








Modelling Financial Markets

Chapter 11: Continuous-Time Models



Daniel Traian PELE

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

Can One Model Describe Prices at Any Time Scale?

> *"It is shown that common-stock prices, and the value of money can be regarded as an ensemble of decisions in statistical equilibrium, with properties quite analogous to an ensemble of particles in statistical mechanics."*
>
> — *Osborne (1959)*

- Log prices behave like Brownian motion
  - geometric Brownian motion (GBM) is the price model behind Black & Scholes (1973)
- Daily S&P 500 (Standard & Poor's) returns, 1990–2026, reject GBM
  - the excess kurtosis is 10.9 instead of 0
  - the ACF (AutoCorrelation Function) of $|r_t|$ is 0.27 at lag 1 and still 0.14 at lag 50, instead of 0
- Merton reproduces the tail, Heston reproduces volatility clustering
  - neither reaches the excess kurtosis of the data: 3.66 (Merton) and 3.21 (Heston), against 10.88

▸ Central chart: Which Model Reproduces the S&P 500?

How do we describe prices in continuous time?

What the models explain

- Can one model describe prices at **every** time scale at once?
  - prices are recorded daily or every few milliseconds
- Part I: Brownian motion, Itô calculus, simulation, geometric Brownian motion (GBM), mean reversion
- Part II: estimating diffusions, change of measure, jumps, stochastic volatility
  - then models against data, two case studies, AI-assisted (Artificial Intelligence) discovery

What you should be able to do

- derive solutions with Itô's lemma and simulate a stochastic differential equation with a known error order
- compute GBM, Ornstein–Uhlenbeck, Merton and Heston estimates on data
  - with standard errors that respect what the sample identifies
- interpret which feature of the data each model reproduces, and which feature none of them does
- interpret why option pricing (Chapter 12) and interest-rate models use this language
- Further Topics and Appendix: reading for familiarity

Indices and Data Sources

- **VIX**: Cboe Volatility Index.
- **BET**: Bucharest Exchange Trading (index).
- **BVB**: Bursa de Valori București — the Bucharest Stock Exchange.
- **CBOE**: Chicago Board Options Exchange.
- **FRED**: Federal Reserve Economic Data.
- **DTB3**: 3-Month Treasury Bill: secondary market rate (FRED series code).

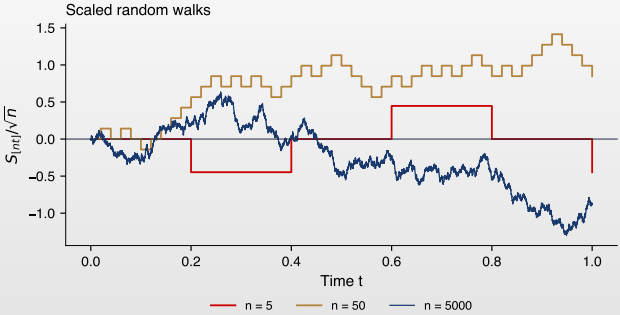
- Detailed examples and advanced topics are in Further Topics, after the Conclusions
  - the footer button leads there; the Back button returns to the starting slide

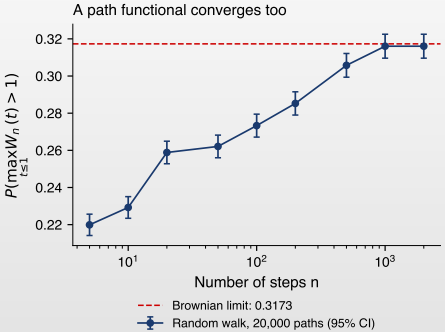
From Random Walk to Brownian Motion: Donsker's Theorem (1/3)

- **Question**: which process do we get as price steps become smaller and more frequent?
  - prices move in discrete steps
- **Random walk**: $S_0 = 0$, $S_k = S_{k-1} + \xi_k = \xi_1 + \cdots + \xi_k$
  - steps ξ_i i.i.d. (independent and identically distributed), mean 0 and variance 1
  - it is the log-price model of Chapter 2
  - ξ_k is the standardised return of day k
- **Scaled random walk**: $W_n(t) = S_{\lfloor nt \rfloor}/\sqrt{n}$, with n steps per unit of time
  - $\lfloor nt \rfloor$: integer part of nt
  - why $\sqrt{n}$: $\mathrm{Var}(W_n(t)) = \lfloor nt \rfloor / n \to t$
  - divided by n the limit would be 0
  - undivided the variance would explode
- Central limit theorem (CLT): $W_n(t) \to N(0, t)$ for each t separately
- Donsker's theorem: $W_n \Rightarrow W$ for the **whole path** Billingsley (1999)
  - $\Rightarrow$: convergence in distribution
  - holds whatever the law of the steps

From Random Walk to Brownian Motion: Donsker's Theorem (2/3)

CI: Confidence Interval.





MFM_ch11_donsker ▶ Interactive

From Random Walk to Brownian Motion: Donsker's Theorem (3/3)

- Left: three scaled random walks on $[0, 1]$ with $n = 5$, 50 and 5 000 steps
  - the vertical axis is $S_{\lfloor nt \rfloor} / \sqrt{n}$
  - for large n the step path becomes continuous and irregular, like a Brownian path
- Right: $P(\max_{t \le 1} W_n(t) > 1)$ against n on a log axis
  - estimated from 20 000 paths (95% CI)
  - the probability rises from 0.220 ($n = 5$) to 0.316 ($n = 2000$)
  - the limit is the reflection-principle value $2(1 - \Phi(1)) = 0.3173$
  - Φ: standard Normal distribution function
- A coarse grid misses the maximum reached between observations
- Convergence is slow, of order $1/\sqrt{n}$
  - daily closes understate the intraday range

MFM_ch11_donsker

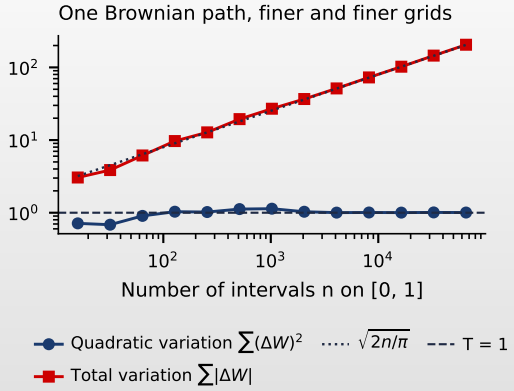
Brownian Motion: Definition and Properties

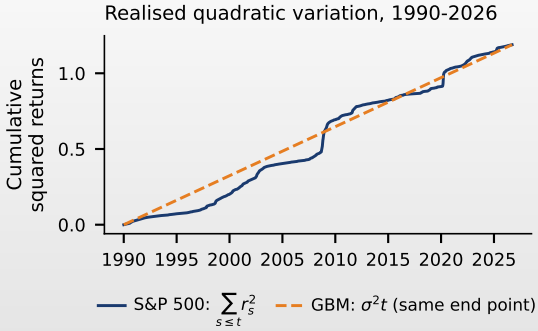
- **Standard Brownian motion**: four defining properties
  - $W_0 = 0$
  - independent increments
  - $W_t - W_s \sim N(0, t-s)$ for $s < t$
  - continuous paths
  - existence: Wiener (1923)
  - simulation on a grid: $W_{t+\Delta t} = W_t + \sqrt{\Delta t}\, Z$, $Z \sim N(0,1)$
- A Gaussian process with $\mathrm{Cov}(W_s, W_t) = \min(s,t)$
- A **martingale**: $\mathbb{E}[W_t \mid \mathcal{F}_s] = W_s$
  - $\mathcal{F}_s$: the information up to s
  - the efficient-market random walk of Chapter 2
- **Self-similarity**: $\{c^{-1/2} W_{ct}\}$ is again a standard Brownian motion for any $c > 0$
- Paths are continuous but **nowhere differentiable**
  - $(W_{t+h} - W_t)/h$ has standard deviation $1/\sqrt{h} \to \infty$
  - ordinary calculus fails

Total and Quadratic Variation

- **Question**: why does ordinary calculus fail on a Brownian path, and what replaces it?
- **Total variation** $\sum_k |\Delta W_k|$
  - grid $0 = t_0 < \cdots < t_n = T$, step $\Delta t = T/n$, $\Delta W_k = W_{t_k} - W_{t_{k-1}}$
  - its mean is $\sqrt{2nT/\pi} \to \infty$
  - a Brownian path has infinite length
- **Quadratic variation** $[W]_T = \lim \sum_k (\Delta W_k)^2 = T$
  - a random sum with a **deterministic** limit
  - $Q_n = \sum_{k=1}^{n} (\Delta W_k)^2$, $\mathrm{Var}(Q_n) = 2T^2/n$
  - one year of daily data gives about 9% relative error ($\sqrt{2/252}$)
  - the idea behind realised variance (Chapter 9)
- Shorthand of Itô calculus: $(\mathrm{d}W_t)^2 = \mathrm{d}t$

Quadratic Variation: Simulation and Data (1/2)





MFM_ch11_quadratic_variation ▶ Interactive

Quadratic Variation: Simulation and Data (2/2)

- Left: one Brownian path on grids with 16 to 65,536 intervals
  - the quadratic variation settles at 1.006 ($T = 1$)
  - the total variation keeps growing: 205 against $\sqrt{2n/\pi} = 204$
- Right: cumulative $\sum r_t^2$ of the S&P 500, 1.188 over 36.7 years
  - average volatility 18.0%
  - under geometric Brownian motion it would follow the straight line $\sigma^2 t$
- The sum rises in steps
  - September 2008 – March 2009: 1.6% of the days, but 15.2% of the quadratic variation
  - February – April 2020: 0.5% of the days, but 7.8% of the quadratic variation
  - volatility is itself random (Heston, Section 10)

MFM_ch11_quadratic_variation

The Itô Integral (1/2)

- Goal: a meaning for $\int_0^T H_t \, \mathrm{d}W_t$
  - the gain of holding H_t units of an asset priced W
  - Riemann–Stieltjes integration fails because the total variation is infinite
- **Itô integral**: the limit in mean square of sums with the integrand at the **left** end point
  - $\int_0^T H_t \, \mathrm{d}W_t = \lim_{n\to\infty} \sum_{k=1}^{n} H_{t_{k-1}} \left(W_{t_k} - W_{t_{k-1}}\right)$
  - for H **adapted** with $\mathbb{E}\left[\int_0^T H_t^2 \, \mathrm{d}t\right] < \infty$
  - H_t uses only information up to t
  - the position is chosen before the next price move

The Itô Integral (2/2)

- Zero mean: $\mathbb{E}\left[\int_0^T H_t \, \mathrm{d}W_t\right] = 0$
- **Itô isometry**: $\mathbb{E}\left[\left(\int_0^T H_t \, \mathrm{d}W_t\right)^2\right] = \mathbb{E}\left[\int_0^T H_t^2 \, \mathrm{d}t\right]$
- $t \mapsto \int_0^t H_s \, \mathrm{d}W_s$ is a martingale
  - no strategy earns money on average
  - weaker conditions: Appendix
- **Stratonovich** integral: the integrand averaged over the end points, $\frac{1}{2}(H_{t_{k-1}} + H_{t_k})$
  - it gives a different number
  - finance uses Itô: the averaged integrand would use the price at t_k, unknown when the position is chosen

Itô Processes and Itô's Lemma

- **Itô process**: $\mathrm{d}X_t = a_t\,\mathrm{d}t + b_t\,\mathrm{d}W_t$
  - drift a_t, diffusion coefficient b_t, both adapted
  - shorthand for $X_t = X_0 + \int_0^t a_s\,\mathrm{d}s + \int_0^t b_s\,\mathrm{d}W_s$
- **Itô's lemma**: for $f(t,x)$ continuously differentiable once in t and twice in x Itô (1944)
  - $\mathrm{d}f(t,X_t) = \left(f_t + a_t f_x + \frac{1}{2}b_t^2 f_{xx}\right)\mathrm{d}t + b_t f_x\,\mathrm{d}W_t$
  - partial derivatives evaluated at (t, X_t)
  - $f_t = \partial f/\partial t$, $f_x = \partial f/\partial x$, $f_{xx} = \partial^2 f/\partial x^2$
- Rule of thumb: Taylor expansion to second order
  - with $\mathrm{d}t \cdot \mathrm{d}t = 0$, $\mathrm{d}t \cdot \mathrm{d}W = 0$, $\mathrm{d}W \cdot \mathrm{d}W = \mathrm{d}t$
  - the second-order term $\frac{1}{2}b_t^2 f_{xx}\,\mathrm{d}t$ does not vanish
  - this is the whole novelty

Stochastic Differential Equations and the Euler–Maruyama Scheme

- **Question**: how do we simulate a model whose transition (the law of $X_{t+\Delta t}$ given X_t) is not easy to simulate, such as Heston?
  - exact simulation exists, but is costly: Broadie & Kaya (2006)
- **Question**: how large is the discretisation error?
- A **stochastic differential equation** (SDE): $\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}t + b(t, X_t)\,\mathrm{d}W_t$, X_0 given
  - drift and diffusion depend on the current state
  - GBM, Ornstein–Uhlenbeck and Merton have exact transitions
  - we simulate them with no discretisation error
  - Heston has only a closed-form characteristic function
  - Existence conditions: ▸ Appendix
- **Euler–Maruyama**: freeze a and b over each step Higham (2001)
  - $X_{k+1} = X_k + a(X_k)\,\Delta t + b(X_k)\,\Delta W_k$, $\Delta W_k = \sqrt{\Delta t}\,Z_k$
  - X_k is the approximation at $t_k = k\Delta t$
  - Z_k i.i.d. $N(0,1)$

The Milstein Scheme and Strong Convergence

- **Milstein**: add the next Itô–Taylor term Milstein (1975)
  - $X_{k+1} = X_k + a(X_k)\,\Delta t + b(X_k)\,\Delta W_k + \frac{1}{2} b(X_k)\,b'(X_k)\left[(\Delta W_k)^2 - \Delta t\right]$
  - $b' = \mathrm{d}b/\mathrm{d}x$
  - the correction has conditional mean zero
  - it raises the strong order (path accuracy), not the weak order (accuracy of means)
- **Strong error**: $\mathbb{E}|X_T - X_T^{\Delta t}| \le C\,\Delta t^{\gamma}$
  - how close each simulated path is to the true one
  - X_T is the exact solution, $X_T^{\Delta t}$ the scheme with step Δt
  - C is a constant and γ the strong order
  - $\gamma = 1/2$ for Euler–Maruyama and $\gamma = 1$ for Milstein
  - it matters for path-dependent quantities (hedging errors, barrier crossings)

Geometric Brownian Motion (GBM) from Itô's Lemma (1/2)

- **Question**: which price model stays positive?
  - it should give percentage returns with the same distribution at every price level
- SDE: $\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t$
  - the **percentage** change has constant drift μ and volatility σ per year
  - proposed by Osborne (1959) and Samuelson (1965)
  - the price model behind Black & Scholes (1973)
- Itô's lemma for $f(x) = \ln x$
  - $f_x = 1/x$, $f_{xx} = -1/x^2$, $a_t = \mu S_t$, $b_t = \sigma S_t$
  - $\mathrm{d} \ln S_t = \left(\mu - \frac{1}{2}\sigma^2\right) \mathrm{d}t + \sigma \,\mathrm{d}W_t$
  - so $S_t = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right)$
  - prices stay positive
  - exact simulation needs only Normal draws

Geometric Brownian Motion (GBM) from Itô's Lemma (2/2)

- The log price grows at the log drift $m = \mu - \sigma^2/2$, not at μ
  - volatility lowers the growth of the median investor's wealth
  - With $\sigma = 18\%$, the cost is $\sigma^2/2 = 1.6$ percentage points a year
  - With $\sigma = 67\%$ (Bitcoin), the cost reaches 22 percentage points a year
- Testable implications: $r_t = \ln(S_t/S_{t-\Delta t})$ i.i.d. $N(m\Delta t, \sigma^2\Delta t)$ at every frequency
  - Skewness and excess kurtosis are zero
  - There is no autocorrelation in r_t, $|r_t|$ or r_t^2
  - The variance is proportional to the horizon

Estimating GBM by Maximum Likelihood (1/2)

- Maximum likelihood estimation (MLE) from n log returns at step Δt, over $T = n\Delta t$ years
  - $\hat{\sigma}^2 = \frac{1}{n\Delta t}\sum_t (r_t - \bar{r})^2$, $\hat{m} = \bar{r}/\Delta t$, $\hat{\mu} = \hat{m} + \hat{\sigma}^2/2$ ($\bar{r}$: mean log return)
  - under GBM, the standard error (SE) $\mathrm{SE}(\hat{\sigma}) = \sigma/\sqrt{2n}$ shrinks with the number of observations
  - $\mathrm{SE}(\hat{m}) = \sigma/\sqrt{T}$ depends only on the **calendar span**
  - reason: $\sum_t r_t = \ln(S_T/S_0)$ uses only the first and the last price

	Years	$\hat{\sigma}$ (%)	$\mathrm{SE}(\hat{\sigma})$	$\hat{m}$ (%)	$\mathrm{SE}(\hat{m})$
S&P 500	36.7	18.0	0.13	8.3	3.0
BET	26.5	22.2	0.19	16.1	4.3
Bitcoin	12.0	66.6	0.71	43.1	19.2

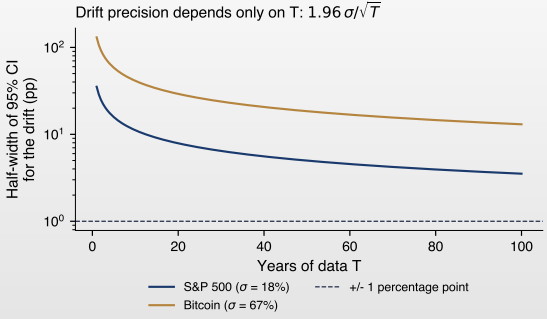
- The SE is 0.1–0.7 percentage points for volatility, but 3–19 points for the log drift
- S&P 500: $\hat{\mu} = 8.3\% + 1.62\% = 10.0\%$ per year

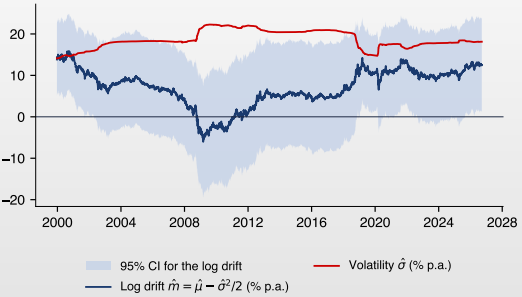
Estimating GBM by Maximum Likelihood (2/2)

- These SE are conditional on GBM
  - with excess kurtosis k, even independent returns inflate $\mathrm{SE}(\hat{\sigma})$ by $\sqrt{(k+2)/2} = 2.5$ (S&P 500)
  - block-bootstrap SE: Seminar 11, B1
  - the drift SE needs only a finite variance
  - the volatility SE requires a finite fourth moment, which Chapter 1 questions

How Precisely Can We Estimate the Drift? (1/2)

p.p. = percentage points.





MFM_ch11_drift_precision ▶ Interactive

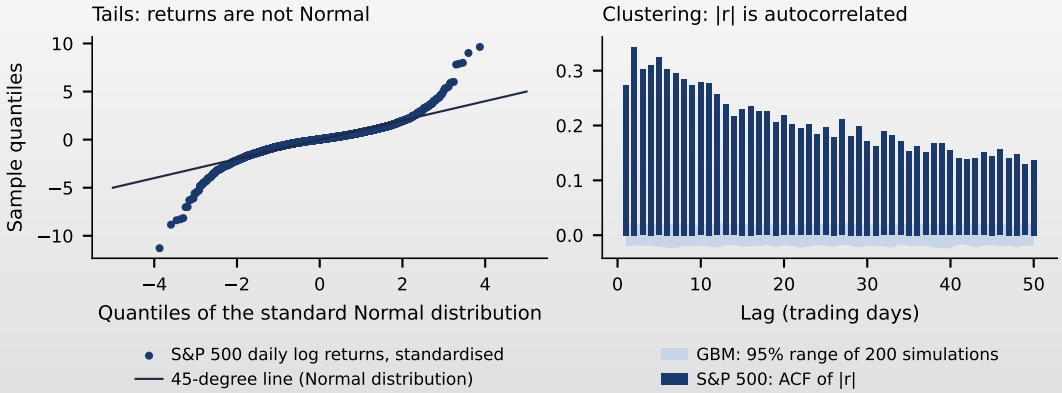
How Precisely Can We Estimate the Drift? (2/2)

- Left: half-width of the 95% CI for the drift against years of data
- A precision of ± 1 point needs $(1.96\,\hat{\sigma}/0.01)^2$ years of data
  - S&P 500: $(1.96\,\hat{\sigma}/0.01)^2 = 1\,243$ years
  - 17 028 years for Bitcoin
- Right: S&P 500 on rolling 10-year windows
  - the log drift ranges from -6.0% (window ending 2009-03) to 15.8% (2000-09)
  - each estimate has a CI of about ± 11 points
  - with this precision, the swings cannot be safely attributed to changes in the drift
  - volatility stays between 14.1% and 22.3%: precise, but not constant
- Option pricing in Chapter 12 avoids μ
  - under the risk-neutral measure only σ enters (Girsanov, Section 8)

MFM_ch11_drift_precision

Does GBM Fit the S&P 500? (1/2)

ACF: AutoCorrelation Function.



MFM_ch11_gbm_vs_data ▶ Interactive

Does GBM Fit the S&P 500? (2/2)

- Left: quantile–quantile (QQ) plot of standardised daily log returns against the Normal distribution
  - excess kurtosis 10.9 instead of 0
  - the sample kurtosis consistently estimates a finite kurtosis only with a finite fourth moment, which Chapter 1 questions
  - inference on the kurtosis usually requires a finite eighth moment
  - worst day at -11.3 standard deviations
  - 62 days beyond ± 4 standard deviations, against 0.59 expected under the Normal distribution
  - Jarque–Bera (JB) statistic: 45 805 against the 5% critical value 5.99 (χ^2_2)
  - JB tests zero skewness and zero excess kurtosis jointly
- Right: autocorrelation function (ACF) of $|r_t|$
  - 0.27 at lag 1 and still 0.14 at lag 50
  - the 95% band from 200 GBM samples of the same length stays below 0.024 in absolute value
- GBM fails on both stylised facts (fat tails, volatility clustering) of Chapter 1 Cont (2001)
  - the rest of the chapter adds mean reversion, jumps and random volatility
- Type of evidence: description (it shows what happened in the data, without testing a cause)
  - the data reject two implications of GBM without identifying the mechanism that produces them

MFM_ch11_gbm_vs_data

The Ornstein–Uhlenbeck (OU) Process

- **Question**: how do we model a variable that returns to a normal level?
  - examples: an interest rate or the VIX
  - unlike a stock price, it does not drift away
- SDE: $\mathrm{d}x_t = \kappa(\theta - x_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, $\kappa > 0$ Uhlenbeck & Ornstein (1930)
  - θ: long-run mean; κ: speed of mean reversion; σ: volatility
  - Above θ the drift is negative, below θ positive
  - the drift is a force pulling x back to θ
  - Used for interest rates Vasicek (1977), volatility, spreads and commodity prices
- **Half-life** $\ln 2/\kappa$: the time for the expected deviation from θ to halve
  - With $\kappa = 0.5$ per year, the half-life is 1.39 years

Exact Discretisation: an AR(1)

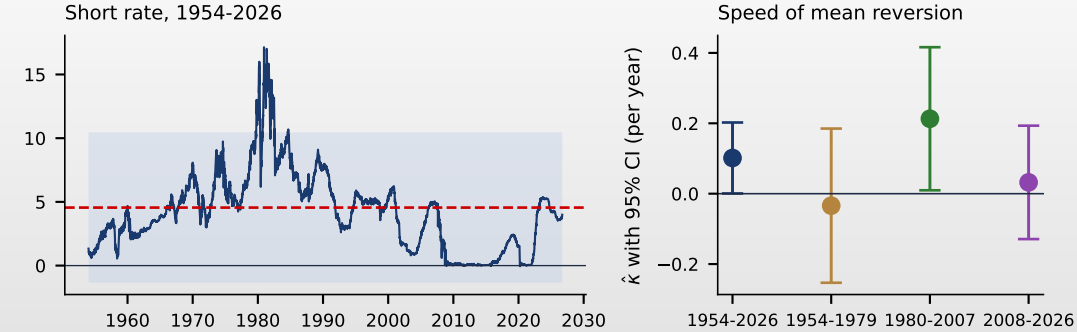
- Sampled at step Δt, the Ornstein–Uhlenbeck (OU) process is exactly an autoregressive model of order 1 (AR(1))
  - $x_{t+\Delta t} = a + b\,x_t + \varepsilon_t$, $b = e^{-\kappa \Delta t}$, $a = \theta(1-b)$, $\mathrm{Var}(\varepsilon_t) = \frac{\sigma^2}{2\kappa}(1-b^2)$
- Maximum likelihood = ordinary least squares (OLS) of $x_{t+\Delta t}$ on x_t
  - OLS gives the intercept $\hat{a}$, the slope $\hat{b}$ and the residual standard deviation $\hat{s}_\varepsilon$
  - then invert the mapping: $\hat{\kappa} = -\ln \hat{b}/\Delta t$, $\hat{\theta} = \hat{a}/(1-\hat{b})$, $\hat{\sigma} = \hat{s}_\varepsilon \sqrt{2\hat{\kappa}/(1-\hat{b}^2)}$
- Standard errors of $\hat{\kappa}$ and $\hat{\theta}$: the delta method (linearising the transformation), from the OLS covariance of $(\hat{a}, \hat{b})$
  - $\hat{\kappa}$ depends on $\hat{b}$ only: $\partial\kappa/\partial b = -1/(b\Delta t)$
  - so $\mathrm{Var}(\hat{b})$ suffices (Seminar 11, A1)
- With daily data b is very close to 1
  - κ is estimated imprecisely, as in unit-root tests

Short-Rate Models: Vasicek and CIR

- **Vasicek**: the short rate r_t is an OU process Vasicek (1977)
  - Normal distribution: negative rates have positive probability
  - the long-run probability is $\Phi(-\theta\sqrt{2\kappa}/\sigma)$
  - Bond prices and yields in closed form (Feynman–Kac, Section 8)
- **Cox–Ingersoll–Ross** (CIR): $\mathrm{d}r_t = \kappa(\theta - r_t)\,\mathrm{d}t + \sigma\sqrt{r_t}\,\mathrm{d}W_t$ Cox, Ingersoll & Ross (1985)
  - Volatility shrinks near zero
  - The rate stays positive if $2\kappa\theta \geq \sigma^2$ (Feller condition) Feller (1951)
- Which is closer to the data is an empirical question Chan, Karolyi, Longstaff & Sanders (1992)
  - tested in Section 7

Vasicek on the 3-Month Treasury Bill Rate (1/2)

s.d. = standard deviation.



MFM_ch11_vasicek ▶ Interactive

Vasicek on the 3-Month Treasury Bill Rate (2/2)

- Left: daily 3-month United States (US) Treasury bill rate (DTB3), 1954–2026
  - with the Vasicek long-run mean $\hat{\theta} = 4.55\%$
  - band: $\hat{\theta}$ plus/minus two stationary standard deviations (s.d.)
  - $\hat{\kappa} = 0.101$ (SE 0.051), a half-life of 6.8 years
  - volatility $\hat{\sigma} = 1.33\%$
  - Daily AR(1) coefficient $\hat{b} = 0.99960$: almost a unit root
- Right: $\hat{\kappa}$ with 95% CI over three periods
  - -0.034 in 1954–1979 (explosive), 0.213 in 1980–2007 and 0.032 in 2008–2026
  - Two of the three CIs include zero
- Test of the random-walk null ($\kappa = 0$): parametric bootstrap
  - 500 random walks of the same length
  - reason: under $\kappa = 0$ the distribution of $\hat{\kappa}$ is not Normal and is shifted upwards
  - p-value $= 0.287$: the sample does not establish mean reversion (Seminar 11, B2)
- Type of evidence: description

MFM_ch11_vasicek

Infill Asymptotics: Effect of the Sampling Frequency

Question: which parameters do we learn from sampling more often, and which only from more years?

Infill: $\Delta \to 0$ on a fixed $[0, T]$

- $\sum_k (\Delta X_k)^2 \to \int_0^T b_t^2 \, \mathrm{d}t$
- Quadratic variation, hence b_t^2 along the path (under regularity): identified without error in the limit
  - if the Itô process is continuous and observed without measurement noise
- The whole diffusion function or its parameters need further conditions
- Sampling more often helps σ and $b(x)$
  - σ: Section 5; $b(x)$: nonparametric, Further Topics

Long-Span Asymptotics: Effect of the Sample Length

Long span: $T \to \infty$

- The drift is not consistently estimable on a fixed span
  - Fisher information $I(m) = -\mathbb{E}[\partial^2 \ell / \partial m^2]$ (ℓ: log-likelihood) bounds $\mathrm{Var}(\hat{m}) \geq 1/I(m)$
  - for n GBM returns $I(m) = n\Delta t/\sigma^2 = T/\sigma^2$
  - so $\mathrm{SE}(\hat{m}) = \sigma/\sqrt{T}$
- OU: $\mathrm{Var}(\hat{\kappa}) \approx (e^{2\kappa\Delta} - 1)/(T\Delta) \to 2\kappa/T$ as $\Delta \to 0$
  - derivation: Seminar 11, A1
- The bias of $\hat{\kappa}$ does not shrink with finer sampling Yu (2012)
  - $b = e^{-\kappa\Delta} \to 1$
  - $\hat{\kappa}$ is far from Normal when κT is small

- The imprecise GBM drift and the biased $\hat{\kappa}$ have the same cause
  - information grows with T, not with n

Why Prices Use a Different Probability: a One-Period Example (1/2)

- A stock costs 100 today and tomorrow is worth 120 (up) or 90 (down)
  - the interest rate is 0
  - the real-world probability of up is $p = 0.6$
- A claim pays 1 if up and 0 if down
  - replicate it with Δ shares and B in cash
  - $120\Delta + B = 1$ and $90\Delta + B = 0$ give $\Delta = 1/30$, $B = -3$
  - the replica costs $1/3$, and any other price is an arbitrage
- The same price is the expectation under $q = (100 - 90)/(120 - 90) = 1/3$
  - q is the only weight with $120q + 90(1 - q) = 100$
  - p never enters

Why Prices Use a Different Probability: a One-Period Example (2/2)

- In general, absence of arbitrage means that discounted traded prices are martingales under an equivalent measure Q
  - fundamental theorem of asset pricing, Shreve (2004), Ch. 5
  - Girsanov builds Q in continuous time
- Check: what is the price of the claim if $p = 0.9$?
- And if the down state is 80 instead of 90?
  - Answer: still 1/3
  - with 80, $q = 20/40 = 1/2$, so the price is 1/2

Girsanov's Theorem (1/2)

- Real-world measure P and an equivalent measure Q, linked by the **Radon–Nikodym density**
  - equivalent: the same events of probability zero
  - $\left.\frac{\mathrm{d}Q}{\mathrm{d}P}\right|_{\mathcal{F}_T} = Z_T = \exp\left(-\int_0^T \Theta_s \,\mathrm{d}W_s - \frac{1}{2}\int_0^T \Theta_s^2 \,\mathrm{d}s\right)$
  - Θ_s is adapted (below: the market price of risk)
  - $Z_T > 0$ with $\mathbb{E}^P[Z_T] = 1$ reweights the probability of each path
  - $Q(A) = \mathbb{E}^P[Z_T 1_A]$
- Girsanov (1960): under Q, $\tilde{W}_t = W_t + \int_0^t \Theta_s \,\mathrm{d}s$ is a Brownian motion
  - condition: $\mathbb{E}\left[e^{\frac{1}{2}\int_0^T \Theta_s^2 \mathrm{d}s}\right] < \infty$ (Novikov condition)

Girsanov's Theorem (2/2)

- For $\mathrm{d}X = a\,\mathrm{d}t + b\,\mathrm{d}W$: $\mathrm{d}X = (a - b\Theta)\,\mathrm{d}t + b\,\mathrm{d}\tilde{W}$
- For continuous Itô processes **only the drift changes**
  - the diffusion and the quadratic variation are the same under P and Q
- With jumps (Section 9), Q may also change the jump intensity and the jump-size distribution
  - the jump intensity $\hat{\lambda}$ estimated from returns is not the one that prices options
- With stochastic volatility, P and Q weigh future variance paths differently
  - historical variance is a benchmark, not a calibration

Market Price of Risk: P versus Q (1/2)

- GBM: $\Theta = (\mu - r)/\sigma$ is the **market price of risk**
  - under Q, $\mathrm{d}S = rS\,\mathrm{d}t + \sigma S\,\mathrm{d}\tilde{W}$
  - $e^{-rt}S_t$ is a Q-martingale (Seminar 11, A4)
- S&P 500, 1990–2026: $\hat{\Theta} = (10.0\% - 2.7\%)/18.0\% = 0.40$
  - r: the mean 3-month T-bill rate
  - SE $1/\sqrt{T} = 0.17$ under GBM, as imprecise as the drift
  - beyond GBM: the delta method, with the long-run covariance of the first and second moments of returns
  - this requires a finite fourth moment, which Chapter 1 questions

Market Price of Risk: P versus Q (2/2)

- Vasicek with risk premium $\lambda\sigma$: $\kappa^Q = \kappa$, $\kappa\theta^Q = \kappa\theta - \lambda\sigma$
  - θ, θ^Q: long-run levels
  - The long-run level that prices bonds differs from the one estimated on data (Section 6)
  - A positive term premium means $\lambda < 0$
  - then $\theta^Q = \theta - \lambda\sigma/\kappa > \theta$ and long yields are higher
- Heston with volatility risk premium λv_t (v_t: the variance, Section 10): $\kappa^Q = \kappa + \lambda$, $\kappa^Q\theta^Q = \kappa\theta$
  - this λ is a separate parameter
  - $\lambda < 0$: investors pay for protection against variance
  - it gives $\theta^Q > \theta$: the variance risk premium of Chapter 9

Feynman–Kac: From Expectations to the Pricing PDE (1/2)

- If $\mathrm{d}X = a(X)\,\mathrm{d}t + b(X)\,\mathrm{d}\tilde{W}$ under Q Kac (1949)
  - $u(t,x) = \mathbb{E}^Q\left[e^{-\int_t^T r_s\,\mathrm{d}s}\, g(X_T) \mid X_t = x\right]$ solves a PDE (partial differential equation)
  - the PDE: $u_t + a\,u_x + \frac{1}{2}b^2 u_{xx} - r\,u = 0$, $u(T,x) = g(x)$
  - u is the price at t of the pay-off $g(X_T)$ at T
  - the pay-off is discounted at the short rate r_s
- Proof idea: $e^{-\int_0^t r_s\mathrm{d}s}\,u(t,X_t)$ is a Q-martingale
  - so its Itô drift is zero
  - by the same route, Chapter 12 derives Black–Scholes

Feynman–Kac: From Expectations to the Pricing PDE (2/2)

- Vasicek bond ($g = 1$, $X = r$): $p(t, r; T) = \exp\left(A(\tau) - B(\tau)\, r\right)$, $\tau = T - t$
  - derivation: Seminar 11, A6
  - $B(\tau) = \frac{1 - e^{-\kappa\tau}}{\kappa}$
  - $A(\tau) = \left(\theta^Q - \frac{\sigma^2}{2\kappa^2}\right)(B(\tau) - \tau) - \frac{\sigma^2 B(\tau)^2}{4\kappa}$
- Yields $y(\tau) = (B(\tau)\, r - A(\tau))/\tau$ tend to $\theta^Q - \sigma^2/(2\kappa^2)$
  - rate uncertainty lowers long yields (convexity)
  - comparison with Treasury yields: ▸ Appendix

Recap: Pricing under Q

- Formulas to keep
  - Girsanov: under Q the drift a becomes $a - b\Theta$, the diffusion b is unchanged
  - GBM: $\Theta = (\mu - r)/\sigma$ turns the drift μ into r
  - so $e^{-rt}S_t$ is a Q-martingale
  - Feynman–Kac: price $= \mathbb{E}^Q$[discounted pay-off], the solution of the pricing PDE
- Check: what are the market price of risk and the drift of S under Q?
  - inputs: $\mu = 9\%$, $\sigma = 20\%$, $r = 3\%$
  - Answer: $\Theta = 0.06/0.20 = 0.30$
  - drift under Q: $\mu - \sigma\Theta = 9\% - 6\% = 3\% = r$
  - $\sigma = 20\%$ is unchanged

Merton's Jump-Diffusion (1/2)

- **Question**: which model explains crash days that are practically impossible under GBM?
- Brownian paths are continuous, but prices gap on news
  - on 19 October 1987 the Dow Jones fell 22.6%
  - about 20 standard deviations under GBM with 18% volatility
  - A jump cannot be hedged by continuous trading Merton (1976)
- Log price = Brownian motion with drift + **compound Poisson** jumps
  - $\mathrm{d}\ln S_t = m\,\mathrm{d}t + \sigma\,\mathrm{d}W_t + \mathrm{d}J_t$, $J_t = \sum_{i \le N_t} Y_i$
  - N_t is a Poisson process with intensity λ
  - the jump sizes $Y_i \sim N(\mu_J, \sigma_J^2)$ are i.i.d.
  - W, N and Y are independent
  - the model has five parameters, $(m, \sigma, \lambda, \mu_J, \sigma_J)$

Merton's Jump-Diffusion (2/2)

- Given k jumps, the return over Δt is $N\left(m\Delta t + k\mu_J,\ \sigma^2\Delta t + k\sigma_J^2\right)$
- Unconditionally, it is a **Poisson mixture** of Normal densities, i.i.d. over time
  - $f(r) = \sum_{k\geq 0} e^{-\lambda\Delta t}\frac{(\lambda\Delta t)^k}{k!}\,\varphi\left(r;\ m\Delta t + k\mu_J,\ \sigma^2\Delta t + k\sigma_J^2\right)$
  - Weight: Poisson probability of k jumps
  - $\varphi(r;\text{mean},\text{variance})$: the Normal density
- Exact simulation: draw $N \sim \text{Poisson}(\lambda\Delta t)$
  - then draw from the Normal distribution with the conditional moments

Merton Estimates for the S&P 500, BET and Bitcoin (1/2)

LR (Likelihood Ratio): the likelihood-ratio statistic.

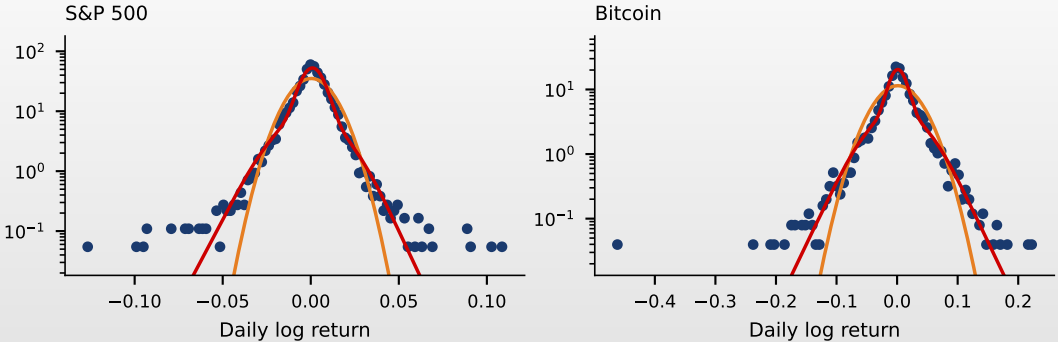
	$\hat{\sigma}$ (%)	$\hat{\lambda}$ (per year)	$\hat{\mu}_J$ (%)	$\hat{\sigma}_J$ (%)	Jump share of variance	LR
S&P 500	9.4	102 (11)	−0.16	1.44	71%	2 401
BET	11.3	79 (7)	−0.09	2.07	73%	2 326
Bitcoin	24.6	239 (21)	0.00	3.85	85%	1 357

- The Poisson-mixture log-likelihood is maximised numerically
  - up to 12 jumps per day
  - in brackets, SE from the Hessian: conditional on the i.i.d. model, around the local maximum
  - sandwich SE of $\hat{\lambda}$ (HAC covariance: heteroskedasticity and autocorrelation consistent): 28 (S&P 500), 12 (BET), 36 (Bitcoin)
- The likelihood is **unbounded**, so the MLE is a local maximum Honoré (1998)
  - with $m\Delta t$ equal to one observed return and $\sigma \to 0$ that density explodes
- The fit prefers **many small** jumps (about 80–240 per year), not a few crashes

Merton Estimates for the S&P 500, BET and Bitcoin (2/2)

- The LR test against GBM ($\lambda = 0$) is defined on $\sigma \geq 5\%$
- The LR statistic is not χ^2_3
  - $\lambda = 0$ lies on the boundary
  - μ_J, σ_J are unidentified
  - bootstrap critical values: Seminar 11, B5; the test is about the presence of jumps, not the precision of the parameters

Merton Fit to the Return Distribution (1/2)



MFM_ch11_merton_density ▶ Interactive

Merton Fit to the Return Distribution (2/2)

- Density of daily log returns on a log scale
  - histogram, the Normal distribution implied by GBM and the fitted Merton mixture
  - Merton follows the peak and the shoulders far better
- S&P 500: Merton versus data
  - excess kurtosis: 3.7 against 10.9 in the data
  - skewness: -0.31 against -0.36
- The most extreme days lie beyond both fitted curves
- About 102 "jumps" a year
  - the model uses jumps to mimic **changing volatility**
- An i.i.d. model cannot tell a jump from a volatile period
  - so a direct test is needed
- Type of evidence: description

MFM_ch11_merton_density

Recap: What Jumps Add

- Formulas to keep
  - $\mathrm{d}\ln S_t = m\,\mathrm{d}t + \sigma\,\mathrm{d}W_t + \mathrm{d}J_t$: mean log return $m + \lambda\mu_J$ per year
  - so m alone is the drift of the diffusion part only
  - Higher cumulants come only from the jumps: $c_k = \lambda\Delta t\,\mathbb{E}[Y^k]$ for $k \geq 3$
  - Excess kurtosis c_4/c_2^2 falls like $1/\Delta t$ with the horizon
- From the data: S&P 500 $\hat{m} + \hat{\lambda}\hat{\mu}_J = 24.8\% + (-16.5\%) \approx 8.3\%$
  - the same as the GBM log drift
- Check: jumps give a daily excess kurtosis of 12
- What is it for 10-day returns, if returns are i.i.d.?
  - Answer: $12/10 = 1.2$
  - over 10 days both c_4 and c_2 grow 10 times, so c_4/c_2^2 falls 10 times

The Heston Model (1/2)

- **Question**: which model generates volatility clustering and negative skewness?
  - here, negative skewness generated by the leverage effect
  - GBM and Merton with constant parameters produce neither volatility clustering nor return–volatility dependence
  - Merton can produce negative marginal skewness, through jumps with $\mu_J < 0$
- Two SDEs, for the log price and for the instantaneous variance v_t Heston (1993)
  - $\mathrm{d} \ln S_t = \left(\mu - \frac{1}{2} v_t\right) \mathrm{d}t + \sqrt{v_t}\, \mathrm{d}W_t^{(1)}$
  - $\mathrm{d}v_t = \kappa(\theta - v_t)\, \mathrm{d}t + \xi \sqrt{v_t}\, \mathrm{d}W_t^{(2)}$, $\mathrm{Corr}(\mathrm{d}W^{(1)}, \mathrm{d}W^{(2)}) = \rho$
  - the variance is a CIR process: level θ, speed κ, volatility of volatility ξ
  - Feller condition $2\kappa\theta \geq \xi^2$: the variance never touches zero Feller (1951)

The Heston Model (2/2)

- $\rho < 0$: the **leverage effect** (Section 11)
  - a price fall raises the variance that scales the next moves
  - so returns are negatively skewed
- Simulation: two schemes
  - Euler scheme with full truncation $v^+ = \max(v, 0)$ Lord, Koekkoek & van Dijk (2010)
  - or the quadratic-exponential scheme Andersen (2008)
- Daily returns need several steps per day
  - with one step, the daily return is conditionally Normal given v at the start of the day
  - one step removes the intraday channel through which ρ shapes skewness; the unconditional skewness can be small, but not necessarily zero
  - Section 11 uses 390 steps per day, one per minute of a 6.5-hour trading session

Heston Parameters from the VIX

- Proxy $y_t = (\mathrm{VIX}_t/100)^2$ in place of v_t
  - discretised CIR regression without an extra intercept
  - $\frac{y_{t+\Delta t}-y_t}{\sqrt{y_t}} = \kappa\theta\frac{\Delta t}{\sqrt{y_t}} - \kappa\Delta t\sqrt{y_t} + \xi\sqrt{\Delta t}\,\varepsilon_t$
- S&P 500 and VIX, 1990–2026: $\hat{\kappa} = 4.13$, $\hat{\xi} = 0.56$ and $\hat{\rho} = -0.69$
  - half-life: 42 days
  - The long-run level from the VIX is $\sqrt{\hat{\theta}} = 20.9\%$
  - the realised volatility is 18.0%
  - The Feller ratio of the proxy is $2\hat{\kappa}\hat{\theta}/\hat{\xi}^2 = 1.16$
- But y_t is **not** v_t, so $\hat{\xi}$ is attenuated (Further Topics)
  - y_t is a risk-neutral expectation of future variance, affine in v_t
  - For the simulated paths we set θ to the realised variance of the S&P 500

Recap: Stochastic Volatility

- Formulas to keep
  - Variance is a CIR process, $\mathrm{d}v_t = \kappa(\theta - v_t)\,\mathrm{d}t + \xi\sqrt{v_t}\,\mathrm{d}W_t^{(2)}$
  - it stays positive if $2\kappa\theta \geq \xi^2$
  - $\rho = \mathrm{Corr}(\mathrm{d}W^{(1)}, \mathrm{d}W^{(2)}) < 0$ gives leverage and negatively skewed returns
  - it also gives a downward implied-volatility skew
  - $(\mathrm{VIX}/100)^2 = a + b\,v_t$ (a, b from the Q-parameters, not the AR(1) ones): κ and ρ carry over
  - ξ is attenuated by b
- Check: with $\kappa = 4$ and $\theta = 0.04$, what is the largest ξ for which the Feller condition holds?
  - Answer: $\xi \leq \sqrt{2 \cdot 4 \cdot 0.04} = \sqrt{0.32} \approx 0.57$
  - a larger vol-of-vol lets the variance touch zero

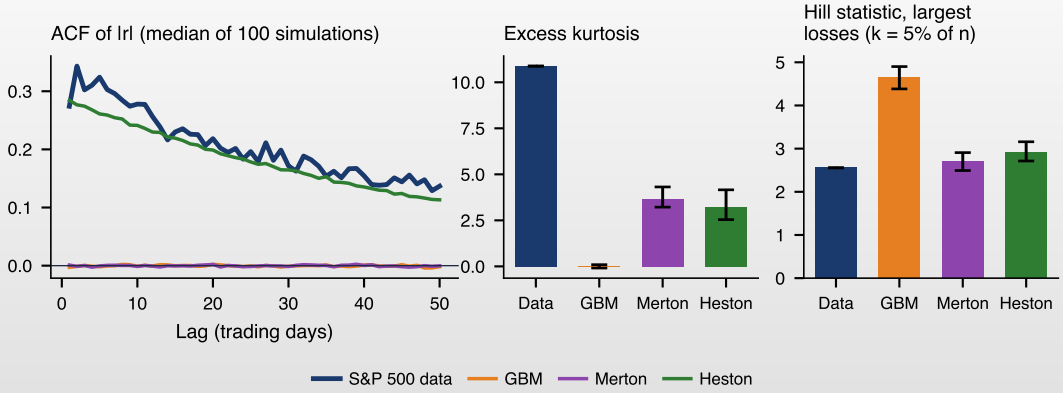
Three Yardsticks for Simulated Returns (1/2)

- **Question**: which estimated model (GBM, Merton, Heston) reproduces S&P 500 returns?
  - three features: the shape, the volatility clustering and the tail
- **Shape**: skewness $c_3/c_2^{3/2}$ and excess kurtosis c_4/c_2^2 of daily returns
  - cumulants: Further Topics
  - the kurtosis yardstick estimates a finite kurtosis consistently only with a finite fourth moment, which Chapter 1 questions
- **Clustering**: the ACF of $|r_t|$ at lags 1–20
  - zero for any i.i.d. model, Merton included
- **Tail**: Hill statistic of the k largest losses $L_{(1)} \geq \cdots \geq L_{(k+1)}$
  - $\hat{\alpha} = \left[\frac{1}{k}\sum_{i=1}^{k} \ln\left(L_{(i)}/L_{(k+1)}\right)\right]^{-1}$

Three Yardsticks for Simulated Returns (2/2)

- The Hill statistic estimates α in $P(L > x) \approx Cx^{-\alpha}$ for large x
  - $P(L > 2x)/P(L > x) \approx 2^{-\alpha}$, so a lower α is a heavier tail
- With $\alpha = 2.56$, doubling the loss leaves $2^{-2.56} \approx 0.17$ of the frequency
  - a Normal tail leaves $P(Z > 4)/P(Z > 2) \approx 0.0014$
- The estimate depends on k
  - for the data: 2.97, 2.56 and 2.04 at $k = 2.5\%$, 5% and 10% of the returns

Which Model Reproduces the S&P 500? (1/2)



MFM_ch11_models_vs_data ▶ Interactive

Which Model Reproduces the S&P 500? (2/2)

- 100 simulated samples of 9 245 days from each fitted model (medians, 90% ranges)
  - statistics compared: ACF of $|r_t|$, excess kurtosis (10.88 in the data)
  - and the Hill statistic at $k = 5\%$ of the returns
- Merton gets the tail and the skewness, but no clustering
  - Hill statistic: 2.70 against 2.56
  - skewness: -0.32 against -0.36
- Heston, with the VIX-based parameters, gets the clustering and part of the skewness
  - skewness: -0.21 against -0.36
  - the leading-order approximation gives about -0.20 at these parameters (Further Topics)
  - simulation: 390 full-truncation Euler steps per day, summed to daily returns
- No fitted model reaches the kurtosis of the data (full table: Appendix)
  - Andersen, Benzoni & Lund (2002) and Eraker, Johannes & Polson (2003) favour jumps **and** stochastic volatility
  - SVJ: stochastic volatility with jumps, Bates (1996)
- Type of evidence: comparison of procedures (it shows which method works better on the same data, not why)
  - the models at their estimated parameters: a Heston model with a lower Feller ratio could reach the kurtosis (Appendix)

MFM_ch11_models_vs_data



Which Model When? Features, Identification and Limits

Feature in the data	Model	What the data identify	Unresolved
Positive prices, $\sqrt{t}$ scaling	GBM	σ precisely; μ only over very long spans	fat tails, clustering
Mean reversion of rates and of the VIX	OU, Vasicek, CIR, CKLS	$b(r)$ precisely; κ biased upwards, needs a long span	one factor too simple for the VIX
Gaps, short-horizon fat tails	Merton	λ, σ_J only under i.i.d. returns; the jump share depends on the frequency	no clustering
Clustering, leverage, implied-volatility skew	Heston	κ, ρ from the VIX; ξ attenuated; $\theta^Q \neq \theta^P$	kurtosis of the data
All of the above	Bates (Heston + jumps)	returns and option surfaces jointly	latent states: filtering needed

- Roughness (irregularity at short scales) and long memory of volatility (next): none of these one-factor models produces them

Background: Fractional Brownian Motion and the Hurst Exponent (deeper) (1/2)

- **Question**: can one process describe volatility that is both irregular and persistent?
  - irregular at small scales, persistent at large scales (the VIX ACF, Section 6)
- **Fractional Brownian motion** B^H, $0 < H < 1$: a Gaussian process with $B_0^H = 0$
  - $\mathrm{Cov}(B_s^H, B_t^H) = \frac{1}{2}\left(s^{2H} + t^{2H} - |t-s|^{2H}\right)$
  - $H = 1/2$ gives $\min(s,t)$, i.e. Brownian motion
  - increments have variance $\mathbb{E}[(B_{t+h}^H - B_t^H)^2] = h^{2H}$
  - $H < 1/2$: successive increments are negatively correlated, paths rougher than Brownian motion
  - $H > 1/2$: positively correlated, smoother

Background: Fractional Brownian Motion and the Hurst Exponent (deeper) (2/2)

- **Fractional OU**: the OU equation driven by B^H instead of W
  - one H sets both ends of its ACF
- Short lags: the variogram $\mathbb{E}[(x_{t+h} - x_t)^2]$ behaves like $h^{2H} = h^{2\alpha+1}$
  - the roughness index is $\alpha = H - 1/2$
  - here α is not the tail index of Section 11
  - $\alpha < 0$: rough
- Long lags: the ACF decays like $h^{2H-2} = h^{-\beta}$
  - so $\beta = 2 - 2H$
  - $\beta < 1$: non-integrable ACF, long memory

Research Question: Can Volatility Be Rough and Persistent?

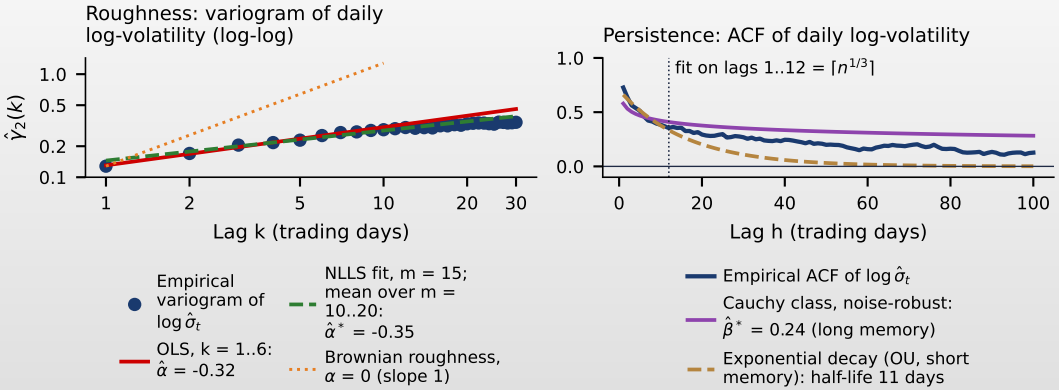
- Check: which H gives $\beta = 0.18$? Is that process rough?
  - Answer: $H = 1 - \beta/2 = 0.91$, so $\alpha = 0.41 > 0$: smooth, not rough
  - one H cannot match both features
- Paper: Bennedsen, Lunde & Pakkanen (2022), *Journal of Financial Econometrics*
  - it treats α and β as separate parameters
  - builds on Gatheral, Jaisson & Rosenbaum (2018): log-volatility as a fractional OU process with $H < 1/2$
  - RFSV: rough fractional stochastic volatility
  - The long-memory model of Comte & Renault (1998) uses $H > 1/2$
- Question 1: are realised volatilities both rough and persistent?
- Question 2: do models that set α and β separately forecast better?
  - model, estimators and results: ▸ Appendix

Case Study: Replication on SPY, $\Delta = 1$ Day (1/3)

- Data: SPY (SPDR S&P 500 ETF Trust) 5-minute bars, 1 468 full days (October 2020 – September 2026)
  - the paper studies E-mini S&P 500 futures; SPY tracks the same index
- $x_t = \log \hat{\sigma}_t = \frac{1}{2} \log \mathrm{BV}_t$
  - BV_t: the daily bipower variation of the 78 five-minute returns
  - robust to jumps Barndorff-Nielsen & Shephard (2004)
- Roughness: variogram $\hat{\gamma}_2(k) = \frac{1}{n-k} \sum_t (x_{t+k} - x_t)^2 \approx c\, k^{2\alpha+1}$ for small k
  - OLS: the slope $\hat{a}_1$ of $\log \hat{\gamma}_2(k)$ on $\log k$, $k = 1, \ldots, 6$
  - it gives $\hat{\alpha} = (\hat{a}_1 - 1)/2$
  - noise-robust NLLS (Non-Linear Least Squares): $\hat{\gamma}_2(k) = b_0 + b_1 k^{2\alpha+1}$
  - b_0 absorbs the measurement noise in x_t
- Persistence: Cauchy-class ACF $\rho(h) = (1 + |h|^{2\alpha+1})^{-\beta/(2\alpha+1)}$
  - fitted on the first $\lceil n^{1/3} \rceil = 12$ lags
- Short-memory benchmark: an OU process, with ACF $e^{-\kappa h}$

Case Study: Replication on SPY, $\Delta = 1$ Day (2/3)

NLLS: Non-Linear Least Squares.



MFM_ch11_rough_vol ▶ Interactive

Case Study: Replication on SPY, $\Delta = 1$ Day (3/3)

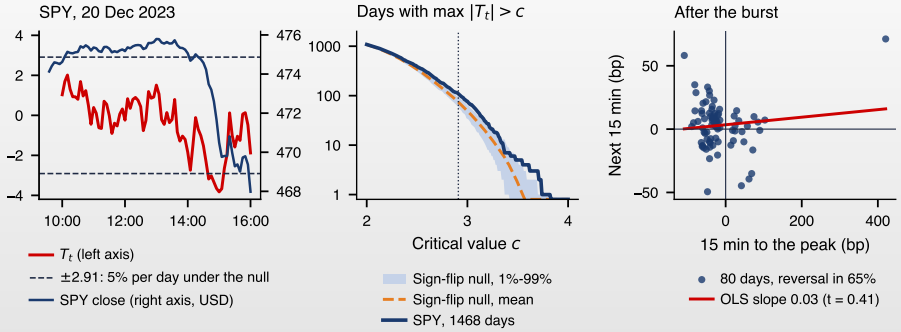
- Left, roughness: empirical variogram $\hat{\gamma}_2(k)$ against k on log axes
  - the dotted line with slope 1 is Brownian roughness ($\alpha = 0$)
  - OLS gives $\hat{\alpha}_{\mathrm{OLS}} = -0.32$
  - the noise-robust NLLS gives $\hat{\alpha}^*_{\mathrm{NLLS}} = -0.35$
  - Bennedsen, Lunde & Pakkanen (2022) at 1 day: -0.30 (OLS) and -0.33 (NLLS)
- Right, persistence: the autocorrelation is 0.13 at lag 100
  - The Cauchy fit on 12 lags gives $\hat{\beta} = 0.26$ and $\hat{\beta}^* = 0.24$ (noise-robust)
  - both below 1: consistent with long memory, not a proof
  - 5-minute bars without pre-averaging; realised measures estimate integrated, not spot, variance
  - An OU fit (half-life 11 days) implies an autocorrelation of 0.002 at lag 100
- In the plot, m is the number of lags used in NLLS, not the GBM log drift
  - 15 for the illustrated fit
  - the reported estimate averages fits over 10–20 lags
- Type of evidence: description

MFM_ch11_rough_vol

Second Case Study (1/6): Drift Bursts

- Paper: Christensen, Oomen & Renò (2022), *Journal of Econometrics*, 227(2), 461–497
  - Data: quotes of six CME (Chicago Mercantile Exchange) futures contracts, 2012–2017 (E-mini S&P 500 from 2010)
  - the contracts: E-mini S&P 500, 10-year T-Note, euro, gold, crude oil, corn
- Question: can the drift μ_t dominate volatility for a few minutes?
  - over long spans the drift is imprecise: ± 11 percentage points over 10 years (Section 5)
  - drift burst: before a time τ, $|\mu_t|$ grows like $(\tau - t)^{-a}$, with $1/2 < a < 1$
  - volatility may rise too, but more slowly; the price stays continuous, with no jump
  - the flash crashes of 6 May 2010 (equities) and 15 October 2014 (US Treasuries) are the paper's examples
- Main claim (the paper's abstract)
  - drift bursts occur once a week on average, in all four asset classes
  - most of them are followed by a price reversal
  - the reversal is stronger after negative drift bursts with large volume

Second Case Study (2/6): Drift Bursts in SPY, 5-Minute Bars



- **Left: the SPY price and the test statistic T_t (defined on the next slide) on 20 December 2023, the day with the largest $|T_t|$ in the sample (-3.83 at 15:00)**
- **Middle: the number of days with $\max_t |T_t| > c$, in the data and under the random-sign null (no drift)**
- **Right: the return in the 15 minutes after the peak of $|T_t|$ against the return in the 15 minutes before it, on days with $\max_t |T_t| > 2.91$**

MFM_ch11_drift_burst ▶ Interactive

Second Case Study (3/6): The Paper's Statistic and the Course Adaptation

- The paper's statistic (Eq. 11–14), with the kernel $K(x) = e^{-|x|}$ for $x \leq 0$ (past only)
  - $\hat{\mu}_t = \frac{1}{h} \sum_i K\left(\frac{t_i - 1 - t}{h}\right) \Delta_i X$ and $\hat{\sigma}_t^2 = \frac{1}{h'} \sum_i K\left(\frac{t_i - 1 - t}{h'}\right) (\Delta_i X)^2$
  - $T_t = \sqrt{h/K_2}\, \hat{\mu}_t / \hat{\sigma}_t$, with $K_2 = \int K^2 = 1/2$
  - without a drift burst, T_t is asymptotically standard Normal; with one, $|T_t| \to \infty$
- The paper: $h = 5$ minutes, $h' = 25$ minutes, pre-averaging and a HAC (heteroskedasticity and autocorrelation consistent) correction for microstructure noise
  - the paper counts drift bursts at fixed thresholds, 4–6, and estimates the false alarms by simulation
- The course: an adaptation to five-minute bars, with the sums starting at the open
  - $\hat{s}_j^2 = \dfrac{\sum_i u_{ij} r_i^2}{\Delta \sum_i u_{ij}}$ and $T_j = \dfrac{\sum_i w_{ij} r_i}{\sqrt{\hat{s}_j^2\, \Delta \sum_i w_{ij}^2}}$
  - the weights: $w_{ij} = e^{-(t_j - t_i - 1)/h}$ and $u_{ij} = e^{-(t_j - t_i - 1)/h'}$
  - the denominator is random and computed from the same returns, so T_j is not exactly Normal
  - with $h/\Delta = 3$ bars the asymptotic approximation is not guaranteed: critical values are calibrated by simulation

Second Case Study (4/6): The SPY Replication

- SPY, 1 468 complete days (October 2020 – September 2026)
  - $h = 15$ minutes (3 bars), $h' = 75$ minutes: the paper's 1:5 ratio
  - 73 tests a day, 10:00–16:00
  - no pre-averaging and no HAC: noise is small at 5 minutes
- The null: returns multiplied by random signs ± 1, 200 replications
  - it keeps the absolute returns and the intraday pattern, but removes the drift
  - it also removes jump asymmetry, sign dependence and the leverage effect
  - so it is narrower than "no drift burst"; the paper uses a jump-diffusion null
- Critical values
  - 2.91: the 95% quantile of $\max_t |T_t|$ under the null
  - Gumbel approximation of the maximum: 3.50, exceeded on only 0.12% of the simulated days
  - a single jump gives at most $|T_t| = 2.74$ when the jump return dominates both kernel sums, below the 2.91 threshold

Second Case Study (5/6): The Result

- 109 candidate days with $\max_t |T_t| > 2.91$, against 73.84 on average under the null
  - none of the 200 replications reaches 109 (maximum 100): $p_{\mathrm{MC}} = 0.005$, under the random-sign null
  - at $c = 3.22$ (the 99% quantile): 23 days against 14.96 ($p_{\mathrm{MC}} = 0.030$)
  - the descriptive excess, 35.16 days, does not estimate the number of true drift bursts
  - the paper, on E-mini S&P 500: 202 drift bursts with $|T_t| > 4.5$ in 2010–2017, about one every two weeks
- 73 of the 109 days peak on the negative side (67%); the random-sign null is symmetric by construction
- The price reverses on 65% of the 80 days with a full window after the peak: a descriptive statistic
  - the peak is chosen in hindsight over the whole day, so the returns after it shape its choice
  - under the null, with the same selection: 56.5% on average; at least 65% in 12.4% of 1 000 replications: the reversal is not distinguishable from the null
  - the size: -43.8 basis points (bp) in the 15 minutes to the peak, then $+4.7$ bp
  - slope of R^+ (after the peak) on R^- (before) (Eq. 36): 0.03 (t = 0.41); the paper, E-mini, $|T_t| > 4$: -0.26 (t $= -12.05$)
  - negative peaks, volume above the median (relative to its mean at the same time of day): -53.9 bp, then $+6.4$ bp (29 days)
  - negative peaks, volume below the median: -34.4 bp, then $+3.1$ bp (31 days)

Second Case Study (6/6): Validating the Procedure; Limits

- The same procedure on data simulated without drift bursts: 50 samples of 1 468 days, 5-minute bars
  - Heston with the VIX-based parameters, leverage ($\rho = -0.69$), one-minute steps
  - mean excess: 1.2 days (95% quantile: 14.8); $p_{\mathrm{MC}} \leq 0.05$ in 8% of the samples
  - plus asymmetric Merton jumps ($\hat{\lambda}$, $\hat{\mu}_J$, $\hat{\sigma}_J$ fitted to the S&P 500)
  - mean excess: 3.8 days (95% quantile: 17.5); $p_{\mathrm{MC}} \leq 0.05$ in 12% of the samples
  - mean reversal after the peak: 56% without jumps and 53% with jumps
- The excess in the data, 35.16 days, lies above the 95% quantile of both simulated models
- Type of evidence: a description, compared with nulls without drift
  - it does not identify the paper's liquidity mechanism
- Limits
  - the power of the test at 5 minutes is not estimated here: low power remains an anticipated limitation
  - with 50 samples, the rejection frequencies carry a Monte Carlo error of about ± 4 percentage points
  - 29 days peak in the last 15 minutes of the session and have no window after the peak

An Open Question: How Much of the Price Path Is Discontinuous? (1/2)

- Model: $\mathrm{d}X_t = \mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t + Y_t\,\mathrm{d}N_t$
  - X_t: the log price
  - N_t: the number of jumps (intensity λ)
  - Y_t: the jump size
- The answer depends on the sampling frequency and on the model
  - Daily Merton MLE (this chapter): jumps carry 71% of S&P 500 variance
  - The 5-minute bipower test (Chapter 9): jumps carry 2.8% of SPY variance

An Open Question: How Much of the Price Path Is Discontinuous? (2/2)

- Christensen, Oomen & Podolskij (2014): at tick frequency most jumps detected at coarser sampling disappear
  - they are rapid but continuous moves
- Christensen, Oomen & Renò (2022): **drift bursts** look like jumps in coarse data
  - drift burst: a short explosive trend in μ_t
- Formal question: H_0: $\lambda = 0$ against H_1: $\lambda > 0$
  - H_0: continuous path with stochastic volatility
  - a non-standard test, with the boundary problem of the Merton LR test (Section 9)

The Discovery Loop with an AI Assistant

The loop with an AI (Artificial Intelligence) assistant

Assistant: an LLM (Large Language Model) such as Claude, ChatGPT or Gemini.

- **1. Literature**: published jump shares by frequency
  - check every Digital Object Identifier (DOI) on Crossref
- **2. Hypotheses**: Merton jumps, Heston volatility, drift bursts
  - each with a distinct testable implication
- **3. Replicate first** the chapter's estimates
- **4. Robustness**: filters and tests fixed before the first run
- **5. Critique**: the assistant challenges the inference, as a referee would
- **6. Report**: every specification, the code, the AI log
- Reusable prompts: Appendix

What the Researcher Checks: Failure and Test

Checks: failure and test

- Boundary, unidentified parameters: parametric bootstrap under H_0
- Look-ahead bias (using future information) in the volatility filter: standardise only with $\mathcal{F}_{t-1}$
- Frequency mismatch: jump shares at every frequency on a common sample
- Multimodal likelihood: many starting points and a profile in λ
- Invented numbers: every claim needs a table and a DOI

Mini-Case: Jumps or Stochastic Volatility in Disguise?

- A typical assistant proposal: "71% of variance comes from jumps"
  - Rival hypothesis: the Poisson mixture absorbs time-varying volatility
  - then standardising by EWMA volatility removes the jumps
  - EWMA: Exponentially Weighted Moving Average, known at $t-1$
  - $\lambda_{\mathrm{EWMA}} = 0.94$, the RiskMetrics value for daily data

	Raw returns	EWMA-standardised
Excess kurtosis	11.0	2.9
LR, Merton vs the Normal distribution	2 403	619
$\hat{\lambda}$ per year (SE)	100 (10)	124 (53)
Jump share of variance	71%	51%

- The same sample: 8 995 days, without the first 250 (the EWMA start-up)
- Standardising cuts the excess kurtosis from 11.0 to 2.9, yet $\hat{\lambda}$ rises from 100 to 124 per year
  - EWMA need not remove heteroskedasticity, and does not separate jumps from stochastic volatility mechanically
- Interpretation: the estimated jump intensity and jump share depend strongly on how volatility is specified
- This exercise does not identify discontinuities structurally
- Type of evidence: comparison of procedures
  - the same Merton model, fitted to raw and to EWMA-standardised returns

Your Turn: A Project Seed

- Questions
  - how much of S&P 500 and Bitcoin variance is discontinuous?
  - how does this share depend on the sampling frequency?
  - are the detected "jumps" drift bursts?
- Data: 5-minute bars for every intraday frequency
  - SPY: October 2020 – September 2026, 1 468 full trading days
  - Bitcoin: September 2024 – September 2026, 659 UTC (Coordinated Universal Time, Temps Universel Coordonné) days with at least 95% of the 288 bars (Chapter 9)
  - S&P 500 and Bitcoin daily closes
- The Bitcoin bars cover two years
  - an intraday comparison of the two assets uses the common window September 2024 – September 2026
- A pre-registration dated before the first run
  - contents: hypotheses, frequencies, filters, tests, bootstrap design

Key Takeaways and Next: Chapter 12 (1/2)

- Rescaled random walks converge to Brownian motion (Donsker)
  - its quadratic variation t drives Itô calculus
- Itô's lemma gives exact solutions (GBM, Ornstein–Uhlenbeck, Merton)
  - other models are simulated with Euler–Maruyama or Milstein at a known error order
- Frequent sampling pins down volatility
  - drift and mean-reversion speed need long spans
  - $\hat{\kappa}$ is biased upwards
- No arbitrage makes prices discounted Q-expectations
  - Girsanov changes only the drift
  - Feynman–Kac turns the expectation into a PDE

Key Takeaways and Next: Chapter 12 (2/2)

- $(\mathrm{VIX}/100)^2$ is a Q-expectation, affine in v_t
- In the fitted models jumps help with the tails, stochastic volatility with the clustering
  - neither reaches the excess kurtosis of the data
- Volatility looks rough at short lags and persistent at long lags
  - a one-factor OU model does not reproduce this
- Next, Chapter 12: pricing options by replication
  - and reading the volatility surface

Check Your Understanding

- Why does a change of variable need an extra term in Itô calculus?
  - Identify the role of quadratic variation.
- Does observing prices more frequently make the drift estimate precise?
  - Separate frequency from the length of the observation period.
- Which feature calls for jumps, and which calls for stochastic volatility?
  - Use the comparison with observed returns.
- Why can a pricing probability differ from a real-world probability?
  - State what changes under the change of measure.

The Practical Decision

- **What I choose**
  - for daily S&P 500 returns, Bates (stochastic volatility with jumps), as the next model to fit and validate
  - the Heston part gives the clustering: ACF of $|r_t|$ at lag 1 is 0.28 against 0.27 in the data
  - the Merton part gives the tail: Hill statistic 2.70 against 2.56
  - $\hat{\lambda}$ describes fat tails, not the number of discontinuities
  - volatility from frequent sampling; drift and κ only from long spans, reported with their CIs
- **Under which conditions**
  - parameters under P for risk and under Q (VIX, option surfaces) for pricing
  - simulation with many steps per day (390 for Heston): one step removes the effect of ρ on daily skewness
- **What result would make me drop it**
  - a simulated Hill statistic or ACF of $|r_t|$ from the fitted model far from the values in the data
  - volatility autocorrelation at long lags above that of one factor, at the horizon that matters
  - SPY already shows this (0.13 at lag 100 against 0.002 for an OU fit): if that horizon matters, I choose a rough, persistent model

Data Used in This Chapter (1/2)

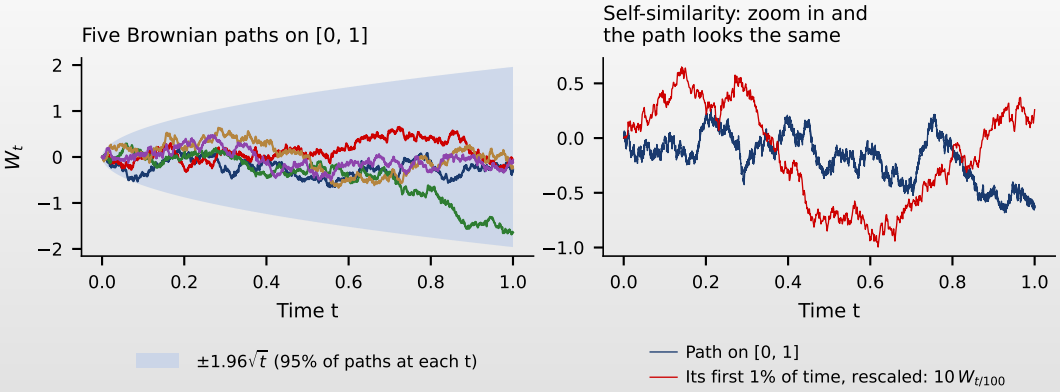
Series	Type	Period	Observations / frequency
S&P 500	US equity index	1990–2026	9 245
BET	BVB equity index	2000–2026	6 670
Bitcoin	crypto-asset, 7 days a week	2014–2026	4 384
VIX	CBOE implied volatility index of the S&P 500	1990–2026	daily
DTB3	3-month US Treasury bill rate (FRED)	1954–2026	daily

- Daily closing values up to 18 September 2026
  - daily log returns $r_t = \ln P_t - \ln P_{t-1}$, each series on its own calendar
  - cleaning rules: Appendix
- Time in years, daily step $\Delta t = 1/252$ (Bitcoin: $1/365$)
  - drifts and volatilities annualised

Data Used in This Chapter (2/2)

- Intraday: SPY 5-minute bars, October 2020 – September 2026, 1 468 full trading days
  - SPY: exchange-traded fund (ETF) tracking the S&P 500
  - used in the case study; see Chapter 9
- Intraday: Bitcoin 5-minute bars, September 2024 – September 2026, 659 UTC days
  - days with at least 95% of the 288 bars, as in Chapter 9
- Notation
  - W_t is standard Brownian motion and S_t (also P_t) the price
  - v_t is the instantaneous variance
  - $\mathcal{F}_t$ is the information (filtration) up to t
  - $\mathrm{d}S/S$ is the instantaneous simple return and $\mathrm{d}\ln S$ the log change
  - Itô's lemma explains their different drifts

Brownian Paths (1/2)



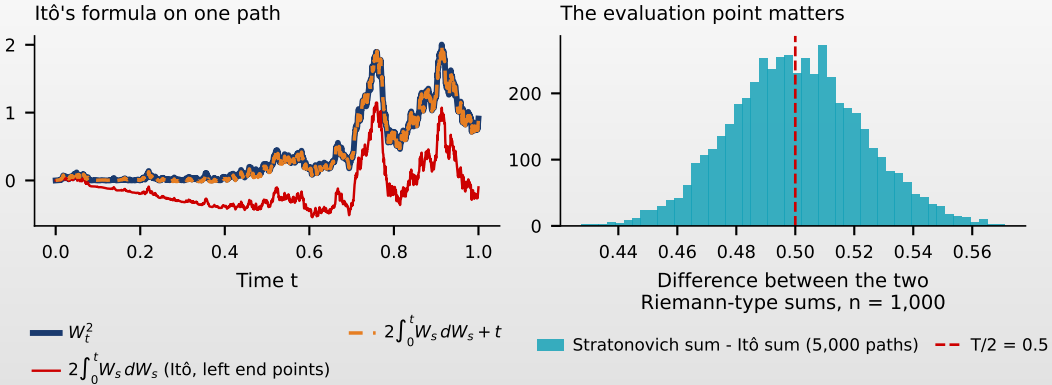
MFM_ch11_bm_paths

Brownian Paths (2/2)

- Left: five paths with the pointwise band $\pm 1.96\sqrt{t}$
  - the band covers 95% for each t separately, not for the whole path
- Right: a path and its first 1% of time, stretched 100 times in time and 10 times in space
- The band widens like $\sqrt{t}$ (the square-root-of-time rule of Chapter 7)
- A zoomed piece looks like the whole path
  - real markets break self-similarity
  - fat tails and clustering depend on the horizon (Chapter 1)

MFM_ch11_bm_paths

Worked Example: $\int_0^T W_t \, \mathrm{d}W_t$ on Simulated Paths (1/2)



MFM_ch11_ito

Worked Example: $\int_0^T W_t \, \mathrm{d}W_t$ on Simulated Paths (2/2)

- Result: $\int_0^T W_t \, \mathrm{d}W_t = \frac{1}{2} W_T^2 - \frac{1}{2} T$
  - each term: $W_{t_{k-1}} \Delta W_k = \frac{1}{2}(W_{t_k}^2 - W_{t_{k-1}}^2) - \frac{1}{2}(\Delta W_k)^2$
  - summed over k, the first parts give $\frac{1}{2}(W_T^2 - W_0^2) = \frac{1}{2} W_T^2$: the intermediate terms cancel
  - the second parts tend to $\frac{1}{2} T$
  - ordinary calculus would give $\frac{1}{2} W_T^2$: the extra $-\frac{1}{2} T$ is the quadratic variation
  - the mean is $\frac{1}{2}(T - T) = 0$
- Left: on one path of 2 000 steps, $2 \int_0^t W \, \mathrm{d}W + t$ lies approximately on W_t^2 (largest gap 0.038)
  - the exact identity on the grid: $W_t^2 = 2 \sum_i W_{t_i} \Delta W_i + \sum_i (\Delta W_i)^2$; t is only the limit of the discrete quadratic variation
- Right: Stratonovich minus Itô sums over 5,000 paths
  - mean 0.500 and standard deviation 0.022
  - theory gives $T/2 = 0.5$
  - Itô sums: mean 0.007 (theory: 0)
  - Itô sums: variance 0.520 (theory, by the isometry: $\int_0^1 t \, \mathrm{d}t = 1/2$)

MFM_ch11_ito

Quadratic Covariation and the Multivariate Itô Formula (1/2)

- **Quadratic covariation** $[X, Y]_t = \lim \sum_k \Delta X_k \, \Delta Y_k$
  - if $\mathrm{Corr}(\mathrm{d}W^{(1)}, \mathrm{d}W^{(2)}) = \rho$, then $\mathrm{d}[W^{(1)}, W^{(2)}]_t = \rho \, \mathrm{d}t$
  - product rule: $\mathrm{d}(X_t Y_t) = X_t \, \mathrm{d}Y_t + Y_t \, \mathrm{d}X_t + \mathrm{d}[X, Y]_t$
- **Multivariate Itô formula** for $X_t \in \mathbb{R}^d$
  - $f(t, x)$ continuously differentiable once in t and twice in x
  - $\mathrm{d}f(t, X_t) = f_t \, \mathrm{d}t + \sum_i f_{x_i} \, \mathrm{d}X_t^i + \frac{1}{2} \sum_{i,j} f_{x_i x_j} \, \mathrm{d}[X^i, X^j]_t$
  - X_t^i is the i-th component

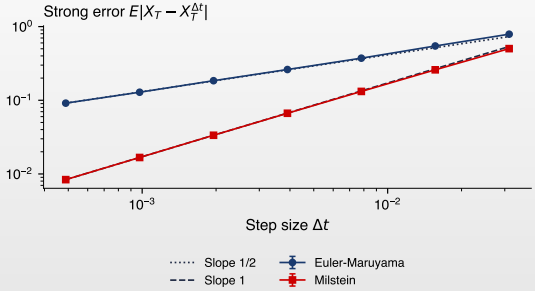
Quadratic Covariation and the Multivariate Itô Formula (2/2)

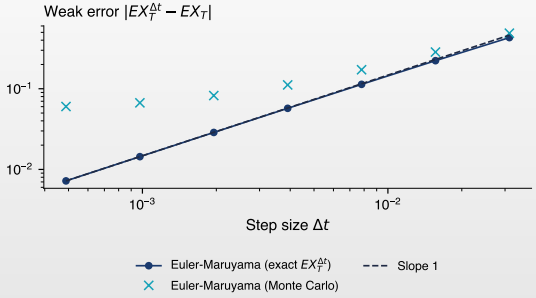
- Example (Heston, Section 10)
  - $\mathrm{d}S = \mu S\,\mathrm{d}t + \sqrt{v}\,S\,\mathrm{d}W^{(1)}$
  - $\mathrm{d}v = \kappa(\theta - v)\,\mathrm{d}t + \xi\sqrt{v}\,\mathrm{d}W^{(2)}$
  - v is the instantaneous rate of quadratic variation of $\ln S$ (the instantaneous variance of returns), not $\mathrm{Var}(S_t)$
  - κ is the speed of mean reversion
  - θ is the long-run variance and ξ the volatility of the variance
- $\mathrm{d}(S_t v_t) = S_t\,\mathrm{d}v_t + v_t\,\mathrm{d}S_t + \rho\,\xi\,v_t S_t\,\mathrm{d}t$
  - ρ enters the drift of the product

Weak Convergence

- **Weak error**: $|\mathbb{E}[g(X_T^{\Delta t})] - \mathbb{E}[g(X_T)]| \leq C\,\Delta t^{\beta}$
  - how close the simulated distribution is to the true one
  - g is a function of the final value (e.g. an option pay-off)
  - β is the weak order
- Both schemes have $\beta = 1$
  - enough for European option prices and other expected values
  - conditions: smooth coefficients with bounded derivatives and a smooth g Kloeden & Platen (1992)
- Slower convergence is possible
  - square-root coefficients (CIR (Cox–Ingersoll–Ross), Heston)
  - quantiles such as Value-at-Risk (VaR)
  - barrier crossings
- Test equation Higham (2001): $\mathrm{d}X = \lambda X\,\mathrm{d}t + \mu X\,\mathrm{d}W$
  - $\lambda = 2$, $\mu = 1$, $X_0 = 1$, $T = 1$, with the exact solution as benchmark
  - on a log-log plot, the order is the slope of the error against the step (next slide)

Convergence of the Schemes (1/2)





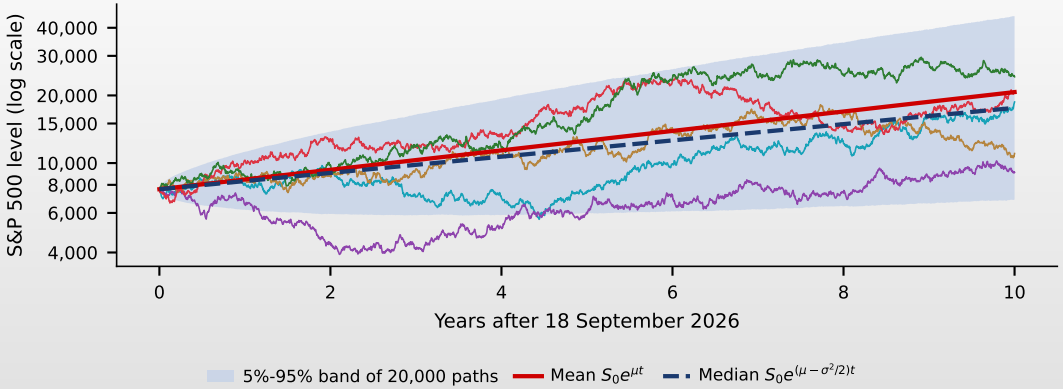
MFM_ch11_convergence

Convergence of the Schemes (2/2)

- 20,000 Brownian paths on a fine grid of 2^{11} steps, coarsened to steps from 2^{-11} to 2^{-5}
- Strong orders: slope of the log-log fit, 95% CI by resampling whole paths
  - Euler–Maruyama 0.518 [0.512, 0.525], Milstein 0.986 [0.984, 0.987]
  - the narrow CIs measure Monte Carlo noise only
  - the small gap to 1/2 and 1 is finite-step curvature, not a different order
  - halving Δt halves the Milstein error, but cuts the Euler error only by 29%
- Weak order of Euler–Maruyama from the exact mean $(1 + \lambda \Delta t)^{T/\Delta t}$: 0.985
  - a deterministic slope
  - the Monte Carlo estimates flatten at small steps
  - their standard error, about 0.068, exceeds the bias
  - the sample size, not the step, is often the binding constraint Glasserman (2003)

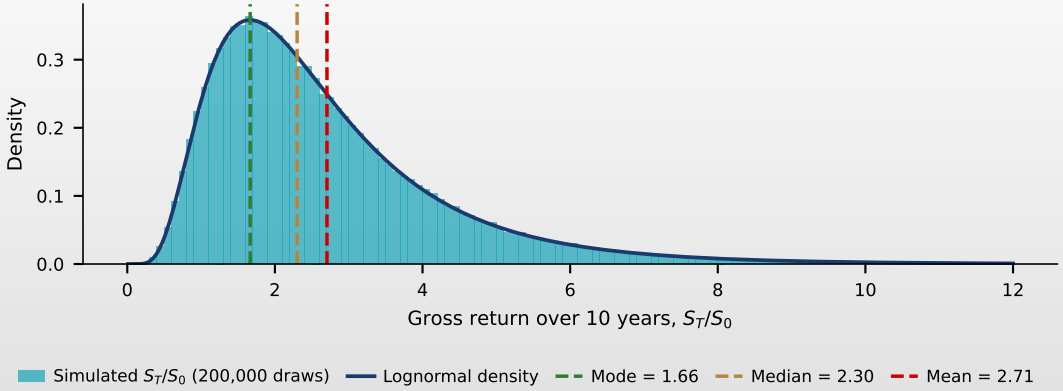
MFM_ch11_convergence

Lognormal Outcomes: Ten Years of Simulated S&P 500 Paths (1/3)



MFM_ch11_gbm_fan

Lognormal Outcomes: Ten Years of Simulated S&P 500 Paths (2/3)



MFM_ch11_lognormal

Lognormal Outcomes: Ten Years of Simulated S&P 500 Paths (3/3)

- GBM with $\hat{\mu} = 10.0\%$ and $\hat{\sigma} = 18.0\%$ (1990–2026), over $T = 10$ years
  - mean $\mathbb{E}[S_T] = S_0 e^{\mu T}$, median $S_0 e^{mT}$, mode $S_0 e^{(m-\sigma^2)T}$
  - density (200 000 draws): the gross return S_T/S_0 is 2.71 at the mean
  - it is 2.30 at the median and 1.66 at the mode
- Fan chart: 20,000 exact paths starting from $S_0 = 7\,650$ (18 September 2026)
  - 90% of the paths end between 6 878 (5% quantile) and 45 017 (95% quantile)
  - probability of a 10-year loss: $P(S_T < S_0) = \Phi(-m\sqrt{T}/\sigma) = 7.1\%$
- $P(S_T < \mathbb{E}[S_T]) = \Phi(\sigma\sqrt{T}/2) = 61.2\%$
  - most investors end below the average
  - so report the median and quantiles
- The fan is too narrow in crises
  - GBM has no fat tails and no volatility clustering (tested on the S&P 500)

MFM_ch11_gbm_fan

Recap: Growth, Lognormality and Drift Precision

- Formulas to keep
  - $\ln(S_T/S_0) \sim N(mT, \sigma^2 T)$ with $m = \mu - \frac{1}{2}\sigma^2$
  - the median grows at m, the mean $\mathbb{E}[S_T]$ at μ
  - $\mathrm{SE}(\hat{\sigma}) = \sigma/\sqrt{2n}$ shrinks with the number of returns
  - $\mathrm{SE}(\hat{m}) = \sigma/\sqrt{T}$ shrinks only with the years covered
- Check: with $\sigma = 20\%$, what is the SE of the log drift after 10 years of data?
- Check: and after 40 years?
  - Answer: $0.20/\sqrt{10} = 6.3$ and $0.20/\sqrt{40} = 3.2$ percentage points
  - four times the span only halves the error, whatever the sampling frequency

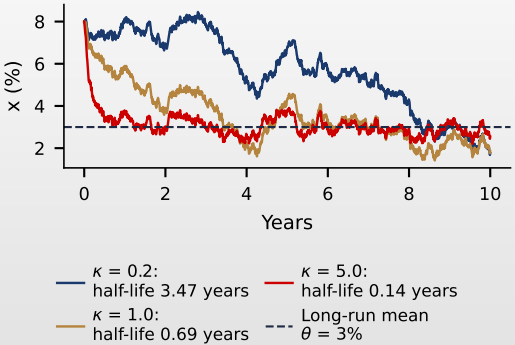
Solving the OU Equation in Three Steps (1/2)

- Step 1: Itô's lemma on $f(t,x) = e^{\kappa t}x$ ($f_{xx} = 0$)
  - result: $\mathrm{d}(e^{\kappa t}x_t) = \kappa\theta e^{\kappa t}\,\mathrm{d}t + \sigma e^{\kappa t}\,\mathrm{d}W_t$
  - the terms in x_t cancel
- Step 2: integrate from 0 to t and multiply by $e^{-\kappa t}$
  - $x_t = \theta + (x_0 - \theta)e^{-\kappa t} + \sigma \int_0^t e^{-\kappa(t-s)}\,\mathrm{d}W_s$
- Step 3: the Itô integral has mean 0
  - by the isometry, its variance is $\sigma^2 \int_0^t e^{-2\kappa(t-s)}\,\mathrm{d}s$
  - so x_t is Normal

Solving the OU Equation in Three Steps (2/2)

- Moments of the Normal x_t
  - $\mathbb{E}[x_t \mid x_0] = \theta + (x_0 - \theta)e^{-\kappa t}$
  - $\text{Var}(x_t \mid x_0) = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa t}\right)$
- Stationary distribution $N\left(\theta, \sigma^2/(2\kappa)\right)$
  - the variance grows with the horizon but, unlike Brownian motion, stays below the limit $\sigma^2/(2\kappa)$
- Check: what are the mean and the standard deviation of x_1?
  - inputs: $x_0 = 8\%$, $\theta = 3\%$, $\kappa = 0.5$, $\sigma = 1\%$
  - Answer, mean: $3\% + 5\% \cdot e^{-0.5} = 6.03\%$
  - standard deviation: $1\% \cdot \sqrt{1 - e^{-1}} = 0.80\%$
  - after one year 61% of the initial gap remains

Three Speeds of Mean Reversion



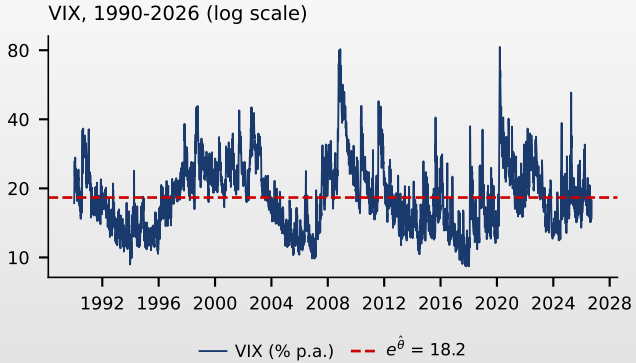
- OU paths driven by the same shocks, starting at 8%
  - $\theta = 3\%$, $\sigma = 1\%$, and speed $\kappa = 0.2$ (slow), 1 or 5 (fast)
  - exact simulation, daily steps
- Half-lives: 3.47 years for $\kappa = 0.2$, 0.69 for $\kappa = 1$ and 0.14 for $\kappa = 5$
  - With $\kappa = 5$ the gap from 8% to 3% closes within a few months
- Stationary standard deviations $\sigma/\sqrt{2\kappa}$
  - 1.58% for $\kappa = 0.2$, 0.71% for $\kappa = 1$ and 0.32% for $\kappa = 5$
  - Slow reversion looks like a random walk for years
- Long-horizon forecasts converge to θ

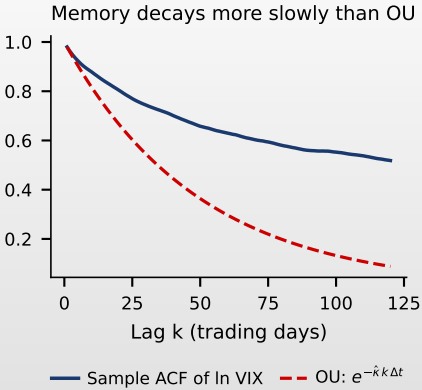
MFM_ch11_ou_paths

Why Estimated Mean Reversion Looks Too Fast

- In finite samples $\hat{\kappa}$ overstates κ by roughly $4/T$ (T in years) Tang & Chen (2009)
  - Simulation: the full-sample estimates as the truth, 500 samples of 72 years
  - The mean $\hat{\kappa}$ is 0.164, a bias of 0.063, close to the approximation $4/T = 0.055$
  - 78% of the estimates exceed the truth (Seminar 11, B2)
- Where the bias comes from: $\hat{b}$ is biased towards zero in a short persistent sample
  - $\hat{b}$: the OLS slope of $x_{t+\Delta t}$ on x_t
  - $\hat{\kappa} = -\ln \hat{b}/\Delta t$ turns a too-small $\hat{b}$ into a too-large $\hat{\kappa}$
- The bias depends on the span T, not on the number of observations
  - sampling daily instead of monthly does not remove it (Section 7)

Is Volatility Mean-Reverting? The VIX (1/2)





MFM_ch11_vix_ou

Is Volatility Mean-Reverting? The VIX (2/2)

- Left: VIX 1990–2026 on a log scale with the OU long-run level $e^{\hat{\theta}} = 18.2$
  - OU fit to $\ln \mathrm{VIX}$: $\hat{\kappa} = 5.10$ (SE 0.53), a half-life of 34 trading days
  - Spikes such as 82.7 (2020-03-16) fade within weeks
- Right: sample ACF of $\ln \mathrm{VIX}$ against the OU-implied $e^{-\hat{\kappa} k \Delta t}$
  - At lag 60, the autocorrelation is 0.63 in the data but 0.30 under OU
  - after the fast decay a slow component remains
- One OU factor is too simple
  - candidates: two-factor or long-memory models
  - rough volatility (Section 12) is a distinct alternative
- Mean-reverting volatility is the building block of Heston

MFM_ch11_vix_ou

Recap: Mean Reversion in Three Formulas

- Formulas to keep
  - Expected path: $\mathbb{E}[x_t \mid x_0] = \theta + (x_0 - \theta)e^{-\kappa t}$
  - Half-life: $\ln 2/\kappa$
  - Long-run spread: stationary standard deviation $\sigma/\sqrt{2\kappa}$
  - Estimation: AR(1) slope $b = e^{-\kappa \Delta t}$, hence $\hat{\kappa} = -\ln \hat{b}/\Delta t$
  - $\hat{\kappa}$ is biased upwards in short samples
- Check: monthly data give $\hat{b} = 0.95$. What are $\hat{\kappa}$ and the half-life?
  - Answer: $\hat{\kappa} = -12 \ln 0.95 = 0.616$ per year
  - half-life: $\ln 2/0.616 = 1.13$ years, about 13.5 months

Likelihood for Nonlinear Diffusions (1/2)

- **Euler pseudo-likelihood**: a Gaussian transition
  - $r_{t+\Delta} \mid r_t \sim N\left(r_t + a(r_t)\Delta,\ b^2(r_t)\Delta\right)$
  - $a(r)$, $b(r)$: the drift and diffusion of the rate r
  - its bias, of order Δ, is small with daily data
- **Exact CIR**: $2c\, r_{t+\Delta} \mid r_t$ is noncentral χ^2 Cox, Ingersoll & Ross (1985)
  - $4\kappa\theta/\sigma^2$ degrees of freedom, $c = 2\kappa/\left(\sigma^2(1 - e^{-\kappa\Delta})\right)$
  - The noncentrality parameter is $2c\, r_t e^{-\kappa\Delta}$
  - The Feller condition is equivalent to at least 2 degrees of freedom (Seminar 11, A7)

Likelihood for Nonlinear Diffusions (2/2)

- General diffusions: closed-form Hermite expansions of the transition density Aït-Sahalia (2002)
- CKLS (Chan–Karolyi–Longstaff–Sanders): $\mathrm{d}r = \kappa(\theta - r)\,\mathrm{d}t + \sigma r^{\gamma}\,\mathrm{d}W$ Chan, Karolyi, Longstaff & Sanders (1992)
  - nests Vasicek ($\gamma = 0$) and CIR ($\gamma = 1/2$)
  - estimated by Gaussian (Euler) quasi-likelihood
  - Tests on γ use sandwich standard errors Newey & West (1987)
  - the score covariance is heteroskedasticity- and autocorrelation-consistent (HAC)

Vasicek, CIR and CKLS on the T-Bill Rate, 1954–2007

Model	$\hat{\kappa}$ daily (SE)	$\hat{\kappa}$ monthly (SE)	γ
Vasicek, exact	0.158 (0.073)	0.201 (0.083)	0
Vasicek, Euler	0.158 (0.073)	0.199 (0.082)	0
CIR, exact	0.113 (0.060)	0.153 (0.070)	1/2
CIR, Euler	0.093 (0.060)	0.144 (0.069)	1/2
CKLS, Euler	0.080 (0.057)	0.144 (0.068)	0.77 (0.033) daily; 0.60 (0.149) monthly

- Data: daily DTB3 ($n = 13\,487$) and month-end values ($n = 647$)
- 2008–2026 is excluded (Appendix)
  - exact zeros and negative quotes lie outside the domain of these likelihoods
- Exact and Euler differ by less than one standard error
  - discretisation error is small next to the long-span uncertainty of $\hat{\kappa}$
- Volatility rises with the level (statistics: Appendix)
  - HAC Wald tests reject $\gamma = 0$ at both frequencies
  - they reject $\gamma = 1/2$ with daily but not with monthly data

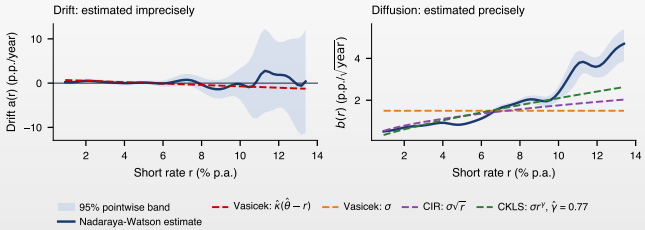
Nonparametric Drift and Diffusion (1/2)

- No functional form: kernel averages of the increments around each level x Stanton (1997)
  - $\Delta r_t = r_{t+\Delta} - r_t$ is the next increment
  - $w_t(x)$ is the weight of day t, large when r_t is close to x
  - $\hat{a}(x) = \sum_t w_t(x)\,\Delta r_t/\Delta$
  - $\hat{b}^2(x) = \sum_t w_t(x)\,(\Delta r_t)^2/\Delta$
  - Nadaraya–Watson weights $w_t(x) = K\left((r_t - x)/h\right) / \sum_s K\left((r_s - x)/h\right)$
  - K: Gaussian kernel; h: bandwidth

Nonparametric Drift and Diffusion (2/2)

- The moment conditions are exact only as $\Delta \to 0$
  - $\mathbb{E}[\Delta r \mid r = x]/\Delta = a(x) + O(\Delta)$
  - $\mathbb{E}[(\Delta r)^2 \mid r = x]/\Delta = b^2(x) + O(\Delta)$
- The diffusion converges at the infill rate, the drift only at the long-span rate
- The results hold for non-stationary but recurrent processes Bandi & Phillips (2003)
- Caution: nonparametric specification tests reject far too often with persistent data Pritsker (1998)

Nonparametric Drift and Diffusion of the T-Bill Rate



- Data: daily DTB3, 1954–2007, bandwidth $h = 0.44$ percentage points
  - Silverman's rule, $h = 1.06\,\hat{s}\,n^{-1/5}$, with $\hat{s}$ the standard deviation of the rate
- The 95% pointwise bands are heteroskedasticity-robust but ignore serial dependence
- Drift: zero lies inside the band at 97% of the levels
- Diffusion: the median relative SE is 8%
  - $b(r)$ rises from 0.5% to 4.7% (log-slope 0.90)
- No parametric shape fits at every level
  - share of levels inside the band: CKLS 35%, CIR 7%, Vasicek 5% (descriptive shares)

MFM_ch11_np_diffusion

Recap: Sampling Frequency and Identification

- Formulas to keep
  - Infill: $\sum_k (\Delta X_k)^2 \to \int_0^T b_t^2 \, \mathrm{d}t$
  - so the diffusion coefficient is learned from frequent sampling
  - Long span: $I(m) = T/\sigma^2$ and $\mathrm{Var}(\hat{\kappa}) \approx 2\kappa/T$
  - so drift and speed are learned only from more years
  - Bias: $\hat{\kappa}$ exceeds κ by roughly $4/T$
- Check: which standard error shrinks, that of $\hat{\sigma}$ or that of $\hat{\kappa}$?
  - setting: the same 20 years sampled daily instead of monthly (21 times more observations)
  - Answer: that of $\hat{\sigma}$, by about $\sqrt{21} \approx 4.6$ times
  - that of $\hat{\kappa}$ barely moves, since $2\kappa/T$ does not depend on the step

Moments of the Merton Return (1/2)

- Cumulants c_k of the return over Δt: $K(u) = \ln \mathbb{E}[e^{ur}]$ and $c_k = K^{(k)}(0)$
  - c_1 mean, c_2 variance, c_3 third central moment
  - cumulants of independent parts add
  - a Normal part has $c_k = 0$ for $k \geq 3$
  - The k-th cumulant of a compound Poisson sum is $\lambda \Delta t \, \mathbb{E}[Y^k]$
- Merton cumulants
  - $c_1 = m\Delta t + \lambda \Delta t \, \mu_J$
  - $c_2 = \sigma^2 \Delta t + \lambda \Delta t (\mu_J^2 + \sigma_J^2)$
  - $c_3 = \lambda \Delta t (\mu_J^3 + 3\mu_J \sigma_J^2)$
  - $c_4 = \lambda \Delta t (\mu_J^4 + 6\mu_J^2 \sigma_J^2 + 3\sigma_J^4)$

Moments of the Merton Return (2/2)

- Skewness $c_3/c_2^{3/2}$ is negative when $\mu_J < 0$
- Excess kurtosis c_4/c_2^2 is positive with any jump
- Excess kurtosis falls like $1/\Delta t$ as the horizon grows
  - jumps matter most at short horizons
- Example: $\sigma = 15\%$, $\lambda = 5$, $\mu_J = -3\%$ and $\sigma_J = 4\%$ (Seminar 11, A3)
  - the daily excess kurtosis is 17.6
  - the skewness is -2.07

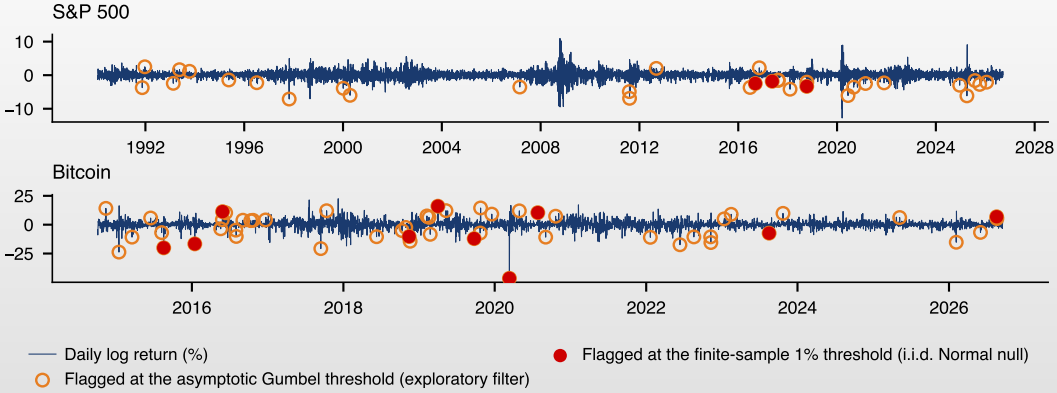
Detecting Jumps: the Lee–Mykland Test (1/2)

- Idea: a jump is a return that is huge **relative to the local volatility** Lee & Mykland (2008)
  - $L_t = r_t/\hat{\sigma}_t$, $\hat{\sigma}_t^2 = \frac{1}{K-2}\sum_{j=t-K+2}^{t-1} |r_j||r_{j-1}|$
  - $\hat{\sigma}_t^2$ is the bipower variation, robust to jumps
  - With daily data the window is $K = 16$ days, as recommended by the authors
- Under no jumps, the maximum of $|L_t|$ over n days has a Gumbel limit
  - Reject if $(|L_t| - C_n)/S_n > -\ln(-\ln(1-\alpha))$
  - the threshold is 4.60 at $\alpha = 1\%$ for the **normalised** statistic, not for $|L_t|$

Detecting Jumps: the Lee–Mykland Test (2/2)

- Normalising constants
  - $C_n = \frac{\sqrt{2 \ln n}}{c} - \frac{\ln \pi + \ln \ln n}{2c\sqrt{2 \ln n}}$, $S_n = \frac{1}{c\sqrt{2 \ln n}}$
  - $c = \sqrt{2/\pi}$: the mean of $|Z|$, which rescales the bipower estimate
  - $n =$ number of tested days
- The test controls, asymptotically, the probability of a false jump over the **whole** sample
  - not day by day

Days Flagged by the Lee–Mykland Test in the S&P 500 and Bitcoin (1/2)



MFM_ch11_jumps

Days Flagged by the Lee–Mykland Test in the S&P 500 and Bitcoin (2/2)

	S&P 500	Bitcoin
Flagged, asymptotic threshold	32	52
Per year	0.87	4.35
Negative (%)	84	52
In the top 10 moves	0	5
Above the calibrated 1% threshold	3	10

- Daily log returns; the asymptotic 1% threshold works here as an exploratory filter
- S&P 500: the largest moves (2008, 2020) are not flagged
  - they came in turbulent periods
  - Instead, a 1.2% move in a calm market is flagged (1993-10-14)
- The flagged days differ from the fitted Merton intensity. Possible explanations:
  - stochastic volatility
  - drift bursts (short explosive drift episodes)
  - low detection power in daily data

MFM_ch11_jumps

The Lee–Mykland Threshold in Finite Samples

- The asymptotic threshold assumes a precisely estimated local volatility; with $K = 16$ days, $\hat{\sigma}_t$ is noisy
  - in 2 000 i.i.d. samples from the Normal distribution, of the S&P 500 length, the threshold 4.60 flags at least one day in 96.6% of the samples
  - on average 3.43 flagged days per sample, although there are no jumps
- Calibrated threshold: the 99% quantile of the sample maximum of the statistic in the same samples (same n, same K)
  - S&P 500: 22.93 ($|L_t| > 11.6$), against 4.60 ($|L_t| > 6.2$); 3 flagged days (2016-09-09, 2017-05-17, 2018-10-10)
  - Bitcoin: 19.61 ($|L_t| > 10.6$); 10 flagged days
- Size under stochastic volatility without jumps: 200 Heston samples (VIX-based parameters, $\rho = -0.69$, 390 steps per day)
  - the asymptotic threshold is exceeded in 100% of the samples
  - the threshold calibrated under i.i.d. Normal returns is exceeded in 44% of the samples
  - with the 99% quantile of the Heston samples (48.7), no S&P 500 day is flagged
- With daily data, the flagged days are candidates for inspection, not jumps detected at the 1% level

What the VIX Identifies in Heston

- Under Q: $y_t = \frac{1}{\tau}\mathbb{E}^Q\left[\int_t^{t+\tau} v_s \, \mathrm{d}s \mid \mathcal{F}_t\right] = a + b\, v_t$
  - $\tau = 30/365$ (Seminar 11, A8)
  - $b = \frac{1-e^{-\kappa^Q \tau}}{\kappa^Q \tau} < 1$, $a = \theta^Q(1-b)$
  - y_t is **affine** in v_t, with Q-parameters
- What the regression on y_t recovers
  - κ: the P-speed, since an affine map keeps the dynamics of v_t
  - ρ: unchanged (correlation is scale-free), but only approximated by daily proxy changes
  - ξ is attenuated (derivation: next slide)
  - the long-run level mixes θ^Q and θ^P
- Proper estimation: one of two routes
  - fit the dynamics of y directly, with diffusion proportional to $\sqrt{y-a}$
  - or fit v_t and the observation equation jointly

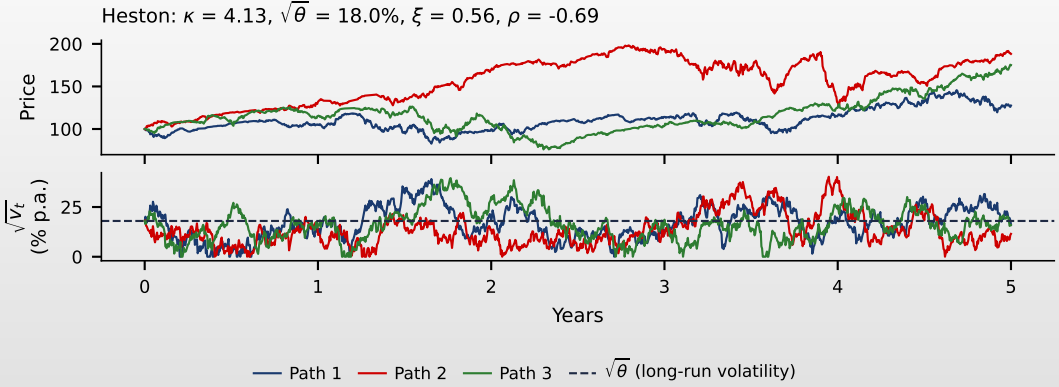
Where the Attenuation of ξ Comes From (1/2)

- Step 1: $y_t = a + b\,v_t$ with constants a, b, so $\mathrm{d}y_t = b\,\mathrm{d}v_t$ under P
  - $\mathrm{d}y_t = b\kappa(\theta^P - v_t)\,\mathrm{d}t + b\,\xi\sqrt{v_t}\,\mathrm{d}W_t^{(2)}$
- Step 2, drift: $b\kappa(\theta^P - v_t) = \kappa(\theta_y - y_t)$ with $\theta_y = a + b\,\theta^P$
  - the speed κ is unchanged
  - $\theta^P = (\theta_y - a)/b$
- Step 3, diffusion: the CIR regression assumes $\xi_y\sqrt{y_t}$
  - matching $\xi_y\sqrt{y_t} = b\,\xi\sqrt{v_t}$ gives $\xi_y = b\,\xi\sqrt{v_t/(a + b\,v_t)}$
  - at $v_t = \theta^Q$, $a + b\,\theta^Q = \theta^Q$, so $\xi_y = b\,\xi$

Where the Attenuation of ξ Comes From (2/2)

- Numbers: with $\kappa^Q = 5$, $b = 0.82$
  - the fitted $\hat{\xi} = 0.56$ corresponds to $\hat{\xi}/b = 0.68$ near $v_t = \theta^Q$
  - other κ^Q: Appendix
  - the correction $\hat{\xi}/b$: Appendix
- κ^Q is not identified from the VIX alone
  - joint P/Q estimation with option prices Chernov & Ghysels (2000)
  - or moments of realised variance Bollerslev & Zhou (2002)

Simulated Heston Paths (1/2)



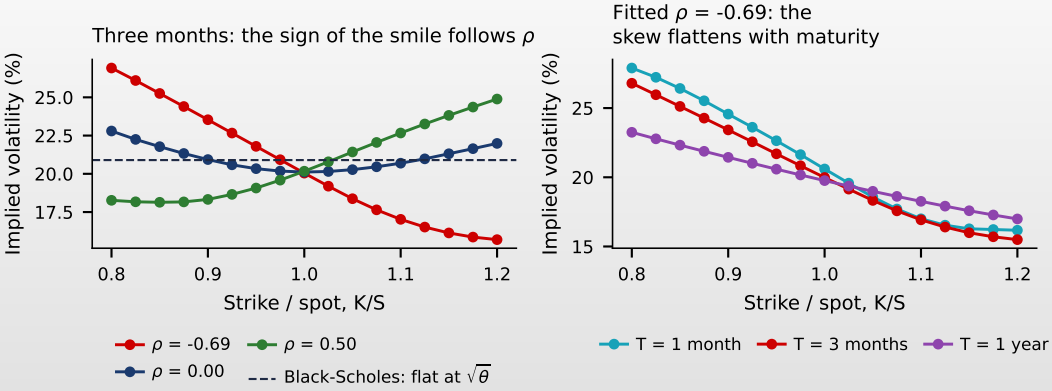
MFM_ch11_heston_paths

Simulated Heston Paths (2/2)

- Three five-year paths: top, the price; bottom, the volatility $\sqrt{v_t}$
  - $\hat{\kappa}$, $\hat{\xi}$ (uncorrected) and $\hat{\rho}$ come from the VIX
  - θ is the realised S&P 500 variance
- Volatility clustering now comes from the model
  - Price falls coincide with volatility spikes
  - Simulated return–variance correlation: -0.69 (leverage through ρ)
- Feller ratio $2\hat{\kappa}\theta/\hat{\xi}^2 = 0.86 < 1$: the variance can reach zero
  - the latent Euler state $\tilde{v}$ is ≤ 0 on 1.4% of the steps
  - full truncation: $\tilde{v}_{k+1} = \tilde{v}_k + \kappa(\theta - \tilde{v}_k^+)\Delta + \xi\sqrt{\tilde{v}_k^+\Delta}\, Z_k$, with $\tilde{v}^+ = \max(\tilde{v}, 0)$
  - $\tilde{v}^+$ enters the drift, the diffusion and the price step; $\tilde{v}^+$ is reported

MFM_ch11_heston_paths

A Preview of Chapter 12: the Volatility Smile (1/2)



MFM_ch11_heston_smile

A Preview of Chapter 12: the Volatility Smile (2/2)

- **Implied volatility**: the σ that makes the Black–Scholes price equal to a given option price
  - flat across strikes under GBM
  - here from Heston call prices (Monte Carlo, 100,000 antithetic paths)
- Left, three months to maturity
  - $\hat{\rho} = -0.69$, a downward **skew**: 26.9% at a strike of 80% of spot, 20.0% at the money, 15.7% at 120%
  - $\rho = 0$, a symmetric **smile**: 22.8% at 80% of spot, 20.1% at the money, 22.0% at 120%
  - $\rho = 0.50$, the skew reverses: implied volatility rises with the strike, from 18.3% to 24.9%
- Right: the skew flattens with maturity
  - one month: 27.9% – 16.2%
  - one year: 23.2% – 17.0%
- Illustration only: pricing needs κ^Q, θ^Q calibrated to option prices (Chapter 12)

MFM_ch11_heston_smile

Affine Jump-Diffusions

Affine jump-diffusions Duffie, Pan & Singleton (2000)

- Drift, variance, jump intensity affine in the state X_t
  - Vasicek, CIR, Merton, Heston and Bates
- $\mathbb{E}\left[e^{iu^\top X_T} \mid X_t\right] = \exp\left(\alpha(\tau) + \beta(\tau)^\top X_t\right)$
  - $i = \sqrt{-1}$, u: a real vector, $\tau = T - t$
  - α, β: solutions of Riccati ODEs
- Option prices by Fourier inversion
  - Simplest case: the Vasicek $A(\tau)$, $B(\tau)$ of Section 8

From GARCH to Diffusions

GARCH (Generalised Autoregressive Conditional Heteroskedasticity) models conditional variance in discrete data.

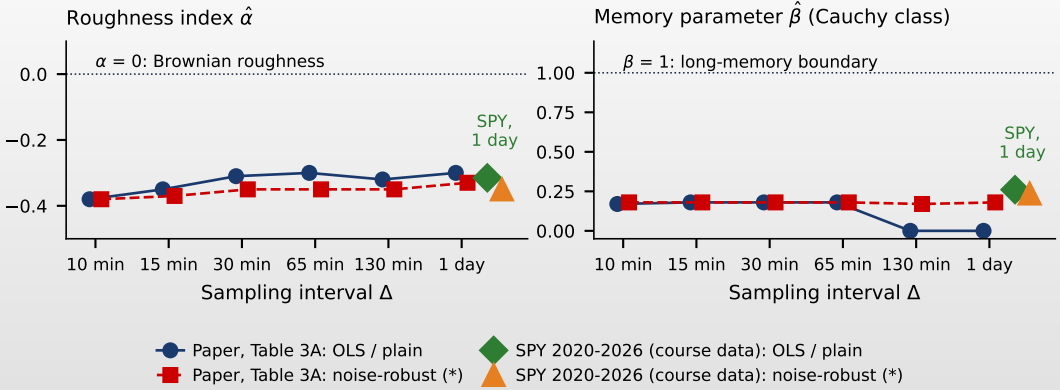
From GARCH to continuous time Nelson (1990)

- As the step shrinks, GARCH(1,1) converges to
  - $\mathrm{d}v_t = \kappa(\theta - v_t)\,\mathrm{d}t + \xi v_t\,\mathrm{d}W_t^{(2)}$
  - Like Heston, with v_t instead of $\sqrt{v_t}$
- $\kappa \approx (1-\alpha-\beta)/\Delta t$
  - α, β: GARCH(1,1) coefficients
  - daily $\alpha+\beta = 0.99$: $\kappa \approx 2.5$ per year (Chapter 5)
  - half-life: about 70 days
- GARCH and its diffusion limit are not asymptotically equivalent statistical experiments Wang (2002)
  - inference can differ even as $\Delta t \to 0$

Why a Negative ρ Skews Daily Returns

- Over one day $h = 1/252$: $R = \int_0^h \sqrt{v_s}\,\mathrm{d}W_s^{(1)}$
  - drift terms of order h are ignored
  - Itô's lemma on $\sqrt{v}$: $\mathrm{d}\sqrt{v_s} = \frac{1}{2}\xi\,\mathrm{d}W_s^{(2)} + O(\mathrm{d}s)$
  - so $\sqrt{v_s} \approx \sqrt{v_0} + \frac{1}{2}\xi W_s^{(2)}$
  - $R \approx \sqrt{v_0}\,W_h^{(1)} + \frac{1}{2}\xi \int_0^h W_s^{(2)}\,\mathrm{d}W_s^{(1)}$
- Third moment: the cross term gives $\mathbb{E}[R^3] \approx \frac{3}{2}\,\xi\rho\,v_0 h^2$
  - while $\mathrm{Var}(R) \approx v_0 h$
  - skewness $\approx \frac{3}{2}\,\rho\,\xi\sqrt{h}/\sqrt{v_0}$, a leading-order approximation
  - with $\hat{\rho} = -0.69$, $\hat{\xi} = 0.56$ and $\sqrt{v_0} = 18.0\%$: about -0.20
  - the simulation with 390 steps per day (Section 11): median -0.21

Case Study: SPY Against Table 3 of the Paper (1/2)



MFM_ch11_rough_vol

Case Study: SPY Against Table 3 of the Paper (2/2)

- Lines: E-mini S&P 500 futures, 2011–2014 (Bennedsen, Lunde & Pakkanen (2022), Table 3, Panel A)
  - Δ from 10 minutes to 1 day
- Markers: SPY at 1 day, 2020–2026, course data
  - * marks the noise-robust estimators
- $\hat{\alpha}$ barely moves with Δ (-0.38 to -0.30)
  - roughness is a property of the process
  - SPY falls in the same band
- $\hat{\beta}^* = 0.17$–0.18 at every Δ, far below 1
  - the plain $\hat{\beta}$ is unstable
- Limits: another contract, period and estimator
  - bars, not every trade
  - no pre-averaging
- Realised measures estimate integrated, not spot, variance

MFM_ch11_rough_vol

Your Turn: Replication and Core Deliverables

- Replicate first
  - the Merton jump share of variance: 71% (daily S&P 500)
  - the days flagged by the Lee–Mykland test at the asymptotic threshold: 0.87 per year
  - the bipower jump share: 2.8% (SPY, 5 minutes, Chapter 9)
- Core deliverables, feasible with the course data
  - SPY: bipower jump share and Lee–Mykland flags, with thresholds calibrated by simulation, at 5, 15 and 30 minutes
  - The EWMA standardisation of the mini-case, repeated on SPY at each frequency

Your Turn: Drift Bursts in Bitcoin and Report

- Do drift bursts occur in Bitcoin, which trades 24/7, more or less often than in SPY?
  - second question: are they followed by a price reversal?
  - Data: Bitcoin 5-minute bars, September 2024 – September 2026, 659 UTC days with at least 95% of the 288 bars (Chapter 9)
  - the SPY 5-minute bars on the same window
  - before the first run: pre-register the bandwidths, the null (sign-randomised returns) and the comparison
  - explain how the absence of a closing auction and of an overnight gap changes the test
- Optional extension: a bootstrap LR test of Heston (null, no jumps) against Bates
  - both likelihoods need filtering of the latent variance
  - e.g. the Markov Chain Monte Carlo (MCMC) approach of Eraker, Johannes & Polson (2003)
  - plan the computing time before committing to it
- Deliverables
  - code and a two-page report with every specification tried
  - a referee-style critique of your own result
- Declare your AI use in `AI_USE.md` and the errors you caught in `AI_ERRORS.md`
  - github.com/danpele/MFM/tree/main/project

From Brown to Bachelier, Einstein and Wiener



Louis Bachelier (1870–1946)

Photo: unknown author; public domain; Wikimedia Commons



Albert Einstein in 1905

Photo: Lucien Chavan / ETH Zürich (1905); public domain; Wikimedia Commons

- 1827: the botanist Robert Brown sees pollen particles move in an irregular zig-zag
  - the observation is made under the microscope
- Bachelier (1900): a random walk in continuous time for prices
  - the prices: French government bonds (rentes) and their options
  - the spread of the price grows with $\sqrt{t}$
  - the result comes five years before physics
- Einstein (1905): molecular collisions explain Brown's motion
  - the mean squared displacement is linear in t
- Wiener (1923): Brownian motion as a probability measure on continuous paths, the **Wiener process**

Back

Itô, Samuelson, Black, Scholes and Merton



Kiyosi Itô (1915–2008)

Photo: Konrad Jacobs (1970); CC BY-SA 2.0 de; Wikimedia Commons



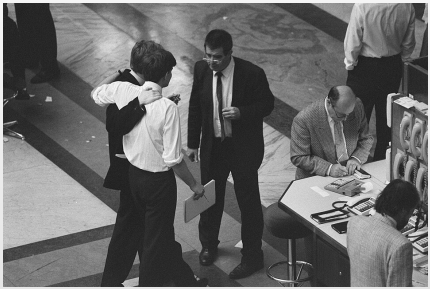
Paul A. Samuelson (1915–2009)

Photo: Bernard Gotfryd (1970–1975); public domain; Wikimedia Commons

- Itô (1944): the stochastic integral and its chain rule
  - the main working tool of this chapter
- Osborne (1959): log prices, not prices, behave like Brownian motion
- Samuelson (1965): geometric Brownian motion
- Black & Scholes (1973) and Merton (1973): an option can be replicated in continuous time
- Merton (1976) adds jumps, Heston (1993) random volatility
  - textbook treatment: Franke, Härdle & Hafner (2019), Part I

Back

A Crash in Pictures, Amsterdam 1987



Amsterdam Stock Exchange, 21 October 1987, after the crash of 19 October

Photo: Bart Molendijk / Anefo (1987); CC0; Wikimedia Commons

- The fall of 19 October 1987 (Section 9): 22.6% in one day
- Under a Normal distribution such a day would not be expected once in the age of the universe
  - prices gap
  - continuous models need jumps or fat-tailed innovations

Back

Data Cleaning and Technical Conditions

- Data: equity indices on weekdays, stale quotes removed
  - stale quote: unchanged close and no intraday range
  - genuine zero returns are kept
  - Bitcoin: 7 days a week
- Itô integral: left-point sums converge for regular integrands
  - e.g. continuous and adapted integrands
  - a general H is reached by approximation with simple processes
  - if only $\int_0^T H_t^2 \, \mathrm{d}t < \infty$ almost surely, $\int H \, \mathrm{d}W$ is merely a local martingale
  - the mean of this local martingale need not be zero
- SDE: unique strong solution if a, b are Lipschitz in x with at most linear growth Kloeden & Platen (1992)
  - square-root coefficients (CIR, Heston) are not Lipschitz at zero
  - they still have unique strong solutions (Yamada–Watanabe conditions)
- Donsker: convergence in distribution in the space of càdlàg paths Billingsley (1999)
  - the space carries the Skorokhod topology

Back

Short Rates, More Numbers (1/2)

- DTB3, 1954–2026: the rate ranged from -0.05% (2020-03-26) to 17.1% (1980-12-11)
  - 4.3% of the days were below 0.05%
  - The fitted Vasicek model gives a 6.1% long-run probability of a negative rate
  - 2008–2026: 13 exact zeros and 8 negative quotes
  - these quotes are outside the domain of the CIR and CKLS likelihoods
  - in this period, 17% of the days form a near-zero regime below 0.05%

Back

Short Rates, More Numbers (2/2)

- CKLS tests on γ, 1954–2007, HAC Wald with 11 (daily) and 6 (monthly) lags
  - $\gamma = 0$ rejected: daily 545, monthly 16
  - $\gamma = 1/2$ rejected daily: 67.0; 22.1 with 55 lags
  - $\gamma = 1/2$ not rejected monthly: 0.5 against 3.84
  - The Gaussian LR for $\gamma = 1/2$ rejects monthly (10.1)
  - the LR test ignores volatility clustering
  - Chan, Karolyi, Longstaff & Sanders (1992) found $\hat{\gamma} \approx 1.5$ for 1964–1989

Back

Vasicek Yield and the Observed Rate (1/2)

- Vasicek bond: 10-year yield 4.05%
  - with the Section 6 estimates, $\lambda = 0$ and $r = 3.99\%$
- The 10-year Treasury yield (FRED series DGS10) on 2026-09-18: 5.01%
  - DGS10 is a semiannual par yield
  - continuously compounded: $2\ln(1 + 5.01\%/2) = 4.95\%$
  - DTB3 is a 91-day discount rate: 3.99% is 4.07% continuously compounded
  - at this rate the model yield rises to 4.10%

Back

Vasicek Yield and the Observed Rate (2/2)

- The gap shrinks only from 0.96 to 0.85 percentage points
  - par versus zero-coupon yield is not adjusted
- The remaining gap cannot be attributed to a term premium
  - only the starting rate was converted, not the parameters estimated on the series
  - DTB3 is a 3-month proxy for the instantaneous rate
  - a structural comparison: convert the series before estimation and compare zero-coupon yields at the same maturity
- In the model, a positive term premium needs $\theta^Q > \hat{\theta}$, i.e. $\lambda < 0$: an implication of the model, not a result identified here
- Nonparametric estimation: Stanton (1997) also estimates the market price of interest-rate risk
- Specification test: Aït-Sahalia (1996)

Back

The VIX Correction by κ^Q and the Full Model Comparison

κ^Q	1	2.5	5	10
b	0.96	0.90	0.82	0.68
Local correction $\hat{\xi}/b$	0.58	0.62	0.68	0.82

	Data	GBM	Merton	Heston
Excess kurtosis	10.88	0.00	3.66	3.21
Skewness	−0.36	0.00	−0.32	−0.21
Hill statistic	2.56	4.64	2.70	2.92
ACF of $\lvert r\rvert$, lag 1	0.27	0.00	0.00	0.28
ACF of $\lvert r\rvert$, sum of lags 1–20	5.32	0.00	0.00	4.79

- $\hat{\xi}/b$ is exact only at $v_t = \theta^Q$
- Model columns: medians of 100 simulations
- A Heston model with a lower Feller ratio could reach the kurtosis
  - short-horizon excess kurtosis $\approx 3\xi^2/(2\kappa\theta)$
- Heston column: 390 Euler steps per day
  - with 780 steps, each median changes by less than 0.01

Back

Further Study: Reading

- Heavier tails: Lévy processes and α-stable distributions Cont & Tankov (2004), Pele (2014)
  - Crypto-assets form a distinct class by the shape of their return distributions Pele et al. (2023)
  - High-frequency information about Bitcoin volatility and tail risk Pele & Mazurencu-Marinescu-Pele (2019)
- Stochastic calculus
  - Shreve (2004), Ch. 5–6
  - Franke, Härdle & Hafner (2019), chapters on the Wiener process and stochastic integration
- Exercises with solutions on all models of this chapter: Borak, Härdle & López-Cabrera (2013)
  - the Heston model: Čížek, Härdle & Weron (2011)
  - applications: Härdle, Chen & Overbeck (2017)
- Monte Carlo methods: Glasserman (2003)
- Numerical SDEs: Kloeden & Platen (1992)

Back

Case Study, Model, Data and Method (1/4)

- **Model**: $\sigma_t = \xi e^{X_t}$ (eq. (1.3)), $X_t = \int_{-\infty}^{t} g(t-s)\, v_s \, \mathrm{d}W_s$ (eq. (1.5))
  - X is a Brownian semistationary (BSS) process with kernel g
  - v_s is the volatility-of-volatility process
  - the forecasts take $v_s \equiv 1$, a Gaussian X (Section 5)
- Gamma kernel $g(x) = x^{\alpha} e^{-\lambda x}$ (Example 1.2): roughness α, short memory
  - memory decays exponentially at rate λ
- Power kernel $g(x) = x^{\alpha}(1+x)^{-\gamma-\alpha}$ (eq. (1.8)): the Power-BSS model
  - $\beta = \gamma$ if $\gamma > 1$ (Remark 1.2)
  - $\beta = 2\gamma - 1$ if $\gamma \in (1/2, 1)$
- Gaussian alternative: the Cauchy class (Section 1.1.1)
  - $\rho(h) = (1 + |h|^{2\alpha+1})^{-\beta/(2\alpha+1)}$: roughness α and memory β set separately

Back

Case Study, Model, Data and Method (2/4)

- **Step 1**: α from the variogram of $\log \hat{\sigma}$, by OLS (eq. (2.1)) and NLLS
  - as in the SPY replication
  - the NLLS estimate averages $m = 10, \ldots, 20$ lags (Sections 2.1 and 3.1)
- **Step 2**: β, γ or λ by matching the model ACF to the empirical one
  - on the first $\lceil n^{1/3} \rceil$ lags (Sections 2.2 and 4.1)
- **Data**: E-mini S&P 500 futures, every trade, 2011–2014 (996 days)
  - trading window: 09:30–16:00 New York time
- Daily data on 1 944 liquid US stocks, 2003–2013 (Sections 4.1 and 4.2)

Back

Case Study, Model, Data and Method (3/4)

- $\hat{\sigma}_t^2$: pre-averaged bipower variation over Δ, divided by Δ
  - Δ from 10 minutes to 1 day (Sections 4 and 4.1)
- Bipower variation of the M returns r_j in a bin: $\mathrm{BV} = \frac{\pi}{2} \sum_{j=2}^{M} |r_j| \, |r_{j-1}|$
  - robust to jumps Barndorff-Nielsen & Shephard (2004)
- Pre-averaging first averages returns over short windows
  - aim: damp microstructure noise (Appendix A of the paper)
- Intraday seasonality removed with the flexible Fourier form of Andersen & Bollerslev (1997)

Back

Case Study, Model, Data and Method (4/4)

- **Forecasts**: variance summed over the next $h = 1, 2, 5, 10, 20$ steps
  - rolling window of 200 observations
  - out of sample: 2 January 2013 – 31 December 2014 (Section 5.1.1)
  - rivals: random walk, AR(5), AR(10), ARMA(1,1), log-HAR Corsi (2009), ARFIMA(0,d,0), RFSV
  - log-HAR: heterogeneous autoregressive model for log volatility
- **Forecast object**: $\mathrm{FO}_t(\Delta, h) = \sum_{k=1}^{h} \mathrm{BV}_{t+k\Delta}$ (eq. (5.3))
  - Gaussian log model: $\mathbb{E}[e^{X_{t+\Delta}} \mid \mathcal{F}_t] = \exp(m + s^2/2)$ (eq. (5.2))
  - m, s^2: the conditional mean and variance
- Losses over the out-of-sample origins
  - $\mathrm{MSE} = \mathrm{mean}\,(\widehat{\mathrm{FO}}_t - \mathrm{FO}_t)^2$
  - $\mathrm{QLIKE} = \mathrm{mean}\,(\log \widehat{\mathrm{FO}}_t + \mathrm{FO}_t / \widehat{\mathrm{FO}}_t)$
- **Model confidence set** (MCS) Hansen, Lunde & Nason (2011): contains the best model with probability 90% (or 75%)
  - membership is not a ranking: the models inside the MCS cannot be told apart at that level
- Scaling BV (Bipower Variation) by $M/(M-1)$ (Seminar 11, C3) shifts x_t by a constant
  - the shift: $\frac{1}{2}\ln(78/77) \approx 0.006$ for $M = 78$
  - the constant cancels in $\hat{\gamma}_2(k)$ and in the ACF

Back

Case Study, Results in the Paper (1/2)

$\log \sigma$, E-mini S&P 500 (Table 3, Panel A)	$\hat{\alpha}_{\mathrm{OLS}}$	$\hat{\alpha}^*_{\mathrm{NLLS}}$	$\hat{\beta}^*_{\mathrm{Cauchy}}$	$\hat{\lambda}^*_{\mathrm{Gamma}}$
$\Delta = 10$ minutes	-0.38	-0.38	0.18	0.02
$\Delta = 65$ minutes	-0.30	-0.35	0.18	0.02
$\Delta = 1$ day	-0.30	-0.33	0.18	0.02

- Rough at every scale and persistent: long memory together with roughness
  - rough: $\hat{\alpha} < 0$, far rougher than Brownian motion ($\alpha = 0$)
  - persistent: $\hat{\beta}^* < 1$
  - * = noise-robust
  - SPY comparison in the main text

Back

Case Study, Results in the Paper (2/2)

- 1 944 US stocks: both features look universal (Figures 4 and 5)
  - $\hat{\alpha}^*$ mostly between -0.45 and -0.30
  - $\hat{\beta}^*$ mostly between 0 and 0.4
  - no systematic difference across sectors
- Noise can fake roughness (Table 1, Panel C)
  - simulation: $\alpha = 0$ observed with noise-to-signal ratio 0.3 at $\Delta = 0.1$
  - the mean OLS estimate is -0.47
  - the mean NLLS estimate is 0.01
- Forecasts: modelling persistence pays intraday and at long horizons
  - 15-minute steps, $h = 20$, MSE ($\times 10^{8}$): 0.038 log-HAR, 0.020 RFSV, 0.019 Power-BSS (Table 4, Panel A)
  - daily steps, $h = 1$: RFSV is best, 0.493 ($\times 10^{9}$) (Table 5, Panel B)

Back

Case Study, Intraday Application, Limits and Open Questions (1/3)

- **Intraday application** on the SPY 5-minute bars of the course data
  - illustrative, not an exact replication
  - $\hat{\sigma}_t^2$ = bipower variation of the 5-minute returns in each bin of $\Delta = 30$ or 65 minutes
  - Divided by Δ and by the flexible Fourier seasonal factor Andersen & Bollerslev (1997)
- OLS and NLLS $\hat{\alpha}$, Cauchy $\hat{\beta}^*$ and Gamma-BSS $\hat{\lambda}^*$ against Table 3, Panel A
- log-HAR, RFSV and BSS forecasts against Table 5
  - losses compared: MSE, QLIKE (Quasi-LIKElihood, loss function) and MCS
- Data limit: bars instead of every trade
  - the pre-averaged estimator (Appendix A of the paper) is not feasible
  - the noise is larger
  - only 13 returns per 65-minute bin

Back

Case Study, Intraday Application, Limits and Open Questions (2/3)

- **Limits stated by the authors**
  - Realised measures are noisy estimates of integrated variance
  - The conclusions concern them, not the latent spot volatility (Sections 4.1, 4.2 and 6)
  - No rule for the NLLS bandwidth m (Sections 2.1 and 3.1)
  - NLLS is too costly for rolling forecasts
  - These forecasts use the OLS $\hat{\alpha}$ (Section 5.1.1)
  - Forecasts assume a Gaussian X
  - A non-Gaussian volatility of volatility is left out of scope (Section 5; footnote 3)

Back

Case Study, Intraday Application, Limits and Open Questions (3/3)

- **Open question**: why does long memory improve even short-horizon forecasts?
  - The authors leave it to further research (Section 5.1.2)
  - Rolling $\hat{\alpha}$ is less rough in 2007–2008 and 2011 (Figure 3)
  - windows: 65 minutes, 2005–2014
  - median $\hat{\alpha}$: -0.36
  - $\hat{\alpha}$ rises with log volume, slope 0.0091 (Figure 6)
  - Does Bitcoin (5-minute bars), a market that never closes, fit this pattern?

Back

The Discovery Loop, Reusable Prompts (1/2)

- Tools: an AI assistant (an LLM such as Claude, ChatGPT, Gemini or GitHub Copilot)
  - literature search: Semantic Scholar, Elicit, Google Scholar
- **1. Literature**: tabulate estimates by frequency
  - check every DOI on Crossref (the title must match)
  - `Tabulate published estimates of the jump share of equity-index variance by sampling frequency (daily, 5-minute, tick) and method; give the exact number, the table it comes from, and the DOI.`
- **2. Competing hypotheses**, with distinct testable implications
  - `For three processes that give a daily excess kurtosis near 11 (Merton jumps, Heston volatility, drift bursts), derive the implications for returns standardised by ex-ante volatility, for bipower versus realised variance, and for the drift just before extreme moves.`
- **3. Data and code**: replicate the chapter's estimate first
  - `Write Python that reads the S&P 500 daily closes of the course and fits the Merton model by maximum likelihood (Poisson mixture truncated at 12 jumps, several starting points).`

Back

The Discovery Loop, Reusable Prompts (2/2)

- **4. Robustness**, pre-registered
  - `Write a pre-registered plan: standardise returns by EWMA and by GARCH volatility known at t-1; split 1990-2007 and 2008-2026; test Heston (no jumps, the null) against Bates (SVJ); obtain LR critical values by parametric bootstrap under the Heston null.`
- **5. Critique**: ask the assistant to attack the inference
  - `Act as a referee: why is the LR test of lambda = 0 non-standard, what is the correct reference distribution, and could a Poisson mixture with frequent small jumps simply be a fat-tailed innovation?`
- **6. Report**: contents
  - all frequencies and filters
  - bootstrap critical values
  - code and the AI log

Back

Further Study on Quantinar

Advanced courses on quantinar.com

- Applied Computational Finance
- XFG (Quantlet prefix of Applied Quantitative Finance, Härdle, Chen & Overbeck) Advanced Methods in Quantitative Finance

Code and data

- Quantlets: github.com/danpele/MFM/tree/main/Quantlets/Ch_11

Back

Acronyms Used in This Material (1/4)

- **ACF**: AutoCorrelation Function
- **AI**: Artificial Intelligence
- **AR**: AutoRegressive model
- **ARFIMA**: AutoRegressive Fractionally Integrated Moving Average (model)
- **ARMA**: AutoRegressive Moving Average (model)
- **BET**: Bucharest Exchange Trading (index)
- **BSS**: Brownian SemiStationary (process)
- **BV**: Bipower Variation
- **BVB**: Bursa de Valori București — *the Bucharest Stock Exchange*
- **CBOE**: Chicago Board Options Exchange
- **CI**: Confidence Interval
- **CIR**: Cox–Ingersoll–Ross (model)

Acronyms Used in This Material (2/4)

- **CKLS**: Chan–Karolyi–Longstaff–Sanders (model)
- **CLT**: Central Limit Theorem
- **CME**: Chicago Mercantile Exchange
- **DGS10**: 10-Year Treasury Constant Maturity Rate (FRED series code)
- **DOI**: Digital Object Identifier
- **DTB3**: 3-Month Treasury Bill: secondary market rate (FRED series code)
- **ETF**: Exchange-Traded Fund
- **EWMA**: Exponentially Weighted Moving Average
- **FRED**: Federal Reserve Economic Data
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GBM**: Geometric Brownian Motion
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors)
- **HAR**: Heterogeneous AutoRegressive (model)

Acronyms Used in This Material (3/4)

- **JB**: Jarque–Bera (test)
- **LLM**: Large Language Model
- **LR**: Likelihood Ratio (test)
- **MCMC**: Markov Chain Monte Carlo
- **MCS**: Model Confidence Set
- **MLE**: Maximum Likelihood Estimation
- **MSE**: Mean Squared Error
- **NLLS**: Non-Linear Least Squares
- **ODE**: Ordinary Differential Equation
- **OLS**: Ordinary Least Squares
- **OU**: Ornstein–Uhlenbeck (process)
- **PDE**: Partial Differential Equation

Acronyms Used in This Material (4/4)

- **QLIKE**: Quasi-LIKElihood (loss function)
- **QQ**: Quantile–Quantile (plot)
- **RFSV**: Rough Fractional Stochastic Volatility (model)
- **SDE**: Stochastic Differential Equation
- **SE**: Standard Error
- **SPY**: ticker of the SPDR S&P 500 ETF Trust
- **SVJ**: Stochastic Volatility with Jumps (model)
- **US**: United States
- **UTC**: Coordinated Universal Time (Temps Universel Coordonné)
- **VaR**: Value-at-Risk
- **VIX**: Cboe Volatility Index
- **XFG**: Quantlet prefix of Applied Quantitative Finance (Härdle, Chen & Overbeck)

References (1/5)

- Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9(2), 385–426.
- Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1), 223–262.
- Andersen, L. (2008). Simple and efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance*, 11(3), 1–42.
- Andersen, T.G., Benzoni, L., Lund, J. (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance*, 57(3), 1239–1284.
- Andersen, T.G., Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2–3), 115–158.
- Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure*, 17, 21–86.
- Bandi, F.M., Phillips, P.C.B. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71(1), 241–283.
- Barndorff-Nielsen, O.E., Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1), 1–37.
- Bates, D.S. (1996). Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options. *Review of Financial Studies*, 9(1), 69–107.
- Bennedsen, M., Lunde, A., Pakkanen, M.S. (2022). Decoupling the short- and long-term behavior of stochastic volatility. *Journal of Financial Econometrics*, 20(5), 961–1006.
- Billingsley, P. (1999). Convergence of Probability Measures, 2nd ed. Wiley.
- Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654.
- Bollerslev, T., Zhou, H. (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109(1), 33–65.

References (2/5)

- Borak, S., Härdle, W.K., López-Cabrera, B. (2013). Statistics of Financial Markets: Exercises and Solutions, 2nd ed. Springer.
- Broadie, M., Kaya, Ö. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2), 217–231.
- Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47(3), 1209–1227.
- Chernov, M., Ghysels, E. (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, 56(3), 407–458.
- Christensen, K., Oomen, R.C.A., Podolskij, M. (2014). Fact or friction: jumps at ultra high frequency. *Journal of Financial Economics*, 114(3), 576–599.
- Christensen, K., Oomen, R.C.A., Renò, R. (2022). The drift burst hypothesis. *Journal of Econometrics*, 227(2), 461–497.
- Čížek, P., Härdle, W.K., Weron, R. (Eds.) (2011). Statistical Tools for Finance and Insurance, 2nd ed. Springer.
- Comte, F., Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4), 291–323.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- Cont, R., Tankov, P. (2004). Financial Modelling with Jump Processes. Chapman & Hall/CRC.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174–196.
- Cox, J.C., Ingersoll, J.E., Ross, S.A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2), 385–407.
- Duffie, D., Pan, J., Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6), 1343–1376.
- Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8), 549–560.
- Eraker, B., Johannes, M., Polson, N. (2003). The impact of jumps in volatility and returns. *Journal of Finance*, 58(3), 1269–1300.

References (3/5)

- Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, 54(1), 173–182.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). Statistics of Financial Markets: An Introduction, 5th ed. Springer.
- Gatheral, J., Jaisson, T., Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6), 933–949.
- Girsanov, I.V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability and Its Applications*, 5(3), 285–301.
- Glasserman, P. (2003). Monte Carlo Methods in Financial Engineering. Springer.
- Hansen, P.R., Lunde, A., Nason, J.M. (2011). The model confidence set. *Econometrica*, 79(2), 453–497.
- Härdle, W.K., Chen, C.Y.-H., Overbeck, L. (Eds.) (2017). Applied Quantitative Finance, 3rd ed. Springer.
- Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2), 327–343.
- Higham, D.J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3), 525–546.
- Honoré, P. (1998). Pitfalls in estimating jump-diffusion models. Working paper, SSRN.
- Itô, K. (1944). Stochastic integral. *Proceedings of the Imperial Academy*, 20(8), 519–524.
- Kac, M. (1949). On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society*, 65(1), 1–13.
- Kloeden, P.E., Platen, E. (1992). Numerical Solution of Stochastic Differential Equations. Springer.

References (4/5)

- Lee, S.S., Mykland, P.A. (2008). Jumps in financial markets: a new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6), 2535–2563.
- Lord, R., Koekkoek, R., van Dijk, D. (2010). A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2), 177–194.
- Merton, R.C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1), 141–183.
- Merton, R.C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1–2), 125–144.
- Milstein, G.N. (1975). Approximate integration of stochastic differential equations. *Theory of Probability and Its Applications*, 19(3), 557–562.
- Nelson, D.B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics*, 45(1–2), 7–38.
- Newey, W.K., West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Osborne, M.F.M. (1959). Brownian motion in the stock market. *Operations Research*, 7(2), 145–173.
- Pele, D.T. (2014). A SAS approach for estimating the parameters of an alpha-stable distribution. *Procedia Economics and Finance*, 10, 68–77.
- Pele, D.T., Mazurencu-Marinescu-Pele, M. (2019). Using high-frequency entropy to forecast Bitcoin's daily value at risk. *Entropy*, 21(2), 102.
- Pele, D.T., Wesselhöfft, N., Härdle, W.K., Kolossiatis, M., Yatracos, Y.G. (2023). Are cryptos becoming alternative assets?. *European Journal of Finance*, 29(10), 1064–1105.

References (5/5)

- Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies*, 11(3), 449–487.
- Samuelson, P.A. (1965). Rational theory of warrant pricing. *Industrial Management Review*, 6(2), 13–39; reprinted in *Henry P. McKean Jr. Selecta* (2015), Birkhäuser.
- Shreve, S.E. (2004). Stochastic Calculus for Finance II: Continuous-Time Models. Springer.
- Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, 52(5), 1973–2002.
- Tang, C.Y., Chen, S.X. (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149(1), 65–81.
- Uhlenbeck, G.E., Ornstein, L.S. (1930). On the theory of the Brownian motion. *Physical Review*, 36(5), 823–841.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), 177–188.
- Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *Annals of Statistics*, 30(3), 754–783.
- Wiener, N. (1923). Differential-space. *Journal of Mathematics and Physics*, 2(1–4), 131–174.
- Yu, J. (2012). Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics*, 169(1), 114–122.