Stochastic Calculus

the calculus of random paths, and why finance is built on it

Martin Burke (with Claude)

Draft

Abstract

Ordinary calculus is the mathematics of smooth curves: things with a slope, a tangent, a well-behaved rate of change. The price of a stock, the level of an interest rate, the value of a hedge—none of these are smooth. They jitter endlessly, and the jitter does not vanish when you look closely; it is the whole point. *Stochastic calculus* is the calculus built for such objects, and it rests on one surprising fact: random noise accumulated over time has a *size* that ordinary calculus throws away. Written compactly that fact is $(\mathrm{d}W)^2 = \mathrm{d}t$, and almost everything in this note is that one idea wearing a different hat. We start from Brownian motion and build the two tools a working quant uses every day—the Itô integral and Itô's lemma—then solve the two canonical models, geometric Brownian motion and the Ornstein–Uhlenbeck process. From there we assemble the engine that turns a model into a price—martingales, Girsanov's change of measure, the Feynman–Kac bridge between expectations and partial differential equations, and risk-neutral pricing—and put it to work pricing options in closed form: Black–Scholes, the exchange option, and the early-exercise (American) option. We then take the *Markovian* property seriously—the difference between a model whose future needs only its present (and so collapses to a fast, low-dimensional PDE or tree) and one that drags its whole history behind it (path dependence, rough volatility). Finally we put it all to work on the term structure of interest rates, building the Heath–Jarrow–Morton framework from scratch, deriving its no-arbitrage drift, and showing exactly when it collapses back to a one-factor Hull–White model. The aim throughout is intuition first, a figure for every idea, and just enough mathematics to make the intuition rigorous—and to leave you able to read the other notes in this series unaided.

Contents

1 Why ordinary calculus breaks

Take a chart of a stock price, an exchange rate, a short-term interest rate. Now imagine asking the question ordinary calculus is built to answer: *what is its slope right now?* You cannot answer it. Zoom in on any point and the curve does not straighten into a tangent line—it keeps wiggling, at every scale, all the way down. There is no instantaneous rate of change because there is no instant at which the path is smooth.

This is not a defect of the data or a failure of resolution. It is the defining feature of the mathematical object we use to model such paths—*Brownian motion*—and Figure 1 shows it directly. Each panel zooms into the middle of the previous one. A smooth function would eventually look like a straight line; a Brownian path never does. It is *statistically self-similar*: the close-up is a smaller, rescaled copy of the whole, with the same ragged character forever.

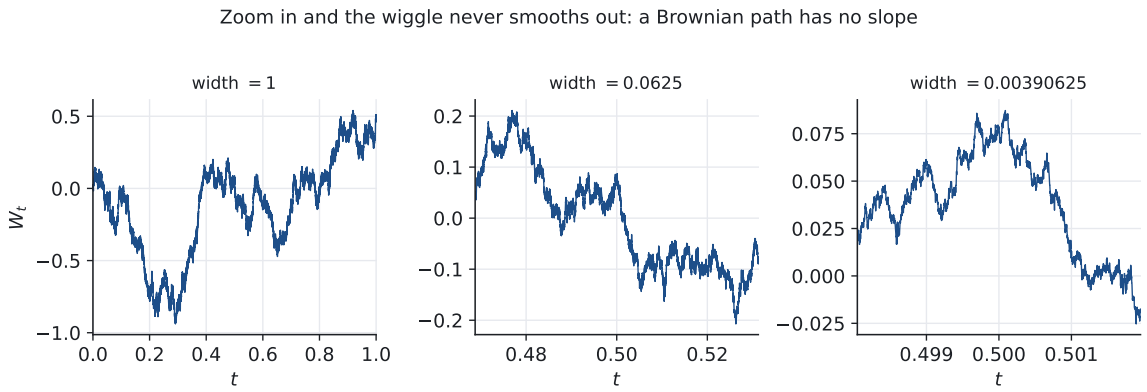

Figure 1: Zoom into a Brownian path and it never smooths out. Each panel is the centre of the one before it, magnified roughly sixteen-fold in time. A differentiable curve would flatten toward its tangent; this one reproduces its own raggedness at every scale. There is no slope to speak of—and so no ordinary derivative, and no ordinary calculus.

So if we cannot talk about $\mathrm{d}W/\mathrm{d}t$, what *can* we talk about? The answer—the seed of everything that follows—is that while the *slope* of a random path is meaningless, its *accumulated squared movement* is not just meaningful but perfectly predictable. Chop the time axis into n little steps and add up the *squares* of the increments. For a smooth function that sum would shrink to zero as the steps get finer (each increment is roughly slope times Δt, so its square is order Δt^2, and there are only $n \sim 1/\Delta t$ of them). For a Brownian path it does not shrink: it converges to the elapsed time itself. Over an interval of length t,

$$\sum_i \left(W_{t_{i+1}} - W_{t_i}\right)^2 \;\longrightarrow\; t \qquad \text{as the mesh} \to 0. \tag{1}$$

This limit is called the *quadratic variation* of W, and the shorthand the whole subject runs on is

$$\boxed{(\mathrm{d}W)^2 = \mathrm{d}t} \tag{2}$$

read as: "over an infinitesimal step, the squared Brownian increment is not negligible—it equals the length of the step." Ordinary calculus is built on throwing away anything of order $\mathrm{d}t^2$ and smaller, including $(\mathrm{d}x)^2$. Stochastic calculus keeps exactly one such term, because for a random path it is not small. That single retained term is the source of every result in this note: the correction in Itô's lemma, the volatility drag in a stock's growth rate, the second-derivative term in every pricing PDE, the no-arbitrage drift in Heath–Jarrow–Morton. Keep (2) in mind and the rest is bookkeeping.

What this note is, and is not. It is a working introduction for the technically educated non-specialist—someone comfortable with calculus, probability and linear algebra who wants to *use* these tools and read the literature that depends on them, including the other notes in this series and the interest-rate models at the end. It is not a course in measure theory. We will name the places where rigour lives—σ-algebras, filtrations, measurable functions, L^2 limits—and then, mostly, get on with the intuition and the calculations. Where a proof would illuminate, we give the idea; where it would only reassure, we cite a textbook and move on.

2 Brownian motion: the raw material

Everything is built from one process. *Brownian motion* (or the *Wiener process*) W_t is the formal limit of a random walk in which the steps become infinitely small and infinitely frequent. It is pinned down by three properties, and it is worth knowing them not as axioms to memorise but as the three things you are allowed to use in a calculation.

1. **It starts at zero and is continuous.** $W_0 = 0$, and the path $t \mapsto W_t$ has no jumps—you can draw it without lifting the pen (even though, as we saw, you can never draw it *smoothly*).

2. **Its increments are independent.** What the path does over $[t, t+s]$ is independent of everything it did up to time t. The process has no memory: the future depends on the present level only, not on how it got there. This is the *Markov* property, and §14 is about how much it buys you.

3. **Its increments are Gaussian, with variance equal to elapsed time.** For $s > 0$, $W_{t+s} - W_t \sim \mathcal{N}(0,\, s)$. The mean drift is zero; the spread grows like $\sqrt{s}$.

Where it comes from: a coin-flip walk, rescaled. The three properties are not arbitrary; they are exactly what you are forced into by shrinking a random walk. Let a walker take a step of size $\pm h$ every Δt seconds, the sign decided by an independent fair coin. After a time $t = n\,\Delta t$ its position is the sum of n independent $\pm h$ steps, with mean zero (the coin is fair) and variance nh^2 (variances of independent things add). Now ask: how must the step size h scale with Δt as we make the steps small, if the walk is to converge to anything non-trivial? If we take the obvious $h = \Delta t$, the variance $nh^2 = t\,\Delta t \to 0$ and the walk freezes. If we take $h = 1$, the variance $n = t/\Delta t \to \infty$ and it explodes. The *only* choice that keeps the variance finite and fixed is

$$h = \sqrt{\Delta t}, \qquad \text{giving} \qquad \text{Var} = nh^2 = n\,\Delta t = t \quad \text{for every } n. \tag{3}$$

With that scaling the variance is exactly t at all resolutions, the central limit theorem makes the position $\mathcal{N}(0, t)$, and disjoint stretches of the walk stay independent—the three properties, derived rather than declared. Letting $\Delta t \to 0$ makes the walk converge to Brownian motion (this convergence is *Donsker's theorem*). And the spine of the whole note is already visible in (3): a step has size $\sqrt{\Delta t}$, so its *square* has size Δt. That is $(\mathrm{d}W)^2 = \mathrm{d}t$ in embryo—forced on us by nothing more than "keep the variance finite."

That same square-root scaling is the engine of (2) in the continuous limit. An increment over a step of length $\mathrm{d}t$ is $\mathcal{N}(0, \mathrm{d}t)$: its typical size is $\sqrt{\mathrm{d}t}$, so its square is typically $\mathrm{d}t$. The increment itself is much larger than the step ($\sqrt{\mathrm{d}t} \gg \mathrm{d}t$ for small $\mathrm{d}t$), which is exactly why the path is so rough; but its *square* is of the same order as the step, which is exactly why squared increments accumulate to something finite. Figure 2 shows the consequence in the large: a fan of paths spreading as $\pm\sqrt{t}$ about a flat mean of zero.

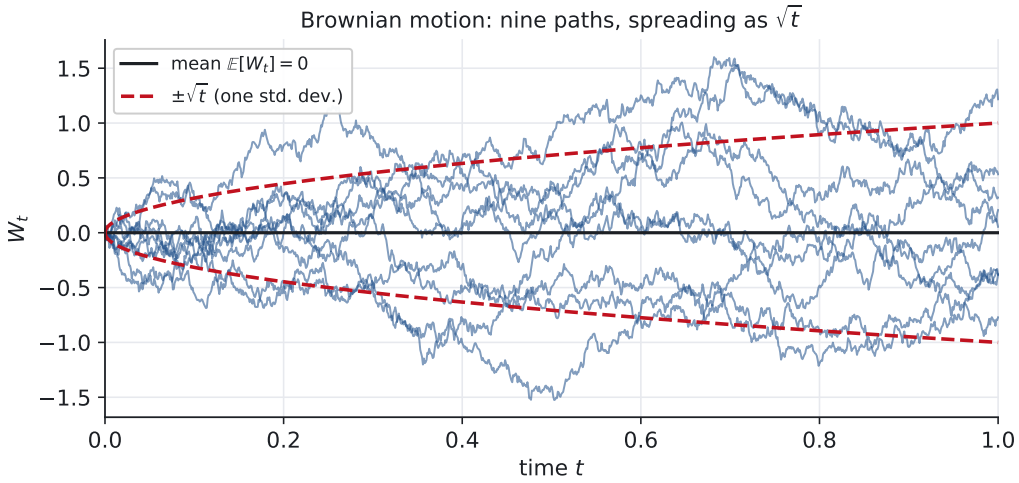

Figure 2: Nine independent Brownian paths. The mean is flat at zero (the increments are unbiased); the cloud widens as $\sqrt{t}$ because variance accumulates linearly in time. The individual paths are continuous but visibly non-smooth—the roughness of Figure 1 seen from further away.

Quadratic variation versus total variation. The claim in (1)—that squared increments sum to t—deserves a picture, because it sits right at the boundary between what survives in the limit and what does not. Figure 3 computes, for a single fixed path and ever finer partitions, two sums: the sum of *squared* increments (left) and the sum of *absolute* increments (right). The squared sum settles down onto the elapsed time $T = 1$. The absolute sum—the *total variation*, the ordinary "arc length" of the wiggle—grows without bound, like $\sqrt{n}$ in the number of steps. The two behaviours come straight from the $\sqrt{\Delta t}$ scaling (3). A single absolute increment is typically of size $\sqrt{\Delta t} = \sqrt{T/n}$, and there are n of them, so the absolute sum is about $n\sqrt{T/n} = \sqrt{nT} \to \infty$. A single squared increment is typically of size $\Delta t = T/n$, and n of those sum to T *regardless of* n—the factors of n cancel. Squaring is exactly the operation that turns the diverging arc length into a finite, stable quantity. The path is so rough that its total length is infinite, yet its total *squared* length is exactly the clock time.

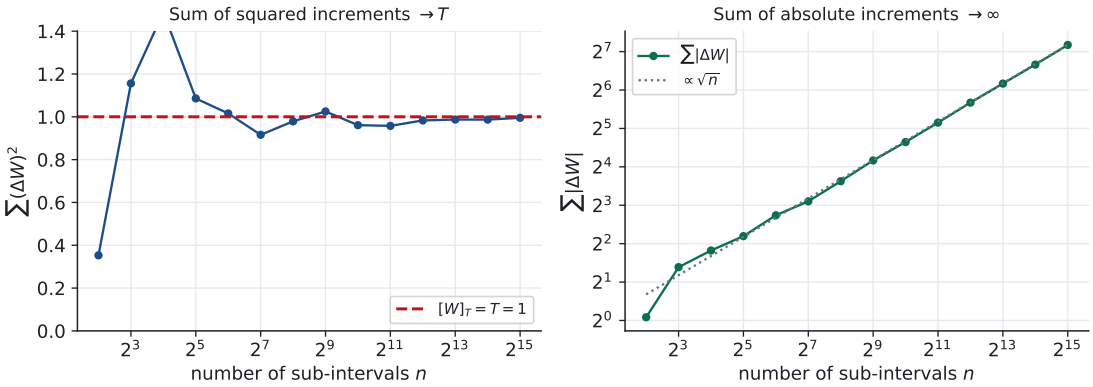

Figure 3: Two sums over a single Brownian path on $[0, 1]$, as the partition is refined. **Left:** the sum of squared increments converges to the elapsed time, $[W]_T = T = 1$. **Right:** the sum of absolute increments diverges like $\sqrt{n}$—infinite total variation. This is the precise sense in which a Brownian path has "infinite length but finite squared length," and it is why the $(\mathrm{d}W)^2$ term cannot be discarded.

One more construction: conditioning on the endpoint. A last fact about Brownian motion itself, useful later when we simulate. Often you know not just where a path starts but where it *ends*—you have observed today and you want to fill in, or simulate, the values in

between. A Brownian motion *conditioned* to pass through a fixed endpoint is a *Brownian bridge*. Pinning $X_0 = a$ and $X_T = b$, the bridge is

$$X_t = a + (b-a)\frac{t}{T} + \Big(W_t - \frac{t}{T}\,W_T\Big), \tag{4}$$

a straight line from a to b with the endpoint-anticipating part of the Brownian motion subtracted off. The mean is that line (the bracket has mean zero). The variance is a short covariance calculation worth doing, because it shows the "tethering" explicitly. Using $\mathrm{Var}(W_t) = t$ and $\mathrm{Cov}(W_t, W_T) = t$ for $t \le T$ (write $W_T = W_t + (W_T - W_t)$; the second piece is an independent future increment, so it adds nothing to the covariance),

$$\mathrm{Var}(X_t) = \mathrm{Var}\Big(W_t - \tfrac{t}{T}W_T\Big) = t - 2\tfrac{t}{T}\,t + \big(\tfrac{t}{T}\big)^2 T = t - \tfrac{t^2}{T} = \frac{t(T-t)}{T}, \tag{5}$$

which is zero at both ends $t = 0, T$ (the path is pinned) and largest at the midpoint $t = T/2$ (Figure 4). Knowing the endpoint has squeezed the free-diffusion variance t down by the factor $(T-t)/T$. Equivalently the bridge solves the mean-reverting SDE $\mathrm{d}X_t = \frac{b - X_t}{T-t}\,\mathrm{d}t + \mathrm{d}W_t$, whose pull toward the target b strengthens without bound as $t \to T$, forcing the path home. Bridges are the workhorse of path simulation: they let you refine a simulated trajectory between coarse dates without disturbing the values already drawn—used for variance reduction, and for checking whether a path *crossed a barrier* between monitoring dates in barrier-option Monte Carlo.

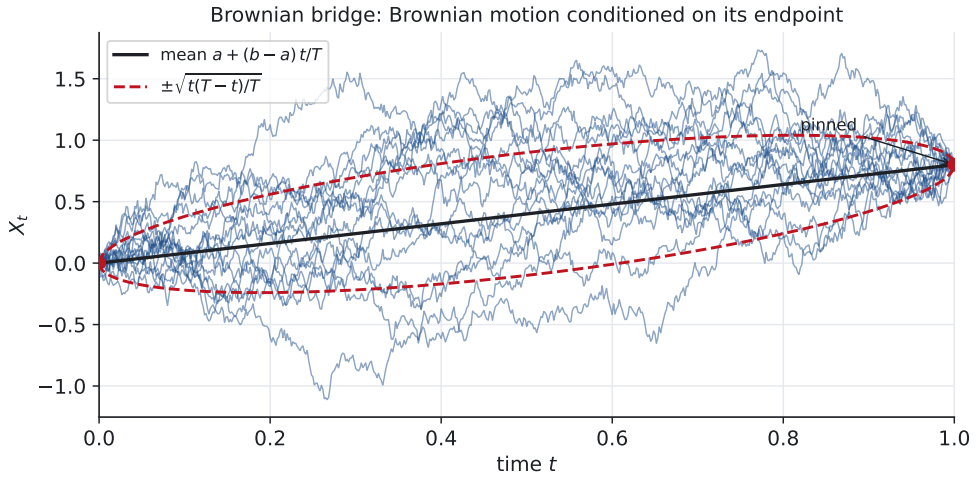

Figure 4: The Brownian bridge (4), sixteen paths pinned at both ends. The mean is the straight line $a + (b-a)t/T$; the spread is $\sqrt{t(T-t)/T}$, vanishing at the pins and maximal at the middle. Conditioning on the endpoint turns free diffusion into a tethered path—the natural object whenever the future is known and the in-between must be filled.

The quadratic-variation/total-variation distinction is the whole reason a *new* integral is needed. The classical way to integrate against a function—Riemann–Stieltjes, $\int g\,\mathrm{d}h$—requires the integrator h to have *finite* total variation, so that the sums converge no matter where inside each sub-interval you sample the integrand. Brownian motion fails that requirement spectacularly. We will have to build the integral $\int g\,\mathrm{d}W$ from scratch, and the place we choose to sample the integrand will turn out to matter enormously.

Notation we will keep. W_t always denotes (standard, one-dimensional) Brownian motion. We write $\mathrm{d}W$ for an infinitesimal increment, $\mathbb{E}[\cdot]$ for expectation, $\mathcal{N}(m, v)$ for a normal with mean m and variance v. Later we use $\mathbb{P}$ for the real-world probability measure and $\mathbb{Q}$ for the risk-neutral one. When several Brownian motions appear we index them $W^1, W^2, \ldots$ and let them be correlated.

Part I — The calculus of random paths

3 The Itô integral

We want to make sense of

$$\int_0^T g_t \, \mathrm{d}W_t, \tag{6}$$

where the integrand g_t may itself be random—typically it is some function of the path so far, like a position size that depends on the current price. In finance this integral is the *gain from a trading strategy*: hold g_t units of an asset whose increments are $\mathrm{d}W_t$, and (6) is your accumulated profit and loss. We need it to make sense, and we need to know its statistics.

Why the sampling point matters. Approximate (6) the obvious way, as a limit of sums $\sum_i g_{\tau_i}\,(W_{t_{i+1}} - W_{t_i})$, where τ_i is some point inside $[t_i, t_{i+1}]$ at which we evaluate the integrand. For an ordinary (finite-variation) integrator it makes no difference where τ_i sits; in the limit you get the same answer. For Brownian motion it makes *all* the difference, because the integrand and the increment can be correlated, and the increment is large ($\sim\sqrt{\Delta t}$). Sample g at the *left* endpoint $\tau_i = t_i$—before the increment happens—and the integrand is independent of the increment it multiplies. Sample it in the middle and it is not.

Itô's choice is the left endpoint, and it is the right choice for finance for a concrete reason: *you must choose your position before you see the next move, not during it.* A trading rule that peeked at $\mathrm{d}W_t$ while choosing g_t would be clairvoyant. The left-endpoint convention encodes "non-anticipating"—g_t may use all information up to time t and not one instant more—directly into the definition of the integral.

What the choice buys: a martingale and a clean variance. Because each left-endpoint term $g_{t_i}(W_{t_{i+1}} - W_{t_i})$ multiplies an independent, mean-zero increment, every term has mean zero, and so

$$\mathbb{E}\left[\int_0^T g_t \, \mathrm{d}W_t\right] = 0. \tag{7}$$

More than that: the running integral $M_t = \int_0^t g_s \, \mathrm{d}W_s$ is a *martingale*—its expected future value, given everything so far, is just its current value (§8 makes this precise). A trading gain driven by fair coin-flips is itself a fair game. And its variance is given by a strikingly clean formula, the *Itô isometry*:

$$\mathbb{E}\left[\left(\int_0^T g_t \, \mathrm{d}W_t\right)^2\right] = \int_0^T \mathbb{E}\big[g_t^2\big] \, \mathrm{d}t. \tag{8}$$

The squared random integral on the left becomes an ordinary integral of the *expected squared integrand* on the right. The proof is one line with (2): square the sum, and the cross terms vanish because distinct increments are independent and mean-zero, leaving $\sum_i \mathbb{E}[g_{t_i}^2]\,\mathbb{E}[(\Delta W_i)^2] = \sum_i \mathbb{E}[g_{t_i}^2]\,\Delta t$, which is the right-hand integral. The isometry is the workhorse behind every variance, every hedge-error calculation, every L^2 convergence argument in the subject; we re-derive it in Appendix B.

The canonical example: $\int_0^T W\,\mathrm{d}W$. Here is the calculation that separates Itô calculus from ordinary calculus, and it is worth doing in full because it shows the quadratic-variation term entering by hand. If W were a smooth variable x, we would write $\int_0^T x\,\mathrm{d}x = \frac{1}{2}x^2\big|_0^T = \frac{1}{2}W_T^2$. Let us instead form the left-endpoint sum carefully and see what survives. The trick is the elementary algebraic identity (just expand $(W_{t_{i+1}})^2 = (W_{t_i} + \Delta W_i)^2$ and solve for the cross term)

$$W_{t_i}\,\Delta W_i = \tfrac{1}{2}\Big(W_{t_{i+1}}^2 - W_{t_i}^2\Big) - \tfrac{1}{2}(\Delta W_i)^2, \qquad \Delta W_i = W_{t_{i+1}} - W_{t_i}. \tag{9}$$

Sum this over the partition. The first bracket *telescopes*—consecutive terms cancel, leaving only the endpoints $\frac{1}{2}(W_T^2 - W_0^2) = \frac{1}{2}W_T^2$. The second piece is exactly the quadratic-variation sum, which we know converges to T. Hence

$$\int_0^T W_t\,\mathrm{d}W_t = \lim \sum_i W_{t_i}\,\Delta W_i = \tfrac{1}{2}W_T^2 - \tfrac{1}{2}\underbrace{\sum_i (\Delta W_i)^2}_{\to T} = \tfrac{1}{2}W_T^2 - \tfrac{1}{2}T. \tag{10}$$

The *extra* $-\frac{1}{2}T$ is the quadratic variation (2) surfacing, and it entered through the one term ordinary calculus discards. The choice of sampling point is what controls it: had we evaluated the integrand at the *midpoint* $\frac{1}{2}(W_{t_i} + W_{t_{i+1}})$ instead of the left end, the identity (9) would have carried no $(\Delta W_i)^2$ correction, the quadratic-variation term would never appear, and the sums would converge to the naive $\frac{1}{2}W_T^2$. That midpoint rule is the *Stratonovich* integral; Figure 5 shows the two conventions converging to answers that differ by exactly the accumulated quadratic variation $\frac{1}{2}T$.

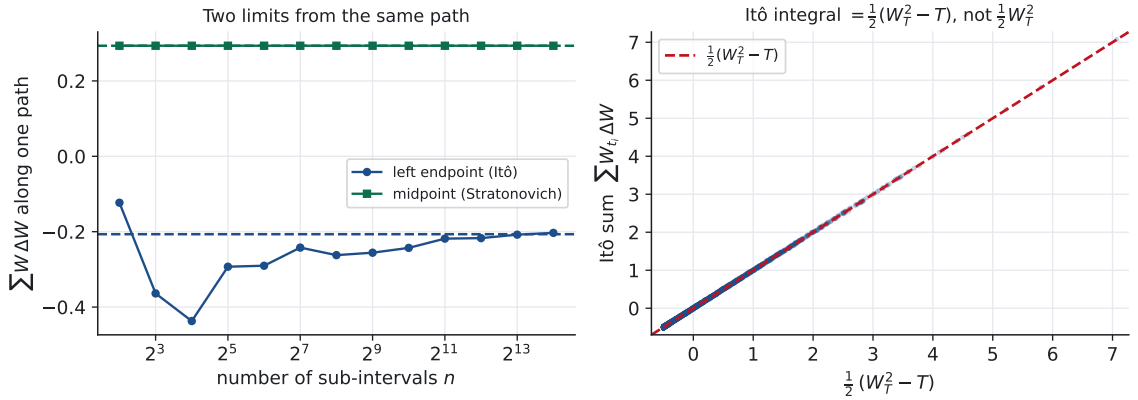

Figure 5: The integral $\int_0^T W\,\mathrm{d}W$ depends on where you sample the integrand. **Left:** along one path, the left-endpoint (Itô) sums converge to $\frac{1}{2}(W_T^2 - T)$, while the midpoint (Stratonovich) sums converge to $\frac{1}{2}W_T^2$; they differ by the quadratic variation $\frac{1}{2}T$. **Right:** across 4000 paths the Itô sum lands exactly on $\frac{1}{2}(W_T^2 - T)$—the diagonal—confirming (10). Ordinary calculus would have predicted the wrong number.

The midpoint convention is called the *Stratonovich* integral, written $\int g \circ \mathrm{d}W$. It obeys the ordinary chain rule, which makes it convenient in physics, but it sacrifices the two properties finance lives on: the Itô integral is a martingale and is non-anticipating, the Stratonovich integral is neither. From here on, *integration means Itô integration*; we mention Stratonovich only to know what we have chosen against.

The integral is a fair game. Figure 6 makes the martingale property of (7) tangible, using the integral we just computed: $M_t = \int_0^t W\,\mathrm{d}W = \frac{1}{2}(W_t^2 - t)$. Sixty paths scatter above and below zero with no drift in their mean, while their spread grows—the variance is $\int_0^t \mathbb{E}[W_s^2]\,\mathrm{d}s = \int_0^t s\,\mathrm{d}s = t^2/2$, exactly the isometry (8) with $g_s = W_s$. A martingale is "all risk, no edge": it goes nowhere on average, but it goes there with ever-increasing uncertainty.

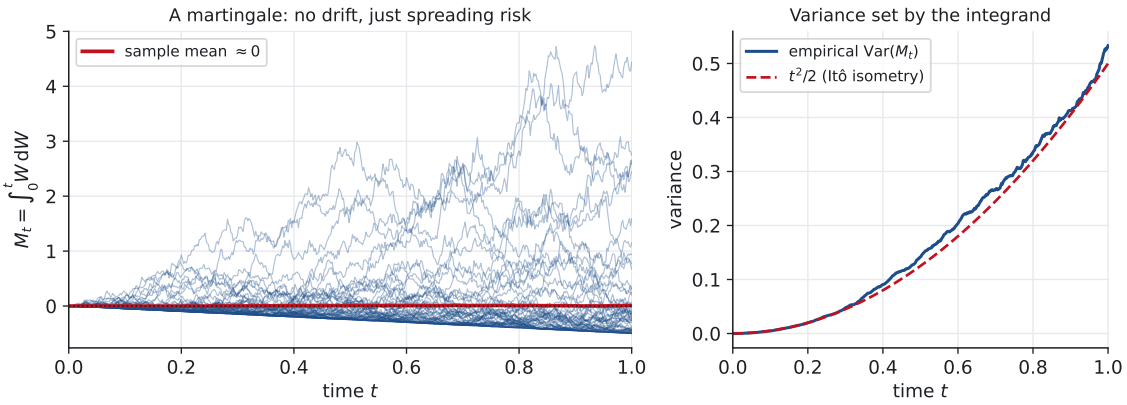

Figure 6: The Itô integral $M_t = \int_0^t W\,\mathrm{d}W = \frac{1}{2}(W_t^2 - t)$ as a martingale. **Left:** sixty paths and their sample mean, which stays at zero—no drift. **Right:** the variance grows as $t^2/2$, matching the Itô isometry exactly. Mean-zero but spreading: a fair game played with rising stakes.

4 Itô's lemma: the chain rule with a tax

We can now state the single most-used result in the subject. It answers: if X_t is driven by Brownian motion, and I look at some function $f(X_t)$ of it, how does f move? In ordinary calculus the chain rule gives $\mathrm{d}f = f'(x)\,\mathrm{d}x$. For a random path there is a second term, and it comes—inevitably by now—from $(\mathrm{d}W)^2 = \mathrm{d}t$.

Derivation in one Taylor expansion. Suppose X follows the stochastic differential equation (SDE)

$$\mathrm{d}X_t = \mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t, \tag{11}$$

read as "over $\mathrm{d}t$, X moves by a deterministic drift $\mu_t\,\mathrm{d}t$ plus a random kick $\sigma_t\,\mathrm{d}W_t$." Expand $f(X_{t+\mathrm{d}t})$ to second order—and here is the one departure from ordinary calculus, we keep the second-order term:

$$\mathrm{d}f \approx f'(X_t)\,\mathrm{d}X_t + \tfrac{1}{2}f''(X_t)\,(\mathrm{d}X_t)^2. \tag{12}$$

Now substitute (11) and square. Here the $\sqrt{\mathrm{d}t}$ scaling of §2 does all the work: a Brownian increment is of size $\mathrm{d}W \sim \sqrt{\mathrm{d}t}$, while a time increment is genuinely of size $\mathrm{d}t$. So the three possible products sort themselves cleanly by *order*:

$$\underbrace{\mathrm{d}t\cdot\mathrm{d}t}_{\sim\mathrm{d}t^2} = 0, \qquad \underbrace{\mathrm{d}t\cdot\mathrm{d}W}_{\sim\mathrm{d}t^{3/2}} = 0, \qquad \underbrace{\mathrm{d}W\cdot\mathrm{d}W}_{\sim\mathrm{d}t} = \mathrm{d}t. \tag{13}$$

Reading these in plain terms: $\mathrm{d}t\cdot\mathrm{d}t$ is of order $\mathrm{d}t^2$ and $\mathrm{d}t\cdot\mathrm{d}W$ is of order $\mathrm{d}t^{3/2}$—both vanishingly small *relative to* $\mathrm{d}t$, so they are discarded just as ordinary calculus discards $(\mathrm{d}x)^2$. Only $\mathrm{d}W\cdot\mathrm{d}W$ is of order $\mathrm{d}t$ itself, and that is the one term that survives, taking the value $\mathrm{d}t$ by (2). The whole departure from ordinary calculus is that the noise has shrunk only half as fast as the clock ($\sqrt{\mathrm{d}t}$, not $\mathrm{d}t$), so its square keeps pace with $\mathrm{d}t$ instead of falling below it. Applying the table, $(\mathrm{d}X_t)^2 = \sigma_t^2\,(\mathrm{d}W_t)^2 = \sigma_t^2\,\mathrm{d}t$ (the $\mu\,\mathrm{d}t$ part of $\mathrm{d}X$ contributes only the discarded orders once squared), and (12) becomes **Itô's lemma**:

$$\mathrm{d}f(X_t) = \underbrace{\Big(f'(X_t)\,\mu_t + \tfrac{1}{2}f''(X_t)\,\sigma_t^2\Big)\mathrm{d}t}_{\text{drift, with the Itô correction}} + \underbrace{f'(X_t)\,\sigma_t\,\mathrm{d}W_t}_{\text{diffusion}}. \tag{14}$$

The diffusion term is what you would have guessed. The new piece is the $\frac{1}{2}f''\sigma^2\,\mathrm{d}t$ inside the drift—the *Itô correction*. It is a tax on *convexity*: whenever f is curved ($f'' \neq 0$), the

symmetric jitter of X does not average out, because f converts the up-moves and down-moves asymmetrically. A convex f ($f'' > 0$) gains from volatility; a concave f loses. This one term is the mathematical root of option convexity, the volatility drag on compounded returns, and the second-derivative term in every pricing PDE.

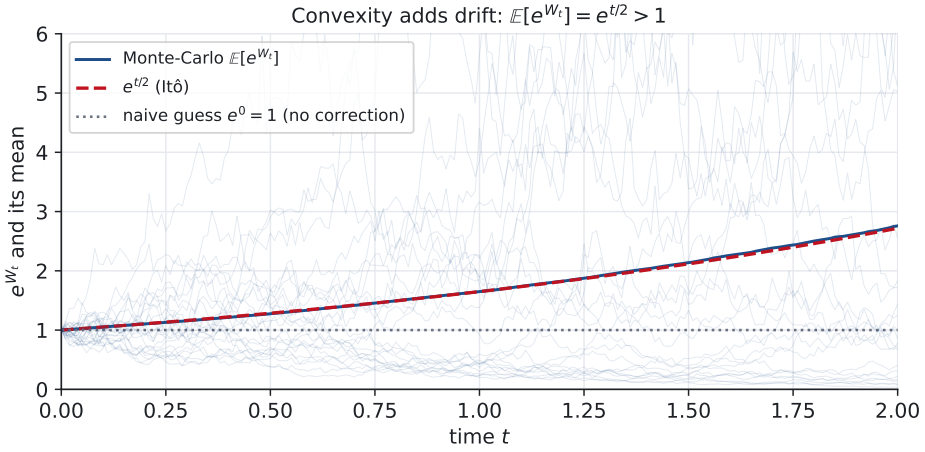

Figure 7: The Itô correction made concrete with $f(W) = e^W$. Naively one might expect $\mathbb{E}[e^{W_t}] = e^{\mathbb{E}[W_t]} = e^0 = 1$, since W_t has mean zero. But e^x is convex, and Itô's lemma predicts a drift of $\frac{1}{2}f'' = \frac{1}{2}e^W$, giving $\mathbb{E}[e^{W_t}] = e^{t/2}$. The Monte-Carlo mean (blue) tracks $e^{t/2}$ (red), climbing away from the naive guess. Convexity manufactures drift out of pure noise.

Itô's lemma (the one to remember). For $\mathrm{d}X = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W$ and smooth f,

$$\mathrm{d}f = \Big(\mu\, f' + \tfrac{1}{2}\sigma^2 f''\Big)\mathrm{d}t + \sigma\, f'\,\mathrm{d}W.$$

Everything new compared with ordinary calculus is the single term $\frac{1}{2}\sigma^2 f''\,\mathrm{d}t$, and it is there because $(\mathrm{d}W)^2 = \mathrm{d}t$.

The time-dependent and multivariate versions. If $f = f(t, X_t)$ depends on time explicitly, add the ordinary $\partial_t f\,\mathrm{d}t$:

$$\mathrm{d}f(t, X_t) = \Big(\partial_t f + \mu\,\partial_x f + \tfrac{1}{2}\sigma^2\partial_{xx} f\Big)\mathrm{d}t + \sigma\,\partial_x f\,\mathrm{d}W. \tag{15}$$

If X is a vector driven by several correlated Brownian motions with $\mathrm{d}W^i\,\mathrm{d}W^j = \rho_{ij}\,\mathrm{d}t$, the correction becomes a sum over second derivatives, $\frac{1}{2}\sum_{i,j}(\sigma\sigma^\top)_{ij}\,\partial_{x_i x_j} f$; we use this in §7. The shape is always the same: first-order terms as usual, plus a half-times-second-derivative tax wherever two diffusion terms multiply.

5 A toolkit of Itô rules

Before solving anything, it pays to collect the handful of algebraic rules that fall straight out of Itô's lemma. They are the random-calculus analogues of the product, quotient and exponential rules you already know—each with one extra term, always the same $(\mathrm{d}W)^2 = \mathrm{d}t$ correction.

The two-process Itô formula. For a function $f(X, Y)$ of two Itô processes, the lemma gains a cross term alongside the two curvature terms:

$$\mathrm{d}f = f_x\,\mathrm{d}X + f_y\,\mathrm{d}Y + \tfrac{1}{2}f_{xx}\,(\mathrm{d}X)^2 + \tfrac{1}{2}f_{yy}\,(\mathrm{d}Y)^2 + f_{xy}\,\mathrm{d}X\,\mathrm{d}Y, \tag{16}$$

where the products are read off the multiplication table: if $\mathrm{d}X = \mu_X \mathrm{d}t + \sigma_X \mathrm{d}W^X$ and $\mathrm{d}Y = \mu_Y \mathrm{d}t + \sigma_Y \mathrm{d}W^Y$ with $\mathrm{d}W^X \mathrm{d}W^Y = \rho\,\mathrm{d}t$, then $(\mathrm{d}X)^2 = \sigma_X^2 \mathrm{d}t$, $(\mathrm{d}Y)^2 = \sigma_Y^2 \mathrm{d}t$, and the *covariation* $\mathrm{d}X\,\mathrm{d}Y = \rho\,\sigma_X\sigma_Y\,\mathrm{d}t$. Everything below is a special case of (16).

Product rule (integration by parts). Take $f(X,Y) = XY$, so $f_{xy} = 1$ and the curvature terms vanish:

$$\mathrm{d}(X_tY_t) = X_t\,\mathrm{d}Y_t + Y_t\,\mathrm{d}X_t + \mathrm{d}X_t\,\mathrm{d}Y_t. \tag{17}$$

It is the ordinary product rule plus the covariation $\mathrm{d}X\,\mathrm{d}Y = \rho\,\sigma_X\sigma_Y\,\mathrm{d}t$—zero in ordinary calculus, and the channel through which *correlation* enters every two-asset hedge, every quanto and convexity adjustment in the rate notes. Rearranged and integrated, it is the *integration-by-parts* formula,

$$\int_0^T X_t\,\mathrm{d}Y_t = X_TY_T - X_0Y_0 - \int_0^T Y_t\,\mathrm{d}X_t - [X,Y]_T, \tag{18}$$

with $[X,Y]_T = \int_0^T \rho\,\sigma_X\sigma_Y\,\mathrm{d}t$ the accumulated covariation—the term a careless transfer of the classical formula would miss.

Quotient rule. Take $f(X,Y) = X/Y$. Then $f_x = 1/Y$, $f_y = -X/Y^2$, $f_{yy} = 2X/Y^3$, $f_{xy} = -1/Y^2$, and (16) gives, after dividing through by X/Y, a clean "logarithmic" form:

$$\frac{\mathrm{d}(X_t/Y_t)}{X_t/Y_t} = \frac{\mathrm{d}X_t}{X_t} - \frac{\mathrm{d}Y_t}{Y_t} - \frac{\mathrm{d}X_t}{X_t}\frac{\mathrm{d}Y_t}{Y_t} + \left(\frac{\mathrm{d}Y_t}{Y_t}\right)^2. \tag{19}$$

The first two terms are the naive answer; the last two are Itô corrections in the denominator. This is exactly the calculation behind a *convexity adjustment*: the expected value of a ratio is not the ratio of expectations, and (19) is the size of the gap. It is also the engine of every change-of-numéraire formula in §11, where the object that matters is precisely a price *divided by* a numéraire.

The stochastic exponential. What process grows in proportion to a driver X—the analogue of $\mathrm{d}y = y\,\mathrm{d}x$ whose solution is e^x? We want $\mathrm{d}Z_t = Z_t\,\mathrm{d}X_t$. The natural guess $Z = e^X$ is wrong, and Itô's lemma says by exactly how much. Apply (14) to $\log Z$: with $f = \log$, $f' = 1/Z$ and $f'' = -1/Z^2$, so

$$\mathrm{d}(\log Z_t) = \frac{\mathrm{d}Z_t}{Z_t} - \tfrac{1}{2}\frac{(\mathrm{d}Z_t)^2}{Z_t^2} = \mathrm{d}X_t - \tfrac{1}{2}\,\mathrm{d}[X]_t, \tag{20}$$

using $\mathrm{d}Z/Z = \mathrm{d}X$ and hence $(\mathrm{d}Z/Z)^2 = (\mathrm{d}X)^2 = \mathrm{d}[X]_t$, the increment of X's quadratic variation. Integrating (20) and exponentiating gives the *stochastic exponential*, written $\mathcal{E}(X)$:

$$Z_t = Z_0\,\exp\Big(X_t - X_0 - \tfrac{1}{2}[X]_t\Big) \;=\; Z_0\,\mathcal{E}(X)_t. \tag{21}$$

The $-\frac{1}{2}[X]_t$ in the exponent is the Itô correction, and it is doing something important: it *exactly cancels* the convexity drift that an ordinary exponential would manufacture (recall Figure 7, where $\mathbb{E}[e^{W_t}] = e^{t/2}$ drifted upward—here the $-\frac{1}{2}[X]_t$ is precisely the $-\frac{1}{2}t$ that pins the mean back down). That cancellation is the whole point, and it gives the single most useful fact about $\mathcal{E}$: *the stochastic exponential of a martingale is a martingale*, with constant mean Z_0. The reason is visible in the defining equation: $\mathrm{d}Z = Z\,\mathrm{d}X$ has *no* $\mathrm{d}t$ *term of its own*, so if X is driftless (a martingale) then Z has no drift either, and a driftless Itô process is a martingale. The plain exponential e^X fails this because squaring manufactures the $+\frac{1}{2}[X]_t$ drift that $\mathcal{E}$ subtracts off. Two objects you already know are stochastic exponentials in disguise: geometric Brownian motion (24) is $S_0\,\mathcal{E}(\mu t + \sigma W)$, and the Radon–Nikodym density that powers Girsanov's change of measure (§9) is $\mathcal{E}(\theta W)$—the same construction, reappearing as the device for re-tilting probabilities.

6 Two worked models

Itô's lemma earns its keep by *solving* SDEs. Two solutions carry an enormous fraction of quantitative finance, and both appear, in some guise, in every other note in this series.

6.1 Geometric Brownian motion: the stock in Black–Scholes

A stock price cannot go negative, and its *returns*, not its absolute changes, are what behave stably. So model the proportional change as constant drift plus constant-volatility noise:

$$\mathrm{d}S_t = \mu\, S_t\, \mathrm{d}t + \sigma\, S_t\, \mathrm{d}W_t. \tag{22}$$

This is *geometric Brownian motion* (GBM). To solve it, apply Itô's lemma to $f(S) = \log S$, the natural move since the noise scales with S. Here $f' = 1/S$, $f'' = -1/S^2$, and with $\mu_t = \mu S$, $\sigma_t = \sigma S$ in (14),

$$\begin{aligned} \mathrm{d}(\log S_t) &= \Big(\tfrac{1}{S}\,\mu S + \tfrac{1}{2}\big(-\tfrac{1}{S^2}\big)\sigma^2 S^2\Big)\mathrm{d}t + \tfrac{1}{S}\,\sigma S\,\mathrm{d}W_t \\ &= \Big(\mu - \tfrac{1}{2}\sigma^2\Big)\mathrm{d}t + \sigma\,\mathrm{d}W_t. \end{aligned} \tag{23}$$

The right-hand side has *constant* coefficients, so $\log S_t$ is just Brownian motion with drift and we can integrate it directly:

$$S_t = S_0\, \exp\Big(\big(\mu - \tfrac{1}{2}\sigma^2\big)t + \sigma W_t\Big). \tag{24}$$

Three things to read off (24), all visible in Figure 8. First, S_t is *lognormal*—its logarithm is Gaussian—so it stays positive and is right-skewed. Second, the Itô correction has struck again: the growth rate of the *logarithm* is $\mu - \frac{1}{2}\sigma^2$, not μ. This gap is the celebrated *volatility drag*. Third, and as a direct consequence, the mean and the median part ways:

$$\mathbb{E}[S_t] = S_0\, e^{\mu t} \qquad \text{but} \qquad \operatorname{median}(S_t) = S_0\, e^{(\mu - \frac{1}{2}\sigma^2)t}. \tag{25}$$

The average outcome grows at μ, but the *typical* outcome grows more slowly, at $\mu - \frac{1}{2}\sigma^2$; the average is dragged up by a thin tail of large winners. Volatility is not free—it costs the typical path $\frac{1}{2}\sigma^2$ per unit time, even though the drift parameter is untouched.

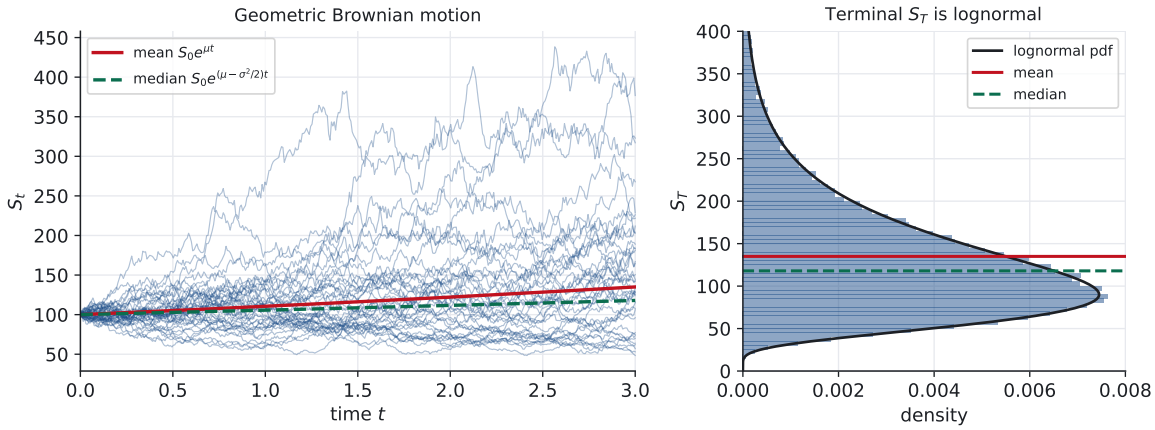

Figure 8: Geometric Brownian motion ($\mu = 10\%$, $\sigma = 30\%$). **Left:** forty paths, the mean $S_0e^{\mu t}$ (red), and the median $S_0e^{(\mu-\sigma^2/2)t}$ (green dashed) pulling apart—the volatility drag. **Right:** the terminal distribution is lognormal, right-skewed, with the mean sitting well above the median in the fat right tail. This is the price process underneath the Black–Scholes formula.

6.2 The Ornstein–Uhlenbeck process: mean reversion

Stock prices wander off to infinity; interest rates, spreads, and volatilities do not—they are tethered. The simplest model of a tethered quantity is the *Ornstein–Uhlenbeck* (OU) process,

$$\mathrm{d}x_t = \theta\,(\mu - x_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \tag{26}$$

in which the drift is a restoring force: when x is above its long-run level μ the drift is negative and pulls it down; below μ it pushes up. The parameter $\theta > 0$ is the *speed* of reversion, σ the size of the shocks fighting against it. For now we treat it purely as a mathematical object, the canonical model of a tethered quantity; it reappears as the workhorse of every mean-reverting model in the series—most importantly the short rate of the interest-rate models in Part V (§16).

To solve it, use the integrating-factor trick—the same one from linear ODEs. Apply Itô's lemma to $f(t,x) = x\,e^{\theta t}$ (here f depends on time explicitly, so we use (15); note $\partial_{xx} f = 0$, so there is no Itô correction this time—the diffusion coefficient does not depend on x):

$$\begin{aligned} \mathrm{d}\big(x_t e^{\theta t}\big) &= \theta x_t e^{\theta t}\,\mathrm{d}t + e^{\theta t}\,\mathrm{d}x_t \\ &= \theta x_t e^{\theta t}\,\mathrm{d}t + e^{\theta t}\big(\theta(\mu - x_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\big) \\ &= \theta\mu\, e^{\theta t}\,\mathrm{d}t + \sigma e^{\theta t}\,\mathrm{d}W_t. \end{aligned} \tag{27}$$

The x_t terms cancel—that is the point of the factor—leaving something we integrate directly from 0 to t:

$$x_t = \mu + (x_0 - \mu)\,e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)}\,\mathrm{d}W_s. \tag{28}$$

Read (28) the same way as before. The deterministic part $\mu + (x_0 - \mu)e^{-\theta t}$ relaxes exponentially from the start x_0 toward μ. The stochastic part is an Itô integral, so by (7) it has mean zero and by the isometry (8) a computable variance. Because the integrand is deterministic, x_t is *Gaussian at every time*, with

$$\mathbb{E}[x_t] = \mu + (x_0 - \mu)e^{-\theta t}, \qquad \mathrm{Var}(x_t) = \frac{\sigma^2}{2\theta}\big(1 - e^{-2\theta t}\big). \tag{29}$$

(The variance integral is done in Appendix C.) As $t \to \infty$ the memory of x_0 fades and the process forgets where it started, settling into a *stationary distribution*

$$x_\infty \sim \mathcal{N}\!\left(\mu,\ \frac{\sigma^2}{2\theta}\right). \tag{30}$$

The stationary spread $\sigma^2/2\theta$ is a tug-of-war: shocks σ widen it, reversion θ narrows it. One more quantity characterises the process—how fast it forgets. The autocorrelation of the stationary process decays exponentially,

$$\mathrm{Cov}(x_t, x_{t+\tau}) = \frac{\sigma^2}{2\theta}\,e^{-\theta\tau}, \tag{31}$$

with a characteristic "half-life" of $\log 2/\theta$. Figure 9 shows all three faces of the process at once.

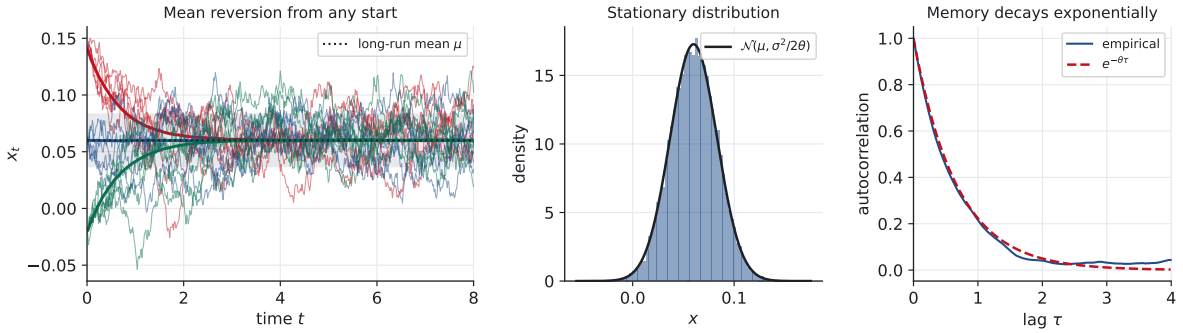

Figure 9: The Ornstein–Uhlenbeck process ($\theta = 1.5$, $\mu = 0.06$, $\sigma = 0.04$). **Left:** paths from three different starts, each pulled to the long-run mean μ (the shaded band is one stationary standard deviation). **Centre:** the long-run histogram matches the stationary normal $\mathcal{N}(\mu, \sigma^2/2\theta)$. **Right:** the autocorrelation decays as $e^{-\theta\tau}$—exponentially fading memory. Every mean-reverting model in this series is a dressed-up version of this picture.

The contrast with GBM is worth stating plainly, because the whole modelling craft lives in it. GBM has a fixed proportional volatility and wanders without bound; its log is a martingale-plus-drift. OU has an absolute volatility and a restoring drift; it is Gaussian, bounded in distribution, and stationary. Real series are chosen to look like one, the other, or a combination—and which one you pick determines everything downstream.

7 SDEs in the round

We have been writing SDEs like (11) and solving particular ones. A few general facts make the tool trustworthy and tell us how to compute with it when no closed form exists—which is most of the time.

When does an SDE have a solution? The SDE $\mathrm{d}X = \mu(t, X)\,\mathrm{d}t + \sigma(t, X)\,\mathrm{d}W$ has a unique solution under essentially the same condition as an ordinary ODE: the coefficients μ and σ should not change too fast (a *Lipschitz* condition) and not grow too fast (a linear-growth bound). Under these the solution exists, is unique, and depends continuously on the start—the usual guarantees. The notes in this series never stray outside this safe zone, so we take existence and uniqueness for granted and spend our attention elsewhere. (Two technical words you will meet: a *strong* solution is a path built on a *given* Brownian motion; a *weak* solution provides only a process with the right distribution. The distinction matters for proofs, rarely for pricing.)

How you actually simulate one: Euler–Maruyama. The most basic numerical scheme just reads the SDE literally—replace $\mathrm{d}t$ by a small step and $\mathrm{d}W$ by a fresh Gaussian draw of variance Δt:

$$X_{t+\Delta t} = X_t + \mu(t, X_t)\,\Delta t + \sigma(t, X_t)\,\sqrt{\Delta t}\;Z, \qquad Z \sim \mathcal{N}(0, 1). \tag{32}$$

This is the *Euler–Maruyama* method, and it is how essentially every Monte Carlo in this series advances a path. It is worth knowing its accuracy, because it is worse than the Euler method you know from ODEs. The *strong* error—how far the simulated path is from the true path driven by the *same* noise—shrinks only like $\sqrt{\Delta t}$ ("order $\frac{1}{2}$"), not Δt. The culprit is, predictably, the $(\mathrm{d}W)^2$ term. A careful Taylor expansion of the true increment over one step turns up a piece the scheme omits, $\frac{1}{2}\sigma\sigma'\big((\Delta W)^2 - \Delta t\big)$, the same $(\mathrm{d}W)^2$ second-order term that drives Itô's lemma. Its expectation is zero—which is why the *weak* error (in averaged quantities like a price) is the better order Δt—but the fluctuation $(\Delta W)^2 - \Delta t$ is itself of size Δt and does not average away *along a single path*, so it leaves a per-path error of order Δt per step that accumulates to $\sqrt{\Delta t}$

overall. Restoring exactly that omitted term is the *Milstein* scheme, which recovers strong order one. Figure 10 measures this on GBM, where we can compare against the exact solution (24).

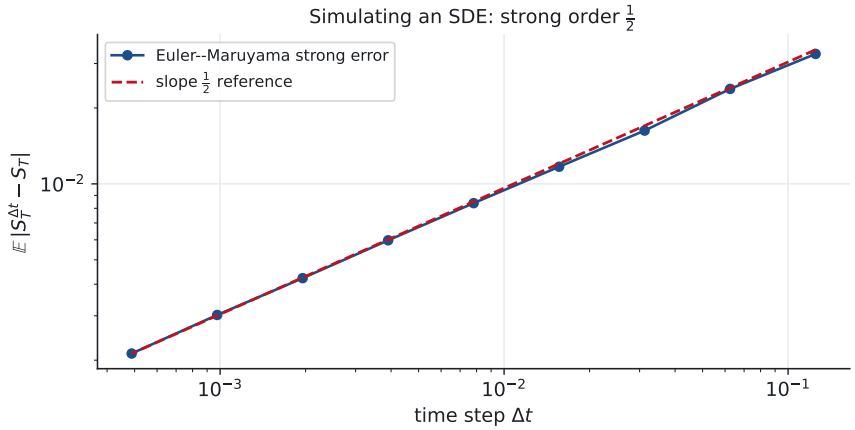

Figure 10: Strong convergence of Euler–Maruyama for GBM, against the exact solution driven by the same Brownian path. The error falls as $\sqrt{\Delta t}$—"strong order $\frac{1}{2}$," the dashed reference slope. Halving the step buys only a factor $\sqrt{2}$ in path accuracy; this is the tax the $(\mathrm{d}W)^2$ term levies on naive simulation, and the reason schemes like Milstein add a correction term to recover order one.

More than one driver: correlation. Most real models have several Brownian motions—a stock and its volatility, two points on a yield curve, a basket of names. Let $W^1, \dots, W^d$ be Brownian motions with instantaneous correlations $\mathrm{d}W^i\,\mathrm{d}W^j = \rho_{ij}\,\mathrm{d}t$ (so $\rho_{ii} = 1$). This one rule is all that changes: the multiplication table (13) now reads $\mathrm{d}W^i\,\mathrm{d}W^j = \rho_{ij}\,\mathrm{d}t$, Itô's lemma picks up a cross term for every pair (the two-process formula (16)), and the product and quotient rules of §5 acquire their covariation terms. That correlation channel—zero in ordinary calculus, the same $(\mathrm{d}W)^2 = \mathrm{d}t$ fact once more—is how every two-asset hedge, quanto and convexity adjustment in the later notes comes about. We put it to work directly on the exchange option of §12.

Part II — The pricing engine

8 Martingales, and why prices are expectations

The word *martingale* has come up three times; it is time to make it the centre of attention, because it is the bridge from "modelling a process" to "pricing a claim."

Definition and meaning. A process M_t is a martingale if it is the mathematical form of a *fair game*: its expected future value, given everything known now, equals its value now,

$$\mathbb{E}\big[M_T \mid \mathcal{F}_t\big] = M_t \qquad (T > t). \tag{33}$$

The symbol $\mathcal{F}_t$ is shorthand for the *information set*—everything that can be observed up to time t (in our case, the Brownian path so far)—so $\mathbb{E}[\,\cdot \mid \mathcal{F}_t]$ reads "the best forecast of this quantity given everything we know by time t." It is the same conditioning you already use, with "what we know" bundled into one symbol. The martingale condition then says: no drift, no predictable

edge—the best forecast of tomorrow is today. We have already met the prototype—any Itô integral $\int g\,\mathrm{d}W$ is a martingale, because each increment multiplies an independent mean-zero kick. The converse is deep and, for finance, decisive.

The martingale representation theorem. On a Brownian filtration, *every* martingale can be written as a constant plus an Itô integral:

$$M_t = M_0 + \int_0^t h_s\,\mathrm{d}W_s \tag{34}$$

for some non-anticipating integrand h. In words: *if a quantity is a fair game driven by this Brownian motion, then it can be exactly tracked by trading the Brownian motion*—there is a position h_s that replicates it. This abstract-sounding fact is the reason hedging works and options have unique prices. A derivative payoff, once expressed as a martingale (by discounting and switching to the right measure, below), is the integral of *something*; that something is the hedge ratio, the delta. Replication and the existence of h in (34) are two views of one theorem. The LSV, local-vol and Hull–White notes all lean on this without always naming it: "the payoff can be hedged" *is* (34).

Why we want prices to be martingales. Here is the whole logic of arbitrage-free pricing in three steps, to be made concrete in §9–§11. (i) If a traded asset's *discounted* price were not a martingale—if it had a predictable drift after accounting for the cost of money—you could on average profit from it risk-adjusted, which the market competes away. So discounted traded prices should be martingales under *some* probability measure. (ii) Pricing a derivative is then just taking an expectation under that measure: its value is the expected discounted payoff. (iii) The catch is that this special measure is *not* the real-world one—the drift you observe is not the drift you price with. The tool that converts between them is Girsanov's theorem, and it is the next section.

9 Girsanov's theorem: changing the drift

Girsanov's theorem answers a question that sounds impossible: can we change the *drift* of a process—make a trending series driftless, or vice versa—without changing the paths themselves, only their probabilities? Yes, and the recipe is to *reweight* the paths. Some paths become more likely, some less, and under the new weighting the average tilt is whatever we chose.

The mechanism: reweight by an exponential. Suppose under the real-world measure $\mathbb{P}$ the process is a driftless Brownian motion W_t. We want a new measure $\mathbb{Q}$ under which $\tilde{W}_t = W_t - \theta t$ is the driftless Brownian motion—equivalently, under which the original W_t has drift θ. Girsanov says: define the new measure by the *Radon–Nikodym density*

$$Z_t = \left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_t = \exp\!\Big(\theta W_t - \tfrac{1}{2}\theta^2 t\Big), \tag{35}$$

which tells you, path by path, how much to up- or down-weight its $\mathbb{P}$-probability to get its $\mathbb{Q}$-probability. If you have ever used *importance sampling*—drawing samples from one distribution and reweighting them to behave like samples from another—this is precisely that: Z_t is the likelihood-ratio weight, and $\mathbb{P}$ and $\mathbb{Q}$ describe the *same* random paths seen through two different weightings. (The notation $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$, the *Radon–Nikodym derivative*, is just "how many $\mathbb{Q}$-units of probability per $\mathbb{P}$-unit"—a density of one measure with respect to another, exactly as an ordinary density is a measure with respect to length.) Two things have to be checked, and both are short.

The weights are genuine ($\mathbb{E}^{\mathbb{P}}[Z_t] = 1$, so $\mathbb{Q}$ assigns total probability one). Since $W_t \sim \mathcal{N}(0, t)$ under $\mathbb{P}$, this is just the mean of a lognormal: $\mathbb{E}^{\mathbb{P}}[e^{\theta W_t}] = e^{\frac{1}{2}\theta^2 t}$ (the convexity drift of §4), so the $-\frac{1}{2}\theta^2 t$ in (35) is exactly the factor that cancels it, $\mathbb{E}^{\mathbb{P}}[Z_t] = e^{\frac{1}{2}\theta^2 t} e^{-\frac{1}{2}\theta^2 t} = 1$. This is no coincidence: $Z_t = \mathcal{E}(\theta W)_t$ is the stochastic exponential of §5, and the whole point of $\mathcal{E}$ is to pin the mean at $Z_0 = 1$.

The reweighting really does shift the drift. Take the simplest case, the terminal value W_T. Its $\mathbb{Q}$-density is its $\mathbb{P}$-density times the weight Z_T, and completing the square does the rest:

$$\underbrace{\frac{1}{\sqrt{2\pi T}}\, e^{-w^2/2T}}_{\mathbb{P}\text{-density of } W_T} \cdot \underbrace{e^{\theta w - \frac{1}{2}\theta^2 T}}_{\text{weight } Z_T} = \frac{1}{\sqrt{2\pi T}}\, e^{-(w-\theta T)^2/2T}, \tag{36}$$

which is the density of $\mathcal{N}(\theta T,\, T)$. The same outcomes, reweighted, are now a Gaussian centred at θT instead of 0: under $\mathbb{Q}$ the process has acquired drift θ, while under $\mathbb{P}$ it had none, and the variance T is untouched. Figure 11 shows exactly this reweighting doing its work.

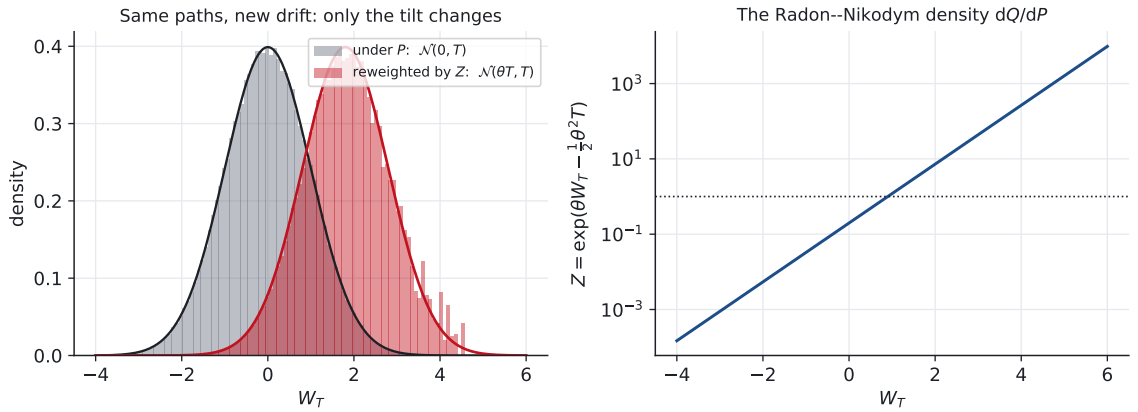

Figure 11: Girsanov's theorem as reweighting. **Left:** the terminal value W_T is $\mathcal{N}(0, T)$ under $\mathbb{P}$ (grey). Weighting each outcome by the Radon–Nikodym density $Z = \exp(\theta W_T - \frac{1}{2}\theta^2 T)$ produces exactly $\mathcal{N}(\theta T, T)$ (red)—the same sample paths, a new drift. **Right:** the density Z itself, which lifts the up-paths and suppresses the down-paths. The paths are untouched; only their probabilities move.

The one thing Girsanov cannot touch. Notice what changed and what did not. The drift moved freely. The *volatility*—the quadratic variation, the σ in $\sigma\, dW$—did *not*, and *cannot*. Changing measure tilts the odds of paths but cannot make a rough path smooth or a calm one wild; the quadratic variation (2) is a property of the path, not of the probability you assign it. This is the deepest expression of the note's spine: *you may negotiate the drift, but the quadratic variation is non-negotiable.* It is why volatility is the fundamental, measure-independent quantity in pricing, while drift is a convention you choose for convenience.

Why quadratic variation is untouchable: the intrinsic clock. There is a beautiful reason measure change cannot reach the quadratic variation: it *is* the process's own clock. *Lévy's characterisation* makes this exact: a continuous martingale X whose quadratic variation is $[X]_t = t$ *is* a Brownian motion—nothing else qualifies. More broadly (Dambis–Dubins–Schwarz), *every* continuous martingale is a Brownian motion run on a stretched clock, $X_t = B_{[X]_t}$: its own quadratic variation tells the time. Re-tilting probabilities, as Girsanov does, can change where the path tends to drift, but not the rate at which its clock ticks. So "how much quadratic variation" is the one question every model is ultimately answering, and the one answer no change of measure can rewrite—the spine of the note in its sharpest form.

Why finance needs it: $\mathbb{P} \to \mathbb{Q}$. Under the real-world measure $\mathbb{P}$, a stock drifts at some hard-to-estimate rate μ reflecting its risk premium. Pricing, by the logic of §8, wants the *discounted* price to be a martingale—to have no drift after financing costs. Girsanov supplies the change of measure that replaces the real drift μ with the risk-free rate r: under the *risk-neutral* measure $\mathbb{Q}$, every traded asset grows at r, discounted prices are martingales, and the unknowable risk premium has been defined away. The GBM of §6.1 becomes, under $\mathbb{Q}$,

$$\mathrm{d}S_t = r\,S_t\,\mathrm{d}t + \sigma\,S_t\,\mathrm{d}W_t^{\mathbb{Q}}, \tag{37}$$

with μ swapped for r and σ untouched—precisely Girsanov, precisely "drift negotiable, volatility not." This single substitution is what lets the next section price an option as an ordinary expectation.

10 Feynman–Kac: expectations are PDEs

We now have two ways to describe the value of a claim that pays $g(X_T)$ at time T. One is probabilistic: *simulate* the SDE many times and average the (discounted) payoff. The other is analytic: *solve a partial differential equation.* The Feynman–Kac theorem says these are the same thing, and the bridge between them is—one last time—the Itô correction.

The bridge. Let X follow $\mathrm{d}X = \mu(t,X)\,\mathrm{d}t + \sigma(t,X)\,\mathrm{d}W$, and define the conditional expectation

$$u(t,x) = \mathbb{E}\big[\, e^{-r(T-t)}\, g(X_T) \;\big|\; X_t = x \,\big]. \tag{38}$$

Then u solves the backward PDE

$$\partial_t u + \mu\,\partial_x u + \tfrac{1}{2}\sigma^2\,\partial_{xx} u - r\,u = 0, \qquad u(T,x) = g(x). \tag{39}$$

The derivation is one application of Itô's lemma, and worth seeing because the PDE simply *is* a no-drift condition. The discounted value process $M_t = e^{-rt}u(t,X_t)$ is a conditional expectation of a fixed (discounted) payoff, so it must be a martingale—its best forecast is its current value. Expand it with the time-dependent Itô lemma (15), remembering the e^{-rt} factor contributes a $-r\,e^{-rt}u\,\mathrm{d}t$ term:

$$\mathrm{d}M_t = e^{-rt}\Big(\underbrace{\partial_t u + \mu\,\partial_x u + \tfrac{1}{2}\sigma^2\,\partial_{xx} u - r\,u}_{\text{drift}}\Big)\mathrm{d}t + e^{-rt}\,\sigma\,\partial_x u\,\mathrm{d}W_t. \tag{40}$$

A martingale has no drift, so the bracket must vanish—and that bracket set to zero *is* the Feynman–Kac PDE (39). The boundary condition $u(T,x) = g(x)$ is just the definition (38) at $t = T$. The $\frac{1}{2}\sigma^2\partial_{xx}u$ term—the diffusion term of the PDE—is the Itô correction in disguise. The quadratic-variation tax we first met in §4 is literally the second-derivative term that makes a pricing equation a (parabolic) PDE rather than a first-order transport equation.

Black–Scholes as a special case. Take the risk-neutral GBM (37) and a call payoff $g(S) = (S-K)^+$. Plug $\mu = rS$, $\sigma = \sigma S$ into (39) and you get the *Black–Scholes PDE*; its solution is the Black–Scholes formula. Equivalently, by Feynman–Kac, the price is the discounted risk-neutral expectation $e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S_T-K)^+]$, which we can evaluate by Monte Carlo. The two must agree, and Figure 12 confirms they do, across strikes, to Monte-Carlo noise.

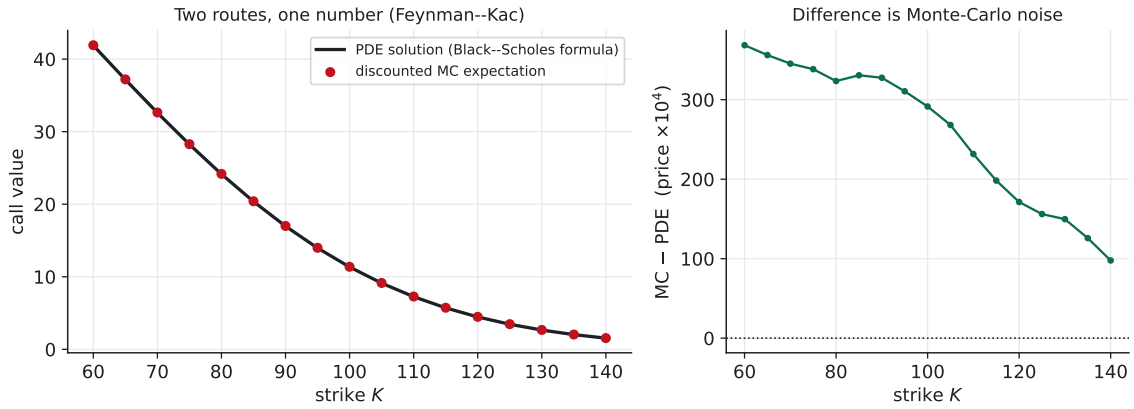

Figure 12: Feynman–Kac, two routes to one number. **Left:** European call values across strikes computed as the Black–Scholes PDE solution (line) and, independently, as a discounted risk-neutral Monte-Carlo expectation (dots). **Right:** their difference is a few parts in 10^4 of the price—pure simulation noise. "Solve the PDE" and "average the simulation" are the same statement, linked by Itô's lemma.

This equivalence is the practical heart of the subject. It is why some models are priced on PDE grids (the local-vol and Hull–White notes) and others by Monte Carlo (the LSV particle method), and why you may switch between the two whenever one is more convenient: they compute the same expectation. It is also where the Markovian property of §14 becomes pricing-critical—a PDE in (t, x) only exists if the future depends on the present state x alone.

11 Discounting, numéraires, and the forward measure

One loose end remains before we can build a term-structure model: *what, exactly, do we discount by, and which measure goes with it?* The answer is the notion of a *numéraire*, and it is the last piece of general machinery we need.

The money-market account and the risk-neutral measure. The natural unit of account is the *money-market account* $\beta_t = \exp(\int_0^t r_s \, \mathrm{d}s)$: one unit of cash rolled over continuously at the (possibly random) short rate r_s. The fundamental theorem of asset pricing says that markets are arbitrage-free precisely when there is a measure $\mathbb{Q}$ under which *every* traded asset, divided by β_t, is a martingale. Pricing is then the one formula

$$V_t = \beta_t \, \mathbb{E}^{\mathbb{Q}}\left[\frac{V_T}{\beta_T} \;\middle|\; \mathcal{F}_t \right] = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_s \, \mathrm{d}s} \, V_T \;\middle|\; \mathcal{F}_t \right]. \tag{41}$$

For a stock with deterministic rates this collapses to the $e^{-rT}\mathbb{E}^{\mathbb{Q}}[\cdot]$ we already used. For *interest-rate* products it does not, and that is the whole difficulty of the rate notes: r_s is itself the random thing being modelled, so it sits *inside* the discount factor and *is* the payoff, all at once.

Changing numéraire: the forward measure. A numéraire is just the *unit you quote prices in*. So far that unit has been rolled-up cash, β_t. But a price is only ever meaningful *relative* to something—a dollar is itself only worth a fixed basket of goods—and nothing forces that something to be cash. We could equally quote the option in units of a zero-coupon bond ("how many bonds is it worth?") or in shares of the stock. No-arbitrage is fundamentally a statement about *ratios* of traded assets, and a ratio does not care what unit sits in the denominator, so the theory survives any choice of unit. This is the deep and useful fact: *any* positive traded asset N_t can serve as numéraire, and each choice comes paired with its own measure $\mathbb{Q}^N$—the one under which *every* traded price divided by N_t is a martingale. (The money-market account β_t, with its

measure $\mathbb{Q}$, is just the most familiar choice.) Switching between two units is a Girsanov-style reweighting whose Radon–Nikodym density is simply the *relative growth of the two numéraires*, $(N'_T/N'_0)/(N_T/N_0)$—change the ruler, and the probabilities re-tilt to keep every ratio a fair game. The choice that tames (41) for rates is the *zero-coupon bond* $P(t,T)$—the value today of \$1 paid at T—used as numéraire. The reason it works is a single special value. By definition $V_t/P(t,T)$ is a martingale under the associated *T-forward measure* $\mathbb{Q}^T$, so

$$\frac{V_t}{P(t,T)} = \mathbb{E}^{\mathbb{Q}^T}\Big[\frac{V_T}{P(T,T)} \;\Big|\; \mathcal{F}_t\Big]. \tag{42}$$

But a bond paying \$1 at its own maturity is worth exactly \$1 then: $P(T,T) = 1$. The denominator on the right collapses, and multiplying back through by the known-today $P(t,T)$ leaves

$$V_t = P(t,T)\,\mathbb{E}^{\mathbb{Q}^T}\big[V_T \;\big|\; \mathcal{F}_t\big]. \tag{43}$$

Compare with the money-market version (41): there the discount factor $e^{-\int_t^T r_s\,\mathrm{d}s}$ is *random* and tangled up *inside* the expectation with the payoff. Here it has been pulled out and replaced by the single deterministic number $P(t,T)$, known at time t, leaving a clean expectation of the payoff alone—that decoupling is the whole reason for changing numéraire. Under $\mathbb{Q}^T$, the T-maturity forward rate is a martingale—a fact we will use directly when we build HJM. The single idea to carry forward: *pick the numéraire that makes your payoff simple, and a matching measure always exists.* Figure 13 collects the three units this note actually uses.

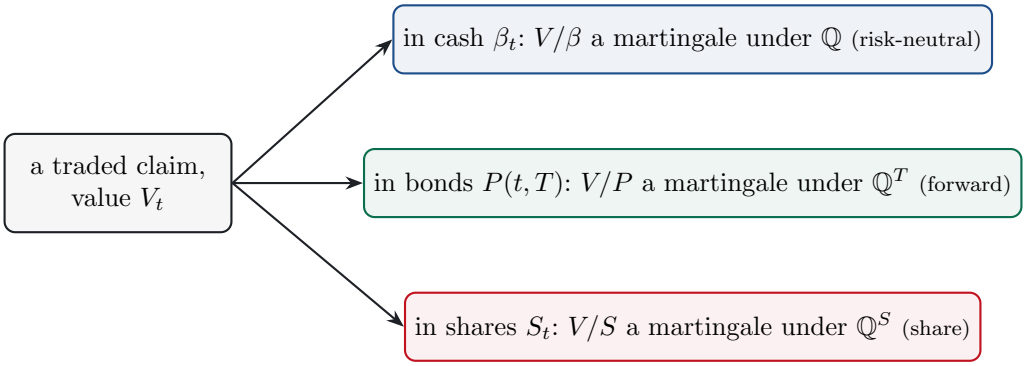

Figure 13: One price, many rulers. The *same* claim can be quoted in any positive traded asset; each choice of unit (numéraire) comes with the probability measure that turns "price divided by that unit" into a martingale. Cash β_t gives the risk-neutral measure $\mathbb{Q}$ (used throughout §10); the bond $P(t,T)$ gives the forward measure $\mathbb{Q}^T$ (which cleans up random discounting for rates); the stock S_t gives the share measure $\mathbb{Q}^S$ (which cracks the asset leg of Black–Scholes in §12.1). Switching ruler is a Girsanov reweighting.

Part III — Pricing in action

12 Analytical pricing: a few options in closed form

The engine of Part II is now complete, and it is enough to price several fundamental options *in closed form*. These formulas are worth deriving once by hand, because each shows a tool from earlier doing real work—and because they are the sanity checks every numerical method is measured against.

12.1 The Black–Scholes call and put, by change of measure

Under the risk-neutral measure $\mathbb{Q}$ the stock is the geometric Brownian motion (37), $\mathrm{d}S = rS\,\mathrm{d}t + \sigma S\,\mathrm{d}W^{\mathbb{Q}}$, and the price of a European call paying $(S_T - K)^+$ is the discounted expectation (41),

$$C = e^{-rT}\,\mathbb{E}^{\mathbb{Q}}\big[(S_T - K)^+\big] = e^{-rT}\,\mathbb{E}^{\mathbb{Q}}\big[(S_T - K)\,\mathbf{1}_{\{S_T > K\}}\big]. \tag{44}$$

Split the payoff into its two pieces and price each:

$$C = \underbrace{e^{-rT}\,\mathbb{E}^{\mathbb{Q}}\big[S_T\,\mathbf{1}_{\{S_T>K\}}\big]}_{\text{the asset leg}} \;-\; \underbrace{K\,e^{-rT}\,\mathbb{E}^{\mathbb{Q}}\big[\mathbf{1}_{\{S_T>K\}}\big]}_{\text{the cash leg}}. \tag{45}$$

The cash leg is immediate: $\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{S_T>K\}}] = \mathbb{Q}(S_T > K)$, and this probability is where d_2 comes from—it is not pulled from a hat. From the risk-neutral solution (24)–(37), $\log S_T = \log S_0 + (r - \tfrac{1}{2}\sigma^2)T + \sigma W_T^{\mathbb{Q}}$ is Gaussian with mean $\log S_0 + (r - \tfrac{1}{2}\sigma^2)T$ and standard deviation $\sigma\sqrt{T}$. So, standardising,

$$\mathbb{Q}(S_T > K) = \mathbb{Q}(\log S_T > \log K) = N\left(\frac{\log(S_0/K) + (r - \tfrac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) = N(d_2), \tag{46}$$

where N is the standard normal CDF and the symmetry $1 - N(x) = N(-x)$ has flipped the inequality. That argument of N *is* d_2. The asset leg will yield $N(d_1)$ by the identical calculation done under the share measure, where (as we are about to see) the stock's drift is raised from $r - \tfrac{1}{2}\sigma^2$ to $r + \tfrac{1}{2}\sigma^2$—which is the only difference between d_2 and d_1. The asset leg, $e^{-rT}\mathbb{E}^{\mathbb{Q}}[S_T\,\mathbf{1}_{\{S_T>K\}}]$, is harder: the payoff is weighted by S_T itself, an inconvenient extra factor multiplying the indicator, so it is not simply a probability. This is where the *change of measure* earns its keep. Quote the leg in *shares* instead of dollars—that is, switch the numéraire from the money-market account to the *stock itself* (§11, Figure 13)—and the troublesome S_T is exactly the numéraire, so it cancels: the leg becomes "one share, worth S_0 today, times a plain probability,"

$$e^{-rT}\,\mathbb{E}^{\mathbb{Q}}\big[S_T\,\mathbf{1}_{\{S_T>K\}}\big] = S_0\,\mathbb{Q}^S(S_T > K) = S_0\,N(d_1),$$

where $\mathbb{Q}^S$ is the "share measure," the one in which the stock is the unit of account (its Radon–Nikodym density is the stochastic exponential $\mathcal{E}(\sigma W)$ of §5). Trading the S_T-weighting for a change of measure has shifted the drift used to compute the probability—which is exactly what replaces $N(d_2)$ with the larger $N(d_1)$. Each leg is thus an exercise probability, *computed under the measure that makes it simple*. The result is the Black–Scholes formula,

$$\boxed{C = S_0\,N(d_1) - K\,e^{-rT}\,N(d_2)}, \qquad \begin{aligned} d_1 &= \frac{\log(S_0/K) + (r + \tfrac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}, \end{aligned} \tag{47}$$

and the put follows from put–call parity, $P = C - S_0 + Ke^{-rT}$, or identically by the same two-measure split. Figure 14 shows the two probabilities side by side: $N(d_2)$ is the exercise probability under the money-market measure $\mathbb{Q}$, $N(d_1)$ the exercise probability under the stock numéraire. They differ—and that difference, the whole content of "change of measure," is what makes the option worth more than its discounted intrinsic value. (The same number came out of the Feynman–Kac PDE and a raw Monte Carlo in Figure 12—three routes, one price.)

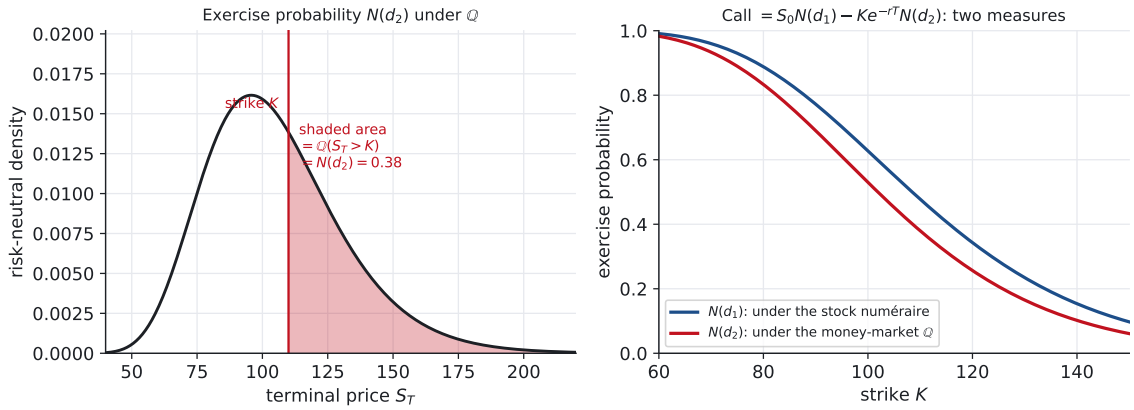

Figure 14: Black–Scholes as two exercise probabilities under two measures. **Left:** under the risk-neutral measure $\mathbb{Q}$, the probability the call finishes in the money is the shaded area $\mathbb{Q}(S_T > K) = N(d_2)$. **Right:** the call $C = S_0N(d_1) - Ke^{-rT}N(d_2)$ pairs that probability with $N(d_1)$, the corresponding probability under the stock-numéraire measure. The gap between the curves is exactly what the change of measure accounts for.

Digital options, for free. The split (45) has already priced two simpler options. A *cash-or-nothing* digital, paying \$1 if $S_T > K$, is worth exactly the cash leg, $e^{-rT}N(d_2)$. An *asset-or-nothing* digital, paying S_T if $S_T > K$, is the asset leg, $S_0N(d_1)$. The vanilla call is just the asset-or-nothing minus K cash-or-nothings—which is all (47) says.

12.2 Margrabe's exchange option, by change of numéraire

Now two risky assets, and the option to swap one for the other: the *exchange option*, paying $(S_T^1 - S_T^2)^+$—the right, at T, to give up asset 2 and receive asset 1. With two stocks, $\mathrm{d}S^i = rS^i\,\mathrm{d}t + \sigma_i S^i\,\mathrm{d}W^i$ and $\mathrm{d}W^1\mathrm{d}W^2 = \rho\,\mathrm{d}t$, this looks like a two-dimensional problem. The change of numéraire collapses it to one dimension. Use *asset 2* as the numéraire: price the option in units of S^2. The relative price $R_t = S_t^1/S_t^2$ is a traded asset divided by the numéraire, hence a martingale under the measure $\mathbb{Q}^2$ associated with S^2. Feed $\mathrm{d}S^i/S^i = r\,\mathrm{d}t + \sigma_i\,\mathrm{d}W^i$ into the quotient rule (19): the drift terms (the two $r\,\mathrm{d}t$'s and the Itô corrections) must cancel, because R is a $\mathbb{Q}^2$-martingale and the quotient rule is exactly the bookkeeping that enforces it. What is left is pure diffusion,

$$\frac{\mathrm{d}R_t}{R_t} = \sigma_1\,\mathrm{d}W^1 - \sigma_2\,\mathrm{d}W^2 + (\text{zero drift}), \tag{48}$$

and the volatility of R is the standard deviation of that single combined increment. Using $\mathrm{Var}(\sigma_1\mathrm{d}W^1 - \sigma_2\mathrm{d}W^2) = (\sigma_1^2 + \sigma_2^2 - 2\rho\,\sigma_1\sigma_2)\,\mathrm{d}t$ (the cross term carries the correlation $\mathrm{d}W^1\mathrm{d}W^2 = \rho\,\mathrm{d}t$),

$$\sigma \;=\; \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\,\sigma_1\sigma_2}\,. \tag{49}$$

So R_t is a driftless geometric Brownian motion with vol σ, and the payoff $(S_T^1 - S_T^2)^+ = S_T^2\,(R_T - 1)^+$ is, in numéraire units, an ordinary call on R struck at 1 with zero interest rate. Black–Scholes (47) applies verbatim and gives *Margrabe's formula*,

$$\boxed{C^{\text{exch}} = S_0^1\,N(d_1) - S_0^2\,N(d_2)}, \qquad \begin{aligned} d_1 &= \frac{\log(S_0^1/S_0^2) + \tfrac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}. \end{aligned} \tag{50}$$

Two features are worth pausing on. The interest rate has vanished—both legs are traded assets, so financing cancels. And the only volatility that matters is the *relative* one (49): the more correlated the two assets, the smaller σ, and the cheaper the option—if they move together there

is little reason to switch. Figure 15 confirms the formula against Monte Carlo and shows the price falling as correlation rises. Margrabe is the parent of a whole family of two-asset products (spread, quanto, outperformance options), and the change-of-numéraire move that cracks it is the same one that underlies the forward-measure pricing of the rate notes.

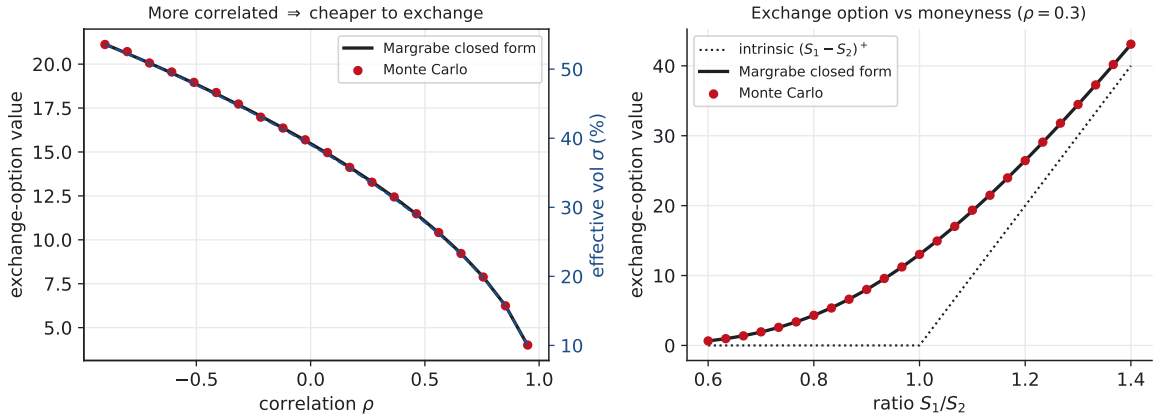

Figure 15: Margrabe's exchange option, closed form (50) against Monte Carlo. **Left:** value versus correlation ρ; the effective volatility (49) (blue, right axis) shrinks as $\rho \to 1$, taking the option value with it—more-correlated assets are cheaper to exchange. **Right:** value versus moneyness S_0^1/S_0^2 at $\rho = 0.3$, the familiar convex option profile sitting above its intrinsic value. The closed form and the simulation agree throughout.

12.3 Path extremes: barriers and lookbacks, via reflection

The vanilla call and the exchange option both pay on the *terminal* value S_T. A second large family pays instead on the *extreme* a path reached along the way—a barrier option that knocks in or out if a level is ever touched, a lookback that pays relative to the running maximum. Pricing these in closed form needs the distribution of that maximum, and a single symmetry argument delivers it.

The reflection principle and the running maximum. How high does a Brownian path get? The *running maximum* $M_T = \max_{0 \le t \le T} W_t$ is governed by a beautiful symmetry argument, the *reflection principle*. The first time the path reaches a level $m \ge 0$ is a *stopping time* τ—a moment defined using only the path so far (the subject of §13, next). After τ, by the symmetry of Brownian motion, the continuation and its mirror image reflected about m are equally likely (Figure 16, left). The argument is a clean split into two cases. A path has $M_T \ge m$ either because it finishes above m, or because it reached m but fell back below:

$$\mathbb{P}(M_T \ge m) = \underbrace{\mathbb{P}(M_T \ge m,\ W_T \ge m)}_{=\mathbb{P}(W_T \ge m)} + \underbrace{\mathbb{P}(M_T \ge m,\ W_T < m)}_{=\mathbb{P}(W_T \ge m)}. \tag{51}$$

The first term equals $\mathbb{P}(W_T \ge m)$ because finishing above m already *implies* the level was reached. The second term equals it too: reflecting the post-τ continuation about m maps every path that reached m and finished *below* to an equally likely path that finishes *above*, a one-to-one correspondence. Adding the two identical terms,

$$\mathbb{P}(M_T \ge m) = 2\,\mathbb{P}(W_T \ge m), \qquad \text{so} \qquad M_T \overset{d}{=} |W_T| \sim \left|\mathcal{N}(0,T)\right| \tag{52}$$

(the symbol $\overset{d}{=}$ reads "has the same distribution as"; the second equality follows because $|W_T| \ge m$ also has probability $2\mathbb{P}(W_T \ge m)$ by the symmetry of the normal). The running maximum has the distribution of the absolute value of the endpoint—a fact you can read straight off the

histogram in Figure 16 (right). This closes the path-extreme family: *barrier* and *lookback* prices both reduce to a statement about M_T, whose distribution we now have in hand.

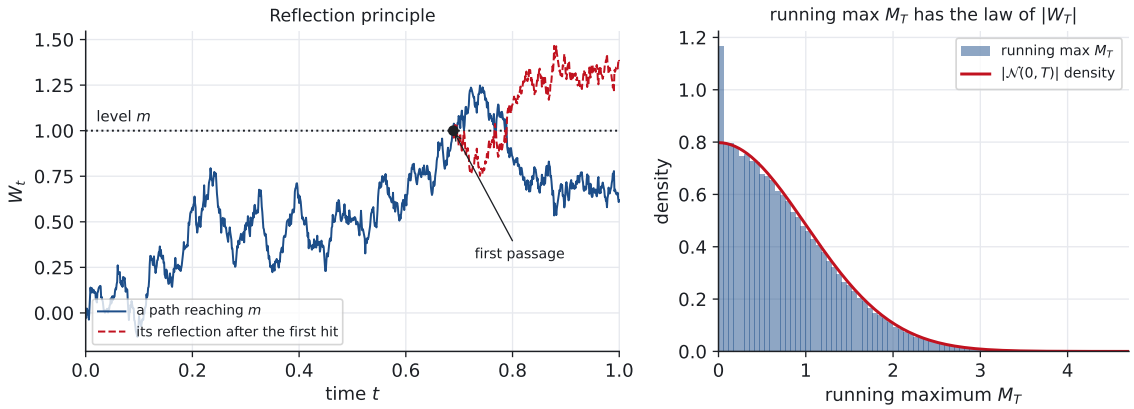

Figure 16: The reflection principle. **Left:** once a path first touches level m (a stopping time), its continuation and the version reflected about m are equally likely—so paths finishing above m are in one-to-one correspondence with reflected paths finishing below. **Right:** the consequence, (52): the running maximum M_T has exactly the law of $|W_T|$, here confirmed against 200,000 simulated paths.

13 Stopping times and optimal stopping

Every option so far had a fixed expiry T. But some of the most important contracts—American options, callable bonds, the decision to abandon or continue a project—let you choose *when* to act. The mathematics of that choice is the theory of *stopping*, and it rests on two ideas: a rule for when you are allowed to stop, and a theorem about what stopping can and cannot achieve.

Stopping times. A *stopping time* τ is a random time whose arrival you can recognise using only the information available so far: for every t, whether $\{\tau \le t\}$ has happened is decided by the path up to t, with no peeking ahead. "The first time the price hits \$120" is a stopping time—you know it when you see it. "The time of this month's high" is *not*—you cannot know today's move is the peak until the month is out. This is the same non-anticipating discipline that defined the Itô integrand in §3, now applied to decisions. The first-passage time behind the reflection principle (§12.3) is the canonical example.

Optional stopping: you cannot out-time a fair game. The fundamental fact about martingales and stopping times is the *optional stopping theorem*: if M_t is a martingale and τ a (suitably bounded) stopping time, then

$$\mathbb{E}[M_\tau] = M_0. \tag{53}$$

Whatever non-anticipating rule you use to decide when to quit, your expected payoff is your starting value. You cannot beat a driftless market by clever timing—no stop-loss, no take-profit, no entry rule changes the expectation of a fair game. This is the precise sense in which a martingale is unbeatable.

It is also a remarkably effective *calculation* tool—"apply it to a cleverly chosen martingale and read off the answer" is best shown than stated. Take a Brownian motion started at $W_0 = x$ inside the band $(0, b)$, and let τ be the first time it hits either wall. What is the probability p that it reaches b before 0? Brownian motion is itself a martingale, so optional stopping gives $\mathbb{E}[W_\tau] = W_0 = x$. But W_τ takes only the two values b (with probability p) and 0 (with

probability $1-p$), so $\mathbb{E}[W_\tau] = p\,b$. Equating,

$$p\,b = x \qquad \Longrightarrow \qquad p = \frac{x}{b}\,: \tag{54}$$

the exit probability is just linear in the starting point—the continuous "gambler's ruin." No PDE, no integral, one line. This is the engine behind hitting-probability and first-passage results throughout the subject (including the reflection principle of §12.3): pick the martingale whose stopped value encodes the quantity you want, then read it off.

Optimal stopping: when timing is the whole problem. Now flip it around. Suppose the payoff is not a fixed claim but something you *collect at a moment of your choosing*, and you want the timing that maximises its expected discounted value:

$$V_0 = \sup_{\tau}\ \mathbb{E}^{\mathbb{Q}}\big[\,e^{-r\tau}\,g(S_\tau)\,\big], \tag{55}$$

the supremum over all stopping times $\tau \le T$. This is the *optimal stopping problem*, and it *is* the American option: g is the exercise payoff, and you may exercise at any moment up to expiry. Its solution has a beautifully intuitive shape. At each instant you compare two numbers—the value of *exercising now*, $g(S_t)$, and the value of *waiting*, the expected discounted continuation. You exercise the moment the first overtakes the second. This dynamic-programming rule (the *Snell envelope*) splits the state space into a *continuation region*, where holding is worth more, and a *stopping region*, where you exercise, separated by a *free boundary* $S^*(t)$—"free" because its location is part of the unknown, not given in advance. Figure 17 shows it for an American put.

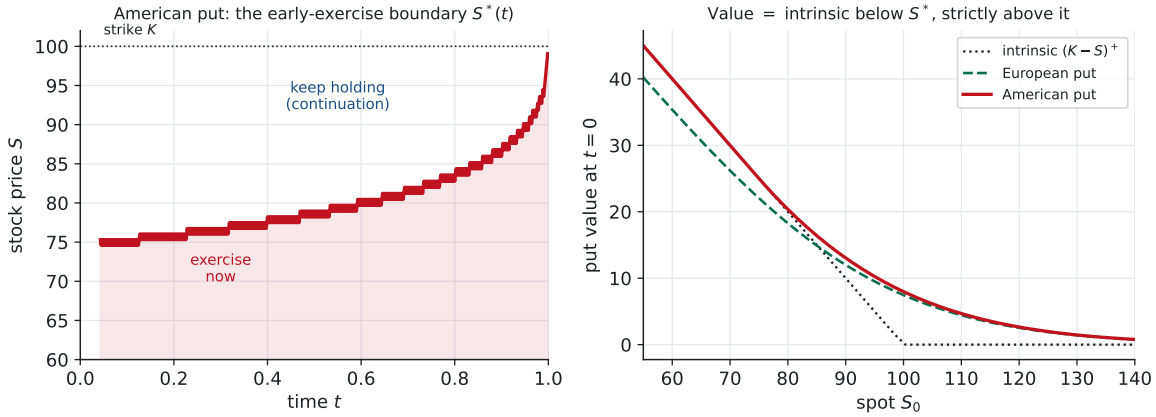

Figure 17: Optimal stopping, made concrete with an American put. **Left:** the early-exercise boundary $S^*(t)$ splits the plane—exercise below it, keep holding above. The boundary rises to the strike at expiry, where the choice collapses to the European one. **Right:** at $t = 0$ the American value sits strictly above both the European value and the intrinsic payoff $(K - S)^+$, meeting the payoff tangentially at $S^*(0)$ ("smooth pasting"), below which immediate exercise is optimal. The early-exercise right is worth the gap between the red and green curves.

How it connects to everything else. Optimal stopping ties the threads of this note together. The value function $V(t, S)$ obeys the Black–Scholes PDE in the continuation region and equals the payoff in the stopping region—a *free-boundary* PDE, solvable on a grid or tree precisely because the dynamics are Markovian (§14); the recombining tree of Figure 17 is exactly such a solver. When the state is higher-dimensional and no grid will do, the standard tool is *regression Monte Carlo* (Longstaff–Schwartz): simulate paths forward, then estimate the continuation value by regressing discounted future payoffs on the current state, and exercise wherever it falls short of the payoff. Either way the object is (55), and the answer is a stopping time—a decision rule, not just a number.

Part IV — Markovian versus non-Markovian

14 What "Markovian" buys you

We have quietly relied on a property that deserves to be dragged into the open, because it silently decides whether a model is fast or slow, tractable or forbidding. A process is *Markovian* if its future depends on the present *state* alone, not on the path that led there: knowing where you are tells you everything knowing your whole history would. Brownian motion is Markovian by its independent-increments property; so is any SDE whose coefficients depend only on the current (t, X_t).

Why it matters: a low-dimensional state means a PDE. The payoff in Feynman–Kac (38) was a function of *the current state* x, and that is exactly what let us write a PDE in (t, x). This is the general rule:

Markov ⇒ the value is a function of the state. If the dynamics are Markovian in a state X_t of dimension n, the price of a claim is a *deterministic function* $V(t, X_t)$ solving a PDE in n space dimensions—and can be computed on a grid, a tree, or a recombining lattice. Small n means fast and exact; this is why low-dimensional Markov models dominate production pricing.

A one-factor Markov model (Hull–White, local vol in one spatial variable) gives a one-dimensional PDE—trivial to solve on a grid, and representable by a *recombining* tree where up-then-down lands where down-then-up does. That recombination, which keeps the tree from exploding to 2^n nodes, *is* the Markov property made computational.

State augmentation: buying back Markovianity. Many models that look non-Markovian in the variable you care about become Markovian once you enlarge the state. The leading example is *stochastic volatility*. The stock price S_t alone is *not* Markov when its volatility v_t is itself random and persistent—tomorrow's distribution of S depends on today's volatility, which you cannot read off S alone. But the *pair* (S_t, v_t) *is* Markov: together they carry everything the future needs. The price is a function $V(t, S, v)$ solving a *two*-dimensional PDE. You have traded one dimension of state for the right to use all the machinery again. The LSV note lives exactly here—its state is (spot, variance)—and Figure 18 draws the idea.

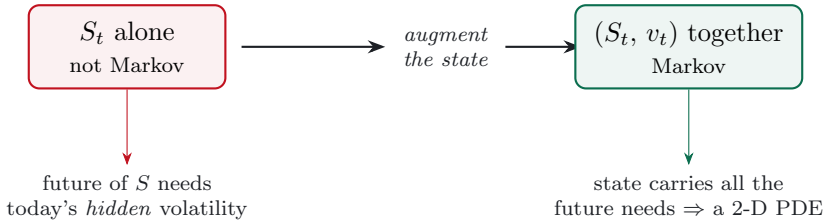

Figure 18: State augmentation. A stock with random, persistent volatility is not Markov in price alone—the future needs the hidden volatility too. Enlarge the state to the pair (price, volatility) and Markovianity returns, at the cost of one extra PDE dimension. This is the standard move that keeps stochastic-volatility models tractable.

The lesson generalises: non-Markovianity is often a sign that you are looking at *too few* variables. Add the right hidden states and the Markov machinery—PDEs, trees, the lot—comes back. The question §15 confronts is what to do when *no finite* set of extra states will do.

15 When memory will not collapse

Sometimes no finite state suffices, and the model genuinely drags its whole past behind it. Two flavours of this are worth distinguishing, because one is benign and the other is structural.

Path-dependent payoffs (benign). An Asian option pays on the *average* price over the life of the trade; a barrier option cares whether a level was *ever* touched; a lookback pays on the *running maximum*. The *dynamics* here are perfectly Markov—it is the *payoff* that depends on the path. The fix is the same state augmentation as before: add the running average, or the running maximum, as an extra state variable, and you are Markov again in the enlarged space. The memory is finite-dimensional; you just had to name it.

Rough volatility and fractional noise (structural). The harder case is when the *driving noise itself* has memory. Replace Brownian motion with *fractional* Brownian motion B_t^H, whose increments are *correlated* across time, governed by a *Hurst exponent* $H \in (0,1)$:

$$\mathbb{E}\big[(B_t^H)^2\big] \propto t^{2H}, \tag{56}$$

so that $H = \frac{1}{2}$ recovers ordinary Brownian motion (independent increments, the memoryless case), $H > \frac{1}{2}$ gives *persistent* increments (a move up tends to be followed by another up—trends), and $H < \frac{1}{2}$ gives *anti-persistent* ones (moves tend to reverse—a jagged, "rough" path). Figure 19 shows all three and verifies the scaling (56).

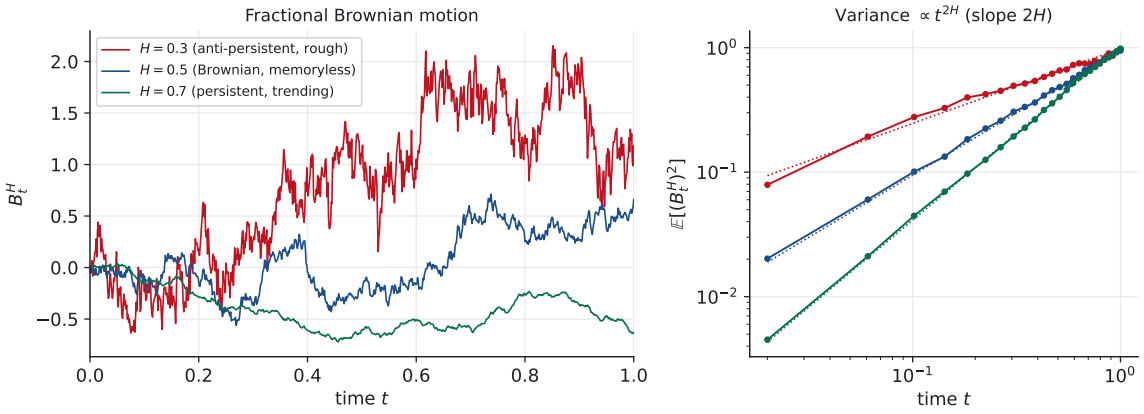

Figure 19: Fractional Brownian motion for three Hurst exponents. **Left:** $H < \frac{1}{2}$ (red) is anti-persistent and visibly rougher; $H = \frac{1}{2}$ (blue) is ordinary Brownian motion; $H > \frac{1}{2}$ (green) is persistent and trends. **Right:** the mean-squared displacement scales as t^{2H}, the straight log-log lines confirming the exponent. For $H \neq \frac{1}{2}$ the increments are correlated—the process has memory and is *not* Markov in any finite state.

The crucial structural fact: for $H \neq \frac{1}{2}$, fractional Brownian motion is *not* a Markov process and *not* a semimartingale, and no finite augmentation of the state will make it one. (A *semimartingale* is precisely a process that splits into a drift plus a martingale—the class Itô's integral of §3 was built to handle. Fractional noise falls outside it, so one cannot even form $\int g\,\mathrm{d}B^H$ in the ordinary Itô way, and the entire toolkit of this note has to be rebuilt or replaced.) The correlation in (56) reaches across all time lags; to know the future you need, in principle, the entire past. This is not a modelling inconvenience to be engineered away—it is the point. Empirically, realised volatility is very *rough*, well described by $H \approx 0.1$, far below $\frac{1}{2}$. *Rough volatility* models exploit exactly this to fit the steep short-maturity smile that Markovian stochastic-vol models struggle with. The price of that fit is the loss of the Markov toolkit: no low-dimensional PDE, no recombining tree, and Monte Carlo schemes that must carry history. It

is the cleanest illustration in the note of the trade-off the whole subject negotiates—*tractability against realism*, with the Markov property as the currency.

The Markov dividing line. *Markov* (Brownian-driven, state of finite dimension): value is $V(t, X_t)$, solvable by PDE/tree, fast. *Path in payoff only*: augment the state (running average, max) and recover Markov. *Memory in the noise* (fractional/rough, $H \neq \frac{1}{2}$): no finite state suffices—realism bought with the loss of the PDE/tree machinery.

Part V — Putting it to work: the term structure and HJM

16 From a short rate to the whole curve

We close by assembling everything—Itô, OU, Girsanov, numéraires, the Markov question—into the modelling of interest rates, ending at the Heath–Jarrow–Morton framework. This is the deep end the other rate notes swim in, and the point of the whole note is to leave you able to follow them.

The objects. Fix today as time 0. The *zero-coupon bond* $P(t,T)$ is the value at t of \$1 paid at T; it is what discounting actually means once rates are random. From the curve of bond prices one reads two equivalent descriptions. The *short rate* r_t is the instantaneous rate for borrowing right now; discounting uses the money-market account $\beta_t = \exp(\int_0^t r_s\,\mathrm{d}s)$ from §11, and risk-neutral pricing (41) gives the bond as an expectation,

$$P(t,T) = \mathbb{E}^{\mathbb{Q}}\Big[\, e^{-\int_t^T r_s\,\mathrm{d}s} \;\Big|\; \mathcal{F}_t \Big]. \tag{57}$$

The *instantaneous forward rate* $f(t,T)$ is the rate, agreed at t, for an infinitesimal loan starting at the future date T; it is defined from the bond by

$$f(t,T) = -\,\partial_T \log P(t,T), \qquad \text{equivalently} \qquad P(t,T) = \exp\Big(-\int_t^T f(t,u)\,\mathrm{d}u\Big), \tag{58}$$

and the short rate is its front end, $r_t = f(t,t)$. The forward rates for all maturities T make up the *forward curve*, the object HJM models directly.

The short-rate approach (Vasicek/Hull–White), in one paragraph. The oldest tractable idea is to model the single short rate r_t with a mean-reverting SDE—and we have already solved it. Vasicek takes r_t to be the OU process (26), $\mathrm{d}r = \theta(\mu - r)\,\mathrm{d}t + \sigma\,\mathrm{d}W$, with all the structure of §6.2: Gaussian rates, exponential reversion, a stationary distribution. Hull–White makes the long-run level time-dependent, $\mathrm{d}r = (\vartheta(t) - a\,r)\,\mathrm{d}t + \sigma\,\mathrm{d}W$, choosing the function $\vartheta(t)$ so that the model reproduces *today's* observed curve exactly while keeping the OU dynamics for everything stochastic. The trick is a degrees-of-freedom count: a constant long-run level can match only one number, but a whole *function* $\vartheta(t)$ carries one free value per future date, exactly enough to pin the model's initial forward curve $f(0,T)$ onto the observed one maturity by maturity. (Indeed $\vartheta(t)$ is read off in closed form by differentiating the observed curve.) Because r_t is a one-dimensional Markov process, the bond (57) is a function $P(t,r_t)$ solving a one-dimensional PDE (Feynman–Kac, §10); in fact it is *affine*—linear in the rate r_t inside the exponent, $P(t,T) = \exp(A(t,T) - B(t,T)\,r_t)$, with A and B functions of time only—which collapses the bond price to two simple deterministic functions and is why the model is so fast. This is the entire content of a short-rate model: *one Markov driver, one PDE.* Its limitation is also its definition—one driver moves the whole curve, so all maturities are perfectly correlated.

17 The Heath–Jarrow–Morton framework

HJM inverts the short-rate philosophy. Instead of modelling one rate and *deriving* the curve, it models the evolution of the *entire forward curve* $f(t,T)$ at once, taking today's curve $f(0,\cdot)$ as the given starting point. Write an SDE for every maturity T simultaneously:

$$\mathrm{d}f(t,T) = \alpha(t,T)\,\mathrm{d}t + \sigma(t,T)\,\mathrm{d}W_t, \tag{59}$$

one Brownian motion W (or several) driving a whole continuum of forward rates, each maturity T with its own drift $\alpha(t,T)$ and volatility $\sigma(t,T)$. Because the model starts *from* the observed curve, it fits today's term structure perfectly by construction—no calibration of an initial condition needed. The state, note, is now the whole curve: an *infinite-dimensional* object. We will return to what that costs.

The no-arbitrage drift condition—the spine, one last time. Here is the elegant heart of HJM, and it is the note's central theme in its purest form. You are *not* free to choose the drifts $\alpha(t,T)$ and the volatilities $\sigma(t,T)$ independently. No-arbitrage forces the drift to be a specific function of the volatility. Concretely, under the risk-neutral measure,

$$\boxed{\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)\,\mathrm{d}u} \tag{60}$$

the famous *HJM drift condition*. The volatility structure $\sigma(t,T)$ is the *only* free input; the drift is then determined. Where does (60) come from? From insisting that the discounted bond $P(t,T)/\beta_t$ be a $\mathbb{Q}$-martingale (§8)—that bonds, like all traded assets, carry no risk-adjusted edge. Applying Itô's lemma to $\log P(t,T) = -\int_t^T f(t,u)\,\mathrm{d}u$ and setting the drift of the discounted bond to zero produces (60) directly. The $\int_t^T \sigma\,\mathrm{d}u$ factor is, once again, an Itô second-order term—the quadratic variation of the bond's own log-price—surfacing as a no-arbitrage constraint. We carry out the derivation in Appendix D.

This is the same lesson as Girsanov, restated for the curve: *choose the volatility and the drift is no longer yours to set.* Volatility is the real input; drift is whatever no-arbitrage makes it. Figure 20 evolves a forward curve under (59)–(60): it moves randomly yet stays arbitrage-free at every step, because the drift is slaved to the chosen vol.

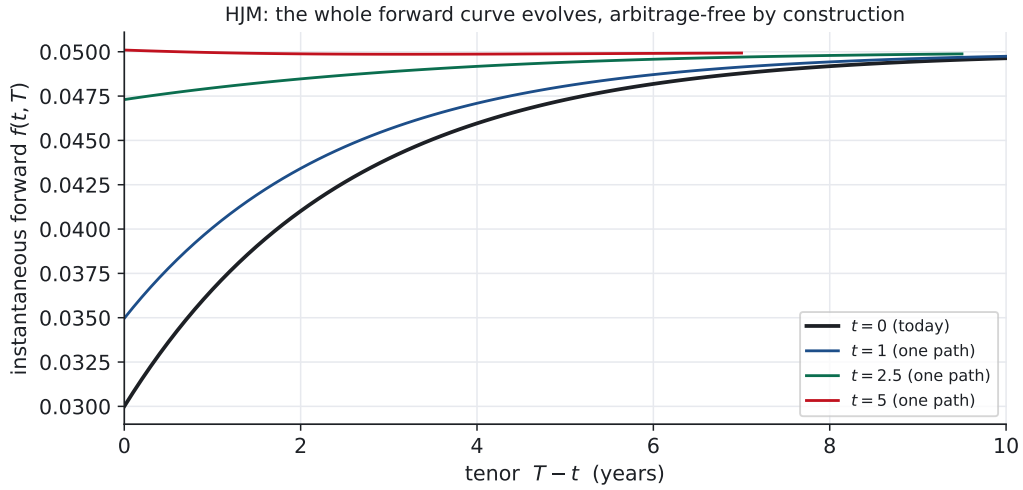

Figure 20: An HJM forward curve evolving under the no-arbitrage drift (60), one simulated path sampled at four calendar times (plotted against tenor $T-t$). The whole curve moves at once—HJM's defining feature—yet every snapshot is arbitrage-free by construction, because the drift is not a free parameter but is forced by the volatility we chose.

Why HJM subsumes the short-rate models. Every short-rate model is a special case of HJM, recovered by reading off the implied forward-rate volatility $\sigma(t,T)$. HJM is the general theory; Vasicek and Hull–White are particular volatility choices within it. The natural question is therefore: *which* choices give back the fast, one-factor, Markovian models—and that is precisely the Markov question of Part IV, now with real stakes.

18 When HJM is Markovian—and when it is not

The forward curve (59) is an infinite-dimensional state. Generically, to evolve it you must carry *the entire curve* forward through time—every maturity, updated together—and the model is *non-Markovian* in any finite set of variables, exactly the situation of §15. This is the default for HJM, and it is why a naive HJM model is simulated by Monte Carlo over the whole curve rather than solved on a small PDE grid.

The collapse: separable (exponential) volatility. But for special volatility structures the infinite-dimensional curve turns out to be a deterministic function of just a *few* state variables, and Markovianity returns—the §14 state-augmentation story, now running in reverse, *shrinking* an infinite state to a finite one. The cleanest case is *exponential* (separable) volatility,

$$\sigma(t,T) = \sigma\, e^{-a(T-t)}, \tag{61}$$

a volatility that decays exponentially with tenor $T-t$. The collapse is not a coincidence; it is one line of algebra. Integrate the forward-rate SDE (59) from 0 to t:

$$f(t,T) = f(0,T) + \underbrace{\int_0^t \alpha(s,T)\,\mathrm{d}s}_{\text{deterministic} = D(t,T)} + \int_0^t \sigma(s,T)\,\mathrm{d}W_s. \tag{62}$$

The drift integral is a fixed function of (t,T)—call it $D(t,T)$. Everything turns on the *stochastic* integral, and here the exponential does its trick. Because the tenor splits as $T-s = (T-t)+(t-s)$, the exponential factorises, $e^{-a(T-s)} = e^{-a(T-t)}\,e^{-a(t-s)}$, and the maturity-dependent piece pulls straight outside the integral:

$$\int_0^t \sigma\, e^{-a(T-s)}\,\mathrm{d}W_s = e^{-a(T-t)}\,\underbrace{\sigma\int_0^t e^{-a(t-s)}\,\mathrm{d}W_s}_{=X_t} = e^{-a(T-t)}\,X_t. \tag{63}$$

The quantity X_t is a *single* scalar—and it is one we already solved: it is exactly the Itô integral in the OU solution (28), so it follows the mean-reverting SDE $\mathrm{d}X_t = -a\,X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$. Substituting (63) back into (62),

$$f(t,T) = f(0,T) + D(t,T) + e^{-a(T-t)}\,X_t, \tag{64}$$

in which the maturity T enters the random part *only* through the deterministic factor $e^{-a(T-t)}$. The entire curve hangs off one number X_t. Knowing X_t you can reconstruct every maturity; the infinite-dimensional model has collapsed to one Markov factor—and that factor's short rate $r_t = f(t,t)$ is exactly the Hull–White process of §16. *Exponential HJM is Hull–White.* Figure 21 demonstrates the collapse numerically: a full HJM simulation that updates every maturity independently lands exactly on the curve rebuilt from the single state X_t via (64).

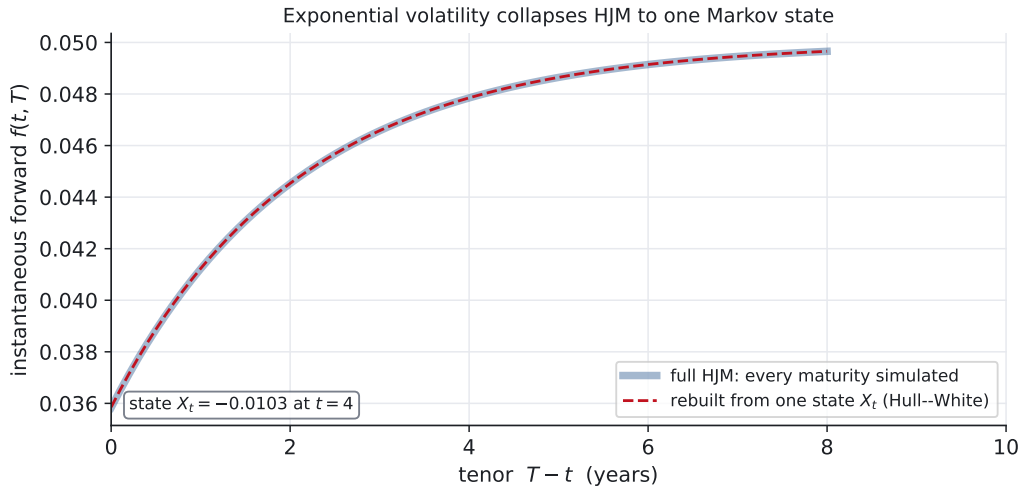

Figure 21: Exponential volatility collapses HJM to one Markov state. The thick band is a full HJM simulation, updating every maturity on the curve independently under the common shocks. The dashed line is the same curve *reconstructed from a single scalar* X_t via (64). They coincide: under (61) the whole forward curve is a function of one number, the model is Markovian in X_t, and it *is* Hull–White. This is the precise sense in which the fast one-factor model is a special case of the general framework.

The general rule, and the cost. More generally, HJM is Markovian in a *finite* state exactly when the volatility $\sigma(t,T)$ is a sum of *separable* terms—products of a function of t and a function of T, of which the exponential (61) is the simplest. Each separable factor adds one state variable; a few of them give the multi-factor Gaussian and quadratic-Gaussian models (e.g. two factors for level and slope). Any volatility structure *outside* this separable family leaves you with the full infinite-dimensional, path-dependent curve—realistic, flexible, but expensive, with no PDE or recombining tree to lean on. This is the Markov dividing line of §15 in its most consequential setting: *the shape of the volatility function alone decides whether your interest-rate model is a fast one-factor PDE or a slow infinite-dimensional Monte Carlo.*

19 How this underwrites the rest of the series

Every tool in this note is load-bearing somewhere in the companion notes. A short map, by way of conclusion:

- **Itô's lemma and GBM** (§4–§6.1) are the ground floor of the local-volatility and Black–Scholes material—Dupire's equation is Itô's lemma applied to an option's value, read backwards to extract the local volatility.
- **Girsanov and the change of measure** (§9) underlie every appearance of the risk-neutral measure, and the forward measure (§11) is the natural setting for the rate notes.
- **Feynman–Kac** (§10) is why the local-vol and Hull–White notes can price on PDE grids while the LSV note prices by Monte Carlo—two faces of one expectation.
- **The Ornstein–Uhlenbeck process** (§6.2) is the short rate in Vasicek and Hull–White, the mean-reverting spread in the time-series and pairs-trading notes, and the hidden state tracked by the Kalman filter.
- **The Markov question** (Part IV) is the hinge of the LSV note (state = spot × variance), the stochastic-volatility literature (augmentation), and the choice of HJM volatility (§18).

- **Change of numéraire** (§11, §12) is what makes the Black–Scholes and Margrabe formulas drop out as two-measure exercise probabilities, and what the forward-measure pricing of the rate notes runs on.
- **Optimal stopping** (§13) is the American-option and early-exercise machinery—the free-boundary PDE behind tree pricers and the regression-Monte-Carlo (Longstaff–Schwartz) approach of the least-squares note.
- **Martingale representation** (§8) is the silent guarantor that the hedges in all of them exist.

A candid closing word on limits. Stochastic calculus as presented here is the calculus of *continuous* paths—no jumps, finite variance, Brownian noise. Real markets gap, crash, and occasionally exhibit the heavy tails and memory that the rough-volatility section gestured at. The continuous theory is not the truth; it is an extraordinarily productive approximation, and knowing where it is thin—jumps, liquidity, parameter uncertainty, the gap between $\mathbb{P}$ and $\mathbb{Q}$—is part of using it well. But the single idea it is built on, that randomness accumulated over time has a size of its own, $(\mathrm{d}W)^2 = \mathrm{d}t$, is as close to a load-bearing truth as this corner of mathematics has. Everything else in this note, and a great deal of modern finance, is that one fact, carefully unpacked.

Appendices

A The quadratic variation limit, a little more carefully

The statement $(\mathrm{d}W)^2 = \mathrm{d}t$ is shorthand for a limit. Partition $[0,t]$ into n equal steps and let $Q_n = \sum_{i=1}^{n} (\Delta W_i)^2$. Each $\Delta W_i \sim \mathcal{N}(0, t/n)$ independently, so $\mathbb{E}[(\Delta W_i)^2] = t/n$ and, using $\mathrm{Var}(Z^2) = 2(\mathbb{E}Z^2)^2$ for a Gaussian, $\mathrm{Var}\left((\Delta W_i)^2\right) = 2(t/n)^2$. Summing the n independent terms,

$$\mathbb{E}[Q_n] = n \cdot \frac{t}{n} = t, \qquad \mathrm{Var}(Q_n) = n \cdot 2\Big(\frac{t}{n}\Big)^2 = \frac{2t^2}{n} \xrightarrow{n\to\infty} 0. \tag{65}$$

So $Q_n \to t$ in mean square: the quadratic variation is not just *on average* the elapsed time, it concentrates on it, its fluctuations vanishing like $1/n$. That mean-square convergence is the rigorous content of (2), and it is why we may treat $(\mathrm{d}W)^2$ as the *deterministic* quantity $\mathrm{d}t$ inside Itô's lemma rather than as a random thing.

B The Itô isometry

For a simple (piecewise-constant, non-anticipating) integrand g, write the integral as $\sum_i g_{t_i} \Delta W_i$. Squaring,

$$\mathbb{E}\Big[\big(\textstyle\sum_i g_{t_i} \Delta W_i\big)^2\Big] = \textstyle\sum_{i,j} \mathbb{E}\big[g_{t_i} g_{t_j}\, \Delta W_i\, \Delta W_j\big]. \tag{66}$$

For $i \neq j$ the later increment is independent of everything else in its term and has mean zero, so those cross terms vanish. The diagonal $i = j$ terms use independence of g_{t_i} (non-anticipating) from ΔW_i and $\mathbb{E}[(\Delta W_i)^2] = \Delta t$:

$$\mathbb{E}\Big[\big(\textstyle\sum_i g_{t_i} \Delta W_i\big)^2\Big] = \textstyle\sum_i \mathbb{E}[g_{t_i}^2]\, \Delta t \xrightarrow{\text{mesh}\to 0} \int_0^T \mathbb{E}[g_t^2]\, \mathrm{d}t, \tag{67}$$

which is the isometry (8). Extending from simple integrands to general ones is the standard density argument that also *defines* the Itô integral as an L^2 limit.

C Ornstein–Uhlenbeck moments

From the solution (28), the only random piece is the Itô integral $I_t = \sigma \int_0^t e^{-\theta(t-s)}\, \mathrm{d}W_s$, whose integrand is deterministic. Its mean is zero, and by the Itô isometry (8),

$$\mathrm{Var}(x_t) = \mathbb{E}[I_t^2] = \sigma^2 \int_0^t e^{-2\theta(t-s)}\, \mathrm{d}s = \sigma^2\, \frac{1 - e^{-2\theta t}}{2\theta}, \tag{68}$$

which is (29); as $t \to \infty$ it tends to $\sigma^2/2\theta$, the stationary variance (30). For the autocovariance at lag $\tau > 0$ in the stationary regime, use $x_{t+\tau} = \mu + (x_t - \mu)e^{-\theta\tau} + \sigma \int_t^{t+\tau} e^{-\theta(t+\tau-s)} \mathrm{d}W_s$; the new integral is independent of x_t, so

$$\mathrm{Cov}(x_t, x_{t+\tau}) = e^{-\theta\tau}\ \mathrm{Var}(x_t) \xrightarrow{\text{stationary}} \frac{\sigma^2}{2\theta}\, e^{-\theta\tau}, \tag{69}$$

the exponential decay (31).

D The HJM drift condition

Write the log bond price as $\log P(t,T) = -\int_t^T f(t,u)\,\mathrm{d}u$ and abbreviate $\Sigma(t,T) = \int_t^T \sigma(t,u)\,\mathrm{d}u$ and $A(t,T) = \int_t^T \alpha(t,u)\,\mathrm{d}u$. Differentiating $\log P$ in t and substituting the forward dynamics (59) (the $f(t,t) = r_t$ boundary term supplies the short rate) gives

$$\mathrm{d}\log P(t,T) = \Big(r_t - A(t,T)\Big)\mathrm{d}t - \Sigma(t,T)\,\mathrm{d}W_t. \tag{70}$$

By Itô's lemma the bond itself, $P = e^{\log P}$, then has drift equal to its log-drift *plus* the half-variance Itô correction $\frac{1}{2}\Sigma(t,T)^2$:

$$\frac{\mathrm{d}P(t,T)}{P(t,T)} = \Big(r_t - A(t,T) + \tfrac{1}{2}\Sigma(t,T)^2\Big)\mathrm{d}t - \Sigma(t,T)\,\mathrm{d}W_t. \tag{71}$$

No-arbitrage (§8, §11) requires the bond to earn the short rate r_t under $\mathbb{Q}$, i.e. the bracketed drift must equal r_t. Hence $A(t,T) = \frac{1}{2}\Sigma(t,T)^2$. Differentiating in T (using $\partial_T A = \alpha$ and $\partial_T \Sigma = \sigma$) gives

$$\alpha(t,T) = \Sigma(t,T)\,\sigma(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)\,\mathrm{d}u, \tag{72}$$

the drift condition (60). The correction $\frac{1}{2}\Sigma^2$—the quadratic variation of the log bond—is the entire source of the constraint: kill the Itô term and the drift would be free, and the model would admit arbitrage.

References

[1] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed., Springer, 2003. The standard gentle-but-rigorous entry point; Chapters 3–5 cover the Itô integral and lemma.

[2] S. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*, Springer, 2004. The finance-facing reference for everything in Parts I–II, including Girsanov, Feynman–Kac and numéraires.

[3] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer, 1991. The rigorous treatment, for when the hand-waving here needs underpinning.

[4] D. Heath, R. Jarrow and A. Morton, *Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation*, Econometrica **60**(1), 1992, 77–105. The original HJM paper and the drift condition of §17.

[5] D. Brigo and F. Mercurio, *Interest Rate Models—Theory and Practice*, 2nd ed., Springer, 2006. The encyclopaedic reference for the short-rate and HJM material, including the Markovian-collapse conditions of §18.

[6] J. Gatheral, T. Jaisson and M. Rosenbaum, *Volatility is Rough*, Quantitative Finance **18**(6), 2018, 933–949. The empirical case for $H \approx 0.1$ and the rough-volatility programme of §15.

[7] W. Margrabe, *The Value of an Option to Exchange One Asset for Another*, Journal of Finance **33**(1), 1978, 177–186. The exchange-option formula of §12.2 and the change-of-numéraire argument.

[8] F. Longstaff and E. Schwartz, *Valuing American Options by Simulation: A Simple Least-Squares Approach*, Review of Financial Studies **14**(1), 2001, 113–147. The regression-Monte-Carlo method for the optimal-stopping problem of §13.