.01)$, h the spacing in $\log x$ (Figure 17.3, left).

As a parametrization, Theorem 17.10(i) needs only $\Sigma \geq 0$: every vector of discrete local volatilities gives a martingale density, hence (with nondecreasing v) a surface strictly free of static arbitrage. Prices are no longer linear in the parameters, so calibration is a smooth nonlinear least-squares or gradient problem. This is the parametrization that DYSANOS uses.

17.5 DYSANOS: generating arbitrage-free surfaces

A *generative model* of option markets simulates future surfaces, for risk management and for training hedging strategies. If each simulated surface is a decoded SANOS surface, static arbitrage is excluded on every day by construction.

Definition 17.11 (DYSANOS with the AR(1) baseline)**.** Fix model expiries, a strike grid containing 1, and bounds $0 < \Sigma_{\min} < \Sigma_{\max}$. The *decoder* (ML-SANOS, DYSANOS Definition 2.10 and eqs. (6)–(17)) maps an unconstrained $u \in \mathbb{R}^p$ to discrete local volatilities with $(\Sigma^i_j)^2 = \min\{\Sigma_{\max}, \Sigma_{\min} + e^{u^{\Sigma,j,i}}\}$ (DYSANOS writes $\Sigma^{j,i}$ for the square), to increasing smoothing total volatilities W^j (forward volatilities squashed into an interval, times smoothing factors), to the martingale density of Theorem 17.10(i), and to the SANOS surface c_u. A

DYSANOS generator with the AR(1) baseline consists of an affine compression $u = u_0 + \Lambda h$, $h \in \mathbb{R}^d$ (eq. (21)), and a Gaussian AR(1) for the state,

$$h_{t+\Delta} = m + A_\Delta(h_t - m) + G_\Delta^\top z_{t,\Delta}, \qquad A_\Delta = e^{-\kappa\Delta}, \qquad \kappa = LL^\top + S + \kappa_{\min} I,$$

with $z_{t,\Delta} \sim \mathcal{N}(0, I)$ independent, S antisymmetric and $\kappa_{\min} > 0$ (eqs. (23)–(24)), together with a driftless spot whose one-day return is the quantile, under the front-expiry SANOS marginal, of a Gaussian score correlated with $z_{t,\Delta}$ (eqs. (28)–(29)). The generated surface on day t is $c_{u_0+\Lambda h_t}$.

For simplicity the definition fixes the strike grid; in DYSANOS the common grid is itself decoded from u, as exponentials of normalized moneyness times total volatilities (eq. (15)).

The paper uses $M = 18$ expiries from 1 to 70 business days, $N = 200$ strikes, $p = 3620$, $d = 200$, and fits u_0, Λ, the states h_t and the AR(1) to 891 days of S&P 500 quotes (2022–2025) by gradient descent in stages, the last one joint, with penalties (eqs. (20)–(27) and Appendix A).

Proposition 17.12 (What the generator guarantees)**.** *(i) For every state h, the surface $c_{u_0+\Lambda h}$ is strictly free of static arbitrage and smooth in the strike.*

(ii) Every eigenvalue λ of κ has $\operatorname{Re}\lambda \geq \kappa_{\min}$, so every eigenvalue of A_Δ has modulus at most $e^{-\kappa_{\min}\Delta} < 1$: the AR(1) is stable for every $\Delta > 0$.

Proof. (i) The decoded Σ are positive, so Theorem 17.10(i) gives a martingale density, and W^j is increasing; apply Theorem 17.8. (ii) For a complex unit eigenvector e, $\lambda = e^*\kappa e = |L^\top e|^2 + e^*Se + \kappa_{\min}$, and e^*Se is purely imaginary because S is real antisymmetric. The eigenvalues of $e^{-\kappa\Delta}$ are the $e^{-\lambda\Delta}$. □

Part (i) is the sense of the paper's "free of static arbitrage by construction"; (ii) is the stability argument of its Section 3.2. Neither says anything about *dynamic* arbitrage: on day t one may buy an option at the generated price and sell it on day $t+1$ at the price of the next generated surface, and nothing forces these prices, together with the spot, to be martingales under some measure equivalent to the generating one. DYSANOS Section 4 studies this for a book of the spot and 48 options, 25 days and 100,000 paths, with one-day returns R_t. Under the generating measure the mean returns are far from zero (largest $|z|$-score 53, Table 8). Reweighting the paths by minimal relative entropy subject to $\mathbb{E}[DR_t] = 0$ gives $\mathbb{Q}_1$ (eqs. (31)–(32)); an exponential-utility problem with transaction costs over neural-network strategies, in the spirit of the minimal entropy martingale measure of Frittelli (2000), gives $\mathbb{Q}_2$ (Theorem 4.2, eqs. (33)–(38)). The test of conditional drift is the following.

Proposition 17.13 (A test of conditional drift)**.** *Let $\mathcal{G}$ be a sub-σ-field, R a random vector in $\mathbb{R}^k$ with $\mathbb{E}[e^{\lambda|R|}] < \infty$ for all $\lambda > 0$, and $\mathcal{A}$ the set of bounded $\mathcal{G}$-measurable vectors. If $a^* \in \mathcal{A}$ minimizes $J(a) = \mathbb{E}\big[\exp(-a^\top R + \frac{1}{2}|a|^2)\big]$ over $\mathcal{A}$, then*

$$\mathbb{E}[a^{*\top} R] \geq \mathbb{E}\big[|a^*|^2\big], \qquad \text{and} \qquad \mathbb{E}[a^{*\top} R] = 0 \iff \mathbb{E}[R \mid \mathcal{G}] = 0 \text{ almost surely.}$$

Proof. For bounded $\mathcal{G}$-measurable b, the function $\varepsilon \mapsto J(a^* + \varepsilon b)$ is minimal at 0 and may be differentiated under $\mathbb{E}$ (dominated convergence, using the exponential moments), so

$$\mathbb{E}\big[e^{-a^{*\top} R + |a^*|^2/2}(a^* - R)^\top b\big] = 0.$$

Taking $b = e_k \mathbf{1}_A$, $A \in \mathcal{G}$, gives $\mathbb{E}[e^{-a^{*\top}R}(R - a^*) \mid \mathcal{G}] = 0$. Take the inner product with the $\mathcal{G}$-measurable a^*: with $\zeta = a^{*\top}R$ and $y_0 = |a^*|^2$, $\mathbb{E}[e^{-\zeta}(\zeta - y_0) \mid \mathcal{G}] = 0$. Since $(y - y_0)(e^{-y} - e^{-y_0}) \le 0$ for every real y, $e^{-\zeta}(\zeta - y_0) \le e^{-y_0}(\zeta - y_0)$, so $0 \le e^{-y_0}\mathbb{E}[\zeta - y_0 \mid \mathcal{G}]$: $a^{*\top}\mathbb{E}[R \mid \mathcal{G}] \ge |a^*|^2$, and taking expectations gives the inequality. If $\mathbb{E}[a^{*\top}R] = 0$, then $a^* = 0$ almost surely, and the first-order condition reads $\mathbb{E}[R \mid \mathcal{G}] = 0$. Conversely, if $\mathbb{E}[R \mid \mathcal{G}] = 0$, the first-order condition holds at $a = 0$; J is convex, so 0 is a minimizer, and J is strictly convex (so is its integrand in a), so $a^* = 0$. □

This is DYSANOS Theorem 4.3 for one date (their objective sums over dates, and they assume finite, twice differentiable conditional Laplace transforms; their Appendix C uses strong convexity). The test fits a^* on half of the paths and evaluates the captured return $a^{*\top}R$ on the other half. Under $\mathbb{Q}_1$ it rejects ($T = 41.4$); under $\mathbb{Q}_2$ it does not ($T = -0.45$, $p = 0.65$; Table 9). The paper does not claim risk-neutrality beyond this book and strategy class (Section 4.4).

Replication. We simulate 250 days of the Heston market with a stochastic initial variance v_t (a square-root process with $\kappa = 8$, $\theta = 0.045$, $\xi = 0.4$ and spot correlation -0.7) and fit each day by the decoder with discrete local volatilities $\Sigma_j^i = 0.01 + \min(3, e^{u_j(z)})$, u_j a polynomial of degree 4 in $z = \log K^i/(0.2\sqrt{T_j})$ (flat beyond the quotes), the smoothing variances from the mid at-the-money forward variances, and a penalty tying u_j to u_{j-1}: 36 parameters a day instead of 3620. The daily fits put 95.4% of the prices inside the spreads. An SVD compression to $d = 6$ keeps 91.1% (DYSANOS: 82.3% and 80.9% with $d = 200$, Table 1), and least squares give an AR(1) with spectral radius 0.967 per day, a half-life of 21 trading days (the variance v_t has $\ln 2/8$ years, 22 days; DYSANOS: 15.4 weeks, Table 5). Generated surfaces pass every static check to rounding, as Theorem 17.12 guarantees (Figure 17.3, right), and in a long simulation from the invariant law their 3-month at-the-money volatility has the historical mean and spread (19.9% and 4.1 points against 20.2% and 4.0). The dynamics are another matter. The synthetic history is itself not free of dynamic arbitrage: its variance has volatility of volatility 0.4 while its surfaces are priced with 0.5, and an equivalent change of measure cannot alter the quadratic variation $\xi^2 \int v\,dt$ of the variance, a limit in probability of sums of squared increments (Theorem 7.12). We also replace the front SANOS marginal of eq. (29) by a lognormal law at the front at-the-money volatility, which creates drift by itself: with the surface frozen, out-of-the-money calls ($x = 1.05$ at one month, 1.1 at three) gain 2.7% a day on average ($z = 20$ and 34), because a lognormal spot ignores the skew. On 4000 antithetic paths of 25 days of the full generator, with standard errors computed on the antithetic pairs, 38 of the 125 mean daily returns (of the spot, and of calls at $x = 0.95, 1$ for one month and $x = 0.9, 1$ for three) have $|z| > 2.576$ (largest 7.7; median 6.6 for the three-month at-the-money call), and the cross-fitted test of Theorem 17.13 rejects decisively ($T = 30$ with constant actions, $T = 32$ on the three-month at-the-money call alone): pooled over 25 dates, even a modest daily drift is unmistakable.

▶ Interactive demo: discrete local volatility and Dupire; the AR(1) generator, static checks of generated surfaces, and the drift tests.

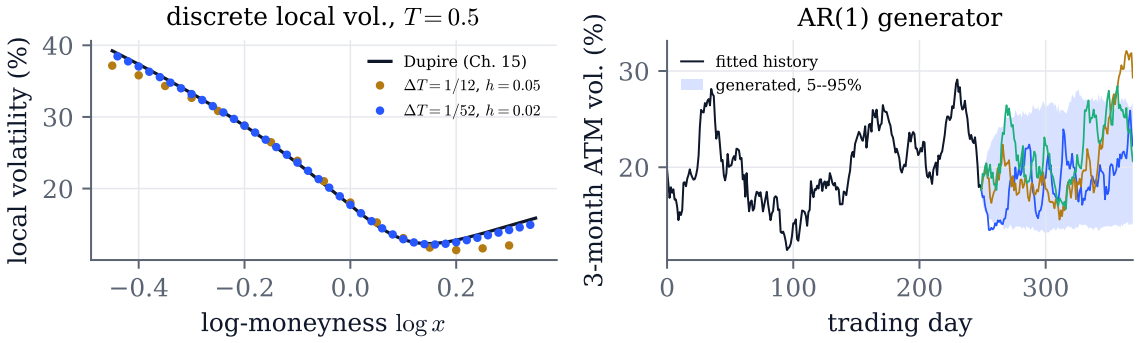

Figure 17.3: Left: discrete local volatilities implied by Heston prices, (17.5), on two grids, against Dupire's local volatility of Figure 15.4 at $T = 0.5$. Right: the 3-month at-the-money volatility of 250 fitted days of the synthetic market, and 120 days generated by the AR(1) baseline from the last state (three paths, and the 5–95% band of 400).

Further reading

Beyond what is covered here, Buehler et al. (2026a) treat different strike grids at different expiries and the generalized models of their Section 3.4, and Buehler et al. (2026b) heavy-tailed innovations (Appendix B) and the data pipeline (Appendix E). Gatheral and Jacquier (2014) give the parametric alternative, SSVI, with explicit no-arbitrage conditions. Davis and Hobson (2007) and Cousot (2007) characterize quotes free of static arbitrage, Carr and Madan (2005) give sufficient conditions, and Buehler (2006) studies the martingales consistent with a surface. Martingale couplings (Theorem 17.5) are also the objects of martingale optimal transport, which bounds the prices of exotics over all models consistent with the vanillas; Brown et al. (2001) give such bounds for barrier options (Section 15.6).

Chapter 18

Hawkes processes and limit order books

At the scale of milliseconds a market is not a diffusion but a stream of events: limit orders are posted and cancelled, market orders execute against them, and the price moves by whole ticks. The events cluster. A trade makes further trades more likely in the next seconds, a cancellation is often followed by a new order, a burst of activity feeds on itself. *Hawkes processes* are point processes in which every event raises the rate of future events. They are tractable (a branching structure, an explicit likelihood, an exact simulation algorithm) and, fitted to high-frequency data, they often turn out to sit close to the boundary of stability.

This chapter builds the toolkit (intensities, the likelihood, the time-change theorem), the statistics (maximum likelihood and goodness of fit), and the application to order flow; it ends with the scaling limits of Jaisson and Rosenbaum. Hawkes processes close to instability look, on a large time scale, like the CIR process of Theorem 14.9, or, when the excitation decays slowly, like the rough processes of Chapter 19. A price driven by such order flow becomes a (rough) Heston model. The survey of Bacry et al. (2015) covers the financial applications, and Daley and Vere-Jones (2003) the theory of point processes.

Setting. $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Every point process starts at time 0 with no events before. For functions on $[0,\infty)$, $(f * g)(t) = \int_0^t f(t-s)g(s)\,\mathrm{d}s$; for matrix-valued functions the products inside are matrix products. $\varrho(G)$ is the spectral radius (largest modulus of an eigenvalue) of a square matrix G.

18.1 Point processes and intensities

A *d-variate point process* is a sequence of event times $0 < t_1 < t_2 < \cdots$ with types $\kappa_k \in \{1, \dots, d\}$; we allow $t_k = \infty$ (then $t_{k+1} = \infty$: there are only $k-1$ events). Its *counting processes* are $N_t^i = \#\{k : t_k \le t,\ \kappa_k = i\}$ and $N_t = \sum_i N_t^i$, and $H_k = (t_1, \kappa_1, \dots, t_k, \kappa_k)$ is the history of the first k events. The process *explodes* if $t_\infty = \lim_k t_k < \infty$.

Definition 18.1 (Conditional intensity)**.** For each type i, let $\lambda^i(t \mid h) \geq 0$ be defined for every finite history $h = (s_1, j_1, \dots, s_k, j_k)$ with $s_1 < \cdots < s_k$ and every $t > s_k$ (with $s_0 = 0$), measurable in (t, h) and integrable in t on bounded intervals, and let $\bar\lambda = \sum_i \lambda^i$. A point process has *conditional intensity* $(\lambda^i)_{i \leq d}$ if, for every $k \geq 0$, on $\{t_k < \infty\}$,

$$\mathbb{P}\big(t_{k+1} \in \mathrm{d}t,\ \kappa_{k+1} = i \,\big|\, H_k\big) = \lambda^i(t \mid H_k) \exp\Big(- \int_{t_k}^{t} \bar\lambda(u \mid H_k)\,\mathrm{d}u\Big)\mathrm{d}t, \qquad t > t_k. \tag{18.1}$$

We write $\lambda^i(t) = \lambda^i(t \mid H_k)$ for $t_k < t \leq t_{k+1}$: the intensity given the events strictly before t.

Summing (18.1) over i and integrating, $\mathbb{P}(t_{k+1} > t \mid H_k) = \exp(-\int_{t_k}^{t} \bar\lambda(u \mid H_k)\mathrm{d}u)$: the next event comes at the *hazard rate* $\bar\lambda$, and it is of type i with probability $\lambda^i/\bar\lambda$ at that time. Such a process always exists: with i.i.d. exponential variables E_k and uniform U_k, set $t_{k+1} = \inf\{t > t_k : \int_{t_k}^{t} \bar\lambda(u \mid H_k)\mathrm{d}u \geq E_{k+1}\}$ and pick the type by U_{k+1} (Daley and Vere-Jones (2003), Algorithm 7.4.III). A Poisson process of rate μ is the case $\lambda \equiv \mu$, with i.i.d. $\mathrm{Exp}(\mu)$ gaps; an inhomogeneous one, with rate $h(t)$, the case $\lambda(t \mid h) = h(t)$.

Remark 18.2 (Intensities and martingales). Heuristically, $\mathbb{P}(\text{a type-}i\text{ event in } (t, t+\delta] \mid \text{past}) = \lambda^i(t)\delta + o(\delta)$. In the general theory the intensity is defined through martingales: $N^i_t - \int_0^t \lambda^i(s)\,\mathrm{d}s$ is a (local) martingale for the filtration generated by the process, and this property determines the law of the process (Daley and Vere-Jones (2003), Lemma 7.2.V and Proposition 7.2.IV). We only use its consequence for expectations, proved directly below.

Lemma 18.3 (Compensator and density)**.** *Let N have conditional intensity (λ^i).*

(i) For $t \geq 0$, $k \geq 1$ and every i, $\mathbb{E}[N^i_{t \wedge t_k}] = \mathbb{E}\int_0^{t \wedge t_k} \lambda^i(s)\,\mathrm{d}s$, and $\mathbb{E}[N^i_t] = \mathbb{E}\int_0^{t \wedge t_\infty} \lambda^i(s)\,\mathrm{d}s$ in $[0, \infty]$.

(ii) If $\mathbb{E}[N_t] < \infty$ for every t, then N does not explode, and $\mathbb{E}\big[\sum_{k:\kappa_k = i} g(t_k)\big] = \int_0^\infty g(s)\,\mathbb{E}[\lambda^i(s)]\,\mathrm{d}s$ for every Borel $g \geq 0$.

(iii) For $T > 0$, $n \in \mathbb{N}$, $0 < s_1 < \cdots < s_n \leq T$ and types $j_1, \dots, j_n$, writing λ^i for the intensity of the configuration $h = (s_1, j_1, \dots, s_n, j_n)$,

$$\mathbb{P}\big(N_T = n,\ t_k \in \mathrm{d}s_k,\ \kappa_k = j_k,\ k \leq n\big) = \prod_{k=1}^{n} \lambda^{j_k}(s_k)\ \exp\Big(- \int_0^T \bar\lambda(u)\,\mathrm{d}u\Big)\,\mathrm{d}s_1 \cdots \mathrm{d}s_n. \tag{18.2}$$

Proof. (i) Fix $k \geq 1$ and write $H = H_{k-1}$. On $(t_{k-1}, t_k]$ the intensity is $\lambda^i(\cdot \mid H)$, and $\mathbb{P}(t_k > s \mid H) = e^{-\int_{t_{k-1}}^{s} \bar\lambda(u \mid H)\mathrm{d}u}$. By (18.1) and Tonelli, on $\{t_{k-1} < t\}$,

$$\mathbb{P}(t_k \leq t,\ \kappa_k = i \mid H) = \int_{t_{k-1}}^{t} \lambda^i(s \mid H)\,\mathbb{P}(t_k > s \mid H)\,\mathrm{d}s = \mathbb{E}\Big[\int_{t_{k-1}}^{t_k \wedge t} \lambda^i(s)\,\mathrm{d}s \,\Big|\, H\Big],$$

and both sides vanish on $\{t_{k-1} \geq t\}$. Take expectations and sum over the first k events: the intervals $(t_{l-1}, t_l]$ partition $(0, t_k]$. Letting $k \to \infty$ gives the second identity by monotone convergence, since $N^i_{t \wedge t_k} \uparrow N^i_t$.

(ii) If $\mathbb{E}[N_t] < \infty$, then $N_t < \infty$ almost surely, so $t_\infty > t$; taking $t = 1, 2, \dots$ gives $t_\infty = \infty$ almost surely. By (i) and Tonelli, $\mathbb{E}[N^i_b - N^i_a] = \int_a^b \mathbb{E}[\lambda^i(s)]\,\mathrm{d}s$ for $a < b$. The measures $A \mapsto \mathbb{E}[N^i(A)]$ and $A \mapsto \int_A \mathbb{E}[\lambda^i]$ agree on intervals and are finite on bounded sets, so they agree on Borel sets (Dynkin's π–λ theorem); approximate g by simple functions and use monotone convergence.

(iii) Apply (18.1) to $k = 0, \dots, n-1$ in turn and multiply by $\mathbb{P}(t_{n+1} > T \mid H_n) = e^{-\int_{s_n}^{T} \bar\lambda}$; the exponentials combine into $e^{-\int_0^T \bar\lambda}$. □

18.2 The Hawkes process

Definition 18.4 (Hawkes process)**.** Let $d \geq 1$, $\mu \in [0,\infty)^d$, and let $\varphi_{ij} : (0,\infty) \to [0,\infty)$ be Borel with $\|\varphi_{ij}\|_1 = \int_0^\infty \varphi_{ij}(t)\,\mathrm{d}t < \infty$. A *$d$-variate Hawkes process* with *baseline* μ and *kernels* φ_{ij} is a point process with conditional intensity

$$\lambda^i(t) = \mu_i + \sum_{j=1}^{d} \sum_{k:\,\kappa_k = j,\ t_k < t} \varphi_{ij}(t - t_k) = \mu_i + \sum_{j=1}^{d} \int_{(0,t)} \varphi_{ij}(t-s)\,\mathrm{d}N_s^j. \tag{18.3}$$

Its *branching matrix* is $G = (\|\varphi_{ij}\|_1)_{i,j}$; for $d = 1$, $\|\varphi\|_1$ is the *branching ratio*. *Exponential kernels* are $\varphi_{ij}(t) = \alpha_{ij}\beta_{ij}e^{-\beta_{ij}t}$ with $\alpha_{ij} \geq 0$, $\beta_{ij} > 0$, so that $G = (\alpha_{ij})$.

Hawkes (1971) introduced these processes on the whole real line and computed their mean rate and second-order structure. An event of type j raises the intensity of type i by φ_{ij}; the entry G_{ij} is the total extra number of type-i events it causes directly (Theorem 18.7).

Remark 18.5 (Exponential kernels). For $d = 1$ and $\varphi(t) = \alpha\beta e^{-\beta t}$, between events $\lambda(t) = \mu + (\lambda(s) - \mu)e^{-\beta(t-s)}$ relaxes towards μ at rate β, and at each event it jumps by $\alpha\beta$. The intensity is therefore a Markov process, and the past enters through one number. For $d \geq 2$ the d^2 sums $R_{ij}(t) = \sum_{\kappa_k = j,\, t_k < t} e^{-\beta_{ij}(t-t_k)}$ play this role. In the stationary regime the covariance density of the process decays like $e^{-\beta(1-\alpha)\tau}$ (Hawkes (1971), Section 3), and the mean intensity of the process started empty approaches its limit at the same rate (Exercise 18.1): the memory becomes long as $\alpha \uparrow 1$.

Proposition 18.6 (Mean intensity and stability)**.** *Let N be a Hawkes process, $\Phi = (\varphi_{ij})$, $\Phi^{*1} = \Phi$, $\Phi^{*(n+1)} = \Phi * \Phi^{*n}$, and*

$$m(t) = \mu + \sum_{n \geq 1} \Big(\int_0^t \Phi^{*n}(s)\,\mathrm{d}s \Big)\mu \in [0,\infty]^d.$$

(i) N does not explode, $\mathbb{E}[N_t] = \int_0^t m(s)\,\mathrm{d}s < \infty$ and $\mathbb{E}[\lambda(t)] = m(t)$ for every t. The function m is finite and nondecreasing, and it is the only locally bounded solution of the renewal equation $m = \mu + \Phi * m$.

(ii) If $\varrho(G) < 1$, then $m(t) \uparrow \Lambda = (I - G)^{-1}\mu = \sum_{n \geq 0} G^n \mu$ as $t \to \infty$, and $\mathbb{E}[N_t]/t \to \Lambda$. The total rate of events tends to $\mathbf{1}^\top (I - G)^{-1}\mu$.

(iii) If $\varrho(G) \geq 1$ and $\mu_i > 0$ for every i, then $\max_i m_i(t) \to \infty$. For $d = 1$: if $\|\varphi\|_1 \geq 1$ and $\mu > 0$, then $\mathbb{E}[\lambda(t)] \to \infty$.

Proof. For $\theta > 0$ let $\hat\Phi(\theta) = \int_0^\infty e^{-\theta u}\Phi(u)\,\mathrm{d}u$. By dominated convergence $\hat\Phi(\theta) \to 0$, so we may fix θ with $c = \max_i \sum_j \hat\Phi_{ij}(\theta) < 1$. The Laplace transform of a convolution is the product of the

transforms (Tonelli), so for every locally bounded $f \geq 0$,

$$e^{-\theta t}(\Phi^{*n} * f)(t) \leq \hat{\Phi}(\theta)^n \sup_{s \leq t} e^{-\theta s} f(s) \quad \text{entrywise}, \qquad \|\hat{\Phi}(\theta)^n\|_\infty \leq c^n. \tag{18.4}$$

Finite mean. Let $n_k(t) = \mathbb{E}[N_{t \wedge t_k}]$ (a vector, with entries at most k). For $s \leq t_k$ the events before s are among the first k, so by Theorem 18.3(i) and Tonelli,

$$n_k^i(t) \leq \mu_i t + \sum_j \mathbb{E} \int_0^t \int_{(0,s)} \varphi_{ij}(s-u)\, \mathrm{d}N^j_{u \wedge t_k}\, \mathrm{d}s = \mu_i t + \sum_j \int_0^t \varphi_{ij}(t-s)\, n_k^j(s)\, \mathrm{d}s,$$

using $\int_{(0,t)} \int_u^t \varphi(s-u)\, \mathrm{d}s\, \mathrm{d}N_u = \int_{(0,t)} \int_u^t \varphi(t-s)\, \mathrm{d}s\, \mathrm{d}N_u = \int_0^t \varphi(t-s) N_s\, \mathrm{d}s$. Multiply by $e^{-\theta t}$, bound $te^{-\theta t} \leq 1/(e\theta)$, and let $a_k(t) = \sup_{s \leq t} \max_i e^{-\theta s} n_k^i(s)$: then $a_k(t) \leq \max_i \mu_i/(e\theta) + c\, a_k(t)$, so $\mathbb{E}[N_t^i] = \lim_k n_k^i(t) \leq e^{\theta t} \max_i \mu_i/(e\theta(1-c))$. By Theorem 18.3(ii), N does not explode.

The mean. Now the same computation holds with equalities: $M(t) = \mathbb{E}[N_t]$ is finite, nondecreasing and solves $M(t) = \mu t + (\Phi * M)(t)$. Iterating n times, $M = \sum_{l<n} \Phi^{*l} * (\mu\, \mathrm{id}) + \Phi^{*n} * M$, and the remainder tends to 0 by (18.4); since $(\Phi^{*l} * \mathrm{id})(t) = \int_0^t \int_0^u \Phi^{*l}$, $M(t) = \int_0^t m(u)\, \mathrm{d}u$, with m finite. By Theorem 18.3(ii) with $g(s) = \varphi_{ij}(t-s)\mathbf{1}_{\{s<t\}}$, $\mathbb{E}[\lambda(t)] = \mu + (\Phi * m)(t)$, and $\mu + \Phi * m = m$ term by term. If $\tilde{m}$ is another locally bounded solution, $m - \tilde{m} = \Phi^{*n} * (m - \tilde{m})$ for all n, and (18.4) applied to $|m - \tilde{m}|$ gives $\tilde{m} = m$.

(ii) By Tonelli, $\int_0^\infty \Phi^{*n} = G^n$, and $\int_0^t \Phi^{*n} \uparrow G^n$. If $\varrho(G) < 1$ the Neumann series $\sum_n G^n$ converges to $(I-G)^{-1}$, so $m(t) \uparrow \Lambda$ by monotone convergence, and $\mathbb{E}[N_t]/t = \frac{1}{t}\int_0^t m \to \Lambda$.

(iii) If $G^n \to 0$, then $G^n w = z^n w \to 0$ for an eigenvector w with eigenvalue $|z| = \varrho(G)$, so $\varrho(G) < 1$. Hence if $\varrho(G) \geq 1$ some entry $(G^n)_{ij}$ exceeds some $\varepsilon > 0$ for infinitely many n, and $\lim_t m_i(t) = \sum_n (G^n \mu)_i \geq \sum_n (G^n)_{ij} \mu_j = \infty$. □

So the process started empty settles at the rate Λ when $\varrho(G) < 1$, and, if every $\mu_i > 0$, its expected rate grows without bound otherwise. On the whole line, Hawkes (1971) computed the mean rate $(I-G)^{-1}\mu$ of a stationary version (his equation (21)), and Hawkes and Oakes (1974) built the stationary univariate process as a cluster process, the construction we now give on $[0, T]$.

Proposition 18.7 (Branching structure)**.** *Let μ and φ_{ij} be as in Theorem 18.4, and $T > 0$. Let* generation 0 *consist of independent Poisson processes on $[0,T]$ with rates μ_i (the* immigrants *of type i); given generations $0, \dots, g$, let every event (s, j) of generation g have, independently, type-i* children *forming a Poisson process on $(s, T]$ with intensity $\varphi_{ij}(\cdot - s)$, for each i; these form generation $g+1$.*

(i) The set of all events is almost surely finite and, forgetting the generations, it has the law of the Hawkes process with baseline μ and kernels φ_{ij} on $[0,T]$.

(ii) On $[0,\infty)$, an event of type j has a Poisson(G_{ij}) number of type-i children. If $\varrho(G) < 1$, its expected number of type-i descendants, itself included, is $[(I-G)^{-1}]_{ij}$, and its cluster *of descendants is almost surely finite.*

Proof. (ii) The number of points of a Poisson process with intensity $\varphi_{ij}(\cdot - s)$ is Poisson with mean G_{ij}. Let $Z_g \in \mathbb{N}^d$ count the descendants of each type in generation g of one type-j event, $Z_0 = e_j$. Given Z_g, the children are independent Poisson counts, so $\mathbb{E}[Z_{g+1} \mid Z_g] = GZ_g$ and

$\mathbb{E}[Z_g] = G^g e_j$. By Tonelli, $\mathbb{E}\sum_g Z_g = \sum_g G^g e_j = (I-G)^{-1}e_j < \infty$, so the cluster is finite almost surely.

(i) Conditioning on generation g, the expected density of generation $g+1$ is $\Phi * \rho_g$ if ρ_g is that of generation g; with $\rho_0 = \mu$, the expected number of events in $[0,T]$ is $\sum_g \int_0^T \mathbf{1}^\top(\Phi^{*g} * \mu) = \mathbf{1}^\top \mathbb{E}[N_T] < \infty$ by Theorem 18.6, so it is finite almost surely. Now fix $0 < s_1 < \cdots < s_n \le T$ and types $j_1, \dots, j_n$. A *genealogy* is a map $\pi : \{1,\dots,n\} \to \{0,\dots,n\}$ with $\pi(k) < k$, where $\pi(k) = 0$ means that event k is an immigrant and $\pi(k) = l \ge 1$ that its parent is event l. The construction uses independent Poisson processes: one per type for the immigrants, and one per type and per event for its children. A Poisson process with intensity h on an interval I has exactly the points $\mathrm{d}r_1, \dots, \mathrm{d}r_q$ with probability $\prod_k h(r_k)\,\mathrm{d}r_k\, e^{-\int_I h}$ (Theorem 18.3(iii) with a deterministic intensity). Multiplying over all these processes, the probability that the events are exactly at $\mathrm{d}s_1, \dots, \mathrm{d}s_n$, with these types and genealogy π, is

$$\prod_{k=1}^{n} w_k(\pi(k))\,\mathrm{d}s_k \;\exp\Big(-T\sum_i \mu_i - \sum_{l=1}^{n}\sum_i \int_{s_l}^{T} \varphi_{ij_l}(u-s_l)\,\mathrm{d}u\Big),$$

with $w_k(0) = \mu_{j_k}$ and $w_k(l) = \varphi_{j_k j_l}(s_k - s_l)$. The exponent is $-\int_0^T \bar\lambda(u)\,\mathrm{d}u$ for the Hawkes intensity (18.3) of the configuration. Each $\pi(k)$ ranges over $\{0,\dots,k-1\}$ independently of the others, so the sum over all genealogies factorizes: $\sum_\pi \prod_k w_k(\pi(k)) = \prod_k \sum_{l<k} w_k(l) = \prod_k \lambda^{j_k}(s_k)$. This is (18.2). The law of a finite point process on $[0,T]$ is the law of $(N_T, t_1, \kappa_1, \dots, t_{N_T}, \kappa_{N_T})$, which these densities determine. □

The proof is the likelihood of Section 18.4 read backwards. Each immigrant starts a Galton–Watson tree in continuous time, with Poisson offspring; the fraction of events that are immigrants is the *exogenous* part of the activity, $1 - \|\varphi\|_1$ for $d = 1$ (Exercise 18.2).

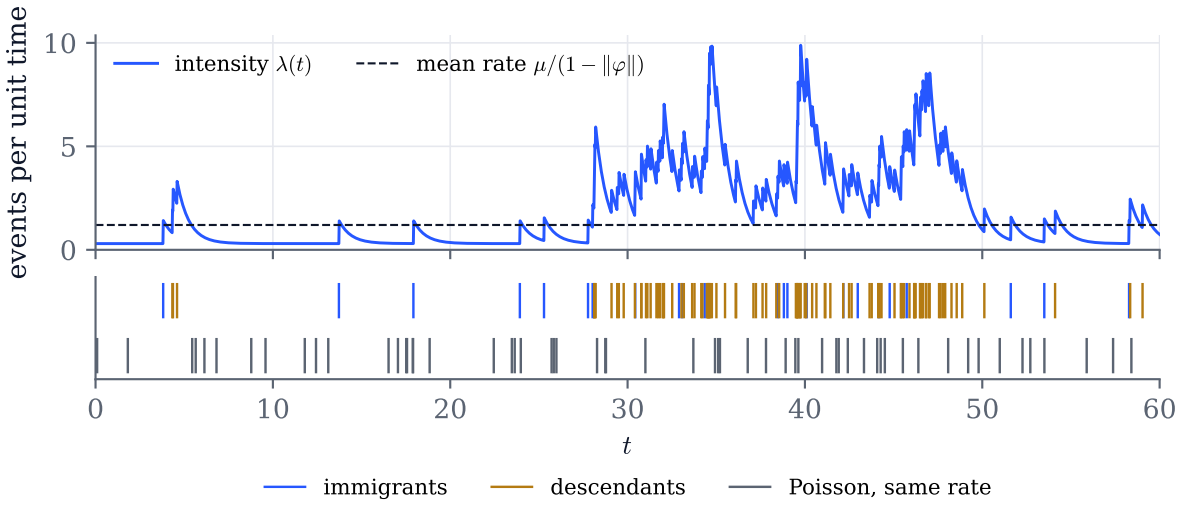

Figure 18.1: A Hawkes process with $\mu = 0.3$ and $\varphi(t) = 0.75 \cdot 1.5\, e^{-1.5t}$ (branching ratio 0.75), simulated by the cluster construction of Theorem 18.7: 23 immigrants and 101 descendants on $[0, 60]$. Top: the intensity and the stationary rate $\mu/(1-\|\varphi\|_1) = 1.2$. Bottom: the events, and a Poisson process with the same rate for comparison.

18.3 Simulation by thinning

Lewis and Shedler (1979) simulated an inhomogeneous Poisson process with rate $h \le M$ by *thinning*: draw a Poisson process of rate M and keep a point at s with probability $h(s)/M$. Ogata (1981) extended the idea to intensities that depend on the past, updating the bound after each candidate.

Proposition 18.8 (Ogata's thinning algorithm)**.** *Let (λ^i) be a conditional intensity such that, for every history h with last event at s_k, $t \mapsto \bar\lambda(t \mid h)$ is nonincreasing on (s_k, ∞), with finite right limits. Start from $s = 0$ and the empty history h, and repeat:*

1. *set $M = \bar\lambda(s+ \mid h)$, draw $E \sim \mathrm{Exp}(1)$ and set $s \leftarrow s + E/M$; stop if $s > T$;*
2. *draw U uniform on $[0,1]$; if $UM \le \bar\lambda(s \mid h)$, accept s as an event of the type i with $\sum_{l<i} \lambda^l(s \mid h) < UM \le \sum_{l \le i} \lambda^l(s \mid h)$, and append (s,i) to h.*

The accepted events, with fresh variables E and U at every step, form a point process on $[0,T]$ with conditional intensity (λ^i).

Proof. Fix a history h with last event s_k, a type i and $t > s_k$, and write $\lambda^l(u) = \lambda^l(u \mid h)$. While no new event is accepted, the algorithm restarts at each candidate s with fresh variables, so the probability $F(s)$ that the next acceptance lies in $(s,t]$ and has type i depends only on the current candidate time $s \in [s_k, t]$. Conditioning on the next candidate $s' = s + E/M$, with $M = \bar\lambda(s+) \ge \bar\lambda$ on (s,∞): it is accepted with type i with probability $\lambda^i(s')/M$, rejected with probability $1 - \bar\lambda(s')/M$, so

$$F(s) = \int_s^t e^{-M(s'-s)} \big(\lambda^i(s') + (M - \bar\lambda(s'))F(s')\big)\,\mathrm{d}s'.$$

The function $\tilde F(s) = \int_s^t \lambda^i(u) e^{-\int_s^u \bar\lambda}\mathrm{d}u$ solves the same equation: it satisfies $\tilde F' = -\lambda^i + \bar\lambda \tilde F$ and $\tilde F(t) = 0$, so $\frac{\mathrm{d}}{\mathrm{d}s'}(e^{-M(s'-s)}\tilde F(s')) = -e^{-M(s'-s)}(\lambda^i + (M-\bar\lambda)\tilde F)(s')$, and integrating from s to t gives the claim. The difference $D = F - \tilde F$ is bounded and satisfies $|D(s)| \le \bar M \int_s^t |D(s')|\,\mathrm{d}s'$ with $\bar M = \bar\lambda(s_k+)$; by Gronwall's lemma (Theorem 9.3, applied to $r \mapsto |D(t-r)|$), $D = 0$. At $s = s_k$ this is (18.1). □

For a Hawkes process with nonincreasing kernels, $\bar\lambda(\cdot \mid h)$ is nonincreasing between events, and $\bar\lambda(s+)$ includes the jump of an event accepted at s. With exponential kernels the state (R_{ij}) of Theorem 18.5 is updated in $O(d^2)$ operations per candidate; with general kernels each evaluation of λ costs $O(n)$, and a path with n events $O(n^2)$. When the kernel is not monotone the bound must be chosen differently (Exercise 18.3). The cluster construction of Theorem 18.7 is another exact sampler, which simulates whole generations at once; the code uses it for long or many paths.

▶ Interactive demo: Ogata's algorithm in one and six dimensions, the intensity it produces, and the cluster view of the same process.

18.4 Likelihood and goodness of fit

Proposition 18.9 (Maximum likelihood)**.** *Let a d-variate Hawkes process be observed on $[0,T]$, with events $(t_k,\kappa_k)_{k\le n}$, $n=N_T$, and let its intensity λ_ϑ depend on a parameter ϑ.*

(i) The log-likelihood is

$$\ell_T(\vartheta)=\sum_{i=1}^{d}\Big(\sum_{k:\,\kappa_k=i}\log\lambda^i_\vartheta(t_k)-\int_0^T\lambda^i_\vartheta(s)\,\mathrm{d}s\Big). \tag{18.5}$$

(ii) For exponential kernels, with $R_{ij}(t_1)=0$ and $R_{ij}(t_k)=e^{-\beta_{ij}(t_k-t_{k-1})}\big(R_{ij}(t_{k-1})+\mathbf{1}_{\{\kappa_{k-1}=j\}}\big)$,

$$\lambda^i(t_k)=\mu_i+\sum_j\alpha_{ij}\beta_{ij}R_{ij}(t_k),\qquad \int_0^T\lambda^i=\mu_iT+\sum_j\alpha_{ij}\sum_{k:\,\kappa_k=j}\big(1-e^{-\beta_{ij}(T-t_k)}\big),$$

so that ℓ_T costs $O(nd^2)$ operations (Ozaki, 1979).

(iii) For fixed decays β_{ij}, ℓ_T is a concave function of $(\mu,\alpha)\in[0,\infty)^{d+d^2}$, and a sum of d concave functions, one of $(\mu_i,\alpha_{i1},\dots,\alpha_{id})$ for each i.

(iv) (Asymptotics; Clinet and Yoshida (2017), Theorem 4.6.) *Let the process start empty, let the parameter $\vartheta=(\mu,\alpha,\beta)$ of exponential kernels range over a bounded open set Θ on which every $\mu_i,\alpha_{ij},\beta_{ij}$ lies between two positive constants, and let the true $\vartheta^*\in\Theta$ satisfy $\varrho(G^*)<1$. Then the maximizer $\hat\vartheta_T$ of ℓ_T over the closure of Θ satisfies $\sqrt{T}(\hat\vartheta_T-\vartheta^*)\to\mathcal{N}(0,\Gamma^{-1})$ in law, together with its moments, where $-\frac{1}{T}\nabla^2\ell_T(\vartheta^*)\to\Gamma$ in probability (Γ is the Fisher information per unit time). Their kernels are $c_{ij}e^{-a_{ij}t}$; the statement transfers through the smooth reparametrisation $c=\alpha\beta$, $a=\beta$, with Γ changed accordingly. Zero entries $\alpha_{ij}=0$ are excluded.*

Proof. (i) is the logarithm of (18.2). (ii) $R_{ij}(t_k)=\sum_{l<k,\,\kappa_l=j}e^{-\beta_{ij}(t_k-t_l)}$: split off $l=k-1$. Integrate (18.3) with $\int_{t_l}^T\beta e^{-\beta(u-t_l)}\mathrm{d}u=1-e^{-\beta(T-t_l)}$. (iii) With β fixed, each $\lambda^i(t_k)$ is a linear function of $(\mu_i,\alpha_{i\cdot})$, the logarithm is concave and nondecreasing, and the compensator $\int_0^T\lambda^i$ is linear. □

In practice the covariance of $\hat\vartheta_T$ is estimated by $(-\nabla^2\ell_T(\hat\vartheta_T))^{-1}$, the inverse *observed* information. Ogata (1978) proved consistency and asymptotic normality of the maximum likelihood estimator for stationary point processes; the result above is for the exponential Hawkes process started empty, as observed in practice; it does not cover the toy order book below, whose G has zero entries. If $N^i_T\ge1$ and the fitted intensities at the events are positive, the fitted compensator matches the count, $\int_0^T\hat\lambda^i=N^i_T$: Exercise 18.4 proves it for fixed decays, and the same scaling argument at $\hat\beta$ gives it for the full fit.

Theorem 18.10 (Random time change)**.** *Let N have conditional intensity (λ^i) with $\int_{s_k}^\infty\bar\lambda(u\mid h)\,\mathrm{d}u=\infty$ for every history h (for a Hawkes process: $\sum_i\mu_i>0$), and let $\Lambda(t)=\int_0^t\bar\lambda(u)\,\mathrm{d}u$.*

(i) The transformed gaps $\tau_k=\Lambda(t_k)-\Lambda(t_{k-1})$, $k\ge1$, are i.i.d. Exp(1): the points $\Lambda(t_1)<\Lambda(t_2)<\cdots$ form a Poisson process of rate 1.

(ii) (Multivariate version.) *If* $\Lambda^i(t) = \int_0^t \lambda^i \to \infty$ *almost surely for each* i*, the points* $\Lambda^i(t)$ *at the type-*i *event times form* d *independent Poisson processes of rate* 1 *(Meyer, 1971; Brown and Nair, 1988). More generally, (i) holds for every simple point process whose compensator is continuous and tends to infinity (Daley and Vere-Jones (2003), Theorem 7.4.I).*

Proof. (i) Given H_{k-1}, the function $L(t) = \int_{t_{k-1}}^t \bar\lambda(u \mid H_{k-1})\,du$ is continuous, nondecreasing, and tends to infinity. For $x \ge 0$ let $\sigma_x = \sup\{t : L(t) \le x\}$, so that $L(\sigma_x) = x$ and $\{\tau_k > x\} = \{L(t_k) > x\} = \{t_k > \sigma_x\}$. By (18.1), $\mathbb{P}(\tau_k > x \mid H_{k-1}) = e^{-L(\sigma_x)} = e^{-x}$. This does not depend on H_{k-1}, of which $\tau_1, \dots, \tau_{k-1}$ are functions, so τ_k is Exp(1) and independent of $(\tau_1, \dots, \tau_{k-1})$; by induction the τ_k are i.i.d. □

For the processes of Theorem 18.1, (i) is the construction with exponential variables read in reverse. Its use is as a *test*: compute $\hat\tau_k$ with the fitted intensity and compare them with Exp(1), for instance by a QQ plot and the Kolmogorov–Smirnov statistic (for each type separately, by (ii)). As the parameters are estimated from the same data, the Kolmogorov–Smirnov p-values are approximate.

Figure 18.2 and Table 18.1 replicate both steps on simulated data with $(\mu, \alpha, \beta) = (0.5, 0.6, 2)$ and $T = 1000$. Maximum likelihood recovers the parameters, the standard errors from the observed information match the spread of the estimates, and the time-change test accepts the Hawkes fit and rejects a Poisson fit by a wide margin.

	true	mean of $\hat\vartheta$	std. dev. of $\hat\vartheta$	mean standard error	coverage of ± 1.96 s.e.
μ	0.5	0.504	0.033	0.035	96%
α	0.6	0.600	0.026	0.031	99%
β	2	2.034	0.197	0.183	92%

Table 18.1: Maximum likelihood on 100 independent paths on $[0, 1000]$ (about 1250 events each), with standard errors from the inverse observed information.

▶ Interactive demo: fit a Hawkes process by maximum likelihood and test the fit by the time change, with the model right or wrong.

18.5 Order flow in a limit order book

Many exchanges match orders in a *limit order book*: a limit order offers to buy (a *bid*) or sell (an *ask*) a quantity at a given price and waits in the queue at that price; a market order buys or sells at once against the best opposite queue; a waiting order may be cancelled. The best bid and ask define the spread and the mid-price, which moves when a best queue is depleted. Message data list every event with a time stamp; the LOBSTER files reconstructed from Nasdaq data, for instance, have the columns (time, event type, order id, size, price, direction).

Two empirical facts make Hawkes processes natural here. Order flow is self- and cross-exciting: trades cluster (large orders are split into many small ones) and the book refills after a trade. Large (2007) fitted a ten-dimensional Hawkes model with exponential kernels to the London Stock Exchange book and measured this *resilience*; Bacry et al. (2015), Section 6, review the

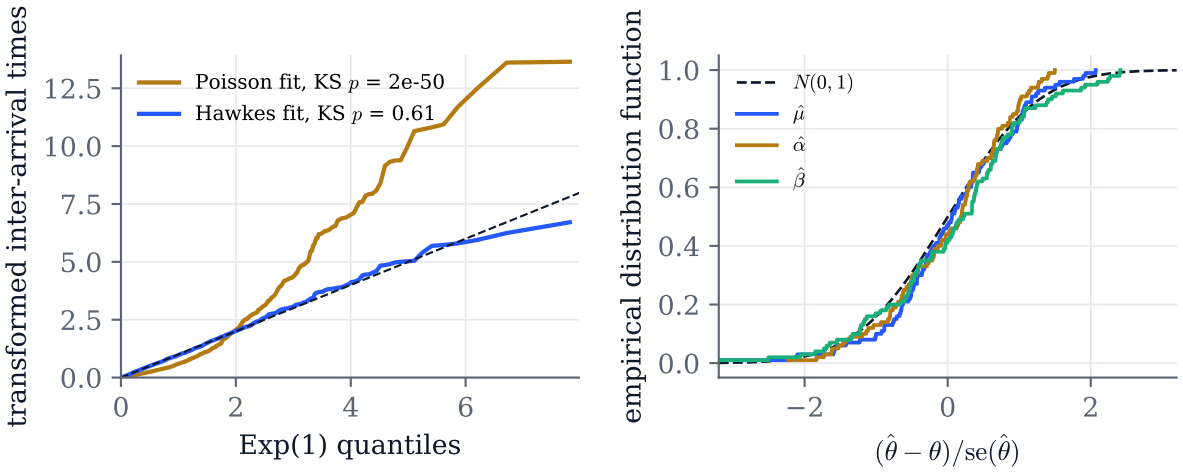

Figure 18.2: Left: one path (1237 events), fitted by maximum likelihood, $(\hat{\mu}, \hat{\alpha}, \hat{\beta}) = (0.509, 0.589, 1.94)$ with standard errors $(0.037, 0.032, 0.19)$; QQ plot of the transformed gaps of Theorem 18.10 under the fitted Hawkes intensity and under a Poisson fit with rate N_T/T. Right: empirical distribution functions of the standardized errors over the 100 paths of Table 18.1, against $\mathcal{N}(0,1)$.

order book models that followed. And the estimated branching ratios are close to one: Hardiman et al. (2013) find kernels with integral close to 1 for mid-price changes of E-mini S&P futures over 1998–2011, and the survey reviews similar findings (Section 3). Most events are then triggered by earlier events rather than by exogenous news.

A toy model. Take $d = 6$ types: market buy (MB) and sell (MS), limit order at the bid (LB) and the ask (LA), cancellation at the bid (CB) and the ask (CA), with exponential kernels of decay $\beta = 4\,\mathrm{s}^{-1}$ and the branching matrix of `code/src/ieor222/hawkes.py`, symmetric between the sides: market orders excite themselves (0.65), a market buy brings new limit orders to the ask (0.30), limit orders excite themselves (0.30) and their own cancellation (0.60), a cancellation is replaced (0.25) or attracts a market order on its side (0.10), and a market buy triggers some ask cancellations (0.05). Then $\varrho(G) = 0.755$, and by Theorem 18.6 the rates are $\Lambda = (0.148, 0.148, 0.284, 0.284, 0.218, 0.218)$ events per second, 1.30 in total, against $\mathbf{1}^\top \mu = 0.34$ from the baselines: three quarters of the activity is endogenous. A price is added by a queue mechanism: the best queues hold four lots, limit orders add one, market orders and cancellations remove one, and when a queue is empty the mid-price moves one tick away from it and both queues restart. Figure 18.3 shows four simulated hours. Fitting the 42 parameters by the concave likelihood of Theorem 18.9(iii), with the decay known, recovers every entry of G within 0.035 (the nonzero ones within 2.3 standard errors), and the time-change test accepts the fit for each type ($p \geq 0.66$).

The model is a toy: the intensities do not see the state of the book, which state-dependent extensions add. What it shows already is the macroscopic effect of endogeneity. On large time scales a Hawkes process is a drift plus a Brownian fluctuation, amplified by $(I - G)^{-1}$.

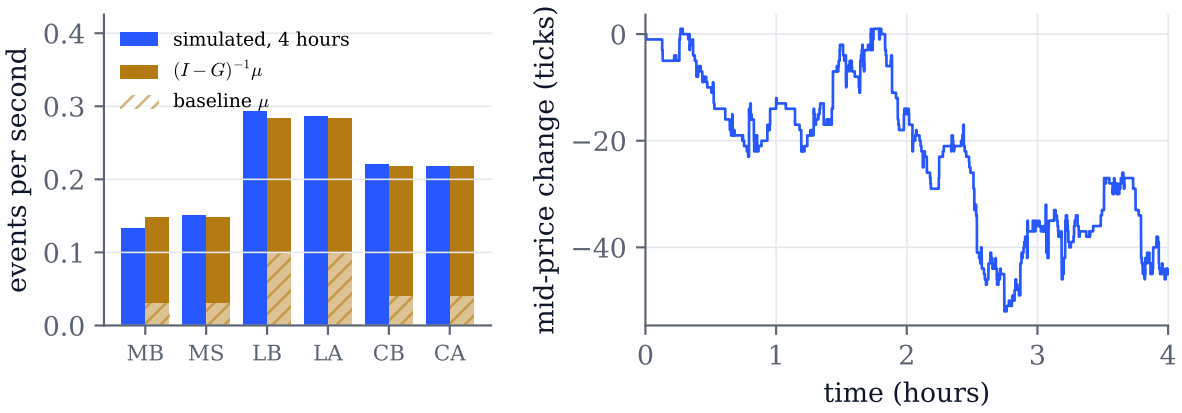

Figure 18.3: Toy order book over four hours (18 759 events, simulated by Theorem 18.8). Left: event rates against $(I-G)^{-1}\mu$, and the exogenous rates μ. Right: the mid-price, which moves 605 times.

Theorem 18.11 (Law of large numbers and central limit theorem)**.** *Let N be a d-variate Hawkes process with $\varrho(G) < 1$, $\Lambda = (I-G)^{-1}\mu$ and $\Sigma = \operatorname{diag}(\Lambda)$. Then $\sup_{v\le 1} |T^{-1}N_{Tv} - v\Lambda| \to 0$ almost surely and in L^2 as $T \to \infty$. If moreover $\int_0^\infty t^{1/2}\varphi_{ij}(t)\,\mathrm{d}t < \infty$ for all i, j, then $\left(\sqrt{T}\,(T^{-1}N_{Tv} - v\Lambda)\right)_{v\in[0,1]}$ converges in law, in the Skorokhod topology, to $(I-G)^{-1}\Sigma^{1/2}W$, where W is a d-dimensional Brownian motion (Bacry et al. (2013), Theorems 1–2 and Corollary 1).*

A price $P = c^\top N$, such as the number of upward minus downward ticks, is thus diffusive on large scales: $(P_{Tv} - Tv\,c^\top\Lambda)/\sqrt{T}$ converges to a Brownian motion with variance $c^\top(I-G)^{-1}\Sigma(I-G)^{-\top}c$ per unit time. The factor $(I-G)^{-1}$ explodes as $\varrho(G) \to 1$: in a nearly unstable book, a few exogenous events generate a large volatility (Exercise 18.5).

▶ Interactive demo: the six-dimensional order book: rates, the mid-price, estimation with known decays, and how to plug in a LOBSTER message file.

18.6 Nearly unstable Hawkes processes and rough volatility

If $\varrho(G)$ is close to 1 and the data span a long period, the relevant asymptotics are not those of Theorem 18.11 but those in which the branching ratio tends to 1 as the horizon grows. Jaisson and Rosenbaum (2015) take $d = 1$, a kernel $a_T\varphi$ with $\|\varphi\|_1 = 1$ and $a_T \uparrow 1$, and look at the process on $[0, T]$ in units of T. We keep the notation of these papers: in this section a_T is the branching ratio, λ (without superscript) a scaling constant and α a tail exponent; λ^T is the intensity.

Remark 18.12 (Where the CIR process comes from). Let $\varphi(t) = \beta e^{-\beta t}$, so $m = \int_0^\infty t\varphi(t)\,\mathrm{d}t = 1/\beta$. By Theorem 18.5, $\mathrm{d}\lambda_t^T = \beta(\mu - \lambda_t^T)\,\mathrm{d}t + a_T\beta\,\mathrm{d}N_t^T = \beta(\mu - (1-a_T)\lambda_t^T)\,\mathrm{d}t + a_T\beta\,\mathrm{d}M_t$, with $M = N^T - \int\lambda^T$. The rescaled intensity $C_t^T = (1-a_T)\lambda_{Tt}^T$ therefore satisfies

$$\mathrm{d}C_t^T = T(1-a_T)\beta\,(\mu - C_t^T)\,\mathrm{d}t + (1-a_T)a_T\beta\,\mathrm{d}M_{Tt}.$$

The martingale part has jumps of size $(1-a_T)a_T\beta \to 0$ at rate $T\lambda_{Tt}^T = TC_t^T/(1-a_T)$, so its

squared jumps accumulate at rate $T(1-a_T)a_T^2\beta^2 C_t^T$. If $T(1-a_T)\to\lambda>0$, the drift tends to $\frac{\lambda}{m}(\mu-C)$ and the squared noise to $\frac{\lambda}{m^2}C$: a CIR process with volatility $\sqrt{\lambda}/m$. The first two moments of C_1^T solve linear differential equations (Exercise 18.6(b)) and converge to those of the limit, whose law at time 1 is a Gamma law (Exercise 18.6(d)); for $\mu=1$, $\lambda=1$, $\beta=1/2$, the variance is 0.0395 at $T=50$ and 0.0387 in the limit.

For heavy-tailed kernels the limit involves the *Mittag-Leffler function* $E_{\alpha,\gamma}(z)=\sum_{n\ge0}z^n/\Gamma(\alpha n+\gamma)$, $E_\alpha=E_{\alpha,1}$, through the kernel $f^{\alpha,\lambda}(t)=\lambda t^{\alpha-1}E_{\alpha,\alpha}(-\lambda t^\alpha)$ and its integral $F^{\alpha,\lambda}(t)=1-E_\alpha(-\lambda t^\alpha)$. Near 0, $f^{\alpha,\lambda}(t)\approx\lambda t^{\alpha-1}/\Gamma(\alpha)$, the kernel of the Riemann–Liouville process of Theorem 19.1 with $H=\alpha-\frac12$.

Theorem 18.13 (Scaling limits of nearly unstable Hawkes processes)**.** *For $T>0$ let N^T be a Hawkes process with $d=1$, baseline $\mu_T>0$ and kernel $a_T\varphi$, where $a_T\in(0,1)$, $a_T\to1$, $\varphi\ge0$ and $\|\varphi\|_1=1$; let λ^T be its intensity.*

(i) (Light tails; Jaisson and Rosenbaum (2015), Theorem 2.2.) *Let $\mu_T=\mu>0$, $m=\int_0^\infty t\varphi(t)\,\mathrm{d}t<\infty$, φ differentiable with φ' bounded and integrable, and $T(1-a_T)\to\lambda>0$. Suppose that $T\psi^T(Tx)/\|\psi^T\|_1$ is bounded uniformly in $x\ge0$ and T, where $\psi^T=\sum_{n\ge1}(a_T\varphi)^{*n}$. Then $C_t^T=(1-a_T)\lambda_{Tt}^T$, $t\in[0,1]$, converges in law, in the Skorokhod topology, to the solution of*

$$\mathrm{d}X_t=\frac{\lambda}{m}(\mu-X_t)\,\mathrm{d}t+\frac{\sqrt{\lambda}}{m}\sqrt{X_t}\,\mathrm{d}B_t,\qquad X_0=0,$$

and $\frac{1-a_T}{T}N_{Tt}^T$ converges in law to $\int_0^t X_s\,\mathrm{d}s$.

(ii) (Heavy tails; Jaisson and Rosenbaum (2016), Proposition 3.2 and Theorems 3.1–3.2.) *Let $\alpha\in(\frac12,1)$ and suppose that $\alpha x^\alpha\int_x^\infty\varphi(s)\,\mathrm{d}s\to K>0$ as $x\to\infty$. Let $\delta=K\Gamma(1-\alpha)/\alpha$ and $\lambda,\mu^*>0$, with $T^\alpha(1-a_T)\to\lambda\delta$ and $T^{1-\alpha}\mu_T\to\mu^*/\delta$. Then $X_t^T=\frac{(1-a_T)\delta}{T^\alpha\mu^*}N_{Tt}^T$, $t\in[0,1]$, is tight in the Skorokhod topology, and every limit point is of the form $X_t=\int_0^t Y_s\,\mathrm{d}s$, where Y has Hölder continuous paths of every order $<\alpha-\frac12$ and solves, for a Brownian motion B,*

$$Y_t=F^{\alpha,\lambda}(t)+\frac{1}{\sqrt{\mu^*\lambda}}\int_0^t f^{\alpha,\lambda}(t-s)\sqrt{Y_s}\,\mathrm{d}B_s. \tag{18.6}$$

The limit in (i) is the CIR process of Theorem 14.9 with $\kappa=\lambda/m$, $\theta=\mu$, $\sigma=\sqrt{\lambda}/m$, started at 0; its law at time 1 is the Gamma law with shape $2\mu m$ and scale $(1-e^{-\lambda/m})/(2m)$ (Exercise 18.6(d)). The limit in (ii) is a *rough* CIR process: equivalently (Exercise 18.7),

$$Y_t=\frac{1}{\Gamma(\alpha)}\int_0^t(t-s)^{\alpha-1}\lambda(1-Y_s)\,\mathrm{d}s+\frac{1}{\Gamma(\alpha)}\int_0^t(t-s)^{\alpha-1}\frac{\lambda}{\sqrt{\mu^*\lambda}}\sqrt{Y_s}\,\mathrm{d}B_s,$$

the variance equation of the rough Heston model with $H=\alpha-\frac12$. In (ii) the rescaled intensity itself is typically not tight; only its integral converges. Heavy tails mean long memory, and long memory with near instability produces roughness.

El Euch et al. (2018) turned this into a model for prices. Let $N^{T,+}$ and $N^{T,-}$ count upward and downward ticks, $P^T=N^{T,+}-N^{T,-}$, and let $(N^{T,+},N^{T,-})$ be a bivariate Hawkes process with

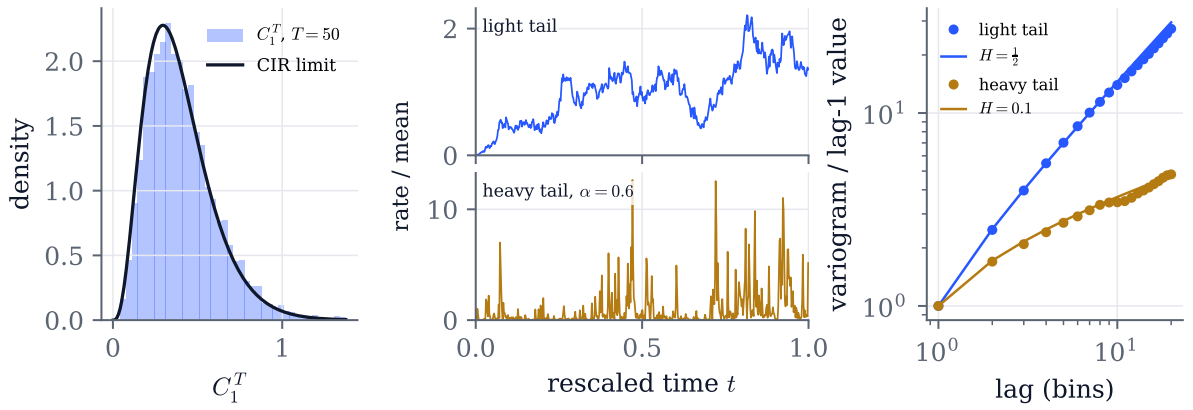

Figure 18.4: Left: C_1^T over 3000 paths with $\varphi(t) = \frac{1}{2}e^{-t/2}$, $\mu = \lambda = 1$, $T = 50$ ($a_T = 0.98$; mean 0.410, exact 0.406), against the Gamma law of the CIR limit (mean 0.393). Middle: event rates on 500 bins of $[0,1]$, light tail as on the left with $T = 2000$, heavy tail as in the text with $T = 10^6$. Right: variograms of the binned rates over eight paths of each kind (dots), against those of bin averages of a process with variogram $|r|^{2H}$, $H = \frac{1}{2}$ and $H = 0.1$ (lines).

baseline $\mu_T(1,1)$ and kernel matrix $a_T\Phi$,

$$\Phi = \begin{pmatrix} \varphi_1 & \beta\varphi_2 \\ \varphi_2 & \varphi_1 + (\beta-1)\varphi_2 \end{pmatrix}, \qquad \beta \geq 1, \quad \|\varphi_1\|_1 + \beta\|\varphi_2\|_1 = 1. \tag{18.7}$$

The rows of Φ have the same sum, $\varphi_1 + \beta\varphi_2$, so upward and downward ticks have the same expected rate (no statistical arbitrage). Here β is an asymmetry parameter, not a decay rate. If $\beta > 1$, a downward tick triggers more downward ticks than an upward tick triggers upward ones ($\varphi_1 + (\beta-1)\varphi_2 > \varphi_1$), and more upward ticks than an upward tick triggers downward ones ($\beta\varphi_2 > \varphi_2$). The matrix $\int_0^\infty \Phi$ has the eigenvalue 1, with right eigenvector $(1,1)$ and left eigenvector $(1,\beta)$, and the eigenvalue $\|\varphi_1\|_1 - \|\varphi_2\|_1$, with left eigenvector $(1,-1)$; so $\varrho(a_T \int \Phi) = a_T$.

Theorem 18.14 (Rough Heston as a scaling limit)**.** (El Euch et al. (2018), Theorem 3.1.) *In the model (18.7), let $\alpha \in (\frac{1}{2}, 1)$ and $C > 0$ with $\alpha x^\alpha \int_x^\infty (\varphi_1 + \beta\varphi_2)(s)\,\mathrm{d}s \to C$, and let $T^\alpha(1-a_T) \to \lambda^* > 0$ and $T^{1-\alpha}\mu_T \to \mu > 0$. Then $\sqrt{(1-a_T)/(\mu T^\alpha)}\, P^T_{tT}$, $t \in [0,1]$, converges in the sense of finite-dimensional distributions to*

$$P_t = \frac{1}{1 - (\|\varphi_1\|_1 - \|\varphi_2\|_1)} \sqrt{\frac{2}{1+\beta}} \int_0^t \sqrt{Y_s}\,\mathrm{d}W_s,$$

where, with $\lambda = \alpha\lambda^*/(C\Gamma(1-\alpha))$, $\theta = 1+\beta$ and $\nu = \sqrt{(1+\beta^2)/(\lambda^*\mu(1+\beta))}$,

$$Y_t = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}\lambda(\theta - Y_s)\,\mathrm{d}s + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}\lambda\nu\sqrt{Y_s}\,\mathrm{d}B_s,$$

and W, B are Brownian motions with $\mathrm{d}[W,B]_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}\,\mathrm{d}t$.

This is the rough Heston model of Theorem 19.15, with $V_0 = 0$, Hurst exponent $H = \alpha - \frac{1}{2}$, and

a correlation that is negative exactly when $\beta > 1$: the leverage effect comes from the asymmetry between buying and selling pressure. The same authors obtain the classical Heston model in the light-tailed case (their Theorem 2.1), and they note that estimated tail exponents α are close to $\frac{1}{2}$, which gives the small values of H measured in Chapter 19. The amplification factor $1/(1-(\|\varphi_1\|_1 - \|\varphi_2\|_1))$ is the one of Theorem 18.11 along the price direction $(1,-1)$.

Figure 18.4 replicates both regimes by simulation. With light tails, C_1^T at $T = 50$ is close to the Gamma law of the CIR limit; the remaining gap (mean 0.410 against 0.393, a Kolmogorov–Smirnov p-value of $5 \cdot 10^{-4}$ with 3000 paths) is the bias of the prelimit, of order $1/T$: its exact mean is 0.406. With a heavy-tailed kernel $\varphi(t) = \alpha(1+t)^{-1-\alpha}$, $\alpha = 0.6$ (so $K = \alpha$, $\delta = \Gamma(0.4)$), $\lambda = \mu^* = 1$ and $T = 10^6$ (about 10^6 events), the rescaled event rate is far rougher than with light tails. Its variogram matches the one of a process with $H = \alpha - \frac{1}{2} = 0.1$, once the averaging over bins is taken into account: the log-log slope of the same variogram would suggest $H \approx 0.24$, as averaging flattens rough variograms at small lags. The mean converges too, slowly: $\mathbb{E}[X_1^T]$, computed from the renewal equation of Theorem 18.6, is 0.462 at $T = 10^3$ (and 0.457 ± 0.012 over 600 simulated paths) and 0.442 at $T = 10^4$, against $\mathbb{E}[X_1] = \int_0^1 F^{0.6,1}(s)\,\mathrm{d}s = 0.431$ in the limit.

▶ Interactive demo: the nearly unstable regime: move T and the tail exponent and watch the CIR limit and the roughness appear.

Further reading

Bacry et al. (2015) survey Hawkes processes in finance: estimation (including non-parametric kernels), market stability, and order book models. Daley and Vere-Jones (2003), Chapter 7, develops conditional intensities, likelihoods and the time change in general. The scaling limits are in Jaisson and Rosenbaum (2015, 2016) and El Euch et al. (2018); Chapter 19 takes up the rough Heston model from the side of option pricing.

Chapter 19

Rough volatility

In the stochastic volatility models of Chapter 15 the variance is a diffusion, driven by a Brownian motion: its paths are Hölder continuous of every order below $\frac{1}{2}$, and its local behaviour looks like that of W. Two observations say that volatility is much rougher. In time series, the increments of log-volatility estimated from high-frequency data scale like Δ^H with H of order 0.1 (Gatheral et al., 2018). In option prices, the at-the-money skew of the S&P 500 decays with the maturity T roughly like $T^{-0.4}$ (Bayer et al., 2016), while in the Heston model it stays bounded as $T \downarrow 0$. Both facts are explained by one parameter: drive log-volatility by a fractional Brownian motion with Hurst index $H \approx 0.1$.

We construct fractional Brownian motion and its Riemann–Liouville variant (Section 19.1), estimate H as Gatheral et al. (2018) do (Section 19.2), price in the rough Bergomi model and derive the power law of the skew (Section 19.3), simulate (Section 19.4), and end with the rough Heston model, whose characteristic function solves a fractional Riccati equation (Section 19.5).

Setting. $K_q = \mathbb{E}|N|^q = 2^{q/2}\Gamma(\frac{q+1}{2})/\sqrt{\pi}$ for $N \sim \mathcal{N}(0,1)$ and $q > 0$. Gaussian vectors and processes are centred unless stated otherwise. For $x \in \mathbb{R}$ and $a \in \mathbb{R}$, $x_+^a = x^a$ if $x > 0$ and $x_+^a = 0$ if $x \le 0$ (so $x_+^0 = \mathbf{1}_{\{x>0\}}$).

19.1 Fractional Brownian motion

Definition 19.1 (Fractional Brownian motion)**.** Let $H \in (0,1)$. A *fractional Brownian motion* (fBM) with *Hurst index* H is a process $W^H = (W_t^H)_{t\ge 0}$ with $W_0^H = 0$ and continuous paths whose finite-dimensional vectors are centred Gaussian with

$$\operatorname{Cov}(W_s^H, W_t^H) = R_H(s,t) := \tfrac{1}{2}(s^{2H} + t^{2H} - |t-s|^{2H}), \qquad s,t \ge 0. \tag{19.1}$$

For $H = \frac{1}{2}$, $R_H(s,t) = \min(s,t)$ and W^H is a Brownian motion (Theorem 5.7). For other H it is not obvious that R_H is a covariance, that is, that $\sum_{i,j} c_i c_j R_H(t_i,t_j) \ge 0$. We prove it by writing down a process with this covariance.

Wiener integrals over the line. Let W and W' be independent Brownian motions and $h \in L^2(\mathbb{R})$. By Theorem 7.10, $\int_T^{T'} h(s)\,\mathrm{d}W_s \sim \mathcal{N}(0, \int_T^{T'} h^2)$ for $T < T'$, so $\int_0^T h(s)\,\mathrm{d}W_s$ is Cauchy in L^2 as

$T \to \infty$; let $\int_{\mathbb{R}} h\,\mathrm{d}B$ be the L^2-limit of $\int_0^T h(s)\,\mathrm{d}W_s + \int_0^T h(-s)\,\mathrm{d}W'_s$. The map $h \mapsto \int_{\mathbb{R}} h\,\mathrm{d}B$ is linear, so every linear combination of such integrals is again one; as L^2-limits of centred Gaussian variables are centred Gaussian (their characteristic functions converge), $(\int_{\mathbb{R}} h_1 \mathrm{d}B, \dots, \int_{\mathbb{R}} h_m \mathrm{d}B)$ is a centred Gaussian vector with $\mathbb{E}[\int h_i \mathrm{d}B \int h_j \mathrm{d}B] = \int_{\mathbb{R}} h_i h_j$. Here $B_t = W_t$ for $t \geq 0$ and $B_t = W'_{-t}$ for $t < 0$ is a *two-sided* Brownian motion.

Proposition 19.2 (Mandelbrot–Van Ness representation)**.** *Let $H \in (0,1)$, $\alpha = H - \frac{1}{2}$, and $k_t(u) = (t-u)_+^\alpha - (-u)_+^\alpha$ for $t \geq 0$, $u \in \mathbb{R}$. Then $k_t \in L^2(\mathbb{R})$, $c_H^2 := \int_{\mathbb{R}} k_1(u)^2\,\mathrm{d}u \in (0,\infty)$, and $X_t = c_H^{-1} \int_{\mathbb{R}} k_t\,\mathrm{d}B$ is a centred Gaussian process with covariance R_H. In particular R_H is a covariance function.*

Proof. On $(0,1)$, $k_1(u)^2 = (1-u)^{2\alpha}$ is integrable because $2\alpha > -1$; near $u = 0^-$, $k_1(u)^2 \leq 2(1-u)^{2\alpha} + 2(-u)^{2\alpha}$ is integrable for the same reason; as $u \to -\infty$, $k_1(u) = \alpha|u|^{\alpha-1}(1+o(1))$ by the mean value theorem, and $|u|^{2\alpha-2}$ is integrable at infinity because $2\alpha - 2 < -1$. So $c_H < \infty$, and $c_H > 0$ since $k_1 \neq 0$. For $s < t$, $k_t(u) - k_s(u) = (t-u)_+^\alpha - (s-u)_+^\alpha = k_{t-s}(u-s)$, and $k_c(cu) = c^\alpha k_1(u)$ for $c > 0$. By the isometry, a shift and a change of scale,

$$\begin{aligned}\mathbb{E}\big[(X_t - X_s)^2\big] &= c_H^{-2} \int_{\mathbb{R}} k_{t-s}(u-s)^2\,\mathrm{d}u = c_H^{-2} \int_{\mathbb{R}} k_{t-s}(u)^2\,\mathrm{d}u \\ &= c_H^{-2}(t-s)^{2\alpha+1} \int_{\mathbb{R}} k_1(v)^2\,\mathrm{d}v = (t-s)^{2H}.\end{aligned}$$

As $k_0 = 0$, $X_0 = 0$ and $\mathbb{E}[X_t^2] = t^{2H}$; polarization, $\mathrm{Cov}(X_s, X_t) = \frac{1}{2}(\mathbb{E}X_s^2 + \mathbb{E}X_t^2 - \mathbb{E}(X_t - X_s)^2)$, gives R_H. □

Mandelbrot and Van Ness (1968) normalize by $\Gamma(H + \frac{1}{2})$ instead of c_H; the constant has the closed form $c_H^2 = \Gamma(H + \frac{1}{2})\Gamma(2-2H)/(2H\,\Gamma(\frac{3}{2} - H))$ (the inverse square of the constant C_H of Bayer et al. (2016), Section 2; checked against quadrature in `code/tests/test_roughvol.py`).

Proposition 19.3 (Properties of fBM)**.** *Let W^H be a fBM, $H \in (0,1)$.*

(i) (Self-similarity) *For $c > 0$, $(c^{-H} W^H_{ct})_{t \geq 0}$ is a fBM with index H.*

(ii) (Stationary increments) *For $s \geq 0$, $(W^H_{s+t} - W^H_s)_{t \geq 0}$ is a fBM with index H.*

(iii) (Characterization) *Conversely, let X have continuous paths, $X_0 = 0$, centred Gaussian finite-dimensional vectors and $\mathrm{Var}\, X_1 = 1$. If $(X_{ct})_t$ and $(c^H X_t)_t$ have the same finite-dimensional laws for every $c > 0$, and so do $(X_{s+t} - X_s)_t$ and $(X_t)_t$ for every $s \geq 0$, then X is a fBM with index H.*

(iv) (Correlated increments) *For $n \in \mathbb{N}$, the increments over $[n, n+1]$ and $[0,1]$ have covariance*

$$\rho_H(n) = \tfrac{1}{2}\big((n+1)^{2H} - 2n^{2H} + |n-1|^{2H}\big). \tag{19.2}$$

For $n \geq 1$, $\rho_H(n)$ is positive if $H > \frac{1}{2}$, zero if $H = \frac{1}{2}$ and negative if $H < \frac{1}{2}$, and $\rho_H(n) \sim H(2H-1)n^{2H-2}$ as $n \to \infty$ when $H \neq \frac{1}{2}$.

(v) (Regularity) *$\mathbb{E}|W_t^H - W_s^H|^q = K_q |t-s|^{qH}$ for $q > 0$. Every centred Gaussian process with covariance R_H has a modification whose paths are locally Hölder continuous of every order $\gamma < H$.*

Proof. The processes in (i) and (ii) have continuous paths, start at 0 and are linear transformations of W^H, so their vectors are centred Gaussian; it suffices to compute covariances: $R_H(cs,ct) = c^{2H}R_H(s,t)$, and $R_H(s+u,s+t) - R_H(s+u,s) - R_H(s,s+t) + R_H(s,s) = R_H(u,t)$ after cancellation. (iii) By the first hypothesis $\operatorname{Var} X_t = t^{2H}\operatorname{Var} X_1 = t^{2H}$, by the second $\operatorname{Var}(X_t - X_s) = \operatorname{Var} X_{t-s} = (t-s)^{2H}$ for $s < t$, and polarization gives (19.1). (iv) For $a,b,c,d \geq 0$, expanding with (19.1) gives $\operatorname{Cov}(W_d^H - W_c^H, W_b^H - W_a^H) = \frac{1}{2}(|d-a|^{2H} + |c-b|^{2H} - |d-b|^{2H} - |c-a|^{2H})$; take $(a,b,c,d) = (0,1,n,n+1)$ (for $n=0$ the formula gives $\operatorname{Var} W_1^H = 1$). For $n \geq 1$, $2\rho_H(n)$ is the second difference $f(n+1) - 2f(n) + f(n-1)$ of $f(x) = x^{2H}$, which is strictly convex on $[0,\infty)$ for $H > \frac{1}{2}$, linear for $H = \frac{1}{2}$ and strictly concave for $H < \frac{1}{2}$. As $x \to 0$, $(1+x)^{2H} - 2 + (1-x)^{2H} = 2H(2H-1)x^2 + O(x^4)$; with $x = 1/n$, $\rho_H(n) = \frac{1}{2}n^{2H}(2H(2H-1)n^{-2} + O(n^{-4}))$. (v) $W_t^H - W_s^H \sim \mathcal{N}(0, |t-s|^{2H})$ by (ii). For $q > 1/H$, $\mathbb{E}|W_t^H - W_s^H|^q = K_q|t-s|^{1+(qH-1)}$, and the Kolmogorov–Chentsov theorem (Karatzas and Shreve (1991), Theorem 2.2.8) gives a continuous modification, locally Hölder of every order below $(qH-1)/q = H - 1/q$; let $q \to \infty$ (modifications of one process with continuous paths are indistinguishable, so one modification serves every q). □

By (v), together with Theorem 19.2, a fBM exists for every $H \in (0,1)$. The index H measures both the roughness of the paths and the sign of the correlation of increments: for $H < \frac{1}{2}$ they are negatively correlated (a rise tends to be followed by a fall), for $H > \frac{1}{2}$ positively, with $\sum_n \rho_H(n) = \infty$ (*long memory*; Exercise 19.1). Comte and Renault (1998) drove volatility by a fBM with $H > \frac{1}{2}$, for long memory; the data of Section 19.2 point to the other side.

Remark 19.4 (fBM is not a semimartingale). For $H \neq \frac{1}{2}$, W^H is not a semimartingale (Rogers, 1997): it cannot be written as a local martingale plus a process of finite variation, and the Itô calculus of Chapters 7 and 8 does not apply to it as an integrator. Along n equal steps of $[0,t]$ the expected sum of squared increments is $t^{2H}n^{1-2H}$: the quadratic variation is infinite for $H < \frac{1}{2}$ and zero for $H > \frac{1}{2}$ (Exercise 19.2), not t as in Theorem 5.16. With continuous trading, a price $e^{\sigma W_t^H}$ admits arbitrage for every $H \neq \frac{1}{2}$ (Rogers (1997); Cheridito (2003) constructs explicit strategies), and the arbitrage disappears if a minimal time between trades is imposed (Cheridito, 2003). Adding an independent Brownian motion can restore the semimartingale property: $W + W^H$ is a semimartingale if and only if $H \in \{\frac{1}{2}\} \cup (\frac{3}{4}, 1]$ (Cheridito, 2001). In the models below, the fBM drives the *volatility*; prices remain Itô integrals against a Brownian motion.

Rough volatility models under the pricing measure use a simpler process, with the same local behaviour: the integral over $(-\infty, 0)$ in Theorem 19.2 is dropped.

Proposition 19.5 (Riemann–Liouville process)**.** *Let $H \in (0,1)$, $\alpha = H - \frac{1}{2}$, W a Brownian motion, and $Y_t = \sqrt{2H}\int_0^t (t-u)^\alpha \, dW_u$, $t \geq 0$.*

(i) (Y,W) is a centred Gaussian process with $\operatorname{Var} Y_t = t^{2H}$ and, for $0 < s \leq t$,

$$\mathbb{E}[Y_s Y_t] = s^{2H}G(t/s), \qquad G(x) = 2H\int_0^1 (1-u)^\alpha (x-u)^\alpha \, du,$$
$$\mathbb{E}[Y_t W_s] = \frac{\sqrt{2H}}{H+\frac{1}{2}}\left(t^{H+\frac{1}{2}} - (t-s\wedge t)^{H+\frac{1}{2}}\right) \quad (s,t \geq 0). \tag{19.3}$$

*(ii) (*Scaling*) For $c > 0$, $(c^{-H}Y_{ct}, c^{-1/2}W_{ct})_{t\geq 0}$ has the finite-dimensional laws of $(Y_t, W_t)_{t\geq 0}$.*

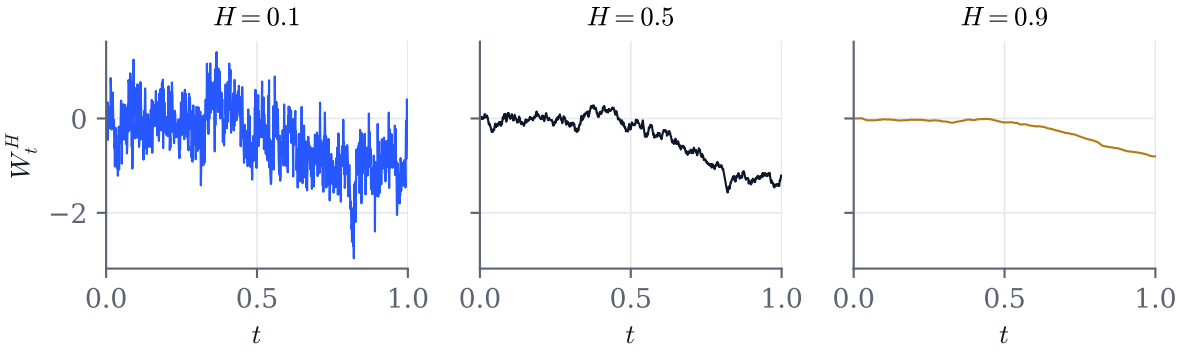

Figure 19.1: Paths of fBM on $[0, 1]$ with $H = 0.1$, 0.5 and 0.9, sampled exactly on 1000 steps from the same normal draws (Cholesky factorization of R_H, Section 19.4).

(iii) For $0 \le s < t$, $(t-s)^{2H} \le \mathbb{E}[(Y_t - Y_s)^2] \le 2Hc_H^2(t-s)^{2H}$. Hence Y has a modification with continuous paths, locally Hölder of every order $\gamma < H$; we always use it.

Proof. (i) $\int_0^t (t-u)^{2\alpha}\mathrm{d}u = t^{2H}/(2H)$, so Y_t is a Wiener integral (Theorem 7.10). A linear combination $\sum_i a_i Y_{t_i} + \sum_j b_j W_{s_j}$ is the Wiener integral of $\sum_i a_i\sqrt{2H}(t_i-u)_+^\alpha + \sum_j b_j \mathbf{1}_{\{u<s_j\}}$, hence centred Gaussian, and the covariances follow from the isometry: $\mathbb{E}[Y_sY_t] = 2H\int_0^s (s-u)^\alpha(t-u)^\alpha \mathrm{d}u$, which is $s^{2H}G(t/s)$ after $u = sv$, and $\mathbb{E}[Y_tW_s] = \sqrt{2H}\int_0^{s\wedge t}(t-u)^\alpha \mathrm{d}u$. (ii) By (19.3), $\mathbb{E}[Y_{cs}Y_{ct}] = c^{2H}\mathbb{E}[Y_sY_t]$ and $\mathbb{E}[Y_{ct}W_{cs}] = c^{H+1/2}\mathbb{E}[Y_tW_s]$, and $\mathbb{E}[W_{cs}W_{ct}] = c\min(s,t)$; centred Gaussian laws are determined by covariances. (iii) $\mathbb{E}(Y_t - Y_s)^2 = 2H\int_0^t \left((t-u)^\alpha - (s-u)_+^\alpha\right)^2 \mathrm{d}u$. The integral over $[s,t]$ alone is $(t-s)^{2H}/(2H)$; the integral over $[0,t]$ is at most the integral of the same integrand over $\mathbb{R}$, which is $\int_{\mathbb{R}}(k_t - k_s)^2 = c_H^2(t-s)^{2H}$ by the proof of Theorem 19.2. Conclude as in Theorem 19.3(v), with $\mathbb{E}|Y_t - Y_s|^q \le K_q(2Hc_H^2)^{q/2}|t-s|^{qH}$. □

Euler's integral for the hypergeometric function gives $G(x) = \frac{2H}{H+1/2}\, x^{H-1/2}\, {}_2F_1(1, \frac{1}{2}-H; H+\frac{3}{2}; \frac{1}{x})$ for $x \ge 1$ (Bayer et al. (2016), Section 4, in another form). If B is two-sided and $W = B$ on $[0,\infty)$, Theorem 19.2 reads $c_H W_t^H = Y_t/\sqrt{2H} + Z_t$ with $Z_t = \int_{-\infty}^0 ((t-u)^\alpha - (-u)^\alpha)\mathrm{d}B_u$, independent of Y and differentiable in mean square for $t > 0$: Y is a fBM with an independent smoother part removed. Its increments are not stationary: $\mathbb{E}(Y_{s+h} - Y_s)^2$ grows from h^{2H} at $s = 0$ to $2Hc_H^2h^{2H}$ as $s \to \infty$ (Exercise 19.6).

19.2 Volatility is rough

The volatility σ_t of an index is not observed, but the integrated variance over a day is estimated well by the *realized variance*, the sum of squared intraday log-returns (Section 15.7). The *Oxford-Man realized library* (Heber et al., 2009), now discontinued, published daily 5-minute realized variances of the main equity indices from 2000. Gatheral et al. (2018) (GJR) treat $x_k = \frac{1}{2}\log(\text{realized variance of day } k)$ as observations of $\log\sigma$ on a daily grid and compute, for $q > 0$ and a lag of Δ days,

$$m(q,\Delta) = \frac{1}{N}\sum_{j=1}^{N} |x_{j\Delta} - x_{(j-1)\Delta}|^q, \tag{19.4}$$

averaged over the Δ possible starting points of the subsampled grid. For each q they regress $\log m(q,\Delta)$ on $\log\Delta$, $\Delta = 1,\dots,30$: the points lie on a line of slope ζ_q, and $\zeta_q = qH$ with one H for all q (their Section 2). For the S&P 500 from 2000 to 2014 they report $H \approx 0.14$ (Section 2.3) and ζ_q/q between 0.124 and 0.128 for $q \in \{0.5, 1, 1.5, 2, 3\}$ (Appendix B), from different versions of the data and the regression; across the indices of the library, H lies between about 0.07 and 0.15. The increments of x are close to Gaussian (Section 2.5). We reproduce this on the public copy of the S&P 500 series in the notebook (3459 days from January 2000 to November 2013): $\zeta_q/q = 0.129, 0.127, 0.126, 0.126, 0.125$ for the same q, and $H = 0.126$ by least squares through the origin. The lag range matters: with $\Delta \le 50$ the same series gives ζ_q/q between 0.126 and 0.134, with $\Delta \le 100$ between 0.127 and 0.140.

Log-volatility thus behaves like a fBM, possibly with a slow mean reversion. Pathwise, the equation $X_t = m - \lambda\int_0^t (X_s - m)\,\mathrm{d}s + \nu W_t^H$ is a linear integral equation, solved by variation of constants.

Definition 19.6 (RFSV model)**.** Let $H \in (0, \frac{1}{2})$, $\nu > 0$, $\lambda \ge 0$ (GJR's α), $m \in \mathbb{R}$, and W^H a fBM under $\mathbb{P}$. In the *rough fractional stochastic volatility* (RFSV) model, $\sigma_t = e^{X_t}$ with

$$X_t = m + \nu\Big(W_t^H - \lambda\int_0^t e^{-\lambda(t-s)} W_s^H\,\mathrm{d}s\Big), \tag{19.5}$$

the solution of $X_t = m - \lambda\int_0^t (X_s - m)\,\mathrm{d}s + \nu W_t^H$ (Fubini's theorem).

GJR use the stationary solution $m + \nu\int_{-\infty}^t e^{-\lambda(t-s)}\mathrm{d}W_s^H$, a fractional Ornstein–Uhlenbeck process (Cheridito et al., 2003), with $\lambda \ll 1/T$ on the horizon T of the data; (19.5) starts it at $X_0 = m$ instead. The next proposition is the identity behind the regression.

Proposition 19.7 (Moment scaling)**.** *Let $q > 0$.*

(i) If $X = m + \nu W^H$ (that is, $\lambda = 0$), then $\mathbb{E}|X_{t+\Delta} - X_t|^q = K_q \nu^q \Delta^{qH}$ for all $t \ge 0$ and $\Delta > 0$. Hence the statistic (19.4) computed from $x_j = X_{j\delta}$ on any grid of step δ satisfies $\log \mathbb{E}[m(q,\Delta)] = \log(K_q\nu^q\delta^{qH}) + qH\log\Delta$ exactly: $\zeta_q = qH$.

(ii) If X is given by (19.5) with $\lambda > 0$, and $0 \le t < t + \Delta \le T$, then $X_{t+\Delta} - X_t$ is centred Gaussian and

$$\mathbb{E}|X_{t+\Delta} - X_t|^q = K_q\nu^q\Delta^{qH}(1+\varepsilon)^q, \qquad |\varepsilon| \le 2\lambda T^H \Delta^{1-H}. \tag{19.6}$$

Proof. (i) $X_{t+\Delta} - X_t = \nu(W^H_{t+\Delta} - W^H_t) \sim \mathcal{N}(0, \nu^2\Delta^{2H})$ by Theorem 19.3(ii), and $\mathbb{E}|cN|^q = K_q|c|^q$. Each term of (19.4) then has mean $K_q\nu^q(\delta\Delta)^{qH}$, and so does their average.

(ii) Let $Y = X - m$. The Riemann integral in (19.5) is the almost sure limit of Riemann sums, which are centred Gaussian, jointly with W^H; almost sure limits of centred Gaussian vectors are centred Gaussian (characteristic functions and dominated convergence). So Y is a centred Gaussian process, and $Y_{t+\Delta} - Y_t = \nu(W^H_{t+\Delta} - W^H_t) - \lambda\int_t^{t+\Delta} Y_s\,\mathrm{d}s$ by the integral equation. By Minkowski's inequality for integrals, $\|Y_s\|_{L^2} \le \nu\big(s^H + \lambda\int_0^s e^{-\lambda(s-r)} r^H \mathrm{d}r\big) \le 2\nu s^H$, so $\big\|\lambda\int_t^{t+\Delta} Y_s\,\mathrm{d}s\big\|_{L^2} \le 2\lambda\nu T^H\Delta$. By the triangle inequality in L^2, $\|Y_{t+\Delta} - Y_t\|_{L^2} = \nu\Delta^H(1+\varepsilon)$ with $|\varepsilon| \le 2\lambda T^H\Delta^{1-H}$, and for a centred Gaussian variable $\mathbb{E}|Z|^q = K_q\|Z\|_{L^2}^q$. □

With GJR's simulation parameters ($\lambda = 5\times10^{-4}$ per day, $H = 0.14$, 2000 days) the bound

(19.6) is 0.054 at $\Delta = 30$ days: mean reversion does not show at these lags, and ν is read off the intercept. The regression uses one path; for fBM, $m(2, \Delta)$ converges in L^2 as the sample grows, for every $H < 1$, at the rate $N^{-1/2}$ when $H < \frac{3}{4}$ (Exercise 19.3 treats that case). Linear ζ_q means that, unlike multifractal models, one exponent describes all moments.

Remark 19.8 (Spot volatility and its proxies). Realized variance averages σ^2 over the trading day, and averaging smooths: it biases the estimate of H upwards. GJR simulate the RFSV model with $H = 0.14$ and find about 0.16 from one-hour averages and 0.18 from whole-day realized variance (their Section 3.4); the rough Bergomi calibrations of Bayer et al. (2016) use H below 0.1. For a moving average of fBM the bias has a closed form (GJR, Appendix C; Exercise 19.4). Figure 19.2 repeats the experiment.

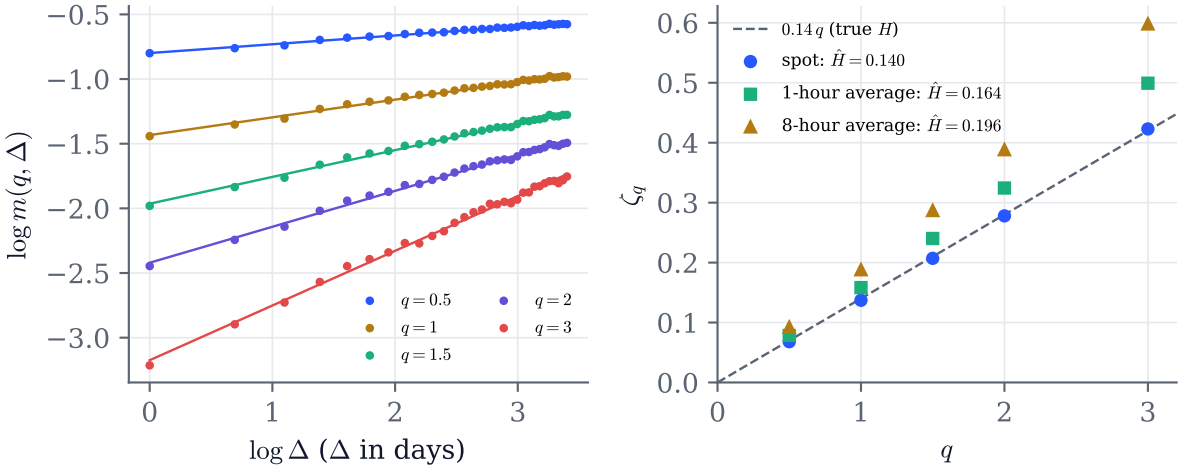

Figure 19.2: The GJR regression on 3500 simulated days of the RFSV model ($H = 0.14$, $\nu = 0.3$, $\lambda = 5 \times 10^{-4}$ per day, $X_0 = m = -5$; 192 steps per day, fBM by circulant embedding). Left: $\log m(q, \Delta)$ for the daily spot log-volatility, with the fitted lines. Right: the slopes ζ_q for the spot values and for the log of σ^2 averaged over the 8 grid points of the first hour and the 64 of the first eight hours of each day. The spot values give $\hat{\nu} = 0.298$ (from the intercept at $q = 2$); over 20 paths the spot estimate of H has standard deviation 0.010 (Exercise 19.8).

▶ Interactive demo: the GJR regression on simulated paths and, when the network is available, on the S&P 500 realized variance.

19.3 Rough Bergomi and the at-the-money skew

Under $\mathbb{P}$, the RFSV model with $\lambda \approx 0$ says $\log v_u - \log v_t = 2\nu(W_u^H - W_t^H)$ for the variance $v = \sigma^2$. Write W^H through Theorem 19.2. Given $\mathcal{F}_t$, the part of W_u^H driven by $\mathrm{d}B$ on $(t, u]$ is the independent Gaussian variable $c_H^{-1} \int_t^u (u - s)^{H-1/2} \mathrm{d}B_s$, and the rest is known. So v_u is its $\mathcal{F}_t$-conditional mean times a lognormal factor of mean one; with a deterministic change of drift to a pricing measure, Bayer et al. (2016), Sections 2–3, obtain the following model, with $\eta = 2\nu/(c_H\sqrt{2H})$.

Definition 19.9 (Rough Bergomi model)**.** Let $H \in (0, \frac{1}{2})$, $\eta > 0$, $\rho \in [-1, 1]$, and let $\xi_0 : [0, \infty) \to (0, \infty)$ be continuous (the *forward variance curve*). Under $\mathbb{Q}$, let $(W, W^\perp)$ be a two-dimensional Brownian motion with a filtration $(\mathcal{F}_t)$, $Z = \rho W + \sqrt{1-\rho^2}\, W^\perp$, and Y the Riemann–Liouville process of W (Theorem 19.5). The *rough Bergomi* (rBergomi) model is

$$v_t = \xi_0(t) \exp\Big(\eta Y_t - \frac{\eta^2}{2} t^{2H}\Big), \qquad S_t = S_0 \exp\Big((r-q)t + \int_0^t \sqrt{v_s}\, \mathrm{d}Z_s - \frac{1}{2}\int_0^t v_s\, \mathrm{d}s\Big). \tag{19.7}$$

Z is a Brownian motion for $(\mathcal{F}_t)$ (Exercise 5.4), v is continuous and adapted, and $\mathbb{E}^{\mathbb{Q}} v_t = \xi_0(t)$ because $\mathbb{E}^{\mathbb{Q}} e^{\eta Y_t} = e^{\eta^2 t^{2H}/2}$; so $\sqrt{v} \in \mathbb{H}^2[0, T]$ and the stochastic integral is that of Theorem 7.7. By Itô's formula, $\mathrm{d}S_t = (r-q)S_t \mathrm{d}t + \sqrt{v_t} S_t\, \mathrm{d}Z_t$. The correlation of the shocks to price and volatility is ρ, the volatility of volatility is η, and the variance v_u depends on the whole past of W through Y. The state that summarizes the past is the forward variance curve $u \mapsto \xi_t(u) = \mathbb{E}^{\mathbb{Q}}[v_u \mid \mathcal{F}_t]$, $u \geq t$ (Bayer et al. (2016), Section 3; Exercise 19.5).

Proposition 19.10 (Properties of rough Bergomi)**.** *In the setting of Theorem 19.9:*

(i) $\mathbb{E}^{\mathbb{Q}}[v_t] = \xi_0(t)$*, and the expected realized variance is* $\frac{1}{T}\mathbb{E}^{\mathbb{Q}} \int_0^T v_t\, \mathrm{d}t = \frac{1}{T}\int_0^T \xi_0(t)\, \mathrm{d}t$*.*

(ii) If $\rho \leq 0$*,* $(e^{-(r-q)t} S_t)$ *is a* $\mathbb{Q}$*-martingale (Gassiat, 2019, Theorem 1); if* $\rho > 0$ *it is not one:* $\mathbb{E}^{\mathbb{Q}}[e^{-(r-q)t} S_t] < S_0$ *for every* $t > 0$ *(same reference).*

(iii) (Scaling) *Let* $r = q = 0$ *and* $\xi_0 \equiv \xi$*. For* $c > 0$*,* $\log(S_{cT}/S_0)$ *in the model with parameters* (ξ, η) *has the law of* $\log(S_T/S_0)$ *in the model with parameters* $(c\xi, c^H \eta)$*, the same* H *and* ρ*. Hence the total implied variances satisfy* $w(k, cT; \xi, \eta) = w(k, T; c\xi, c^H \eta)$*.*

Proof. (i) was shown above; then use Tonelli's theorem. (iii) Write $v^{\xi,\eta}$ for the variance with parameters (ξ, η). By Theorem 19.5(ii), the scaling of Brownian motion (Theorem 5.9) and the independence of $W^\perp$ from W, the triple $(v^{\xi,\eta}_{c\cdot}, W_{c\cdot}, W^\perp_{c\cdot})$ has the finite-dimensional laws of $(c^{-1} v^{c\xi, c^H \eta}, \sqrt{c} W, \sqrt{c} W^\perp)$, since $\xi e^{\eta c^H Y_s - \eta^2 c^{2H} s^{2H}/2} = c^{-1} \cdot c\xi e^{(c^H \eta) Y_s - (c^H \eta)^2 s^{2H}/2}$. Now $\sqrt{v}$ is continuous in mean square with $\sup_{t \leq T} \mathbb{E}^{\mathbb{Q}} v_t < \infty$ ($v_u \to v_t$ almost surely and $\sup_{t \leq T} \mathbb{E}^{\mathbb{Q}} v_t^2 < \infty$ give uniform integrability), so by Theorems 7.4 and 7.9 $\int_0^{cT} \sqrt{v_u}\, \mathrm{d}Z_u$ is the L^2-limit of the sums $\sum_j \sqrt{v_{cs_j}}(Z_{cs_{j+1}} - Z_{cs_j})$ along partitions $\{s_j\}$ of $[0, T]$, and $\int_0^{cT} v_u \mathrm{d}u = c\int_0^T v_{cs} \mathrm{d}s$ is the limit of Riemann sums. Jointly, these sums have the law of the same sums for the model $(c\xi, c^H \eta)$ on $[0, T]$, because the factors $c^{-1/2}$ and $c^{1/2}$ cancel. Limits in probability of sequences with equal laws have equal laws. The last claim holds because normalized call prices are expectations of functions of $\log(S_T/S_0)$. □

The *at-the-money skew* is $\psi(T) = \partial_k \sigma_{\mathrm{imp}}(k, T)|_{k=0}$, with $k = \log(K/F_T)$ as in Chapter 15. In the Heston model of Theorem 15.11, with volatility of variance ξ_{Heston}, the zero-maturity smile has the finite slope $\rho\xi_{\mathrm{Heston}}/(4\sqrt{v_0})$ at the money (Forde et al. (2012), Corollary 4.3), and Figure 19.3 shows $\psi(T)$ approaching it. For the S&P 500, ψ is well fitted by $AT^{-0.407}$ (Bayer et al. (2016), Figure 1.2).

Proposition 19.11 (At-the-money skew)**.** *In the rough Bergomi model with* $r = q = 0$*,* $\xi_0 \equiv \xi > 0$*,* $H \in (0, \frac{1}{2})$*,* $\eta > 0$ *and* $\rho \in (-1, 1)$*,*

$$\lim_{T \downarrow 0} T^{\frac{1}{2} - H}\, \psi(T) = \frac{\rho\eta}{2} E_H, \qquad E_H = \frac{\sqrt{2H}}{(H + \frac{1}{2})(H + \frac{3}{2})}. \tag{19.8}$$

In particular $|\psi(T)| \to \infty$ like $T^{H-1/2}$ if $\rho \neq 0$ (El Euch et al. (2019): Theorem 2.4 for the general expansion, Theorem 5.1 for the rough Bergomi constant).

That a volatility driven by fractional noise produces a skew of order $T^{H-1/2}$ was shown by Alòs et al. (2007) and Fukasawa (2017). The constant has a transparent origin: to first order in η, the expansion of Bergomi and Guyon gives, at every maturity (Bayer et al. (2016), equation (1.3) and Section 7.1),

$$\partial_k \sigma_{\mathrm{imp}}(k,T) \approx \frac{C_T}{2\xi^{3/2}T^2}, \qquad C_T = \int_0^T \int_t^T \frac{\mathrm{d}[\log S, \xi_\cdot(u)]_t}{\mathrm{d}t}\,\mathrm{d}u\,\mathrm{d}t, \tag{19.9}$$

where $\xi_t(u) = \mathbb{E}^{\mathbb{Q}}[v_u \mid \mathcal{F}_t]$ is the forward variance. In rBergomi, $\mathrm{d}[\log S, \xi_\cdot(u)]_t \approx \rho\eta\sqrt{2H}\,\xi^{3/2}(u-t)^{H-1/2}\mathrm{d}t$ at first order, and the double integral is $\rho\eta\xi^{3/2}E_H T^{H+3/2}$, which gives (19.8) at every maturity (Exercise 19.5). With an exponential kernel $e^{-\kappa(u-t)}$ in place of $(u-t)^{H-1/2}$ (the Bergomi and Heston models) the same computation gives a skew with a finite limit at $T=0$. Finally, Theorem 19.10(iii) explains the power law: since $\psi(cT;\xi,\eta) = c^{-1/2}\psi(T;c\xi,c^H\eta)$ and the first-order skew is linear in η and does not depend on ξ, it scales like $c^{H-1/2}$.

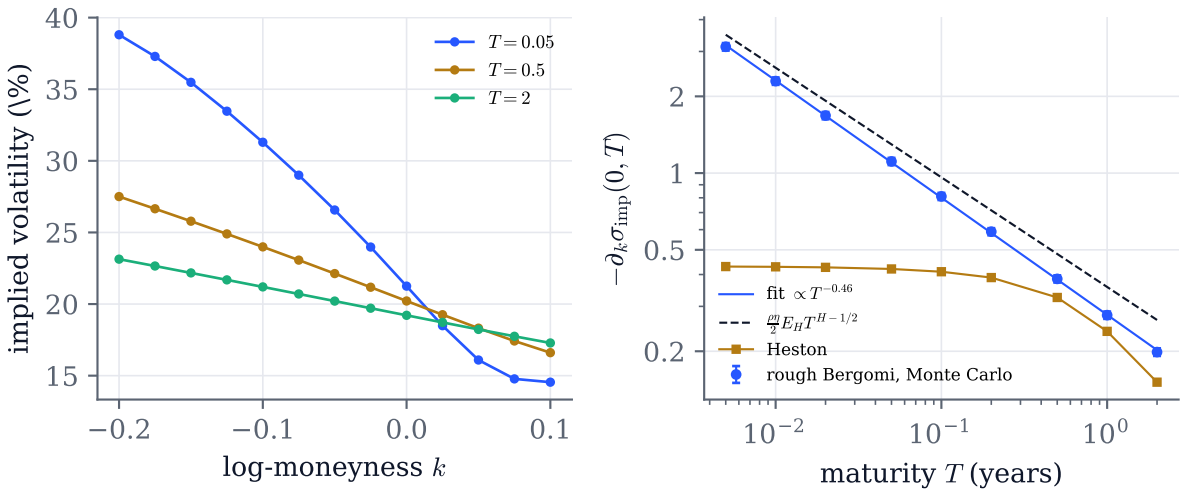

Figure 19.3: Rough Bergomi with the parameters of Bayer et al. (2016) for the S&P 500 on 4 February 2010, $H = 0.07$, $\eta = 1.9$, $\rho = -0.9$, and a flat forward variance $\xi = 0.235^2$: Monte Carlo with 2^{15} paths and 256 steps (hybrid scheme, conditional sampling). Left: smiles. Right: the at-the-money skew with 95% intervals, the least-squares power law (exponent -0.46; $H - \frac{1}{2} = -0.43$), the limit (19.8), which the Monte Carlo approaches as $T \downarrow 0$ (ratio 0.90 at $T = 0.005$), and a Heston model with the same ρ, $v_0 = \theta = \xi$, $\kappa = 2$, $\xi_{\mathrm{Heston}} = 0.45$, whose skew tends to $\rho\xi_{\mathrm{Heston}}/(4\sqrt{v_0}) = -0.43$.

19.4 Simulation

On a grid $0 < t_1 < \cdots < t_n$, a centred Gaussian vector G with covariance matrix Σ is simulated exactly as LN, with $LL^\top = \Sigma$ the Cholesky factorization and N a standard normal vector (Exercise 5.4). For fBM, $\Sigma_{ij} = R_H(t_i,t_j)$ (Figure 19.1); for rBergomi one factors the $2n \times 2n$

covariance of $(Y_{t_i}, W_{t_i})_{i\le n}$ from (19.3) (Bayer et al. (2016), Section 4). The factorization costs $O(n^3)$ and each path $O(n^2)$: fine for n in the hundreds. On a uniform grid the increments of fBM are stationary, their covariance matrix is Toeplitz, and *circulant embedding* samples them exactly with the fast Fourier transform in $O(n\log n)$ per path (Davies and Harte, 1987), provided the embedding has nonnegative eigenvalues, which holds when $H \le \frac12$ (Craigmile, 2003) (our code checks it for every H). The Riemann–Liouville process has no stationary increments, and the *hybrid scheme* of Bennedsen et al. (2017) takes its place.

Definition 19.12 (Hybrid scheme)**.** Let $H \in (0,1)$, $H \neq \frac12$, $\alpha = H - \frac12$, $n \ge 1$, $t_i = i/n$, $\Delta W_j = W_{t_{j+1}} - W_{t_j}$, $\ell \in \{0,1\}$, and evaluation points $b_k \in (k-1, k]$, $k \ge \ell+1$ (for $H = \frac12$ the kernel is constant and the scheme is exact for any b_k). The *hybrid scheme* approximates Y_{t_i} by

$$\hat Y_{t_i} = \sqrt{2H}\Big(\mathbf{1}_{\{\ell=1\}} \int_{t_{i-1}}^{t_i} (t_i - u)^\alpha \, \mathrm{d}W_u + \sum_{k=\ell+1}^{i} \Big(\frac{b_k}{n}\Big)^\alpha \Delta W_{i-k}\Big) :$$

the kernel is integrated exactly on the ℓ cells next to t_i, where it is singular, and frozen at $t_i - u = b_k/n$ on the cell $[t_{i-k}, t_{i-k+1}]$ otherwise. The pairs $(\Delta W_j, \int_{t_j}^{t_{j+1}} (t_{j+1}-u)^\alpha \mathrm{d}W_u)$ are independent Gaussian vectors with covariance matrix $\left(\begin{smallmatrix} 1/n & n^{-\alpha-1}/(\alpha+1) \\ n^{-\alpha-1}/(\alpha+1) & n^{-2H}/(2H) \end{smallmatrix}\right)$ (Theorem 7.10), and the sum is a discrete convolution, computed in $O(n\log n)$ by the fast Fourier transform.

Proposition 19.13 (Error of the hybrid scheme)**.** *For $i \ge 1$,*

$$\mathbb{E}\big[(Y_{t_i} - \hat Y_{t_i})^2\big] = 2H\, n^{-2H} \sum_{k=\ell+1}^{i} \int_{k-1}^{k} \big(y^\alpha - b_k^\alpha\big)^2 \, \mathrm{d}y. \tag{19.10}$$

Each term is minimal for $b_k^\alpha = \int_{k-1}^k y^\alpha \mathrm{d}y$, that is $b_k^* = \big((k^{\alpha+1} - (k-1)^{\alpha+1})/(\alpha+1)\big)^{1/\alpha}$. The series $J(\alpha,\ell,b) = \sum_{k>\ell} \int_{k-1}^k (y^\alpha - b_k^\alpha)^2 \mathrm{d}y$ converges, and at a fixed time $t>0$, along n with $nt \in \mathbb{N}$, $\mathbb{E}[(Y_t - \hat Y_t)^2] \sim 2HJ(\alpha,\ell,b)\, n^{-2H}$ as $n \to \infty$.

Proof. Y_{t_i} and $\hat Y_{t_i}$ are Wiener integrals against the same W on $[0, t_i]$; their difference has the integrand $\sqrt{2H}\big((t_i-u)^\alpha - (b_k/n)^\alpha\big)$ on the cell $k \ge \ell+1$ and 0 on the cells handled exactly. By the isometry and the substitution $t_i - u = y/n$, the cell k contributes $2Hn^{-2\alpha-1}\int_{k-1}^k (y^\alpha - b_k^\alpha)^2 \mathrm{d}y$, and $2\alpha + 1 = 2H$. For fixed k, $c \mapsto \int (y^\alpha - c)^2 \mathrm{d}y$ is minimal at the mean $c = \int_{k-1}^k y^\alpha \mathrm{d}y$. For $k \ge 2$, y and b_k lie in $[k-1, k]$, so by the mean value theorem $|y^\alpha - b_k^\alpha| \le |\alpha|(k-1)^{\alpha-1}$; the terms are $O(k^{2\alpha-2})$, summable since $2\alpha - 2 < -1$; and $J > 0$. At $t = i/n$ the sum in (19.10) runs to $i = nt \to \infty$ and tends to J. □

The rate n^{-H} of the root-mean-square error is slow for H near 0.1, but the constant is not: for $H = 0.1$, $2HJ = 0.538$ with $b_k = k$ and $\ell = 0$ (left-point Riemann sums), 0.446 with b_k^*, and 0.00136 with b_k^* and $\ell = 1$, so that the relative error at $n = 256$ falls from 42% to 2% (Figure 19.4). Bennedsen et al. (2017) prove the asymptotics for general Brownian semistationary processes (their Theorems 2.5 and 2.12, optimal points in Proposition 2.8).

Remark 19.14 (Monte Carlo in rough Bergomi). Given (Y, W) on the grid, $\int_0^T v\,\mathrm{d}t$ and $\int_0^T \sqrt{v}\,\mathrm{d}W$ are replaced by left-point sums (Theorem 7.9). The variance of the estimator drops sharply by conditioning on W (Romano and Touzi, 1997; McCrickerd and Pakkanen, 2018): as $W^\perp$ is independent of W and v is a functional of W, given W the variable $\log(S_T/S_0)$ (with

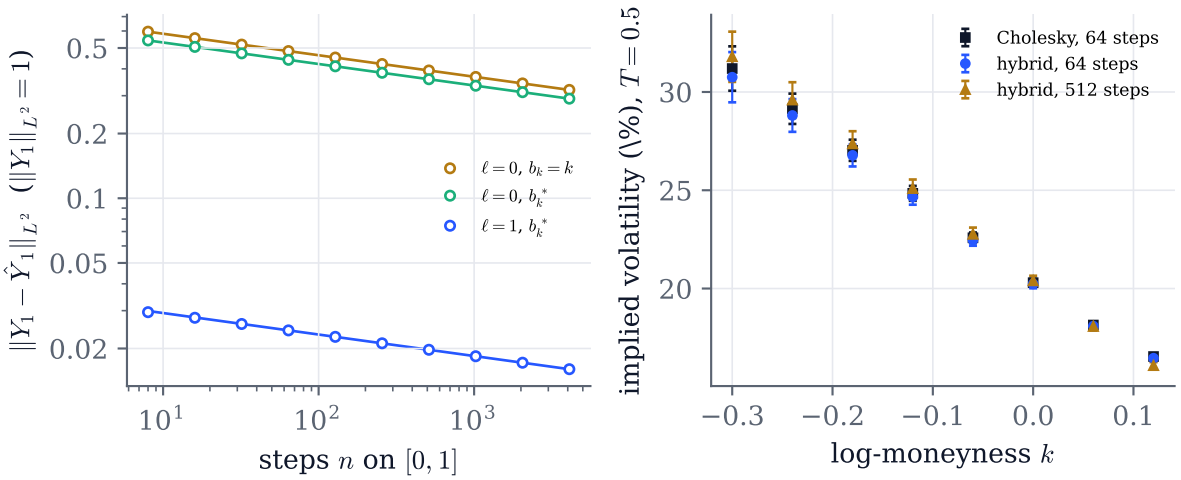

Figure 19.4: Left: root-mean-square error of the hybrid scheme for Y_1, $H = 0.1$, from (19.10) (open markers) and the asymptotics $\sqrt{2HJ}\, n^{-H}$ of Theorem 19.13 (lines). Right: rBergomi smiles at $T = 0.5$ (parameters of Figure 19.3), with exact Cholesky sampling and with the hybrid scheme on the same grid, and with the hybrid scheme on a grid eight times finer; bars are 95% intervals.

$r = q = 0$) is normal with mean $A - \frac{1}{2}(1-\rho^2)V$ and variance $(1-\rho^2)V$, where $V = \int_0^T v_t \mathrm{d}t$ and $A = \rho \int_0^T \sqrt{v_t}\, \mathrm{d}W_t - \frac{\rho^2}{2} V$. So, with b the normalized Black–Scholes price (15.2),

$$\mathbb{E}^{\mathbb{Q}}\big[(S_T/S_0 - e^k)^+\big] = \mathbb{E}^{\mathbb{Q}}\Big[e^A\, b\big(k - A, \sqrt{(1-\rho^2)V}\big)\Big],$$

an average of Black–Scholes prices that is smooth in k, so the skew can be computed by differentiating under the expectation. By Theorem 19.10(iii), one simulation on $[0,1]$ serves every maturity.

▶ Interactive demo: exact and hybrid simulation; rBergomi smiles and the skew power law as H, η and ρ move.

19.5 The rough Heston model

No closed-form prices are known for rough Bergomi. The Heston model of Theorem 15.11 has an affine structure that survives roughening: write its variance as $v_t = v_0 + \int_0^t \kappa(\theta - v_s)\mathrm{d}s + \int_0^t \xi\sqrt{v_s}\, \mathrm{d}B_s$ and put the fractional kernel $(t-s)^{H-1/2}/\Gamma(H+\frac{1}{2})$ inside both integrals. For f continuous on $[0,T]$ and $r > 0$, let $I^r f(t) = \frac{1}{\Gamma(r)} \int_0^t (t-s)^{r-1} f(s)\, \mathrm{d}s$ be the *fractional integral*.

Theorem 19.15 (Rough Heston model). *Let $H \in (0, \frac{1}{2}]$, $a = H + \frac{1}{2}$, $\lambda, \theta, \nu, V_0 > 0$, $\rho \in [-1, 1]$, and $K(t) = t^{a-1}/\Gamma(a)$.*

(i) There are a filtered probability space with a measure $\mathbb{Q}$, Brownian motions W and B with

$[W,B]_t = \rho t$, *and a continuous, adapted, nonnegative process* V *such that*

$$V_t = V_0 + \int_0^t K(t-s)\lambda(\theta - V_s)\,\mathrm{d}s + \int_0^t K(t-s)\lambda\nu\sqrt{V_s}\,\mathrm{d}B_s, \qquad t \geq 0. \tag{19.11}$$

The paths of V *are Hölder continuous of every order* $\gamma < H$, $S_t = S_0 \exp\big(\int_0^t \sqrt{V_s}\,\mathrm{d}W_s - \frac{1}{2}\int_0^t V_s\,\mathrm{d}s\big)$ *is a* $\mathbb{Q}$*-martingale, and the law of* (S,V) *is unique.*

(ii) Let $z \in \mathbb{C}$ *with* $-1 \leq \operatorname{Im} z \leq 0$ *and* $F(z,x) = -\frac{1}{2}(z^2+iz) + \lambda(i\rho\nu z - 1)x + \frac{1}{2}(\lambda\nu)^2x^2$. *The fractional Riccati equation* $h(z,\cdot) = I^a\big[F(z,h(z,\cdot))\big]$ *has a unique continuous solution on* $[0,\infty)$, *and*

$$\mathbb{E}^{\mathbb{Q}}\big[e^{iz\log(S_t/S_0)}\big] = \exp\Big(\theta\lambda\int_0^t h(z,s)\,\mathrm{d}s + V_0\int_0^t F(z,h(z,s))\,\mathrm{d}s\Big), \qquad t \geq 0. \tag{19.12}$$

(iii) For $H = \frac{1}{2}$, (S,V) *is the Heston model of Theorem 15.11 with* $\kappa = \lambda$ *and* $\xi = \lambda\nu$, $h(z,t)$ *is the function* D *of (15.12) with* $u = z$ *and* $T = t$, *and (19.12) is (15.12).*

El Euch and Rosenbaum (2019) prove (ii) for real z and $\rho \in (-1/\sqrt{2}, 1/\sqrt{2}]$ (their Theorem 4.1, written as $D^a h = F(z,h)$ with $I^{1-a}h(z,0) = 0$ and the term $V_0 I^{1-a}h(z,t)$; these agree with ours since $I^{1-a}I^a = I^1$), from the limit theorem for Hawkes processes discussed below. Abi Jaber et al. (2019) prove (i) and (ii) in the stated generality (their Theorem 7.1 and Example 7.2). For (iii), I^1 is integration, so $h' = F(z,h)$, $h(0) = 0$: the Riccati equation solved by D (Exercise 15.7), and $\theta\lambda\int_0^t h$ and $V_0\int_0^t F(z,h) = V_0 h(z,t)$ are the terms A and Dv_0. The strip in (ii) contains $z = u - \frac{i}{2}$, so Theorem 15.10 prices calls.

Through the kernel, V_t depends on the whole past of B, as Y_t does in rough Bergomi; yet the characteristic function costs one deterministic equation, as for Heston. No explicit solution is known for $H < \frac{1}{2}$. The *fractional Adams method* (Diethelm et al., 2002), a predictor–corrector scheme for $h = I^a[F(h)]$ with product-integration weights, solves it on n steps in $O(n^2)$ operations, used by El Euch and Rosenbaum (2019), Section 5. At $H = \frac{1}{2}$ our implementation agrees with (15.12) to 10^{-4} with 200 steps (`code/tests/test_roughvol.py`). Figure 19.5 reproduces the experiment of their Figure 1: with $H = 0.1$ the skew explodes at short maturities, with $H = \frac{1}{2}$ it stays flat. The mean variance is a Mittag-Leffler function of t^a (Exercise 19.7).

Remark 19.16 (Rough Heston from order flow). Rough Heston is not only convenient: it is the scaling limit of a microstructure model. If buy and sell market orders arrive as a nearly unstable Hawkes process with heavy-tailed kernels (Theorem 18.4), the rescaled price converges to the rough Heston price, with H determined by the tail of the kernel (Jaisson and Rosenbaum, 2016; El Euch et al., 2018); see Theorem 18.13. An asymmetry in how buy and sell orders excite each other, measured by a parameter $\beta \geq 0$, produces the leverage $\rho = (1-\beta)/\sqrt{2(1+\beta^2)}$, which lies in $(-1/\sqrt{2}, 1/\sqrt{2}]$ and is negative for $\beta > 1$; this is why El Euch and Rosenbaum (2019) state their theorem for that range.

Further reading

Gatheral et al. (2018) is the empirical starting point and very readable; Bayer et al. (2016) introduce rough Bergomi and explain why it fits the S&P 500 surface with three parameters.

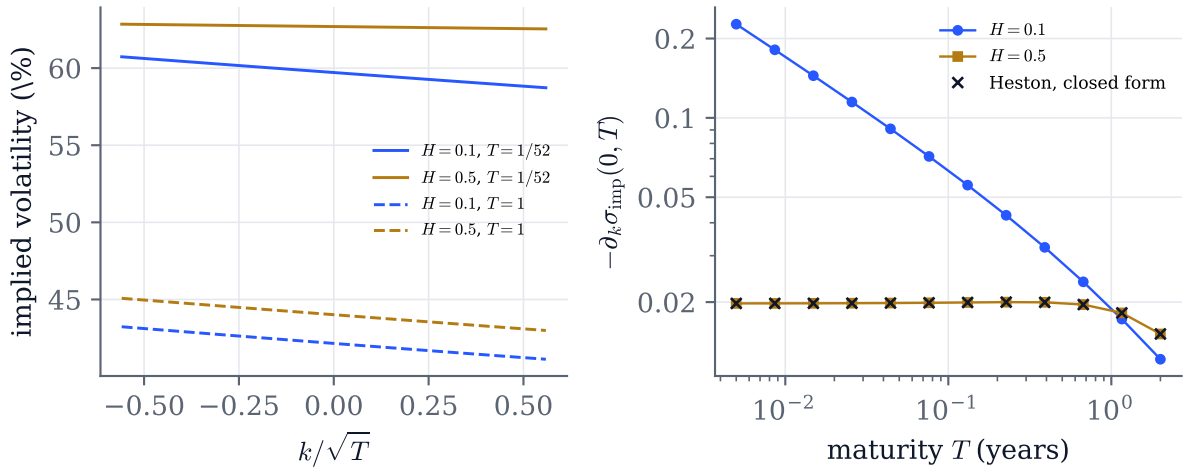

Figure 19.5: Rough Heston with the parameters of El Euch and Rosenbaum (2019), Section 5.2: $\lambda = 2$, $\rho = -0.5$, $V_0 = 0.4$, $\nu = 0.05$, $\theta = 0.04$, for $H = 0.1$ and $H = \frac{1}{2}$ (fractional Adams method, 200 steps, and Theorem 15.10). Left: smiles at $T = 1/52$ and $T = 1$ against $k/\sqrt{T}$. Right: the at-the-money skew; crosses are the closed form (15.12).

Bennedsen et al. (2017) and McCrickerd and Pakkanen (2018) are the references for simulation, and El Euch and Rosenbaum (2019) and Abi Jaber et al. (2019) for rough Heston and affine Volterra processes. Mandelbrot and Van Ness (1968) remains the classic introduction to fBM.

Chapter 20

Signatures and signature-based models

A path-dependent payoff, or the volatility of a model with memory, is a function of a whole path. This chapter is about a systematic set of coordinates for paths, the *signature*: the collection of all iterated integrals of the path. Chen (1957, 1958) studied it in geometry, and it is the central object of the theory of rough paths of Lyons (1998). Three facts make it useful in finance. The signature determines the path up to "tree-like" detours (Theorem 20.8). Products of its coordinates are linear combinations of coordinates (the shuffle identity, Theorem 20.7). Together, they make linear functionals of the signature universal approximators of continuous functionals of paths (Theorem 20.10). Cuchiero et al. (2023) use this to define *signature-based models*, in which the price is a linear functional of the signature of a simple *primary process*, such as a Brownian motion with time; calibration then splits into a simulation done once and an optimization over the coefficients. Cuchiero et al. (2025) model the volatility in the same way and fit S&P 500 (SPX) and VIX options jointly.

The algebra is proved for piecewise C^1 paths, where it is calculus. Brownian paths need stochastic integrals, and the choice between Itô and Stratonovich of Theorem 7.11 reappears: the signature is a Stratonovich object, and the conversion to Itô integrals is what makes the models martingales and their prices computable (Sections 20.2 and 20.3).

Words and tensors. Let $e_1, \dots, e_d$ be the standard basis of $\mathbb{R}^d$. A *word* is a finite sequence $I = (i_1, \dots, i_k)$ of *letters* $i_j \in \{1, \dots, d\}$, of length $|I| = k$; $\emptyset$ is the empty word, IJ the concatenation, and $I = I'i$ means that i is the last letter of I. The tensors $e_I = e_{i_1} \otimes \cdots \otimes e_{i_k}$ with $|I| = k$ form a basis of $(\mathbb{R}^d)^{\otimes k} \cong \mathbb{R}^{d^k}$, and $e_\emptyset = 1$. A *tensor series* is a formal sum $a = \sum_I a^I e_I$ over all words, that is, a sequence $(a_0, a_1, \dots)$ with $a_k \in (\mathbb{R}^d)^{\otimes k}$; their space $T((\mathbb{R}^d))$ is an algebra for

$$a \otimes b = \sum_{I,J} a^I b^J \, e_{IJ}, \qquad \text{that is,} \qquad (a \otimes b)^K = \sum_{K=IJ} a^I b^J,$$

the sum running over the $|K| + 1$ ways of cutting K in two. The *truncation* T^N keeps the words of length at most N. Finite sums $\ell = \sum_I \ell_I e_I$ form the space $T(\mathbb{R}^d)$ and act on series by $\langle \ell, a \rangle =$

$\sum_I \ell_I a^I$, so that $\langle e_I, a\rangle = a^I$. On $(\mathbb{R}^d)^{\otimes k}$ we use the Euclidean norm $|a_k| = (\sum_{|I|=k}(a^I)^2)^{1/2}$, for which $|u \otimes v| = |u|\,|v|$. For a with $a^{\emptyset} = 0$, $\exp(a) = \sum_{k\geq 0} a^{\otimes k}/k!$ is a finite sum at each level.

20.1 The signature of a path

A path $X : [0,T] \to \mathbb{R}^d$ is *piecewise* C^1 if it is continuous and there is a partition $0 = \tau_0 < \cdots < \tau_p = T$ such that X is continuously differentiable on each $[\tau_j, \tau_{j+1}]$, with one-sided derivatives at the ends. Piecewise linear paths are the main example. For such X and f continuous (or piecewise continuous), $\int_s^t f(u)\,\mathrm{d}X^i_u = \int_s^t f(u)\dot{X}^i_u\,\mathrm{d}u$, and $L(X;s,t) = \int_s^t |\dot{X}_u|\,\mathrm{d}u$ is the length of X on $[s,t]$.

Definition 20.1 (Signature)**.** Let $X : [0,T] \to \mathbb{R}^d$ be piecewise C^1 and $0 \leq s \leq t \leq T$. The *signature* of X on $[s,t]$ is the series $\mathrm{Sig}(X)_{s,t} = \sum_I \mathrm{Sig}(X)^I_{s,t}\, e_I$ with $\mathrm{Sig}(X)^{\emptyset}_{s,t} = 1$ and, for $I = (i_1, \ldots, i_k)$,

$$\mathrm{Sig}(X)^I_{s,t} = \int_{s<u_1<\cdots<u_k<t} \mathrm{d}X^{i_1}_{u_1} \cdots \mathrm{d}X^{i_k}_{u_k} = \int_{s<u_1<\cdots<u_k<t} \dot{X}^{i_1}_{u_1} \cdots \dot{X}^{i_k}_{u_k}\,\mathrm{d}u_1 \cdots \mathrm{d}u_k.$$

Equivalently, by Fubini's theorem, $\mathrm{Sig}(X)^{Ii}_{s,t} = \int_s^t \mathrm{Sig}(X)^I_{s,u}\,\mathrm{d}X^i_u$. Its *level* k is $\sum_{|I|=k} \mathrm{Sig}(X)^I_{s,t} e_I = \int_{s<u_1<\cdots<u_k<t} \mathrm{d}X_{u_1} \otimes \cdots \otimes \mathrm{d}X_{u_k}$, and the *truncated signature* $\mathrm{Sig}^N(X)_{s,t}$ keeps the levels $0, \ldots, N$.

Level 1 is the increment, $\mathrm{Sig}(X)^i_{s,t} = X^i_t - X^i_s$. At level 2, $\mathrm{Sig}(X)^{ij}_{s,t} = \int_s^t (X^i_u - X^i_s)\,\mathrm{d}X^j_u$, and by integration by parts $\mathrm{Sig}(X)^{ij}_{s,t} + \mathrm{Sig}(X)^{ji}_{s,t} = (X^i_t - X^i_s)(X^j_t - X^j_s)$: the symmetric part is known from level 1. The new information is the *Lévy area* $A^{ij}_{s,t} = \frac{1}{2}(\mathrm{Sig}(X)^{ij}_{s,t} - \mathrm{Sig}(X)^{ji}_{s,t})$. For a planar path, Green's theorem shows that $A^{12}_{s,t}$ is the signed area enclosed by the path and the chord from X_t back to X_s (on which $x\,\mathrm{d}y - y\,\mathrm{d}x = 0$ in coordinates centred at X_s), positive when this loop turns anticlockwise (Figure 20.1).

Lemma 20.2 (Factorial decay)**.** *For X piecewise C^1 and $k \geq 1$, the level k of $\mathrm{Sig}(X)_{s,t}$ has norm at most $L(X;s,t)^k/k!$.*

Proof. Level k is the integral of $\dot{X}_{u_1} \otimes \cdots \otimes \dot{X}_{u_k}$ over the simplex $\Delta_k = \{s < u_1 < \cdots < u_k < t\}$, so its norm is at most $\int_{\Delta_k} |\dot{X}_{u_1}| \cdots |\dot{X}_{u_k}|\,\mathrm{d}u$. The $k!$ orderings of the coordinates split the cube $[s,t]^k$, up to a null set, into $k!$ copies of Δ_k on which this symmetric integrand has the same integral, so the bound is $\frac{1}{k!}\big(\int_s^t |\dot{X}_u|\,\mathrm{d}u\big)^k$. □

Proposition 20.3 (A straight segment)**.** *If $X_u = X_s + \frac{u-s}{t-s}\,v$ on $[s,t]$ for some $v \in \mathbb{R}^d$, then $\mathrm{Sig}(X)_{s,t} = \exp(v)$, that is, $\mathrm{Sig}(X)^I_{s,t} = v^{i_1} \cdots v^{i_k}/k!$ for $I = (i_1, \ldots, i_k)$.*

Proof. $\dot{X} = v/(t-s)$ is constant, so the integrand is $v^{i_1} \cdots v^{i_k}/(t-s)^k$, and Δ_k has volume $(t-s)^k/k!$. □

Proposition 20.4 (Invariances)**.** *Let X be piecewise C^1 on $[s,t]$.*

*(i) (*Translation.*)* $\mathrm{Sig}(X + c)_{s,t} = \mathrm{Sig}(X)_{s,t}$ *for every $c \in \mathbb{R}^d$.*

(ii) (Reparametrization.) *If $\varphi : [s', t'] \to [s, t]$ is nondecreasing, onto and piecewise C^1, then* $\mathrm{Sig}(X \circ \varphi)_{s',t'} = \mathrm{Sig}(X)_{s,t}$.

Proof. (i) Only $\dot{X}$ enters. (ii) $Y = X \circ \varphi$ is piecewise C^1 (the preimage of each breakpoint of X is a point or an interval), with $\dot{Y}_r = \dot{X}_{\varphi(r)}\varphi'(r)$. We show $\mathrm{Sig}(Y)^I_{s',r} = \mathrm{Sig}(X)^I_{s,\varphi(r)}$ for all r by induction on $|I|$. It holds for $I = \emptyset$, and if it holds for I, the substitution $u = \varphi(r)$, on intervals where φ is C^1 and X is C^1 on φ's range, gives

$$\mathrm{Sig}(Y)^{Ii}_{s',r} = \int_{s'}^{r} \mathrm{Sig}(X)^I_{s,\varphi(q)} \dot{X}^i_{\varphi(q)} \varphi'(q)\,\mathrm{d}q = \int_{s}^{\varphi(r)} \mathrm{Sig}(X)^I_{s,u} \dot{X}^i_u\,\mathrm{d}u = \mathrm{Sig}(X)^{Ii}_{s,\varphi(r)}. \qquad \square$$

The signature records the order of the increments, not the speed. When the speed matters, as for prices observed in calendar time, we add time as a coordinate (Theorem 20.9).

Theorem 20.5 (Chen's identity)**.** *For X piecewise C^1 and $s \le u \le t$,*

$$\mathrm{Sig}(X)_{s,t} = \mathrm{Sig}(X)_{s,u} \otimes \mathrm{Sig}(X)_{u,t}, \qquad \textit{that is,} \qquad \mathrm{Sig}(X)^K_{s,t} = \sum_{K=IJ} \mathrm{Sig}(X)^I_{s,u}\,\mathrm{Sig}(X)^J_{u,t}$$

for every word K. In particular, the signature of the concatenation of two paths is the product of their signatures.

Proof. By induction on $|K|$; for $K = \emptyset$ both sides are 1. Let $K = K'i$. Splitting the integral at u and applying the induction hypothesis to $\mathrm{Sig}(X)^{K'}_{s,r}$, $r \ge u$,

$$\mathrm{Sig}(X)^K_{s,t} = \mathrm{Sig}(X)^K_{s,u} + \int_u^t \mathrm{Sig}(X)^{K'}_{s,r}\,\mathrm{d}X^i_r = \mathrm{Sig}(X)^K_{s,u} + \sum_{K'=IJ'} \mathrm{Sig}(X)^I_{s,u} \int_u^t \mathrm{Sig}(X)^{J'}_{u,r}\,\mathrm{d}X^i_r.$$

The last integral is $\mathrm{Sig}(X)^{J'i}_{u,t}$. The cuts of K are $(K, \emptyset)$, which gives the first term, and $(I, J'i)$ for the cuts (I, J') of K', which give the sum. $\square$

Corollary 20.6 (Piecewise linear paths)**.** *If X is linear on each $[t_j, t_{j+1}]$, $0 = t_0 < \cdots < t_m = T$, with increments $v_j = X_{t_{j+1}} - X_{t_j}$, then* $\mathrm{Sig}(X)_{0,T} = \exp(v_0) \otimes \exp(v_1) \otimes \cdots \otimes \exp(v_{m-1})$.

Proof. Theorem 20.5 at $t_1, \ldots, t_{m-1}$, and Theorem 20.3 on each piece. $\square$

In code

`code/src/ieor222/signatures.py` stores level k of a truncated series as a vector of length d^k, the word $(i_1, \ldots, i_k)$ at position $i_1 d^{k-1} + \cdots + i_k$, with letters numbered from 0 (for a time-augmented path, letter 0 is time, as in the notes). `signature(path, depth)` applies Theorem 20.6, one segment at a time: $S \leftarrow S \otimes \exp(v)$, whose level n is evaluated by Horner's scheme $(\cdots((S_0 \otimes \frac{v}{n} + S_1) \otimes \frac{v}{n-1} + S_2)\cdots) \otimes \frac{v}{1} + S_n$, at a cost of order d^N per segment ($d \ge 2$). The truncated signature has $(d^{N+1} - 1)/(d - 1)$ coordinates: 40 for $d = N = 3$, 127 for $d = 2$, $N = 6$. The tests check Theorems 20.2 to 20.5 and 20.7 on random paths.

Shuffles. A *shuffle* of the words I and J is a word of length $|I|+|J|$ obtained by interleaving them, keeping the order of the letters of each. The *shuffle product* $e_I \sqcup\!\sqcup e_J$ is the sum of the $\binom{|I|+|J|}{|I|}$ shuffles, counted with multiplicity (one for each choice of the positions of the letters of I); recursively, $e_I \sqcup\!\sqcup e_\emptyset = e_\emptyset \sqcup\!\sqcup e_I = e_I$ and

$$e_{Ii} \sqcup\!\sqcup e_{Jj} = (e_I \sqcup\!\sqcup e_{Jj}) \otimes e_i + (e_{Ii} \sqcup\!\sqcup e_J) \otimes e_j,$$

according to the word that provides the last letter. It extends bilinearly to $T(\mathbb{R}^d)$. For example, $e_1 \sqcup\!\sqcup e_2 = e_{12} + e_{21}$, $e_1 \sqcup\!\sqcup e_1 = 2e_{11}$ and $e_{12} \sqcup\!\sqcup e_3 = e_{123} + e_{132} + e_{312}$.

Proposition 20.7 (Shuffle identity)**.** *For X piecewise C^1, words I, J and $s \le t$, $\mathrm{Sig}(X)^I_{s,t}\,\mathrm{Sig}(X)^J_{s,t} = \langle e_I \sqcup\!\sqcup e_J, \mathrm{Sig}(X)_{s,t}\rangle$. Hence $\langle \ell, \mathrm{Sig}(X)_{s,t}\rangle \langle m, \mathrm{Sig}(X)_{s,t}\rangle = \langle \ell \sqcup\!\sqcup m, \mathrm{Sig}(X)_{s,t}\rangle$ for all $\ell, m \in T(\mathbb{R}^d)$, and $\mathrm{Sig}(X)^{(i,\dots,i)}_{s,t} = (X^i_t - X^i_s)^k/k!$ for the word of k letters i.*

Proof. By induction on $|I|+|J|$; the case $I = \emptyset$ or $J = \emptyset$ is trivial. Let $I = I'i$, $J = J'j$, $f(r) = \mathrm{Sig}(X)^I_{s,r}$ and $g(r) = \mathrm{Sig}(X)^J_{s,r}$. These are piecewise C^1, vanish at s, and $\mathrm{d}f = \mathrm{Sig}(X)^{I'}_{s,\cdot}\,\mathrm{d}X^i$, $\mathrm{d}g = \mathrm{Sig}(X)^{J'}_{s,\cdot}\,\mathrm{d}X^j$. Integration by parts on each piece and the induction hypothesis give

$$f(t)g(t) = \int_s^t f\,\mathrm{d}g + \int_s^t g\,\mathrm{d}f = \int_s^t \langle e_I \sqcup\!\sqcup e_{J'}, \mathrm{Sig}(X)_{s,r}\rangle \mathrm{d}X^j_r + \int_s^t \langle e_{I'} \sqcup\!\sqcup e_J, \mathrm{Sig}(X)_{s,r}\rangle \mathrm{d}X^i_r.$$

Since $\int_s^t \langle \ell, \mathrm{Sig}(X)_{s,r}\rangle\,\mathrm{d}X^j_r = \langle \ell \otimes e_j, \mathrm{Sig}(X)_{s,t}\rangle$ by Theorem 20.1, the right side is $\langle (e_I \sqcup\!\sqcup e_{J'}) \otimes e_j + (e_{I'} \sqcup\!\sqcup e_J) \otimes e_i, \mathrm{Sig}(X)_{s,t}\rangle$, which is the recursion for $e_I \sqcup\!\sqcup e_J$. For the last claim, $e_i \sqcup\!\sqcup \cdots \sqcup\!\sqcup e_i$ (k factors) is $k!$ times the word $(i, \dots, i)$. □

Algebraically, Ree (1958) showed that the formal series whose coefficients multiply by the shuffle rule are exactly the exponentials of Lie series; with Theorem 20.7, the logarithm of a signature is a Lie series, as Chen (1957) had shown. The coordinates of a truncated signature are therefore far from independent: by Theorem 20.7, products of coordinates of level at most N are linear combinations of coordinates of level at most $2N$.

What the signature determines. A continuous path $X : [0,T] \to \mathbb{R}^d$ is *tree-like* if there is a continuous $h : [0,T] \to [0,\infty)$ with $h(0) = h(T) = 0$ and

$$|X_t - X_s| \le h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u) \qquad \text{for all } s \le t$$

(Hambly and Lyons, 2010, Definition 1.3). A path followed by its reversal is tree-like (take for $h(u)$ the distance travelled from the start up to u on the way out, and the distance left to travel on the way back), and so is the detour in Figure 20.1.

Theorem 20.8 (Uniqueness of the signature, Hambly and Lyons (2010))**.** *Let X be a continuous path of bounded variation in $\mathbb{R}^d$. Then $\mathrm{Sig}(X)_{0,T} = 1$ if and only if X is tree-like (their Theorem 1). Consequently, two such paths X and Y have the same signature if and only if X followed by the reversal of Y, translated to start at X_T, is tree-like (their Corollary 1.5, which says "Lipschitz tree-like", the same thing here by their Theorem 1 and Chen's identity); and among the paths with a given signature there is, up to translation and reparametrization, a unique one of minimal length (their Corollary 1.7).*

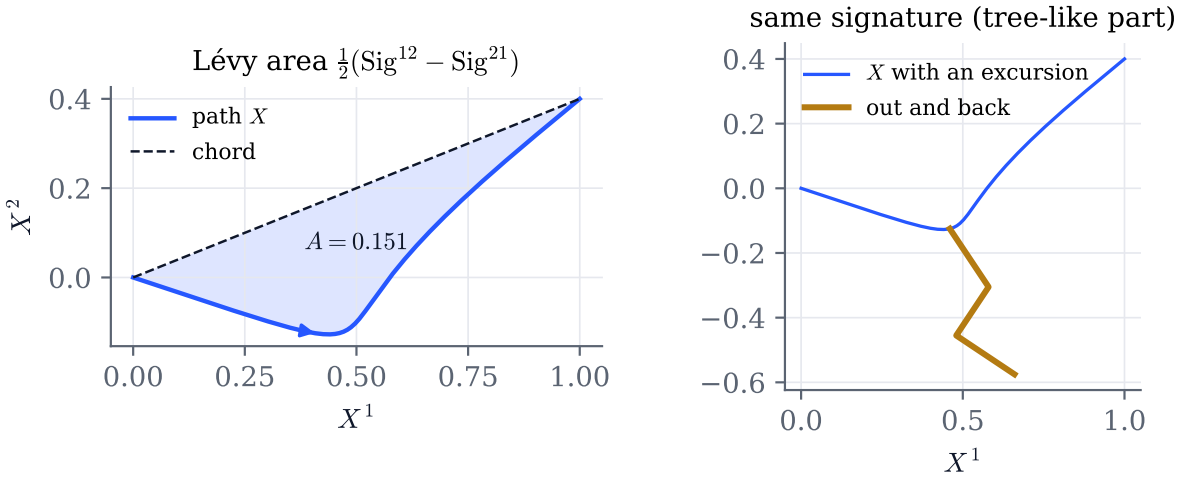

Figure 20.1: Left: a planar path and its chord; the Lévy area $\frac{1}{2}(\mathrm{Sig}^{12} - \mathrm{Sig}^{21})$ is the shaded signed area. Right: the same path with a detour that goes out along a zig-zag branch and comes back along it; its signature agrees with that of the path in every coordinate up to level 4, to rounding error (Theorem 20.8).

For paths of bounded variation the signature is defined with Riemann–Stieltjes integrals, which agree with ours for piecewise C^1 paths. Chen (1958) had proved the uniqueness for piecewise regular paths without back-tracking (irreducible paths). Adding time removes all ambiguity, with an elementary proof.

Proposition 20.9 (Time-augmented paths)**.** *Let $X : [0,T] \to \mathbb{R}^d$ be piecewise C^1, and $\widehat{X}_t = (t, X_t) \in \mathbb{R}^{1+d}$, with letter 0 for the time coordinate. For $i \in \{1,\dots,d\}$ and $k \geq 1$, writing $i0^k$ for the word $(i,0,\dots,0)$ with k zeros,*

$$\langle e_{i0^k}, \mathrm{Sig}(\widehat{X})_{0,T} \rangle = \int_0^T (X_u^i - X_0^i)\, \frac{(T-u)^{k-1}}{(k-1)!}\, \mathrm{d}u.$$

Hence, if Y is another piecewise C^1 path and $\mathrm{Sig}(\widehat{X})_{0,T} = \mathrm{Sig}(\widehat{Y})_{0,T}$, then $X - X_0 = Y - Y_0$ on $[0,T]$.

Proof. Let $F_k(t) = \langle e_{i0^k}, \mathrm{Sig}(\widehat{X})_{0,t} \rangle$. Then $F_0(t) = X_t^i - X_0^i$ and, as the letter 0 integrates against $\mathrm{d}t$, $F_k(t) = \int_0^t F_{k-1}(u)\,\mathrm{d}u$. The formula follows by induction, exchanging the integrals: $\int_0^T \int_0^v F_0(u) \frac{(v-u)^{k-2}}{(k-2)!} \mathrm{d}u\,\mathrm{d}v = \int_0^T F_0(u) \frac{(T-u)^{k-1}}{(k-1)!} \mathrm{d}u$. If the signatures agree, then $D = (X^i - X_0^i) - (Y^i - Y_0^i)$ is continuous and $\int_0^T D(u)p(u)\,\mathrm{d}u = 0$ for every polynomial p, since the polynomials $(T-u)^{k-1}$, $k \geq 1$, span them all. By the Weierstrass theorem (Rudin (1976), Theorem 7.26) there are polynomials $p_n \to D$ uniformly, so $\int_0^T D^2 = \lim_n \int_0^T D p_n = 0$, and $D = 0$. □

Theorem 20.10 (Universal approximation)**.** *Let $\mathcal{P}$ be the set of piecewise C^1 paths $X : [0,T] \to \mathbb{R}^d$ with $X_0 = 0$, with the distance $\delta(X,Y) = \int_0^T |\dot{X}_u - \dot{Y}_u|\,\mathrm{d}u$. Let $K \subset \mathcal{P}$ be compact and $f : K \to \mathbb{R}$ continuous. For every $\varepsilon > 0$ there is $\ell \in T(\mathbb{R}^{1+d})$, a finite combination of*

words, with

$$\sup_{X\in K}\left|f(X)-\langle \ell, \mathrm{Sig}(\widehat{X})_{0,T}\rangle\right|<\varepsilon.$$

Proof. Let $\mathcal{A}$ be the set of functions $X\mapsto\langle \ell,\mathrm{Sig}(\widehat{X})_{0,T}\rangle$ on K, $\ell\in T(\mathbb{R}^{1+d})$. *They are continuous.* Let $a=\dot{\widehat{X}}$, $b=\dot{\widehat{Y}}$, so that $|a-b|=|\dot{X}-\dot{Y}|$ and $\int_0^T|a|\le T+L(X;0,T)$. Writing $a^{\otimes k}-b^{\otimes k}=\sum_{j=1}^{k}a^{\otimes(j-1)}\otimes(a-b)\otimes b^{\otimes(k-j)}$ (with the variables $u_1<\cdots<u_k$) and bounding the integral over the simplex by that over $[0,T]^k$, every coordinate of level k changes by at most $k(T+\max(L(X),L(Y)))^{k-1}\delta(X,Y)$, and $L(Y)\le L(X)+\delta(X,Y)$. *$\mathcal{A}$ is an algebra containing the constants* by Theorem 20.7, with $\ell=e_\emptyset$ for the constants. *$\mathcal{A}$ separates the points of K* by Theorem 20.9, since all paths start at 0. By the Stone–Weierstrass theorem (Rudin (1976), Theorem 7.32), $\mathcal{A}$ is dense in the continuous functions on the compact metric space K for the uniform norm. □

This Stone–Weierstrass argument is standard in the signature literature (see the primer of Chevyrev and Kormilitzin (2016) and, for semimartingale lifts, the theorem of Cuchiero et al. below). Time matters here: without it, the signature of a one-dimensional path is $\exp(X_T-X_0)$ (Theorem 20.7) and sees only the increment (Exercise 20.2). Compact sets in δ are small, and Brownian paths have infinite length (Theorem 5.17); the theory of rough paths (Lyons, 1998; Lyons et al., 2007) extends the signature, Theorem 20.5 and Theorem 20.7 to paths of finite p-variation enriched with their first iterated integrals, and Cuchiero et al. (2023), Theorem 2.12, prove the analogue of Theorem 20.10 on compact sets of time-extended continuous semimartingales lifted with their Stratonovich integrals, for a suitable distance on these lifts, such as a p-variation rough-path distance with $p\in(2,3)$ (their Remark 2.13). In practice the approximation is used statistically: regress the payoff on the truncated signature of simulated paths. Figure 20.2 does this for an Asian call and for the maximum of a Brownian path: on 4000 test paths, R^2 rises from 0.52 at depth $N=1$ to 0.99 at $N=6$ for the Asian call, and from 0.69 to 0.98 for the maximum. The kink of the call and the non-smooth maximum slow the convergence.

▶ Interactive demo: signatures of planar paths, Chen and shuffle checks, and regression of path functionals on the signature, for any depth.

20.2 Signatures of Brownian paths

Throughout, $W=(W^1,\dots,W^d)=LB$ for a d-dimensional Brownian motion B for $(\mathcal{F}_t)$ and a matrix L with $LL^\top=\rho$, a correlation matrix, as in Chapter 8: $[W^i,W^j]_t=\rho_{ij}t$. Brownian paths are not of bounded variation, so Theorem 20.1 does not apply; simulated paths are piecewise linear, and level 2 shows what their signatures converge to.

Proposition 20.11 (Level two of Brownian paths)**.** *Let $\Pi=\{0=t_0<\cdots<t_m=T\}$, W^Π the piecewise linear interpolation of W on Π, and $i,j\in\{1,\dots,d\}$. For every outcome,*

$$\mathrm{Sig}(W^\Pi)^{ij}_{0,T}=\sum_{k=0}^{m-1}W^i_{t_k}(W^j_{t_{k+1}}-W^j_{t_k})+\frac{1}{2}C(W^i,W^j;\Pi),$$

with C the covariation sum of Theorem 7.12, and $\mathbb{E}\big[(C(W^i,W^j;\Pi)-\rho_{ij}T)^2\big]=(1+\rho_{ij}^2)\sum_k h_k^2$, $h_k=t_{k+1}-t_k$. As $\|\Pi\|\to0$, $\mathrm{Sig}(W^\Pi)^{ij}_{0,T}\to\int_0^T W^i_u\,\mathrm{d}W^j_u+\frac{1}{2}\rho_{ij}T$ in L^2. In particular

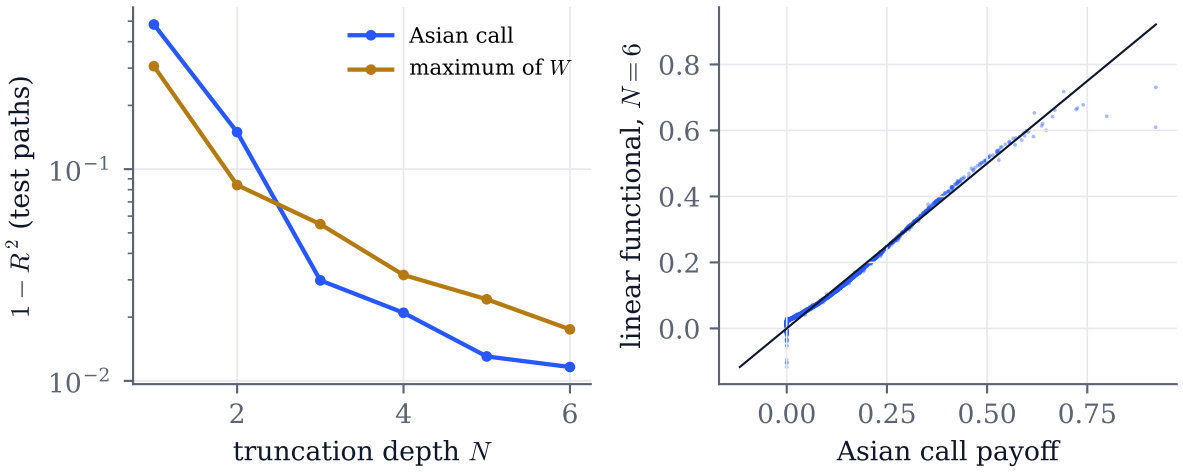

Figure 20.2: Least squares on $\mathrm{Sig}^N(\widehat{W})_{0,1}$, $\widehat{W}_t = (t, W_t)$, with 4000 training and 4000 test paths on 100 steps. Left: $1 - R^2$ on the test paths for the Asian call $(\int_0^1 S_t \,\mathrm{d}t - 1)^+$ with $S_t = e^{0.3W_t - 0.045t}$, and for $\max_{t \le 1} W_t$. Right: Asian payoffs against their linear predictor at $N = 6$.

$\mathrm{Sig}(W^\Pi)^{ii}_{0,T} = \frac{1}{2}(W^i_T)^2$, *and the Lévy area* A^{ij} *of* W^Π *converges to* $\frac{1}{2}\int_0^T (W^i \,\mathrm{d}W^j - W^j \,\mathrm{d}W^i)$.

Proof. By Theorem 20.6, level 2 of $\exp(v_0) \otimes \cdots \otimes \exp(v_{m-1})$ is $\sum_{k<l} v_k \otimes v_l + \frac{1}{2}\sum_k v_k \otimes v_k = \sum_l (W_{t_l} - W_0) \otimes v_l + \frac{1}{2}\sum_l v_l \otimes v_l$, with $v_l = W_{t_{l+1}} - W_{t_l}$; take the coordinate ij. The first sum converges in L^2 to the Itô integral by Theorem 7.9. The vectors v_k are independent and Gaussian with covariance $h_k \rho$, so the terms $v^i_k v^j_k$ are independent with mean $\rho_{ij} h_k$ and variance $\mathbb{E}[(v^i_k)^2]\mathbb{E}[(v^j_k)^2] + 2\mathbb{E}[v^i_k v^j_k]^2 - \rho^2_{ij} h^2_k = (1 + \rho^2_{ij}) h^2_k$ (write $v^j_k = \rho_{ij} v^i_k + z$ with z Gaussian and independent of v^i_k, and use $\mathbb{E}[(v^i_k)^4] = 3h^2_k$), which gives the variance of C, at most $(1 + \rho^2_{ij})T\|\Pi\|$. The diagonal case is Theorem 20.7 with $k = 2$. The Itô and covariation parts of the Lévy area cancel, as $C(W^i, W^j; \Pi) = C(W^j, W^i; \Pi)$. □

So the piecewise linear signature computes *Stratonovich* integrals: for $\rho_{ij} = 1$ this is the midpoint limit $\frac{1}{2}W^2_T$ of Theorem 7.11, not the Itô integral (Figure 20.3). The same holds at every level (a Wong–Zakai theorem; Friz and Victoir (2010), Chapter 13). We take the Stratonovich signature as the definition. For Itô processes Y, Z driven by B, let

$$\int_s^t Y_u \circ \mathrm{d}Z_u = \int_s^t Y_u \,\mathrm{d}Z_u + \frac{1}{2}([Y, Z]_t - [Y, Z]_s),$$

so that, by the product rule (Theorem 8.18), $Y_t Z_t - Y_s Z_s = \int_s^t Y \circ \mathrm{d}Z + \int_s^t Z \circ \mathrm{d}Y$: Stratonovich integrals follow the rules of ordinary calculus.

Definition 20.12 (Signature of Brownian motion with time)**.** Let $\widehat{X}_t = (t, W_t)$, with letters 0 (time) and $1, \ldots, d$, and write $\mathbb{X}_{s,t} = \mathrm{Sig}(\widehat{X})_{s,t}$ for $0 \le s \le t$: $\mathbb{X}^\emptyset_{s,t} = 1$ and $\mathbb{X}^{Ii}_{s,t} = \int_s^t \mathbb{X}^I_{s,u} \circ \mathrm{d}\widehat{X}^i_u$, with $\widehat{X}^0_u = u$ (an ordinary integral) and $\widehat{X}^i = W^i$ for $i \ge 1$.

By induction on $|I|$, each $\mathbb{X}^I_{s,\cdot}$ is an Itô process on $[s, \infty)$, so the definition makes sense. Let

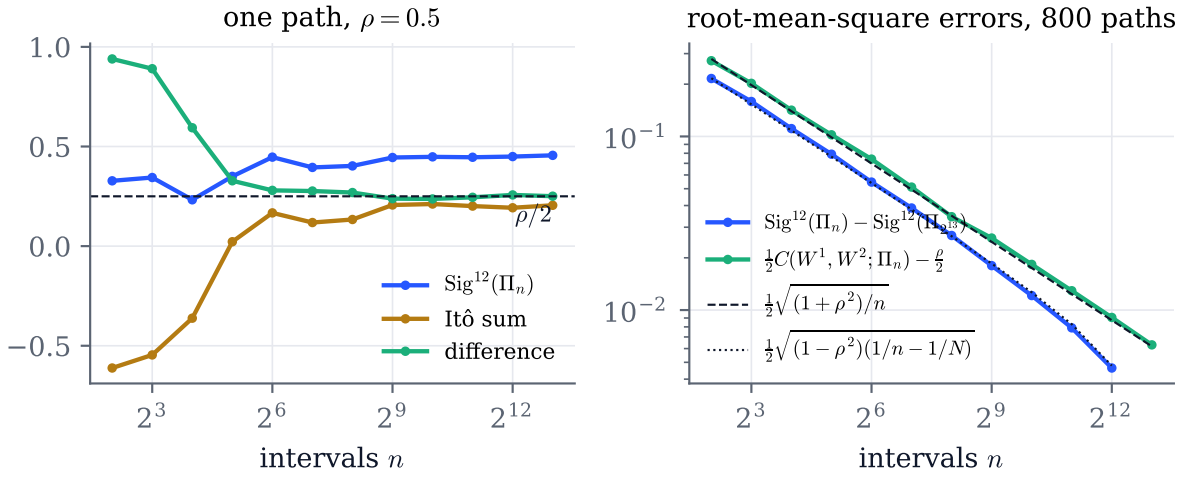

Figure 20.3: Two Brownian motions with $\rho = 0.5$, interpolated on n equal intervals of $[0,1]$ (computed with `code/src/ieor222/signatures.py`). Left: for one path, the level-2 coordinate Sig^{12}, the Itô sum, and their difference $\frac{1}{2}C(W^1, W^2; \Pi_n) \to \rho/2$. Right: over 800 paths, the root-mean-square distance of $\frac{1}{2}C$ to $\frac{\rho}{2}$, against its exact value $\frac{1}{2}\sqrt{(1+\rho^2)/n}$ (Theorem 20.11), and of $\mathrm{Sig}^{12}(\Pi_n)$ to its value on the finest grid Π_N, $N = 2^{13}$. The symmetric parts agree, so the latter is the sum of the Lévy areas of the fine path over the n intervals, of exact root-mean-square value $\frac{1}{2}\sqrt{(1-\rho^2)(1/n - 1/N)}$ (a computation as in Exercise 20.4).

$\rho_{0k} = \rho_{k0} = 0$, and for a word $I = I'i$ and $k \in \{1, \dots, d\}$ let

$$\tilde{e}^k_I = e_{Ik} - \tfrac{1}{2}\rho_{ik}\, e_{I'0}, \qquad \tilde{e}^k_\emptyset = e_k, \qquad A = e_0 + \frac{1}{2}\sum_{i,j=1}^{d} \rho_{ij}\, e_{ij}. \tag{20.1}$$

Proposition 20.13 (Brownian signatures)**.** *For $0 \le s \le u \le t$, words I, J and $k \in \{1, \dots, d\}$:*

(i) $\mathbb{E}\big[\sup_{s \le v \le t} |\mathbb{X}^I_{s,v}|^p\big] < \infty$ *for every* $p \ge 1$*;*

(ii) almost surely, $\mathbb{X}_{s,t} = \mathbb{X}_{s,u} \otimes \mathbb{X}_{u,t}$ *and* $\mathbb{X}^I_{s,t}\mathbb{X}^J_{s,t} = \langle e_I \sqcup\!\sqcup e_J, \mathbb{X}_{s,t}\rangle$*;*

(iii) (Itô integrals) $\int_s^t \mathbb{X}^I_{s,v}\, \mathrm{d}W^k_v = \langle \tilde{e}^k_I, \mathbb{X}_{s,t}\rangle$*;*

(iv) (expected signature) $\mathbb{E}[\mathbb{X}_{s,t} \mid \mathcal{F}_s] = \exp((t-s)A)$*, and* $\mathbb{E}[\mathbb{X}_{0,t} \mid \mathcal{F}_s] = \mathbb{X}_{0,s} \otimes \exp((t-s)A)$*.*

Proof. (iii) By definition $\mathbb{X}^{Ik} = \int \mathbb{X}^I\, \mathrm{d}W^k + \frac{1}{2}[\mathbb{X}^I_{s,\cdot}, W^k]$. If $I = I'i$ with $i \ge 1$, the $\mathrm{d}W$ part of $\mathbb{X}^I_{s,\cdot}$ is $\mathbb{X}^{I'}_{s,\cdot}\, \mathrm{d}W^i$, so by Theorem 8.11 and (8.5) $[\mathbb{X}^I_{s,\cdot}, W^k]_{s,t} = \rho_{ik}\int_s^t \mathbb{X}^{I'}_{s,v}\, \mathrm{d}v = \rho_{ik}\mathbb{X}^{I'0}_{s,t}$; if $i = 0$ or $I = \emptyset$, the covariation vanishes. (i) By induction on $|I|$, with (iii): $\mathbb{X}^{Ik}$ is an Itô integral plus a time integral, and the Burkholder–Davis–Gundy inequality (Revuz and Yor (1999), Chapter IV, Section 4) bounds $\mathbb{E}\sup|\int H\, \mathrm{d}W^k|^p$ by a constant times $\mathbb{E}(\int_s^t H^2 \mathrm{d}v)^{p/2}$. (ii) The proofs of Theorems 20.5 and 20.7 use only the recursion of Theorem 20.12, the product rule, which holds for Stratonovich integrals, $\int_u^t \xi H \circ \mathrm{d}\widehat{X}^i = \xi \int_u^t H \circ \mathrm{d}\widehat{X}^i$ for $\mathcal{F}_u$-measurable ξ, which holds for Itô integrals (for simple integrands by definition, in general by approximation and localization) and for covariations, and the associativity $\int Y \circ \mathrm{d}(\int H \circ \mathrm{d}Z) = \int YH \circ \mathrm{d}Z$, which follows from the same rule for Itô integrals and $\mathrm{d}[YH, Z] = Y\, \mathrm{d}[H, Z] + H\, \mathrm{d}[Y, Z]$. In

Chen's induction, the identity for $\mathbb{X}_{s,r}^{K'}$ is needed for all $r \geq u$ at once: it holds almost surely for each rational r, hence for all r, both sides being continuous in r. (iv) Fix $G \in \mathcal{F}_s$ and let $\psi^I(t) = \mathbb{E}[\mathbf{1}_G \mathbb{X}_{s,t}^I]$, finite by (i). The Itô integrals in (iii) are martingales vanishing at s (Theorem 7.7), so, with Fubini's theorem, $\psi^{Ik}(t) = \frac{1}{2}\rho_{ik}\int_s^t \psi^{I'}(v)\,\mathrm{d}v$ for $I = I'i$, $\psi^k = 0$, and $\psi^{I0}(t) = \int_s^t \psi^I(v)\,\mathrm{d}v$. This is the system $\psi(t) = \mathbb{P}(G) + \int_s^t \psi(v) \otimes A\,\mathrm{d}v$, which determines each coordinate from shorter words; its solution is $\mathbb{P}(G)\exp((t-s)A)$, since $\frac{\mathrm{d}}{\mathrm{d}t}\exp(tA) = \exp(tA) \otimes A$. As G is arbitrary, $\mathbb{E}[\mathbb{X}_{s,t} \mid \mathcal{F}_s] = \exp((t-s)A)$. The last claim follows from (ii), taking out what is known (Theorem 1.16). □

For a standard Brownian motion without time, (iv) reduces to $\mathbb{E}[\mathrm{Sig}(W)_{0,t}] = \exp(\frac{t}{2}\sum_i e_{ii})$, computed by B. Fawcett (Oxford thesis, 2003) and by T. Lyons and N. Victoir (*Proc. R. Soc. A*, 2004). Examples: $\mathbb{E}[\mathbb{X}_{0,t}^{11}] = \mathbb{E}[\frac{1}{2}(W_t^1)^2] = \frac{t}{2}$, $\mathbb{E}[\mathbb{X}_{0,t}^{1111}] = \frac{t^2}{8} = \mathbb{E}[(W_t^1)^4]/4!$, and $\mathbb{E}[\mathbb{X}_{0,t}^{12}] = \frac{1}{2}\rho_{12}t$. The piecewise linear signature on a grid of step h has expectation $(\mathbb{E}\exp(h\,e_0 + v_0))^{\otimes T/h}$, by Theorem 20.5 and independence, within $O(h)$ of $\exp(TA)$ (computed exactly in the code).

20.3 Signature-based models

Prices are under a pricing measure $\mathbb{Q}$, with zero interest rate (or, equivalently, for discounted prices), and W is as in Section 20.2 under $\mathbb{Q}$. The primary process $\widehat{X}_t = (t, W_t)$ and its parameters are fixed; only the coefficients ℓ are fitted. Cuchiero et al. (2023) allow general continuous semimartingales, and Cuchiero et al. (2025) polynomial diffusions, such as Ornstein–Uhlenbeck processes; we stay with Brownian motion, where every step can be proved here.

Definition 20.14 (Signature-based models)**.** Let $n \geq 1$ and $\ell = \sum_{|I|\leq n} \ell_I e_I$, with letters $0, \ldots, d$.

(i) The *linear signature model* of Cuchiero et al. (2023), Definition 3.3, is the price $S_t = \langle \ell, \mathbb{X}_{0,t}\rangle = \ell_\emptyset + \sum_{0<|I|\leq n} \ell_I \mathbb{X}_{0,t}^I$.

(ii) The *signature volatility model* of Cuchiero et al. (2025), equations (3.1)–(3.2), is the price $S_t = S_0 \exp\big(\int_0^t \sigma_u\,\mathrm{d}W_u^{\mathrm{B}} - \frac{1}{2}\int_0^t \sigma_u^2\,\mathrm{d}u\big)$, that is $\mathrm{d}S_t = S_t\sigma_t\,\mathrm{d}W_t^{\mathrm{B}}$, with volatility $\sigma_t = \langle \ell, \mathbb{X}_{0,t}\rangle$ and W^{B} a further Brownian motion with $[W^i, W^{\mathrm{B}}]_t = \rho_{i\mathrm{B}}t$; the process (W, W^{B}) is again as in Section 20.2, and $\mathbb{Z}$ denotes the signature of $(t, W_t, W_t^{\mathrm{B}})$.

In (ii), S is a positive local martingale, hence a supermartingale; whether it is a martingale depends on ℓ and the correlations. In (i), the martingale property is a linear constraint on ℓ.

Proposition 20.15 (Martingale linear models)**.** *If $\ell = \ell_\emptyset + \sum_{k=1}^d \sum_{|I|\leq n-1} \ell_I^k \tilde{e}_I^k$, then $S_t = \ell_\emptyset + \sum_{k=1}^d \int_0^t \langle \ell^k, \mathbb{X}_{0,u}\rangle\,\mathrm{d}W_u^k$, with $\ell^k = \sum_I \ell_I^k e_I$, is a square-integrable $\mathbb{Q}$-martingale with $S_0 = \ell_\emptyset$. Conversely, if S is a local martingale, then S admits a representation of this form (Cuchiero et al. (2023), Corollary 3.12). The coefficients ℓ themselves need not have this form when ρ is singular: for $d = 2$, $\rho_{12} = 1$ and $n = 2$, $\ell = e_{10} - e_{20}$ gives $S \equiv 0$.*

Proof. Theorem 20.13(iii) and linearity give the representation. The integrands are continuous, adapted and in $\mathbb{H}^2$ by Theorem 20.13(i), so the integrals are square-integrable martingales (Theorem 7.7). □

Then $\mathbb{Q}$ is a risk-neutral measure for S (Theorem 11.6), and European options priced by $\mathbb{Q}$-expectations admit no static arbitrage among themselves (Theorem 3.6). The price is not kept positive, however: S_T is a polynomial in iterated integrals of Brownian motion, which can be negative. Theorem 20.17 in Section 20.4 has $\mathbb{Q}(S_T \le 0) \approx 4 \times 10^{-4}$ at one year. Black–Scholes itself is a linear signature model of infinite order (Exercise 20.7).

Proposition 20.16 (Prices through expected signatures)**.** *Let $E_t = \exp(tA)$, with A from (20.1).*

(i) In Theorem 20.15, $S_T = \langle \tilde{\ell}, \mathbb{X}_{0,T} \rangle$ with $\tilde{\ell} = \ell_\emptyset + \sum_{k,I} \ell_I^k \tilde{e}_I^k$, and $\mathbb{E}^{\mathbb{Q}}[S_T^m] = \langle \tilde{\ell}^{\sqcup\!\sqcup m}, E_T \rangle$ for every $m \ge 1$. In particular, $\mathbb{E}^{\mathbb{Q}}[S_T^2] = \theta^\top M \theta$ for the vector θ of the coefficients $(\ell_\emptyset, \ell_I^k)$ and the matrix M of the numbers $\langle f \sqcup\!\sqcup g, E_T \rangle$, $f, g \in \{e_\emptyset, \tilde{e}_I^k\}$: a quadratic form in the parameters.

(ii) In Theorem 20.14(ii), with $\tilde{e}_I^{\mathrm{B}} = e_{I\mathrm{B}} - \frac{1}{2}\rho_{i\mathrm{B}} e_{I'0}$ ($I = I'i$) and $\tilde{e}_\emptyset^{\mathrm{B}} = e_{\mathrm{B}}$,

$$\log \frac{S_T}{S_0} = \sum_{|I| \le n} \ell_I \langle \tilde{e}_I^{\mathrm{B}}, \mathbb{Z}_{0,T} \rangle - \frac{1}{2}\, \ell^\top Q^0(T)\, \ell,$$

where $Q^0_{IJ}(T) = \langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, \mathbb{X}_{0,T} \rangle = \int_0^T \mathbb{X}^I_{0,u} \mathbb{X}^J_{0,u}\, \mathrm{d}u$. The realized variance is $\int_0^T \sigma_u^2\, \mathrm{d}u = \ell^\top Q^0(T) \ell$, and its expectation, the fair strike of a variance swap on it, is $\frac{1}{T} \ell^\top \mathbb{E}^{\mathbb{Q}}[Q^0(T)] \ell$, with $\mathbb{E}^{\mathbb{Q}}[Q^0_{IJ}(T)] = \langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, E_T \rangle$.

(iii) For $\Delta > 0$, $\mathbb{E}^{\mathbb{Q}}[\int_T^{T+\Delta} \sigma_u^2\, \mathrm{d}u \mid \mathcal{F}_T] = \ell^\top Q(T) \ell$, with the $\mathcal{F}_T$-measurable, positive semidefinite matrix $Q_{IJ}(T) = \langle (e_I \sqcup\!\sqcup e_J) \otimes e_0,\ \mathbb{X}_{0,T} \otimes (E_\Delta - 1) \rangle$.

Proof. (i) The representation is Theorem 20.15. By Theorem 20.13(ii), $S_T^m = \langle \tilde{\ell}^{\sqcup\!\sqcup m}, \mathbb{X}_{0,T} \rangle$; take expectations with Theorem 20.13(iv) at $s = 0$, the coordinates being integrable by (i) of the same proposition. (ii) $\log(S_T/S_0) = \int_0^T \sigma\, \mathrm{d}W^{\mathrm{B}} - \frac{1}{2}\int_0^T \sigma^2\, \mathrm{d}u$. Apply Theorem 20.13(iii) to the primary process (W, W^{B}), whose signature restricted to words without the letter B is $\mathbb{X}$: $\int_0^T \mathbb{X}^I_{0,u}\, \mathrm{d}W^{\mathrm{B}}_u = \langle \tilde{e}_I^{\mathrm{B}}, \mathbb{Z}_{0,T} \rangle$. By the shuffle identity, $\sigma_u^2 = \langle \ell \sqcup\!\sqcup \ell, \mathbb{X}_{0,u} \rangle = \sum_{I,J} \ell_I \ell_J \langle e_I \sqcup\!\sqcup e_J, \mathbb{X}_{0,u} \rangle$, and $\int_0^T \langle a, \mathbb{X}_{0,u} \rangle\, \mathrm{d}u = \langle a \otimes e_0, \mathbb{X}_{0,T} \rangle$ by Theorem 20.12. The strike is the expectation of $\frac{1}{T} \ell^\top Q^0(T) \ell$, by Theorem 20.13(iv). (iii) As in (ii), $\int_T^{T+\Delta} \sigma_u^2\, \mathrm{d}u = \langle (\ell \sqcup\!\sqcup \ell) \otimes e_0, \mathbb{X}_{0,T+\Delta} - \mathbb{X}_{0,T} \rangle$, and $\mathbb{E}^{\mathbb{Q}}[\mathbb{X}_{0,T+\Delta} \mid \mathcal{F}_T] = \mathbb{X}_{0,T} \otimes E_\Delta$ by Theorem 20.13(iv). The quadratic form is a conditional expectation of a square, hence nonnegative. □

Here σ is unbounded and S may be a strict local martingale, so Theorem 15.13 does not apply as it stands. But $\sigma \in \mathbb{H}^2$ by Theorem 20.13(i), so $\int \sigma\, \mathrm{d}W^{\mathrm{B}}$ is a martingale and (ii) gives $\mathbb{E}^{\mathbb{Q}}[\int_T^{T+\Delta} \sigma_u^2\, \mathrm{d}u \mid \mathcal{F}_T] = -2\, \mathbb{E}^{\mathbb{Q}}[\log(S_{T+\Delta}/S_T) \mid \mathcal{F}_T]$. We therefore define the model VIX through the log contract, as in Cuchiero et al. (2026), equation (24): $\mathrm{VIX}_T^2 = -\frac{2 \cdot 100^2}{\Delta} \mathbb{E}^{\mathbb{Q}}[\log(S_{T+\Delta}/S_T) \mid \mathcal{F}_T] = \frac{100^2}{\Delta} \ell^\top Q(T) \ell$, with $\Delta = 30$ days: by (iii), a quadratic form in ℓ whose matrix is a linear function of the signature at T, up to level $2n+1$. Its link with the CBOE formula of Theorem 15.14, through the strip of options that replicates the log contract, needs in addition that S be a true martingale. This is Cuchiero et al. (2025), Theorem 5.1, where the primary process is a polynomial diffusion and $\mathbb{X}_{0,T} \otimes E_\Delta$ is replaced by a matrix exponential of the generator of its truncated signature (their Theorem 4.4); (ii) is their Proposition 6.4. No Euler scheme is needed: S_T and VIX_T are explicit functions of ℓ and of the signature at T. More generally,

the price of a *sig-payoff*, a linear functional of the signature of (t, S_t) up to level m, is linear in E_T and a polynomial of degree m in the parameters (Cuchiero et al. (2023), Corollary 3.23 and Remark 3.24).

20.4 Calibration

A call payoff is not a polynomial of S_T, so its price is computed by Monte Carlo; the point of the linearity is that the parameters enter last. The procedure of Cuchiero et al. (2023), Section 4.2, and Cuchiero et al. (2025), Sections 5.3 and 7, is the following.

(1) Fix the primary process, its parameters and the truncation level n.

(2) Simulate N paths once, and store the signature coordinates needed at each maturity T: the features $\langle \tilde{e}_I^k, \mathbb{X}_{0,T}(\omega_j)\rangle$ in model (i); $\mathbb{X}_{0,T}(\omega_j)$ up to level $2n+1$ and the words of $\mathbb{Z}$ in model (ii).

(3) The model price of the call (T, K) is the function $\pi(\ell; T, K) = \frac{1}{N}\sum_j \left(S_T(\ell, \omega_j) - K\right)^+$ of ℓ, with $S_T(\ell, \omega_j) = \ell_\emptyset + \sum_{k,I} \ell_I^k \langle \tilde{e}_I^k, \mathbb{X}_{0,T}(\omega_j)\rangle$ in model (i), or Theorem 20.16(ii) in model (ii). VIX options are priced by the same average of $(\mathrm{VIX}_T(\ell, \omega_j) - K)^+$, with VIX_T from Theorem 20.16(iii): no nested simulation.

(4) Minimize a weighted least-squares loss such as $\sum_i \left((\pi(\ell; T_i, K_i) - \pi_i^{\mathrm{mkt}})/\mathcal{V}_i\right)^2$ with the Black–Scholes vegas $\mathcal{V}_i$, an approximation of the squared implied volatility errors (Theorem 15.4); the second paper penalizes implied volatilities outside the bid–ask spreads, and fits SPX and VIX jointly through $\lambda L_{\mathrm{SPX}} + (1-\lambda) L_{\mathrm{VIX}}$, with BFGS from `scipy`.

(5) Reduce the variance with control variates whose prices are known in closed form, such as sig-payoffs (Theorem 20.16(i)).

The survey of Cuchiero et al. (2026), Section 3.5.3, which we follow for these figures, reports the linear model with two Brownian motions of correlation -0.5, $n = 3$ (13 parameters) and $N = 10^6$, fitted to synthetic Heston surfaces (seven maturities from 30 days to two years, thirteen strikes from 80% to 120%) and to S&P 500 implied volatilities of 17 March 2021. On the market data, one set of parameters misses the steep short-maturity smiles; with a loss close to the maximal error, the largest relative implied-volatility errors beyond 60 days are about 5.6% for $n = 3$ and 2.3% for $n = 4$, and fitting the shortest maturity separately does better; Cuchiero et al. (2023), Section 4.2.4, also let the parameters depend on time. In the SPX–VIX model, the same survey (Section 4.3, summarizing Cuchiero et al. (2025)) reports that the volatility is driven by a three-dimensional Ornstein–Uhlenbeck process with slow and fast mean reversion, $n = 3$ (85 parameters, the words of length at most 3 in 4 letters) and $N = 80{,}000$, on the trading day 02/06/2021 (2 June), with SPX maturities of 14, 44 and 58 days and VIX maturities of 14 and 28 days: the model implied volatilities of both fall within the bid–ask spreads, and the VIX futures are matched to relative errors below 10^{-5}. A two-dimensional Brownian primary process fits poorly at this truncation level (Appendix A of the paper, as cited there). The authors describe the result as a joint fit of SPX and VIX smiles without jumps or rough volatility.

Example 20.17 (Replication: Heston smiles by a linear signature model)**.** We fit model (i) of Theorem 20.15, driven by W^1, with primary process (t, W^1, W^2), correlation -0.5 and

$n = 3$ (the 13 words of length at most 2 in letters $0, 1, 2$), to call prices of the Heston model of Figure 15.3 with $r = 0$ (Theorem 15.11; $v_0 = \theta = 0.04$, $\kappa = 2$, $\xi = 0.5$, $\rho = -0.7$), at maturities $\frac{1}{12}, \frac{1}{4}, \frac{1}{2}, 1$ and strikes $K = S_0 e^{0.2\sqrt{T}x}$, $x \in \{-2, -1.5, \dots, 2\}$: 36 quotes. We use $N = 50{,}000$ antithetic paths on 120 steps per year, the vega-weighted loss with its exact gradient (`scipy.optimize.least_squares`), and price below S_0 through the put and parity, which is the control variate $S_T - S_0$, a sig-payoff of price 0. Simulation and fit take a few seconds. The fit to the calibration sample has a root-mean-square implied-volatility error of 6.1 basis points (19 at most). Repriced on 60,000 independent paths, the error is 19.7 basis points (41.5 at most), against a combined Monte Carlo standard error of 16.2 basis points for the two samples (Figure 20.4). Over the five seeds $7, 9, \dots, 15$ the in-sample error ranges from 2.7 to 6.1 basis points and the out-of-sample error from 11.6 to 19.7; the shift of about +25 basis points of the one-month smile in the figure is sampling noise, since its mean ranges from -2 to $+28$ basis points over these seeds. So the Monte Carlo error accounts for most of the out-of-sample error at these sample sizes; the papers use $N = 10^6$. With $n = 2$ (4 parameters, 30,000 paths for the fit and for the check), the error is 64 to 71 basis points in sample and 64 to 74 out of sample. The grid discretization changes $\mathbb{E}^{\mathbb{Q}}[S_1]$ by 7×10^{-5}, computed exactly from the expected signature of the piecewise linear path.

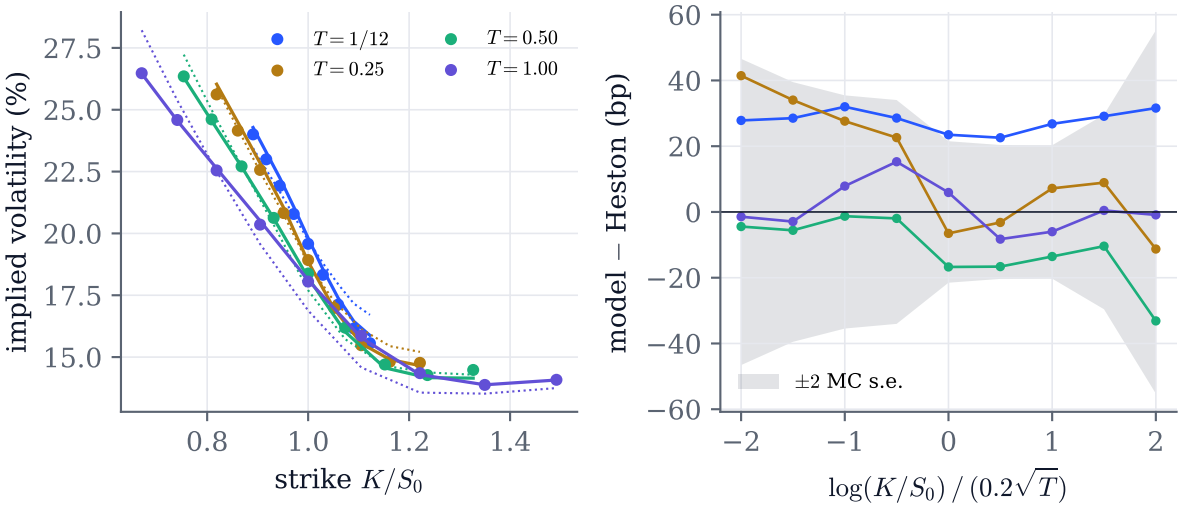

Figure 20.4: Theorem 20.17. Left: Heston implied volatilities (markers) and the calibrated linear signature model with $n = 3$ (lines; 50,000 and 60,000 paths) and $n = 2$ (dotted; 30,000 and 30,000), repriced on independent paths. Right: errors of the $n = 3$ model in basis points against standardized log-strike, with twice the combined Monte Carlo standard error of the calibration and repricing samples (grey).

▶ Interactive demo: calibrate the linear signature model to a Heston or Black–Scholes surface with any truncation level, number of paths and correlation, and compare in- and out-of-sample errors.

The appeal is a separation: the expensive part, simulating the primary process and its signature, is done once, and the optimization over ℓ is then cheap and smooth, while the universality results say that, on compact sets of paths, enough coordinates approximate any continuous functional.

The costs are visible too: the number of parameters grows like $(d+1)^n$, the linear price may become negative, and the fit is only as good as the Monte Carlo sample.

Further reading

Chevyrev and Kormilitzin (2016) is an accessible introduction to signatures, with the proofs of their basic properties and applications to learning; Lyons et al. (2007), Chapter 2, and Friz and Victoir (2010) develop the theory of rough paths. Signature-based models, their pricing formulas and their calibration are surveyed by Cuchiero et al. (2026), in a volume on signature methods in finance.

Bibliography

Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. Affine Volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200, 2019.

Hansjörg Albrecher, Philipp Mayer, Wim Schoutens, and Jurgen Tistaert. The little Heston trap. *Wilmott Magazine*, (1):83–92, 2007.

Elisa Alòs, Jorge A. León, and Josep Vives. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, 11(4):571–589, 2007.

Leif B. G. Andersen and Vladimir V. Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50, 2007.

Louis Bachelier. Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure, 3e série*, 17:21–86, 1900.

Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499, 2013.

Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(1):1550005, 2015.

Guy Barles, Julien Burdeau, Marc Romano, and Nicolas Samsoen. Critical stock price near expiration. *Mathematical Finance*, 5(2):77–95, 1995.

Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

Mikkel Bennedsen, Asger Lunde, and Mikko S. Pakkanen. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965, 2017.

Dimitris Bertsimas, Leonid Kogan, and Andrew W. Lo. When is time continuous? *Journal of Financial Economics*, 55(2):173–204, 2000.

Patrick Billingsley. *Convergence of Probability Measures*. Wiley, New York, 2nd edition, 1999.

Tomas Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, Oxford, 4th edition, 2019.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

Douglas T. Breeden and Robert H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651, 1978.

Menachem Brenner and Marti G. Subrahmanyam. A simple formula to compute the implied standard deviation. *Financial Analysts Journal*, 44(5):80–83, 1988.

Damiano Brigo and Fabio Mercurio. *Interest Rate Models: Theory and Practice*. Springer, Berlin, 2nd edition, 2006.

Mark Broadie, Paul Glasserman, and Steven Kou. A continuity correction for discrete barrier options. *Mathematical Finance*, 7(4):325–349, 1997.

Haydyn Brown, David Hobson, and L. C. G. Rogers. Robust hedging of barrier options. *Mathematical Finance*, 11(3):285–314, 2001.

Timothy C. Brown and M. Gopalan Nair. A simple proof of the multivariate random time change theorem for point processes. *Journal of Applied Probability*, 25(1):210–214, 1988.

Hans Buehler. Expensive martingales. *Quantitative Finance*, 6(3):207–218, 2006.

Hans Buehler, Blanka Horvath, Anastasis Kratsios, Yannick Limmer, and Raeid Saqur. SANOS: Smooth strictly arbitrage-free non-parametric option surfaces. arXiv preprint 2601.11209 [q-fin.CP], version 4, 2026a.

Hans Buehler, Blanka Horvath, Anastasis Kratsios, and Magnus Wiese. DYSANOS: Generative dynamic smooth arbitrage-free non-parametric option surfaces. arXiv preprint 2608.12587 [q-fin.MF], version 4, 2026b.

Robert H. Cameron and William T. Martin. Transformations of Wiener integrals under translations. *Annals of Mathematics*, 45(2):386–396, 1944.

Peter Carr and Dilip Madan. Towards a theory of volatility trading. In Robert Jarrow, editor, *Volatility: New Estimation Techniques for Pricing Derivatives*, pages 417–427. Risk Books, 1998.

Peter Carr and Dilip B. Madan. A note on sufficient conditions for no arbitrage. *Finance Research Letters*, 2(3):125–130, 2005.

Kuo-Tsai Chen. Integration of paths, geometric invariants and a generalized Baker–Hausdorff formula. *Annals of Mathematics*, 65(1):163–178, 1957.

Kuo-Tsai Chen. Integration of paths—a faithful representation of paths by non-commutative formal power series. *Transactions of the American Mathematical Society*, 89(2):395–407, 1958.

Patrick Cheridito. Mixed fractional Brownian motion. *Bernoulli*, 7(6):913–934, 2001.

Patrick Cheridito. Arbitrage in fractional Brownian motion models. *Finance and Stochastics*, 7 (4):533–553, 2003.

Patrick Cheridito, Hideyuki Kawaguchi, and Makoto Maejima. Fractional Ornstein–Uhlenbeck processes. *Electronic Journal of Probability*, 8:Paper no. 3, 1–14, 2003.

Ilya Chevyrev and Andrey Kormilitzin. A primer on the signature method in machine learning. arXiv:1603.03788, 2016. Revised 2025; also Chapter 1 of *Signature Methods in Finance*, Springer Finance, 2026.

Chicago Board Options Exchange. The CBOE volatility index – VIX. White paper, 2009.

Simon Clinet and Nakahiro Yoshida. Statistical inference for ergodic point processes and application to limit order book. *Stochastic Processes and their Applications*, 127(6):1800–1839, 2017.

Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.

Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

Laurent Cousot. Conditions on option prices for absence of arbitrage and exact calibration. *Journal of Banking & Finance*, 31(11):3377–3397, 2007.

John C. Cox, Stephen A. Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979.

John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

Peter F. Craigmile. Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes. *Journal of Time Series Analysis*, 24(5): 505–511, 2003.

John Crank and Phyllis Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Proceedings of the Cambridge Philosophical Society*, 43(1):50–67, 1947.

Colin W. Cryer. The solution of a quadratic programming problem using systematic overrelaxation. *SIAM Journal on Control*, 9(3):385–392, 1971.

Christa Cuchiero, Guido Gazzani, and Sara Svaluto-Ferro. Signature-based models: Theory and calibration. *SIAM Journal on Financial Mathematics*, 14(3):910–957, 2023. doi: 10.1137/22M1512338.

Christa Cuchiero, Guido Gazzani, Janka Möller, and Sara Svaluto-Ferro. Joint calibration to SPX and VIX options with signature-based models. *Mathematical Finance*, 35(1):161–213, 2025. doi: 10.1111/mafi.12442.

Christa Cuchiero, Guido Gazzani, Janka Möller, and Sara Svaluto-Ferro. Signature-based models in finance. In Christian Bayer, Gonçalo dos Reis, Blanka Horvath, and Harald Oberhauser, editors, *Signature Methods in Finance: An Introduction with Computational Applications*, Springer Finance, pages 227–264. Springer, Cham, 2026. doi: 10.1007/978-3-031-97239-3_7.

Robert C. Dalang, Andrew Morton, and Walter Willinger. Equivalent martingale measures and no-arbitrage in stochastic securities market models. *Stochastics and Stochastic Reports*, 29(2): 185–201, 1990.

D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer, New York, 2nd edition, 2003.

R. B. Davies and D. S. Harte. Tests for Hurst effect. *Biometrika*, 74(1):95–101, 1987.

Mark H. A. Davis and David G. Hobson. The range of traded option prices. *Mathematical Finance*, 17(1):1–14, 2007.

Freddy Delbaen and Walter Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(3):463–520, 1994.

Kresimir Demeterfi, Emanuel Derman, Michael Kamal, and Joseph Zou. A guide to volatility and variance swaps. *The Journal of Derivatives*, 6(4):9–32, 1999.

Kai Diethelm, Neville J. Ford, and Alan D. Freed. A predictor-corrector approach for the numerical solution of fractional differential equations. *Nonlinear Dynamics*, 29:3–22, 2002.

Lester E. Dubins and Leonard J. Savage. *How to Gamble If You Must: Inequalities for Stochastic Processes*. McGraw-Hill, New York, 1965.

Darrell Duffie and Peter Glynn. Efficient Monte Carlo simulation of security prices. *The Annals of Applied Probability*, 5(4):897–905, 1995.

Darrell Duffie and Rui Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4): 379–406, 1996.

Bruno Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge, 5th edition, 2019.

Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29(1):3–38, 2019.

Omar El Euch, Masaaki Fukasawa, and Mathieu Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22(2):241–280, 2018.

Omar El Euch, Masaaki Fukasawa, Jim Gatheral, and Mathieu Rosenbaum. Short-term at-the-money asymptotics under stochastic volatility models. *SIAM Journal on Financial Mathematics*, 10(2):491–511, 2019.

Nicole El Karoui, Monique Jeanblanc-Picqué, and Steven E. Shreve. Robustness of the Black and Scholes formula. *Mathematical Finance*, 8(2):93–126, 1998.

William Feller. Two singular diffusion problems. *Annals of Mathematics*, 54(1):173–182, 1951.

Richard P. Feynman. Space-time approach to non-relativistic quantum mechanics. *Reviews of Modern Physics*, 20(2):367–387, 1948.

Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time*. De Gruyter, Berlin, 4th edition, 2016.

Martin Forde, Antoine Jacquier, and Roger Lee. The small-time smile and term structure of implied volatility under the Heston model. *SIAM Journal on Financial Mathematics*, 3(1): 690–708, 2012.

Avner Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliffs, NJ, 1964.

Marco Frittelli. The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance*, 10(1):39–52, 2000.

Peter K. Friz and Nicolas B. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

Masaaki Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17(2):189–198, 2017.

Mark B. Garman and Steven W. Kohlhagen. Foreign currency option values. *Journal of International Money and Finance*, 2(3):231–237, 1983.

Paul Gassiat. On the martingale property in the rough Bergomi model. *Electronic Communications in Probability*, 24(33):1–9, 2019. doi: 10.1214/19-ECP239. Paper no. 33.

Jim Gatheral. *The Volatility Surface: A Practitioner's Guide*. Wiley, Hoboken, NJ, 2006.

Jim Gatheral and Antoine Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.

Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

Stefan Geiss and Emmanuel Gobet. Fractional smoothness and applications in finance. In Giulia Di Nunno and Bernt Øksendal, editors, *Advanced Mathematical Methods for Finance*, pages 313–331. Springer, Berlin, 2011.

Hélyette Geman, Nicole El Karoui, and Jean-Charles Rochet. Changes of numéraire, changes of probability measure and option pricing. *Journal of Applied Probability*, 32(2):443–458, 1995.

Michael B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

Michael B. Giles and Rebecca Carter. Convergence analysis of Crank–Nicolson and Rannacher time-marching. *Journal of Computational Finance*, 9(4):89–112, 2006.

Igor V. Girsanov. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability and its Applications*, 5(3):285–301, 1960.

Paul Glasserman. *Monte Carlo Methods in Financial Engineering*, volume 53 of *Applications of Mathematics*. Springer, New York, 2004.

Emmanuel Gobet. A correction note to "Discrete time hedging errors for options with irregular payoffs". *Finance and Stochastics*, 18(2):483–485, 2014.

Emmanuel Gobet and Emmanuel Temam. Discrete time hedging errors for options with irregular payoffs. *Finance and Stochastics*, 5(3):357–367, 2001.

Anja Göing-Jaeschke and Marc Yor. A survey and some generalizations of Bessel processes. *Bernoulli*, 9(2):313–349, 2003.

M. Barry Goldman, Howard B. Sosin, and Mary Ann Gatto. Path dependent options: "buy at the low, sell at the high". *Journal of Finance*, 34(5):1111–1127, 1979.

Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, 4th edition, 2020.

István Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields*, 71(4):501–516, 1986.

Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward. Managing smile risk. *Wilmott Magazine*, pages 84–108, September 2002.

Ben Hambly and Terry Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of Mathematics*, 171(1):109–167, 2010.

Stephen J. Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B*, 86:442, 2013.

J. Michael Harrison and David M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3):381–408, 1979.

J. Michael Harrison and Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260, 1981.

J. Michael Harrison and Stanley R. Pliska. A stochastic calculus model of continuous trading: Complete markets. *Stochastic Processes and their Applications*, 15(3):313–316, 1983.

Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

Alan G. Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60(1): 77–105, 1992.

Gerd Heber, Asger Lunde, Neil Shephard, and Kevin Sheppard. Oxford-Man institute's realized library, version 0.2. Oxford-Man Institute, University of Oxford, 2009.

Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

Thomas S. Y. Ho and Sang-Bin Lee. Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, 41(5):1011–1029, 1986.

John Hull and Alan White. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987.

John Hull and Alan White. Pricing interest-rate-derivative securities. *Review of Financial Studies*, 3(4):573–592, 1990.

John C. Hull. *Options, Futures, and Other Derivatives*. Pearson, 11th edition, 2021.

Patrick Jaillet, Damien Lamberton, and Bernard Lapeyre. Variational inequalities and the pricing of American options. *Acta Applicandae Mathematicae*, 21(3):263–289, 1990.

Thibault Jaisson and Mathieu Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, 25(2):600–631, 2015.

Thibault Jaisson and Mathieu Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals of Applied Probability*, 26(5):2860–2882, 2016.

Farshid Jamshidian. An exact bond option formula. *The Journal of Finance*, 44(1):205–209, 1989.

Mark Kac. On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society*, 65(1):1–13, 1949.

Raul Kangro and Roy Nicolaides. Far field boundary conditions for Black–Scholes equations. *SIAM Journal on Numerical Analysis*, 38(4):1357–1368, 2000.

Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 2nd edition, 1991.

Hans G. Kellerer. Markov-Komposition und eine Anwendung auf Martingale. *Mathematische Annalen*, 198:99–122, 1972.

A. G. Z. Kemna and A. C. F. Vorst. A pricing method for options based on average asset values. *Journal of Banking & Finance*, 14(1):113–129, 1990.

Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.

Jeremy Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.

Peter D. Lax and Robert D. Richtmyer. Survey of the stability of linear finite difference equations. *Communications on Pure and Applied Mathematics*, 9(2):267–293, 1956.

Jean-François Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Springer, Cham, 2016.

Roger W. Lee. The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, 14(3):469–480, 2004.

Hayne E. Leland. Option pricing and replication with transactions costs. *Journal of Finance*, 40 (5):1283–1301, 1985.

P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

Vadim Linetsky. Spectral expansions for Asian (average price) options. *Operations Research*, 52 (6):856–867, 2004.

Roger Lord, Remmert Koekkoek, and Dick van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2):177–194, 2010.

Terry J. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998.

Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential Equations Driven by Rough Paths: École d'Été de Probabilités de Saint-Flour XXXIV – 2004*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

Benoit B. Mandelbrot and John W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

William Margrabe. The value of an option to exchange one asset for another. *The Journal of Finance*, 33(1):177–186, 1978.

Ryan McCrickerd and Mikko S. Pakkanen. Turbocharging Monte Carlo pricing for the rough Bergomi model. *Quantitative Finance*, 18(11):1877–1886, 2018.

Robert C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

Paul-André Meyer. Démonstration simplifiée d'un théorème de Knight. In *Séminaire de Probabilités V*, volume 191 of *Lecture Notes in Mathematics*, pages 191–195. Springer, Berlin, 1971.

Siddhartha Mishra and Roberto Molinaro. Estimates on the generalization error of physics-informed neural networks for approximating PDEs. *IMA Journal of Numerical Analysis*, 43 (1):1–43, 2023.

Peter Mörters and Yuval Peres. *Brownian Motion*. Cambridge University Press, Cambridge, 2010.

Alexander A. Novikov. On an identity for stochastic integrals. *Theory of Probability and its Applications*, 17(4):717–720, 1972.

Yosihiko Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30:243–261, 1978.

Yosihiko Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin, 6th edition, 2003.

Tohru Ozaki. Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.

David M. Pooley, Kenneth R. Vetzal, and Peter A. Forsyth. Convergence remedies for non-smooth payoffs in option pricing. *Journal of Computational Finance*, 6(4):25–40, 2003.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Rolf Rannacher. Finite element solution of diffusion problems with irregular data. *Numerische Mathematik*, 43(2):309–327, 1984.

Rimhak Ree. Lie elements and an algebra associated with shuffles. *Annals of Mathematics*, 68 (2):210–220, 1958.

Eric Renault and Nizar Touzi. Option hedging and implied volatilities in a stochastic volatility model. *Mathematical Finance*, 6(3):279–302, 1996.

Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*. Springer, Berlin, 3rd edition, 1999.

L. C. G. Rogers. Arbitrage with fractional Brownian motion. *Mathematical Finance*, 7(1):95–105, 1997.

Marc Romano and Nizar Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4):399–412, 1997.

Michael Roper. Arbitrage free implied volatility surfaces. Preprint, School of Mathematics and Statistics, The University of Sydney, 2010.

Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, 3rd edition, 1976.

Steven E. Shreve. *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*. Springer, New York, 2004a.

Steven E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, New York, 2004b.

Justin Sirignano and Konstantinos Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364, 2018.

J. Michael Steele. *Stochastic Calculus and Financial Applications*. Springer, New York, 2001.

Hans R. Stoll. The relationship between put and call option prices. *The Journal of Finance*, 24 (5):801–824, 1969.

Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

John C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. SIAM, Philadelphia, 2nd edition, 2004.

Michael R. Tehranchi. Uniform bounds for Black–Scholes implied volatility. *SIAM Journal on Financial Mathematics*, 7(1):893–916, 2016.

Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.

Jan Večeř. A new PDE approach for pricing arithmetic average Asian options. *Journal of Computational Finance*, 4(4):105–113, 2001.

Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43 (5):A3055–A3081, 2021.

David Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.

Paul Wilmott, Sam Howison, and Jeff Dewynne. *The Mathematics of Financial Derivatives: A Student Introduction*. Cambridge University Press, Cambridge, 1995.

R. Zhang. *Couverture approchée des options européennes*. PhD thesis, École Nationale des Ponts et Chaussées, 1999.