

Modelling Financial Markets

Chapter 2: Market Efficiency — From EMH to Adaptive Markets



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 Past Prices and Public Information Predict Returns?



Eugene F. Fama. Photo: Bengt Nyman (2013); CC BY 2.0; Wikimedia Commons

"A market in which prices always 'fully reflect' available information is called 'efficient.'"
— *Fama (1970)*

- Lag-1 autocorrelation of daily returns: -0.099 for the S&P 500 (Standard & Poor's) and 0.109 for the BET (Bucharest Exchange Trading, index), both significant
  - Bitcoin: -0.022, inside the robust test band, so we do not reject $\rho_1 = 0$
- Yesterday's return explains 0.98% of the variance of today's return for the S&P 500 and 1.19% for the BET
- Dividend–price ratio, 680 monthly forecasts for the S&P 500: out-of-sample $R^2 = -0.51\%$
  - the forecast is worse than the historical mean
- Of 2 000 random trading rules, 5.2% are "significant" at 5%
  - the best one has $|t| = 3.30$, although it is pure chance

▸ Central chart: Lag-1 Autocorrelation Across Markets

Chapter outline

- Can past prices, news or other public information predict returns after adjusting for risk and costs?
- Sections 1–5 contain the core material and end with a short check.
- The further-study topics contain the derivations, the other tests and the project topics.
- Detailed examples and advanced topics are in Further Topics, after the Synthesis
  - the footer button leads there; the Back button returns to the starting slide

- 1. Returns and information
- 2. Autocorrelation and variance ratios
- 3. Out-of-sample predictability
- 4. Event studies
- 5. Multiple testing and anomalies
- Topics for further study

What you should be able to do

- derive the random-walk restrictions and the variance ratio
- compute the lag-1 autocorrelation and the variance ratio robustly
- compute the out-of-sample R^2 and the Clark–West test
- compute the cumulative abnormal return in an event study
- interpret an anomaly after the Holm correction
- Further Topics and Appendix: reading for familiarity

1. Returns and information

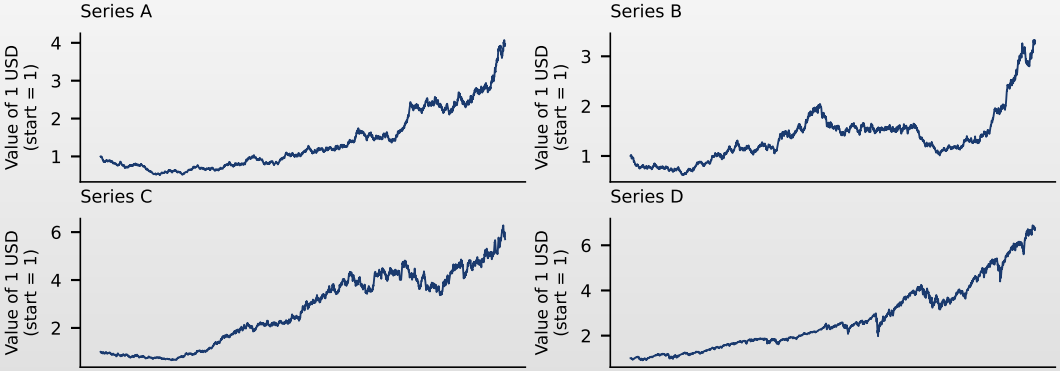
- **RW1**: Random Walk 1 — independent, identically distributed increments.
- **RW2**: Random Walk 2 — independent increments, possibly with different distributions.
- **RW3**: Random Walk 3 — uncorrelated increments, possibly dependent.

- The question: what does it mean, precisely, that past prices contain no information?
  - information about future returns
- Which random-walk version we test matters: a test built under RW1 can reject only because volatility varies over time.
- Every test also needs a model of expected returns, so a rejection does not by itself prove inefficiency.

A Real Price Series among Simulated Random Walks (1/2)

S&P 500: Standard & Poor's 500.

One of these is the S&P 500 (2010-2026); three are simulated random walks with the same drift and volatility



MFM_ch2_random_walk_game ▶ Interactive

A Real Price Series among Simulated Random Walks (2/2)

- Which panel is the S&P 500? Vote: A, B, C or D?

MFM_ch2_random_walk_game

A Real Price Series among Simulated Random Walks (2/2)

- Which panel is the S&P 500? Vote: A, B, C or D?
- Answer: **D**.
- Roberts (1959): simulated random walks produce "patterns" by chance
- We need statistical tests, not visual inspection
- Bachelier (1900) and Samuelson (1965): properly anticipated prices fluctuate randomly
  - so unpredictability is the mark of a well-functioning market

MFM_ch2_random_walk_game

Efficiency and the Information Set

- Fama (1970): a market is **efficient** if prices "fully reflect" all available information $\mathcal{I}_t$
  - the Efficient Market Hypothesis (EMH)
  - portrait and surveys: ▸ Appendix
  - precursors: ▸ Appendix
  - formally: no strategy using $\mathcal{I}_t$ earns **expected** returns above the fair (risk-adjusted) return

Form	Information set $\mathcal{I}_t$	Typical test (this chapter)
Weak	Past prices and returns	Autocorrelation, variance ratio; in the appendix: runs and long-memory tests
Semi-strong	All public information (news, reports)	Event studies
Strong	Public and private (insider) information	Returns of insiders and fund managers

- The strong form implies the semi-strong form, which implies the weak form
- Risk-adjusted returns with factor models follow in Chapter 3

The Joint-Hypothesis Problem

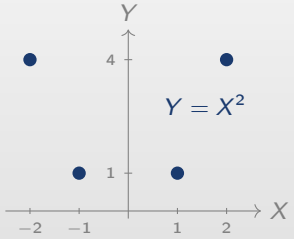
CAPM: Capital Asset Pricing Model.

- "Fair return" requires a model of expected returns
  - e.g. CAPM, Chapter 3
- Every test of efficiency is a **joint** test of efficiency and of the asset-pricing model Fama (1991)
  - A rejection may mean an inefficient market **or** a wrong model of risk
  - Predictable **expected** returns are allowed if they compensate time-varying risk
- At daily horizons expected returns are tiny
  - so the choice of model matters little
  - weak-form tests are the cleanest
- Efficiency can never be "proved", only not rejected by a given test on given data

Correlation and dependence in returns

- "The past contains no information" can mean zero correlation or independence: the two differ.
- Lag-k autocorrelation is the correlation between a return and the return k periods earlier.
- Zero correlation excludes linear association, but may leave dependence in squared returns.

- Toy example: $X = -2, -1, 1, 2$, equally likely; $Y = X^2$.
- $\mathbb{E}[X] = 0$ and $\mathbb{E}[XY] = (-8 - 1 + 1 + 8)/4 = 0$, so the correlation is zero.
- Yet knowing X determines Y exactly: they are dependent.



The Random Walk: the Model

$$p_t = \mu + p_{t-1} + \varepsilon_t, \qquad p_t = \ln P_t$$

- P_t: the index level at the close of day t.
  - p_t: the log price.
  - μ: the drift (mean growth per day).
  - ε_t: the day's shock, with mean zero and variance σ^2.
- By repeated substitution: $p_t = p_0 + \mu t + \sum_{s=1}^{t} \varepsilon_s$.
  - the log price is a linear trend plus the sum of all past shocks.
- Equivalently, the log return $r_t = p_t - p_{t-1} = \mu + \varepsilon_t$.
- The best forecast: $\mathbb{E}[p_{t+h} \mid \mathcal{I}_t] = p_t + \mu h$, with $\mathcal{I}_t = \{p_t, p_{t-1}, \dots\}$.
  - it holds if the shocks have conditional mean zero, $\mathbb{E}[\varepsilon_{t+1} \mid \mathcal{I}_t] = 0$: today's price plus the drift.
  - RW3 (zero autocorrelations only) is not enough.
  - Under RW3, the past does not, in general, improve the best *linear* forecast.
- With uncorrelated shocks of variance σ^2: $\mathrm{Var}(p_{t+h} - p_t) = h\sigma^2$.
  - the variance grows linearly with the horizon.
  - the basis of the variance ratio in section 2.
- RW1, RW2 and RW3 differ only in what they assume about the shocks ε_t Campbell, Lo & MacKinlay (1997).

RW1, RW2, RW3: the restrictions on returns

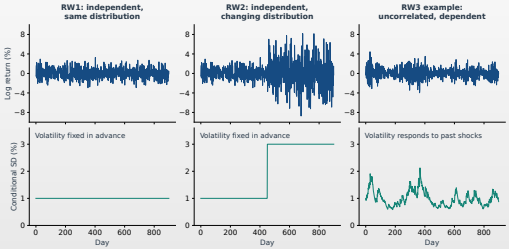
- **i.i.d.**: independent and identically distributed.
- **GARCH**: Generalized AutoRegressive Conditional Heteroskedasticity.
- z_t: a shock with the standard Normal distribution $N(0,1)$.
  - in GARCH(1,1), $\omega > 0$ and $\alpha, \beta \geq 0$.

	Condition on the shocks	Illustrative return model
RW1	ε_t i.i.d., mean 0, variance σ^2	$r_t = \mu + 0.01 z_t$, z_t i.i.d. $N(0,1)$
RW2	ε_t independent; the variance σ_t^2 may vary over time	$r_t = \mu + \sigma_t z_t$, $\sigma_t = 1\%$ then 3%, fixed in advance
RW3	$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$ for every $k \geq 1$	GARCH: $\varepsilon_t = \sigma_t z_t$, $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$

- RW1 implies RW2: i.i.d. returns are in particular independent.
- RW2 implies RW3: independent returns are in particular uncorrelated, provided their variance exists.
- A GARCH model keeps zero correlation but violates RW2.
  - the size of today's move depends on past shocks.
- RW3 does not make tomorrow's mean unpredictable: a nonlinear function of the past can forecast it.

Three models for returns and volatility

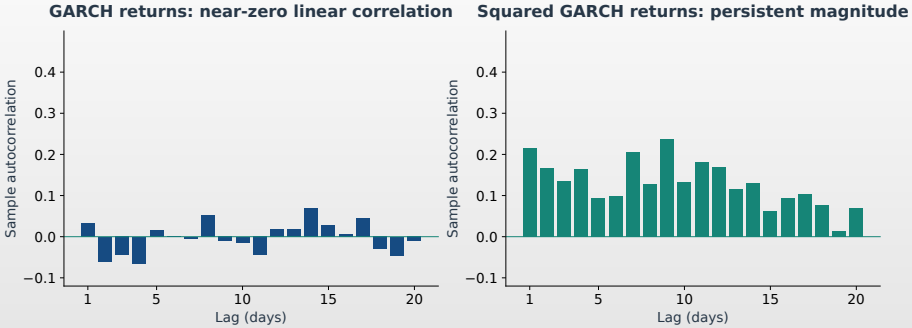
SD: Standard Deviation.



- Three simulated series with mean zero.
  - daily returns on top; below, the conditional standard deviation used in the simulation.
- RW2 volatility is fixed in advance, while GARCH volatility responds to past shocks.
  - 1% then 3%, from the middle of the sample.
- Volatility clustering: large moves tend to be followed by large moves.
- The RW2 series also has a turbulent period although its returns are independent.
  - The chart alone cannot tell a regime change fixed in advance from volatility clustering.

▶ Interactive

GARCH: returns versus squared returns



- Left: the autocorrelations of simulated GARCH returns are close to zero at all lags.
- Right: the autocorrelations of squared returns are positive up to lag 20.
  - a day with a large move is followed by other days with large moves.

▶ Interactive

Check: returns and information

- Can returns remain independent if their standard deviation moves from 1% to 3% on a date fixed in advance?
- Does zero autocorrelation establish independence?
- Why does rejecting an efficiency test not, on its own, prove that the market is inefficient?

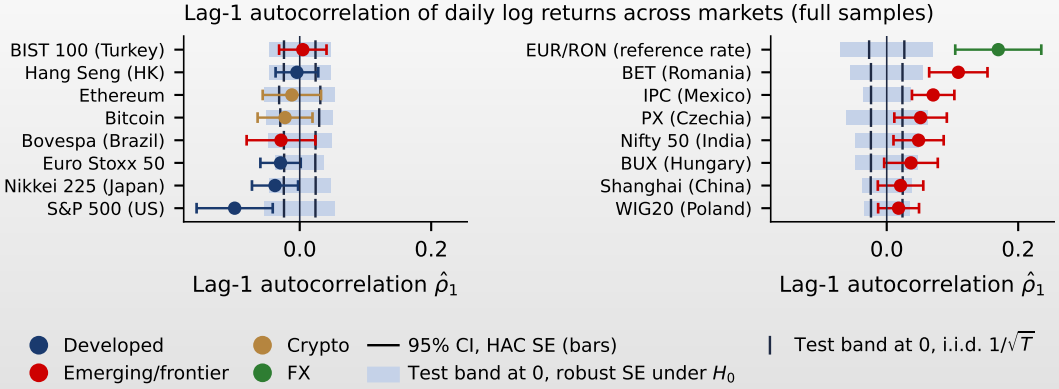
Answers: returns and information

- Yes. The series satisfies RW2 but violates RW1, because the distribution is no longer identical.
- No. In the GARCH example, return autocorrelation is zero, but squared returns are dependent.
- Every efficiency test also tests the model of expected returns.
- A rejection can come from a wrong risk model (the joint-hypothesis problem).

2. Autocorrelation and variance ratios

- The question: does today's return predict tomorrow's in equity, crypto and currency markets?
- The lag-1 autocorrelation ρ_1 measures linear dependence between consecutive days.
  - The variance ratio accumulates the autocorrelations up to horizon q.
- Markets: 13 stock indices, Bitcoin, Ethereum and the EUR (euro)/RON reference rate.
  - developed and emerging, including the BET.
- Why this selection: predictability is more likely:
  - where liquidity is thin (BET, crypto).
  - or where the exchange rate is managed (EUR/RON).
- Samples: all to 18 September 2026.
  - equities from January 2000, EUR/RON from July 2005.
  - Bitcoin from September 2014 and Ethereum from January 2016.

Lag-1 Autocorrelation Across Markets (1/3)



MFM_ch2_autocorrelation ▶ Interactive

Lag-1 Autocorrelation Across Markets (2/3)

SE: Standard Error; **HAC**: Heteroskedasticity and Autocorrelation Consistent; **CI**: Confidence Interval.

- Each dot is the lag-1 autocorrelation $\hat{\rho}_1$ of one market's daily log returns, over its full sample.
- The coloured bar is the 95% confidence interval $\hat{\rho}_1 \pm 1.96\,\mathrm{SE}_{\mathrm{HAC}}$.
  - $\mathrm{SE}_{\mathrm{HAC}}$: the standard error of the estimator, valid also when $\rho_1 \neq 0$ (formula).
- The light band is the robust test band $0 \pm 1.96\,\mathrm{SE}_0$; the ticks mark the i.i.d. band $0 \pm 1.96/\sqrt{T}$.
  - SE_0: derived under H_0 (unpredictable mean), test construction; T: the number of daily returns.
- Reading rule: we reject $\rho_1 = 0$ at 5% if the dot lies outside the robust test band.
  - Nikkei 225 and PX (the Prague Stock Exchange index): the HAC interval excludes 0, but $\hat{\rho}_1$ stays inside the test band, so we do not reject.
- In all 16 series, SE_0 exceeds the i.i.d. value; for the S&P 500 ($T = 6\,717$) it is 0.027 against 0.012.
- BIST (Borsa İstanbul, BIST 100 index), WIG20 (Warszawski Indeks Giełdowy 20, the 20 largest companies), BUX (Budapesti Értéktőzsde index), PX, IPC (Índice de Precios y Cotizaciones) and BET: the stock indices of Istanbul, Warsaw, Budapest, Prague, Mexico and Bucharest.
  - FX is foreign exchange (the EUR/RON reference rate).

Lag-1 Autocorrelation Across Markets (3/3)

- BET ($\hat{\rho}_1 = 0.109$), IPC (0.071), Nifty 50 (0.048) and EUR/RON (0.170): positive and significant autocorrelations.
- S&P 500: $\hat{\rho}_1 = -0.099$, negative and significant.
- Bitcoin ($\hat{\rho}_1 = -0.022$) and Ethereum (-0.012): inside the robust test band, so we do not reject $\rho_1 = 0$.
- For the BET, persistence can come from thin trading in some stocks, whose prices absorb news with a delay.
- For EUR/RON, it can come from the managed float, in which the central bank smooths exchange-rate moves.
- For the S&P 500, the reversal can come from short-term overreaction and liquidity provision.
- The effect is small: $\hat{\rho}_1^2$ is 0.98% for the S&P 500 and 1.19% for the BET.
  - the share of today's return variance explained by yesterday's return.
- Both robust standard errors require a finite fourth moment, which Chapter 1 questions.
- Type of evidence: description (it shows what happened in the data, without testing a cause).
  - The proposed mechanisms (thin trading, managed float, overreaction) are not tested here.

MFM_ch2_autocorrelation

The Variance Ratio: Why and What It Measures

VR: Variance Ratio.

- Why another test: $\hat{\rho}_1$ looks at a single lag.
  - small autocorrelations at many lags can each pass the test yet accumulate over longer horizons.
- **Variance ratio** over a horizon of q days Lo & MacKinlay (1988): $\mathrm{VR}(q) = \dfrac{\mathrm{Var}(r_t + \cdots + r_{t-q+1})}{q\,\mathrm{Var}(r_t)}$.
  - equal to 1 without autocorrelation.
  - $q = 5$: one trading week.
- Without autocorrelation, $\widehat{\mathrm{VR}}(q)$ tends to 1 even when volatility varies over time (RW2, RW3).
- For a covariance-stationary series, $\mathrm{VR}(q) = 1 + 2\sum_{k=1}^{q-1}\left(1 - \frac{k}{q}\right)\rho_k$.
  - a weighted sum of autocorrelations, with weights that fall with the lag k.
- $\mathrm{VR}(q) > 1$ signals persistence, $\mathrm{VR}(q) < 1$ reversal (mean reversion).
- For $q = 2$, $\mathrm{VR}(2) = 1 + \rho_1$: on the S&P 500, $1 - 0.099 = 0.901$, equal to $\widehat{\mathrm{VR}}(2)$ estimated directly.

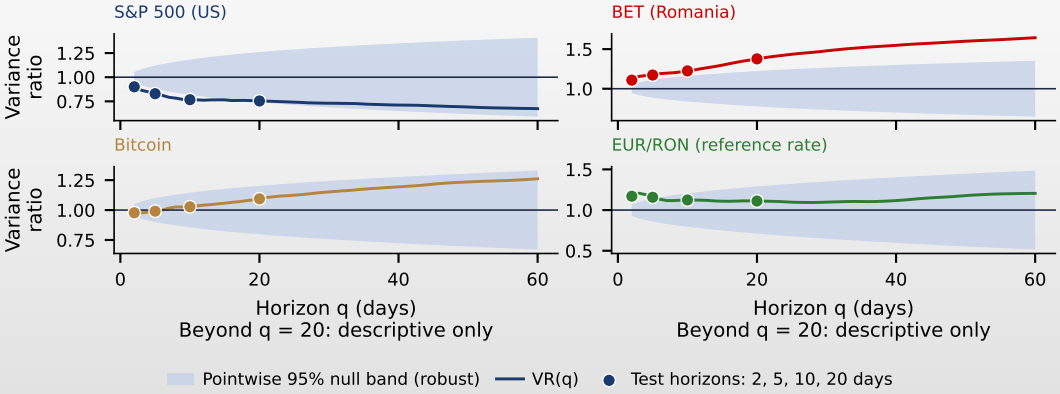
The robust variance-ratio band

- The robust 95% band is centred at 1, the value under the null.
  - if $\widehat{\mathrm{VR}}(q)$ falls outside it, we reject the null at 5%.
- The band is not a confidence interval.
  - a confidence interval would be centred at the estimate $\widehat{\mathrm{VR}}(q)$, not at 1.
- The robust null band is $1 \pm 1.96\sqrt{\hat{\theta}(q)/T}$ (test construction).
  - T: the number of daily returns.
  - $\hat{\theta}(q)$: the estimated variance of $\sqrt{T}\,(\widehat{\mathrm{VR}}(q) - 1)$, robust to heteroskedasticity.
- When large moves follow large moves (volatility clustering), $\hat{\theta}(q)$ grows and the band widens.
- $\hat{\theta}(q)$ also grows with q, so the band widens at long horizons.
  - a sample contains fewer independent q-day periods.

Which variance-ratio horizons do we test?

- Why several horizons: reversal or persistence can appear at one day, a week or a month.
  - we do not know in advance at which.
- We test four horizons, $q = 2, 5, 10, 20$ trading days.
  - two days, about one week, two weeks and one trading month.
- The choice follows Lo & MacKinlay (1988), who use $q = 2, 4, 8, 16$ on weekly data.
  - other horizons can give a different result.
- The robust null band has level 5% at each horizon separately.
  - with four tests, a false rejection occurs more often than 5% of the time.
- The Chow–Denning test Chow & Denning (1993) corrects this.
  - it uses the largest $|z^*(q)|$ across the four horizons.
  - The critical value is 2.49 instead of 1.96.
- The Chow–Denning test rejects RW3 for 5 of the 16 series, including the BET and the S&P 500.
- For Bitcoin, the Chow–Denning test does not reject RW3 (p-value $= 0.835$).
- z^* and the Chow–Denning test require a finite fourth moment, which Chapter 1 questions.

Variance-Ratio Profiles (1/2)



MFM_ch2_variance_ratio ▶ Interactive

Variance-Ratio Profiles (2/2)

- Each curve shows the estimated variance ratio $\widehat{\mathrm{VR}}(q)$ for horizons q from 2 to 60 days.
  - $q = 60$, about three trading months, is only descriptive.
- The shaded area is the robust 95% band around 1, the value without autocorrelation.
- S&P 500: $\widehat{\mathrm{VR}}(q)$ is below 1 (reversal) and leaves the band at 2, 5 and 10 days
  - at 20 days it is just inside the band: 0.754 against the edge 0.740
  - $\widehat{\mathrm{VR}}(2) = 0.901$, $\widehat{\mathrm{VR}}(5) = 0.829$, $\widehat{\mathrm{VR}}(10) = 0.767$, $\widehat{\mathrm{VR}}(20) = 0.754$; $\widehat{\mathrm{VR}}(60) = 0.674$
- BET: $\widehat{\mathrm{VR}}(q)$ is above 1 (persistence), rises with q and leaves the band at all four tested horizons
  - $\widehat{\mathrm{VR}}(2) = 1.108$, $\widehat{\mathrm{VR}}(5) = 1.173$, $\widehat{\mathrm{VR}}(10) = 1.225$, $\widehat{\mathrm{VR}}(20) = 1.377$; $\widehat{\mathrm{VR}}(60) = 1.644$
- Bitcoin stays inside the band at all horizons: the null is not rejected.
- EUR/RON: $\widehat{\mathrm{VR}}(q)$ stays above 1 at every horizon.
  - but leaves the band only at 2 and 5 days: $\widehat{\mathrm{VR}}(2) = 1.170$, $\widehat{\mathrm{VR}}(5) = 1.157$.
- Type of evidence: description.

MFM_ch2_variance_ratio

Check: autocorrelation and variance ratios

- Why can an i.i.d. band be misleading under volatility clustering?
- Does positive autocorrelation imply profit after costs?
- What does a variance ratio below one suggest?
- Transfer: a test statistic has magnitude 2.4.
  - If the VR estimate stays the same but its standard error doubles, what does the magnitude become?
- Is the null still rejected at 5% (two-sided)?

Answers: autocorrelation and variance ratios

- The i.i.d. standard error ignores dependence in squared returns, so it is too small.
  - The test rejects the null too often.
- No: effect size, risk, implementation and costs still matter.
- A variance ratio below 1 suggests reversal over the chosen horizon.
- To call the deviation significant, we need a robust standard error.
- The statistic halves to 1.2.
- No: 1.2 does not cross 1.96.
  - the estimated effect is unchanged, its evidence is weaker.

3. Out-of-sample predictability

- The question: does a variable known today forecast next month's market return?
- The classic candidate is *dp*, the log dividend–price ratio.
  - dividends over the last 12 months divided by the average index level in that month.
- In sample, such regressions appear to work.
  - on new data, most published predictors fail Welch & Goyal (2008).
- The proper test: every forecast uses only the data available when it is made.

How we build the forecasts

OS: Out-of-Sample.

- r: the monthly S&P 500 log return, with dividends, minus the 3-month US (United States) Treasury bill rate.
- Each month t we estimate $r_{s+1} = \alpha + \beta\, dp_s + u_{s+1}$ on all months up to t (α, β: intercept and slope, not the GARCH parameters).
  - at the next step the window grows by one month (expanding window).
- The dp forecast for month $t+1$: $\hat{r}_{t+1} = \hat{\alpha} + \hat{\beta}\, dp_t$.
- The competing forecast: the historical mean $\bar{r}_{t+1}$.
  - The mean of returns up to month t.
  - it corresponds to no predictability.
- The first forecast is for January 1970, giving 680 monthly forecasts.
- Monthly dividends are interpolated from quarterly totals.
  - so dp_t was not known exactly in month t.
  - the exercise is pseudo-out-of-sample.

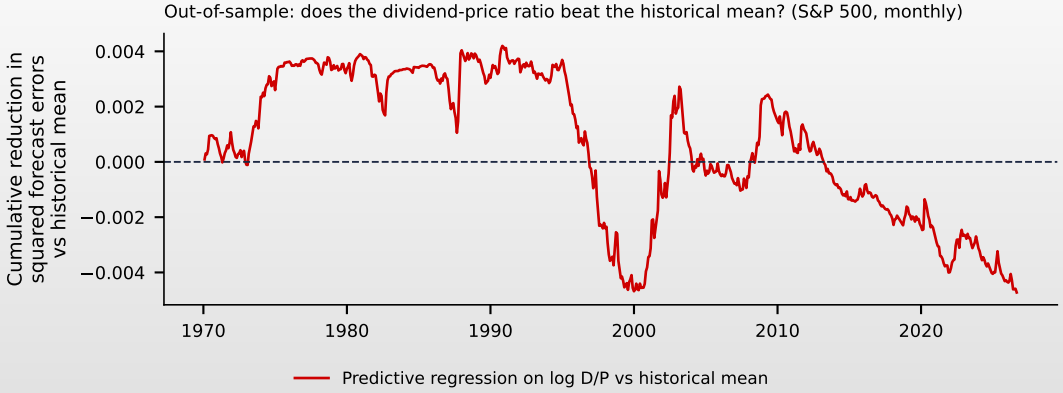
How we score the result: R^2_{OS}

- $R^2_{OS} = 1 - \sum_t (r_t - \hat{r}_t)^2 / \sum_t (r_t - \bar{r}_t)^2$, with the sums over the forecast months.
- $R^2_{OS} > 0$: the *dp* forecast has smaller squared errors than the historical mean.
- $R^2_{OS} < 0$: the historical mean forecasts better.
  - unlike the in-sample R^2, R^2_{OS} can be negative.
- Why the two differ: in sample, $\hat{\beta}$ also fits the noise of the months it explains.
- A positive R^2_{OS} can also arise by chance.
  - we test its significance separately (the Clark–West test, in the results).

Reading the cumulative squared-error chart

- Each month from 1970 adds $(r_t - \bar{r}_t)^2 - (r_t - \hat{r}_t)^2$.
  - the squared error of the historical mean minus that of the *dp* forecast.
- The curve rises in months in which *dp* forecasts better and falls in months in which it forecasts worse.
- The final value of the curve has the sign of R^2_{OS}.
- Vertical axis: the running sum, with returns as fractions (0.01 = 1%), squared.

The Dividend–Price Ratio: Out-of-Sample Forecasts



MFM_ch2_predictability ▶ Interactive

The Dividend–Price Ratio: Results for the S&P 500 (1/2)

- The curve ends below zero: over 680 monthly forecasts, $R^2_{OS} = -0.51\%$.
- The Clark–West test Clark & West (2007) compares the squared errors and corrects a handicap of the *dp* model.
  - even if $\beta = 0$, estimating β adds noise to the forecast.
- On the S&P 500 the test gives $t = 0.42$ (one-sided $p = 0.34$).
  - we do not reject the absence of predictability.
  - the test on squared errors requires a finite fourth moment, which Chapter 1 questions.
- Until 1990 the curve rises: *dp* forecast better than the historical mean.
- In the 1990s boom, *dp* fell to record lows while prices kept rising: the advantage was lost.
- In the 2000s the advantage partly returned, and after 2010 it was lost again.
- Type of evidence: comparison of procedures (it shows which method works better on the same data, not why).

The Dividend–Price Ratio: Results for the S&P 500 (2/2)

- Does a small R^2_{OS} matter economically? Campbell & Thompson (2008) show that it does: even 0.5% is a lot for an investor.
- A mean–variance investor would raise the expected return by about 20%.
  - who raises equity exposure when the forecast is high and cuts it when it is low.
  - formula: Appendix.
- The calculation uses the monthly market Sharpe ratio, 0.160.
- Campbell and Thompson also impose two restrictions.
  - $\hat{\beta}$ with the sign theory predicts.
  - a non-negative equity-premium forecast.
- On the S&P 500 the restrictions bind in 5.7% of months and leave R^2_{OS} at -0.51%.
  - so the economic gain is absent: R^2_{OS} stays negative.

Check: out-of-sample predictability

- What does a negative R^2_{OS} mean?
- Why can a regression with a good in-sample R^2 fail on new data?
- Can a profitable strategy have a negative R^2_{OS}?
- What does the loss of the *dp* advantage in the 1990s say about the stability of the relation?

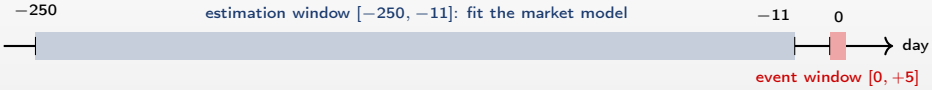
Answers: out-of-sample predictability

- The historical mean forecast better than the model over the months evaluated.
- In sample, $\hat{\beta}$ also fits the noise.
  - on new data the noise does not repeat.
- Yes: profit can come from constant market exposure or from a few correct sign calls.
  - not from small squared errors.
- The relation is not stable.
  - a predictor that was useful until 1990 loses all its advantage when *dp* falls to record lows.

4. Event studies

- The question: how fast and how fully does the price absorb a precisely dated piece of public news?
- To answer it, we need the return the stock would have earned without the news.
- The difference between the observed return and this expected return is the abnormal return (AR).
- An event study measures AR around the news and so tests semi-strong efficiency Fama, Fisher, Jensen & Roll (1969).
- The section's example: the bank tax announced in December 2018.
  - unanticipated news that hit the listed banks directly.

The Event-Study Method (1/2)



- Market model: $r_{i,t} = a_i + b_i r_{m,t} + \varepsilon_{i,t}$.
  - a_i: the intercept, b_i: the market beta of firm i.
  - $r_{m,t}$: the market-index return on day t.
- The model removes from the firm's return the part explained by the same day's market move.
- With daily data, multi-factor models reduce the variance of AR only slightly MacKinlay (1997).
- Estimation window $[-250, -11]$: about one trading year in which we estimate a_i and b_i.
- It stops 10 days before the event, so that any information leak does not enter the estimation.
- Event window $[0, +5]$: the announcement day and five more days.
  - to see whether the reaction continues or reverses.

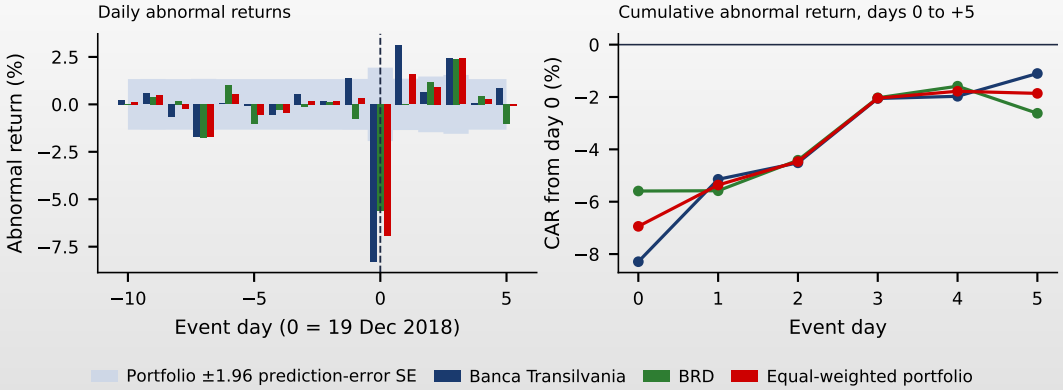
The Event-Study Method (2/2)

- Abnormal return: $AR_{i,t} = r_{i,t} - \hat{a}_i - \hat{b}_i r_{m,t}$, with $\hat{a}_i$, $\hat{b}_i$ estimated in the estimation window.
- Cumulative abnormal return (CAR): $CAR_i[\tau_1, \tau_2] = \sum_{t=\tau_1}^{\tau_2} AR_{i,t}$, where $[\tau_1, \tau_2]$ are the days summed.
- First-pass test: $t = CAR_i / (\hat{\sigma}_\varepsilon \sqrt{L})$ MacKinlay (1997).
  - $\hat{\sigma}_\varepsilon$: the standard deviation of the residuals in the estimation window.
  - L: the number of days summed.
- The formula ignores the uncertainty in $\hat{a}_i$ and $\hat{b}_i$.
  - its correction appears in the worked example.
- For unanticipated news, efficiency predicts a jump on day 0 and no predictable drift or reversal afterwards
- Daily data can only show absorption within a day.
  - speed at the microsecond scale needs intraday data (Chapter 10).
  - literature: Appendix.
- Application: the price response of Ethereum to network development milestones
  - Quantinar (ETH: Ether, the cryptocurrency of Ethereum): Market Responses to ETH Development Milestones

The Bank Tax: the Data and How to Read the Chart

- The government announced the tax on bank assets on the evening of 18 December 2018.
  - event day 0 is 19 December.
- We study Banca Transilvania (ticker TLV), BRD–Groupe Société Générale (ticker BRD) and their equal-weighted portfolio.
  - $r_{m,t}$ is the return of the BET index.
- Left panel: daily *AR*, in percent, from day -10 to day $+5$.
- Light-blue band: ± 1.96 standard deviations of the portfolio's prediction error.
  - a portfolio bar outside it is significant at 5%.
- Right panel: *CAR* from day 0 up to the day on the horizontal axis, in percent.
- A curve that returns towards zero after day 0 indicates a reversed reaction.

Worked Example: the Romanian Bank Tax of December 2018



MFM_ch2_seminar ▶ Interactive

The Bank Tax of December 2018: Results (1/2)

- For TLV, the estimated market beta is 1.18.
  - the residual standard deviation in the estimation window is 0.83% per day.
- On day 0, the abnormal return of TLV is $AR_0 = -8.3\%$, with $t = -10.0$ by the first-pass formula.
- Also on day 0: $AR_0 = -5.6\%$ for BRD and -6.9% for the equal-weighted portfolio.
- That day the log return of the BET was -11.9%, a fall in the index of 11.2%.
- With such an extreme market return the uncertainty in $\hat{b}$ matters.
  - the prediction-error standard deviation for TLV rises from 0.83% to 1.21%.
- With this prediction error, $t = -6.85$.
  - the effect remains strongly significant.
- TLV and BRD carry large weights in the BET.
  - so $r_{m,0}$ already contains part of the shock.
  - AR_0 understates the fall.

The Bank Tax of December 2018: Results (2/2)

- Over days $+1, \ldots, +5$, the cumulative abnormal return of TLV is $+7.2\%$ ($t = 3.82$).
- Over days $0, \ldots, +5$, the cumulative abnormal return is only -1.1% ($t = -0.54$).
  - the abnormal component of TLV's fall, measured against the BET, is almost offset over the next five sessions.
  - this does not show that the share price recovered its loss: over days 0 to $+5$ the log return is -15.6% for TLV and -12.7% for the BET benchmark.
- Explanation 1: an **overreaction** on day 0, against semi-strong efficiency.
- Explanation 2: later news about the final form of the ordinance.
- Explanation 3: the benchmark moved with the event.
  - TLV and BRD carry large weights in the BET, which fell by 11.9% (log) on day 0.
  - so *AR* and *CAR* measure TLV against an index that already contains the shock.
  - the BET did not rebound: over days $+1$ to $+5$ its log return is -0.8%.
- A single event does not let us choose between the three explanations.
- Type of evidence: causal identification, of the day-0 reaction only.
  - If no other news arrived that day; the later offset stays unexplained.

MFM_ch2_seminar

Check: event studies

- Why must you specify the event window?
- Could another announcement explain the same movement?

Answers: event studies

- CAR adds abnormal returns over the days of the window only.
- A window chosen after seeing the data can create or hide an effect.
- Yes: overlapping news is an alternative explanation, so the reaction alone is not causal proof.

5. Multiple testing and anomalies

- The question: does a calendar anomaly remain significant?
  - once we account for how many patterns were searched
- The literature has tested thousands of patterns on the same data, and some pass the test by chance alone Harvey, Liu & Zhu (2016).
- A small p-value means something different after 15 tests than after one.
- First we measure how many false anomalies chance produces.
  - then we correct the real tests with Bonferroni and Holm.

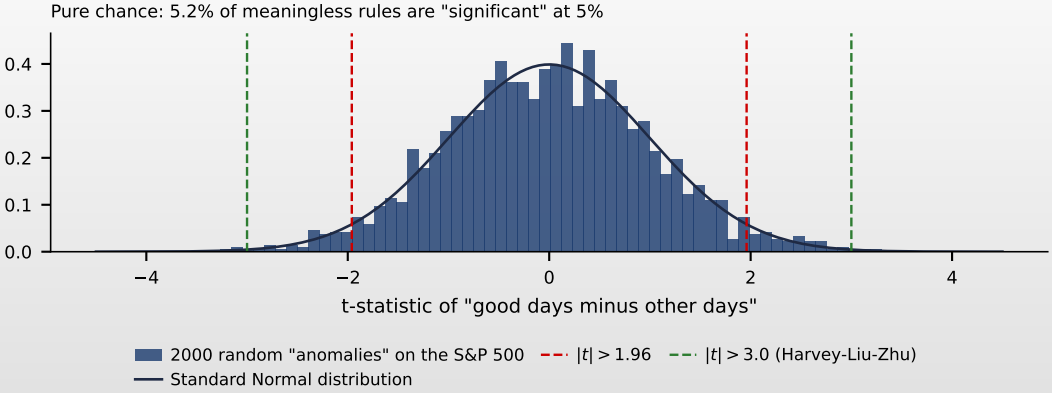
Calendar effects and data snooping

- A calendar effect is a difference in mean returns associated with a day or a month.
- Day-of-the-week effect: mean returns differ across Monday, Tuesday, . . . , Friday French (1980).
- January effect: the mean January return differs from that of the other months Rozeff & Kinney (1976).
- Turn-of-the-month (TOM) effect: higher mean returns on the last trading day of a month and the first three of the next Ariel (1987).
- Data snooping: selecting a pattern after searching many alternatives on the same data.

How Many "Anomalies" Does Pure Chance Produce? (1/3)

- Why we simulate: here every null hypothesis is true, so every rejection is false.
- Experiment: 2 000 trading rules with no economic content, on daily S&P 500 returns.
- Each rule picks 20% of the days at random and calls them "good days".
- For each rule, a t test compares the mean return on "good days" with the mean on the other days.
- The chart: the histogram of the 2 000 t statistics, over the standard Normal density.
- The vertical lines: the thresholds $|t| = 1.96$ (5%) and $|t| = 3.0$, the bar proposed by Harvey, Liu & Zhu (2016).

How Many "Anomalies" Does Pure Chance Produce? (2/3)



MFM_ch2_multiple_testing ▶ Interactive

How Many "Anomalies" Does Pure Chance Produce? (3/3)

- Result: 5.2% of the rules are "significant" at 5% ($|t| > 1.96$), as the test level implies.
- The most "spectacular" rule has $|t| = 3.30$, although it is pure chance.
  - reported alone, it would pass even the strict bar $|t| > 3.0$.
- The bar $|t| > 3.0$ is passed by 0.30% of the random rules, i.e. 6 rules out of 2 000.
- How many rules were tried (the family of tests) matters as much as the bar chosen.
- Sullivan, Timmermann & White (1999) show how to correct for data snooping in trading rules.

MFM_ch2_multiple_testing

Reading the calendar-effects table

HAC: Heteroskedasticity and Autocorrelation Consistent.

- We apply the same logic to 15 real tests, on daily data from 2000.
  - five markets $\times$ three effects (weekday, January, TOM).
- We compare developed markets (S&P 500, Euro Stoxx 50) with emerging markets of the region (BET, WIG20, BUX).
- January extra: the mean January return minus the mean of the other months, in percent per month.
- TOM extra: the mean return on TOM days minus the mean of the other days.
  - in basis points per day (bp; 1 bp $= 0.01$ percentage points).
- We use HAC errors because returns show volatility clustering.
  - HAC errors remain valid under it; ordinary ones do not.
  - In these tests the HAC errors are 0.92–1.54 times the ordinary ones: smaller as well as larger.

January and Turn-of-the-Month Effects

Market	January extra (%/month)	t_{HAC}	TOM extra (bp/day)	t_{HAC}
S&P 500 (US)	−0.66	−0.80	3.2	0.92
BET (Romania)	3.48	1.49	17.3	3.62
Euro Stoxx 50	−0.30	−0.29	5.1	1.21
WIG20 (Poland)	0.53	0.40	8.2	1.72
BUX (Hungary)	2.66	2.25	15.1	3.63

- TOM: the extra return is +17.3 bp per day for the BET and +15.1 bp for the BUX
  - both with $|t| > 3.5$, above the 3.0 bar of Harvey, Liu & Zhu (2016)
- For the S&P 500, the TOM extra return is only +3.2 bp per day ($t = 0.92$): we do not reject a zero effect
- For the BET, January exceeds the other months by 3.48 percentage points
  - but the estimate is imprecise ($t = 1.49$)
- The BUX, with $t = 2.25$, is the only market with a January-effect p-value below 0.05
- For the third effect, the day of the week, we do not reject the null in any market
  - The smallest p-value is 0.2265 (WIG20)
- We ran 15 tests (5 markets × 3 effects): which survive a correction for multiple testing?
- Type of evidence: association (the variables move together; this does not prove causation)

Multiple Testing: Bonferroni and Holm

FWER: Family-Wise Error Rate — the probability of at least one false rejection in the family of tests.

- Why FWER: we want every anomaly we report to be real, not only most of them.
- With M independent tests at significance level α: FWER $= 1 - (1-\alpha)^M$.
- For $M = 15$ and $\alpha = 5\%$, FWER would be 53.7% if the tests were independent.
- **Bonferroni**: reject only if $p_i < \alpha/M$.
  - it guarantees FWER $\leq \alpha$ under any dependence between the tests.
- **Holm**: sort $p_{(1)} \leq \cdots \leq p_{(M)}$ and reject while $p_{(j)} < \alpha/(M-j+1)$, stopping at the first failure.
- Holm has the same guarantee and rejects at least everything Bonferroni rejects.
- Bootstrap methods that use the correlation between tests (StepM): Appendix.

Check: multiple testing and anomalies

- Why does testing many rules change the meaning of a small p-value?
- Does non-rejection prove efficiency?
- Which new sample or robustness check would you request?

Answers: multiple testing and anomalies

- Searching many rules creates more opportunities for false rejection.
- No. A test with low power may fail to reject even if the market is inefficient.
- The conclusion concerns only the alternative tested, not every form of predictability.
- Example: the same rule on post-publication data or on another market.
  - a data-snooping artefact disappears out of sample.
- A real effect usually weakens after publication but does not vanish McLean & Pontiff (2016).

Main results of the chapter

- S&P 500: return reversal at horizons of 2–10 days (variance ratio below 1).
- BET: persistence (variance ratio above 1) at all four tested horizons.
- The dividend–price ratio does not beat the historical mean out of sample ($R^2_{OS} < 0$).
- Bank tax: the abnormal component of TLV's fall, measured against the BET, was almost offset over the next five sessions.
  - a single event cannot separate overreaction from later news.
- After the Holm correction only the TOM effects for the BET and the BUX remain significant.

Topics for further study

- ⊡ Data and conventions
- ⊡ Fair games, pricing and rational bubbles
- ⊡ Test construction and alternatives
- ⊡ Predictive regressions and bias
- ⊡ Long memory and changing efficiency
- ⊡ Event-study inference
- ⊡ Anomalies and behavioural explanations
- ⊡ Machine learning and research design
- ⊡ Case study: time-series momentum in Bitcoin
- ⊡ AI for scientific discovery
- ⊡ Synthesis and technical reference

Data Used in This Chapter

BNR: Banca Națională a României — the National Bank of Romania.

- 16 daily series up to 18 September 2026
  - sample sizes: Appendix
  - 13 equity indices from 2000: 4 developed (S&P 500, Euro Stoxx 50, Nikkei 225, Hang Seng), 9 emerging or frontier (BET, WIG20, BUX, PX, BIST 100, Bovespa, IPC, Nifty 50, Shanghai)
  - Bitcoin from 2014 in the full-sample and rolling tests; the yearly crypto analysis starts in 2011
  - Ethereum from 2016 and EUR/RON, the BNR reference rate, from July 2005
- Daily log returns $r_t = \ln P_t - \ln P_{t-\mathbf{1}}$, each series on its own calendar
  - P_t: the closing price (index level, exchange rate) on day t
- Equity indices use trading days only
  - a repeated close is removed only if the day shows no trading (zero volume, or open = high = low = close)
  - or if its whole row copies the previous day's; such a day is a holiday filled with the previous price
  - a genuine unchanged close, with volume and an intraday range, stays as a zero return (S&P 500: 3 of about 6 700 days)
  - the BET file has closes only: none of its days is removed
  - IPC: 46 such days (44 in 2023–2026) with near-identical volume, likely provider duplicates, are kept
  - dropping them moves $\hat{\rho}_{\mathbf{1}}$ by 0.001, $z^*(2)$ by 0.04, $\hat{d}$ by 0.014, AQ by 0.29
- Crypto-assets trade 7 days a week: 365 observations per year
- Analyses that combine several series align prices on common trading days
  - e.g. event studies
  - before computing returns

What restriction does efficiency impose on prices?

- The question: what exact restriction does efficiency impose on prices, and what does it leave free?
- The standard answer is the fair game.
  - The return in excess of the required return cannot be forecast.
- The autocorrelation tests of section 2 check only one consequence of this condition.
- The required return comes from a pricing model, so every fair-game test is also a test of that model Fama (1991).
- A fair game allows rational bubbles.
  - but rules out deviations that revert predictably to fundamental value (fads).
- Deviations do not always disappear Shleifer & Vishny (1997).
  - arbitrage is risky when price convergence is uncertain (Further Topics).

Moments, tails and asymptotics

- The question: which assumptions make an efficiency test valid, and which alternatives can it detect?
- Robust standard errors of autocorrelations (Further Topics) need a finite fourth moment, $\mathbb{E}[r^4] < \infty$
  - heavy tails: Student t with four degrees of freedom has finite variance but $\mathbb{E}[r^4] = \infty$, and then the robust standard error no longer converges
- Critical values (± 1.96, χ^2 quantiles) are asymptotic approximations
  - correct as the sample grows
  - possibly inaccurate in a short sample such as a 500-day window
- Limit theorems also need weak dependence (mixing)
  - observations far enough apart in time become nearly independent
  - mixing allows volatility clustering and dependence between neighbouring days

Predetermined variables, bias and consistency

- Why does a persistent predictor distort inference in a finite sample?
- A predetermined predictor is known before the return shock it predicts
  - strict exogeneity additionally rules out correlation with shocks at every date.
- Today's dividend–price ratio can forecast next month's return.
  - yet next month's price shock also changes the ratio.
  - this feedback obstructs strict exogeneity.
- Consistent is not unbiased.
  - with a merely predetermined predictor, OLS (Ordinary Least Squares) converges to the true parameter.
  - but is biased in a finite sample.

Short memory, long memory and scaling

- Does dependence persist across horizons and market regimes?

AR: AutoRegressive model.

- Short memory has absolutely summable autocovariances
  - an AR(1) with coefficient 0.5 has correlations 0.5^k at lag k.
- Long memory here means positive dependence decaying so slowly that the autocovariance sum diverges
  - power-law decay can produce this.
- Scaling: a measure of fluctuations over n days, $F(n)$, grows like n^H
  - e.g. the standard deviation of a sum of n returns
  - H is the Hurst exponent, estimated in Further Topics
  - doubling the horizon multiplies fluctuations by 2^H: about 1.41 at $H = 0.5$ (no long memory) and 1.68 at $H = 0.75$
- Anti-persistence describes offsetting increments, as in fractional Gaussian noise (fGn) with $H < 0.5$
  - fGn is a stationary Gaussian increment model with partial-sum variance proportional to n^{2H}.

Variance Ratio and Hurst Exponent: Long or Short Memory?

- For fGn, the variance of a sum of q returns is $\mathrm{Var}(r_t + \cdots + r_{t-q+1}) = q^{2H}\mathrm{Var}(r_t)$.
  - fGn: the exactly self-similar process defined above.
- Hence $\mathrm{VR}(q) = q^{2H-1}$, and an observed $\mathrm{VR}(q)$ implies a Hurst exponent $H(q) = 0.5 + \ln \mathrm{VR}(q)/(2 \ln q)$.
- At $q = 20$ days, the implied exponent is 0.553 for the BET and 0.453 for the S&P 500.
- The problem: a process without long memory can give $H(q) \neq 0.5$
  - Example: an AR(1) with $\rho_1 = -0.1$ has autocorrelations $\rho_k = (-0.1)^k$, so only short memory.
  - For this AR(1), $\mathrm{VR}(20) = 0.826$, so $H(20) = 0.468$: the process looks "anti-persistent" without any long memory.
  - At $q = 250$, the same process gives $H(250) = 0.482$: for short memory, $H(q)$ approaches 0.5 as q grows.
- An AR(1) with BET's autocorrelation, $\hat{\rho}_1 = 0.109$, gives $H(20) = 0.535$.
- So most of the gap between BET's $H(20) = 0.553$ and 0.5 is short memory.
- Separating long from short memory needs a statistic built for it: Lo's modified R/S (Further Topics).

Event-study inference

- The question: how reliable are the t statistics of the Event studies section?
  - when their assumptions fail
- There we assumed:
  - $\hat{a}_i$ and $\hat{b}_i$ are known exactly.
  - firms are independent.
  - the abnormal-return variance does not rise on the event day.
- Estimation error: $\hat{a}_i$ and $\hat{b}_i$ come from a finite window.
  - the predicted normal return is also uncertain.
- Cross-sectional dependence: two banks hit by the same news have abnormal returns that move together.
- Event-induced variance: the event itself raises the variance.
  - the calm days before understate it.

Anomalies and behavioural explanations

- Can behavioural biases explain the anomalies that survive?

Momentum

- Cross-sectional momentum ranks assets.
  - if A gained 10% and B gained 2%, a relative strategy may buy A and short B.
- Time-series momentum uses each asset's own past sign.
  - both A and B would receive positive signals in that example.

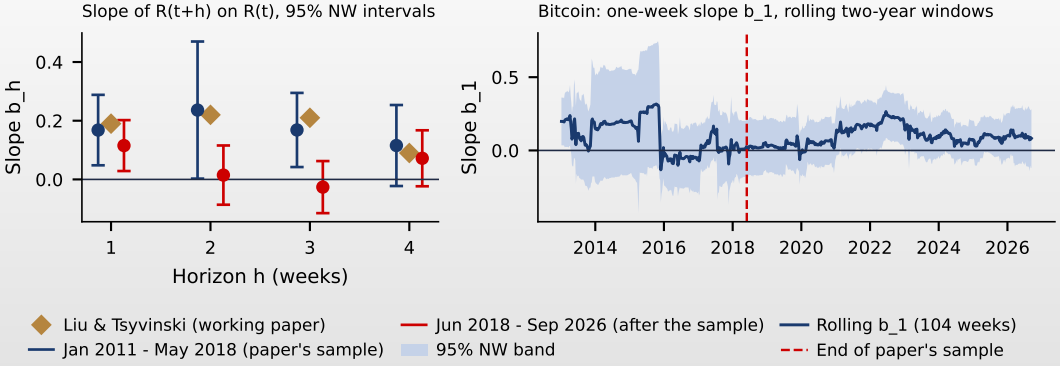
Machine learning and research design

- The question: what does an in-sample rejection mean?
  - when we test efficiency with hundreds of predictors
- Anomalies are found by searching among many predictors: Harvey, Liu & Zhu (2016) count more than 300 published factors.
- The Multiple testing and anomalies section showed that 2 000 random rules produce in-sample "anomalies".
- Here:
  - how to evaluate a flexible model properly.
  - why an in-sample rejection can arise even with rational investors Martin & Nagel (2022).

The Question: Do Past Returns Predict Bitcoin's Return?

- **Question**: does this week's Bitcoin return predict the returns of the following weeks?
  - if so, past prices contain unused information: a departure from weak-form efficiency
- Liu & Tsyvinski (2021), *Review of Financial Studies*
  - cryptocurrency returns have no exposure to the usual stock market factors or to macroeconomic factors
  - they are predicted by factors specific to crypto markets: time-series momentum (TSMOM) and investor attention
- TSMOM: a week with a high return is followed, on average, by high returns in the following weeks
- The authors' numbers (working paper of the National Bureau of Economic Research, NBER w24877; Bitcoin, January 2011 – May 2018)
  - slope of the regression of next week's return on this week's return: 0.19 ($t = 3.73$)
  - one more standard deviation today (16.6%) adds 3.2% to next week's return
  - the quintile of weeks with the highest returns is followed by 11.2% a week on average, the one with the lowest by 2.6%
- Why it matters: an effect that investors exploit should weaken after publication McLean & Pontiff (2016)

Bitcoin: The Momentum Slope in the Paper's Sample and After



- Left: the slope b_h in $R_{t+h} = a + b_h R_t + e_{t+h}$, with 95% Newey–West (NW) intervals; diamonds: the authors' numbers
- Right: the slope b_1 over rolling windows of 104 weeks (two years), dated at the last week of the window

MFM_ch2_crypto_momentum ▶ Interactive

The Method in Brief and How the Replication Differs

- R_t: the simple return of Bitcoin in week t, from the Sunday close to the next Sunday close
  - daily closing prices in UTC (Coordinated Universal Time), as in the paper
- Predictive regression $R_{t+h} = a + b_h R_t + e_{t+h}$, for $h = 1, \ldots, 4$ weeks
  - the paper's sample: 9 Jan 2011 – 27 May 2018, 386 weeks
  - after the sample: 3 Jun 2018 – 13 Sep 2026, 433 weeks
  - the authors' robustness check: the same regression from 2013
- Sorting: the weeks are split into five quintiles by R_t; we compare the mean of R_{t+1} in the top and the bottom quintile
- Differences from the paper
  - EODHD (EOD Historical Data) prices from the course data instead of CoinDesk prices
  - NW standard errors with 5 lags instead of the authors' OLS (ordinary least squares) and bootstrap standard errors
  - no attention factors (Google searches, Twitter) and no Ripple

The Result: Only the One-Week Effect Remains, Three Times Smaller

- In the paper's sample, the course data reproduce the result
  - $b_1 = 0.168$ (OLS $t = 3.33$, NW $t = 2.75$), against 0.19 in the paper
  - the top quintile is followed by 9.9% a week, the bottom one by 2.6% (t of the difference $= 2.18$)
- After the sample, predictability remains only at the one-week horizon
  - $b_1 = 0.115$ (NW $t = 2.61$); it survives the Holm correction over the four horizons ($p = 0.009 < 0.05/4$)
  - at $h = 2, 3, 4$: slopes between -0.026 and 0.072, all with $|t| \leq 1.48$
  - one more standard deviation today adds 0.96% to next week's return, against 2.98% in the paper's sample
  - weekly volatility fell from 17.7% to 8.3%; Ethereum after the sample: $b_1 = 0.107$ (NW $t = 2.45$)
- Type of evidence: statistical predictability of the one-week return, not a profitable strategy
  - no trading costs, no risk adjustment and no out-of-sample forecasts
- Limits
  - from 2013, within the paper's sample, b_1 is no longer significant (NW $t = 1.06$): the extreme weeks of 2011–2012 weigh heavily
  - b_1 is significant in only 13% of the rolling two-year windows
  - the NW t-test requires a finite fourth moment, which Chapter 1 questions

An Open Question: Does a Market Become More Efficient When Arbitrageurs Enter?

- The question: does Bitcoin's informational efficiency rise when new arbitrageurs enter?
  - the new arbitrageurs come in through regulated instruments
  - such as spot ETFs
- Tran & Leirvik (2020) find that the efficiency of crypto-assets varies over time and rises.
- Kia, Liu, Li, Song & Xu (2026) find that, after the spot ETFs, the spot market again became the dominant venue of price discovery.
- We do not know whether the change comes from the ETFs or from market maturity.
  - this is the question tested in this section.

Mini-Case: Did Spot ETFs Make Bitcoin More Efficient?

- An artificial-intelligence (AI) assistant proposes a typical hypothesis:
  - "US spot Bitcoin ETFs (trading since 11 January 2024) brought arbitrageurs, so Bitcoin became more efficient"
- The discovery loop
  - reusable prompts: Appendix
  - **literature**: Kia, Liu, Li, Song & Xu (2026) and Tran & Leirvik (2020) (the topic's opening slide)
  - **hypothesis**: $H_0 : \rho_{1,\text{post}} = \rho_{1,\text{pre}}$, two windows of equal length around the launch date (first trading day)
  - **data and code**: daily Bitcoin closing price (USD); first replicate the chapter's full-sample $\hat{\rho}_1 = -0.022$
- Checks fixed before seeing results
  - rationale for each: Appendix
  - HAC standard errors of the estimates (not $1/\sqrt{T}$); equal windows; the launch date only, plus placebo dates
  - a control asset for confounders (the April 2024 halving, rates, inflows)
  - a multiple-testing correction (Holm or Chow–Denning) for every statistic tried, and the minimum detectable effect (MDE)

Further Topics: mini-case results

Your Turn: A Project Seed

- **Question**: did the spot ETFs change the time-varying efficiency of Bitcoin?
  - relative to a control asset
  - rolling VR and DFA (this chapter) for Bitcoin and Ethereum; Ethereum is untreated until July 2024
  - if parallel trends, no anticipation and no spillovers are not credible, report descriptive, not causal, differences
- **Replicate first**: the chapter's yearly Bitcoin autocorrelations and the mini-case numbers
  - record which Bitcoin price series and first date each window uses, placebo windows included
- **Pre-register**: H_0, H_1, dates, windows and corrections, committed before the results
- **Report**: robust inference, placebo tests and the MDE
- Publish the code as a Quantlet
- **Log**: record the prompts used in `AI_USE.md`
- Record in `AI_ERRORS.md` every error of the AI assistant you caught
  - for example an i.i.d. standard error or an invented reference

Synthesis, Check Yourself, and Next: Chapter 3

Five steps for every efficiency claim

- **Information set**: which information $\mathcal{I}_t$ do we assume prices reflect (past prices, public news)?
- **Null**: RW1, RW3 or the martingale difference ($\mathbb{E}[r_t \mid \mathcal{I}_{t-1}] = \mu$, details in Further Topics), stated together with the risk model
- **Statistic**: its variance valid under volatility clustering
- **Uncertainty**: horizons, markets and windows tried, and out-of-sample evidence
- **Economics**: a statistical rejection is not an exploitable profit after costs
  - a profit does not prove an irrational market

Check yourself and next steps

- Why can RW3 be rejected for the S&P 500 while the market remains efficient in the economic sense?
- Why does BET's $H(20) > 0.5$ (the Hurst exponent implied by VR(20), Long memory) not prove long memory?
- Which of the two results would survive a change of test, window and inference?
  - BET's persistence or Bitcoin's non-rejection
- Next, Chapter 3: risk-adjusted returns with factor models
  - the factor zoo
- Textbooks: Campbell, Lo & MacKinlay (1997), ch. 2, and Franke, Härdle & Hafner (2019)
  - advanced courses: ▸ Appendix

The Practical Decision

- **What I choose**
  - weak-form tests with robust standard errors, not i.i.d. ones (S&P 500: robust SE of $\hat{\rho}_1$ 0.027, against 0.012)
  - Chow–Denning across the four variance-ratio horizons, Holm across the 15 calendar tests
  - for a predictor: R^2_{OS} and the Clark–West test against the historical mean, not the in-sample R^2
- **Under which conditions**
  - the robust standard errors require a finite fourth moment, which Chapter 1 questions
  - a rejection also tests the model of expected returns and is not a profit after costs
- **What result would make me drop it**
  - a predictor: $R^2_{OS} < 0$ with an insignificant Clark–West test, as *dp* on the S&P 500 (-0.51%, $t = 0.42$)
  - an anomaly: a rejection that does not survive Holm, as the BUX January effect ($t = 2.25$)
  - the BET persistence: a variance ratio that stays inside the robust band on new windows

Why the robust band is wider: a numerical check

- For $q = 2$, the classical statistic is $z = \sqrt{T}\,(\widehat{\mathrm{VR}}(2) - 1)$.
  - it assumes i.i.d. returns (RW1).
- The robust statistic z^* divides the same numerator by $\sqrt{\hat{\theta}(2)}$.
- Without volatility clustering, $\hat{\theta}(2) \approx 1$ and $z^* \approx z$.
  - with volatility clustering, $\hat{\theta}(2) > 1$ and $|z^*| < |z|$.
- Toy example: $T = 400$ days and $\widehat{\mathrm{VR}}(2) = 0.90$ give $z = 20 \times (-0.10) = -2.0$: the null is rejected at 5%.
- If volatility clustering makes $\hat{\theta}(2) = 4$, then $z^* = -2.0/\sqrt{4} = -1.0$ and the null is no longer rejected.
- The robust band around 1 is then $\sqrt{4} = 2$ times wider.
- For the S&P 500, $z = -8.15$ and $z^* = -3.66$.
  - the robust statistic is about 2.2 times smaller.
  - the null is still rejected.

Why the dividend–price ratio?

- The Campbell & Shiller (1988) identity: $dp_t \approx c + \sum_{j \geq 0} \rho^j \left(r_{t+1+j} - \Delta d_{t+1+j} \right)$.
  - ρ just below 1.
  - Δd: log dividend growth.
- The identity follows from the definition of the return through a log-linear approximation.
  - with no economic model.
- Consequence: if dp varies, it must forecast future returns, future dividend growth or both.
- In the data dp forecasts dividend growth poorly, so its variation should forecast returns Cochrane (2008).
- That is why dp is the most studied market predictor Fama & French (1988).
- A successful forecast would not prove inefficiency.
  - it can reflect a risk premium that varies over time.

Holm on the 15 calendar p-values

- Two-sided p-values:
  - the HAC t statistic against the Normal distribution (January, TOM).
  - the Wald χ^2_4 test that the five daily means are equal (weekdays).

Market	Weekday, equal means	January	TOM
S&P 500 (US)	0.8263	0.4257	0.3585
BET (Romania)	0.3052	0.1368	0.0003
Euro Stoxx 50	0.6319	0.7738	0.2266
WIG20 (Poland)	0.2265	0.6865	0.0853
BUX (Hungary)	0.9327	0.0245	0.0003

- The two smallest p-values are TOM for the BET and TOM for the BUX, both 0.0003.
- They are below the Holm thresholds $0.05/15 = 0.0033$ and $0.05/14 = 0.0036$, so both nulls are rejected.
- The third p-value, January for the BUX (0.0245), exceeds the threshold $0.05/13 = 0.0038$, so Holm stops here.
- Without correction, January for the BUX would have been the third rejection, since $0.0245 < 0.05$.

Market abbreviations

- **US**: United States
- **HK**: Hong Kong
- **BIST**: Borsa İstanbul — the Istanbul Stock Exchange (BIST 100 index).
- **WIG20**: Warszawski Indeks Giełdowy 20 — the Warsaw Stock Exchange index of the 20 largest companies.
- **BUX**: Budapesti Értéktőzsde index — the Budapest Stock Exchange index.
- **PX**: Prague Stock Exchange index (index Burzy cenných papírů Praha).
- **IPC**: Índice de Precios y Cotizaciones — the Mexican Stock Exchange index.
- **BET**: Bucharest Exchange Trading (index).
- **FX**: Foreign eXchange.

Fair games and martingales

- **Fair game** for excess returns: $\mathbb{E}[r_{t+1} - \mu_t \mid \mathcal{I}_t] = 0$
  - $\mathcal{I}_t$: the information available at the end of day t; $\mathbb{E}[\cdot \mid \mathcal{I}_t]$: the expectation given that information
  - μ_t: the required (equilibrium) return, the compensation for risk
  - in words: the return beyond μ_t cannot be forecast from $\mathcal{I}_t$
- **Martingale**: a process whose expected next value, given the past, equals its current value
  - with a constant required simple return m, the value V_t of a position with dividends reinvested, discounted at m, $V_t/(1+m)^t$, is a martingale Campbell, Lo & MacKinlay (1997)
  - discounting at the risk-free rate gives a martingale only under the risk-neutral measure (slide on the risk-neutral measure)
  - the price P_t itself is a martingale only if no dividends are paid and $m = 0$; the expected log return is then negative (the Jensen term, next slide)

The Jensen term: why averages differ

- The logarithm is concave.
  - averaging log outcomes gives less than taking the log of their average, unless outcomes are constant.
- Equally likely simple returns of $+10\%$ and -10% have mean zero.
- The mean log return is $\frac{1}{2}\ln(1.1) + \frac{1}{2}\ln(0.9) = \frac{1}{2}\ln(0.99) \approx -0.00503$.
- Along a path with one rise and one fall, wealth 100 becomes $100(1.1)(0.9) = 99$.
  - a zero arithmetic mean does not preserve compounded wealth.

Prices, log prices and information

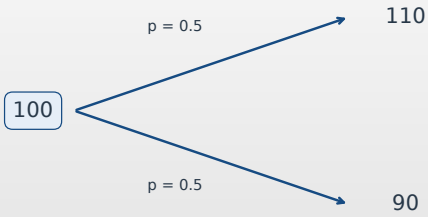
- A martingale in prices and a constant mean log return are different restrictions
  - a martingale in prices: $\mathbb{E}[P_{t+1} \mid \mathcal{I}_t] = P_t$
  - a constant mean log return: $\mathbb{E}[\ln(P_{t+1}/P_t) \mid \mathcal{I}_t] = \mu$
  - the difference is the Jensen term: $\mathbb{E}[\ln(1+R)] \approx \mathbb{E}[R] - \mathrm{Var}(R)/2$, with $R = P_{t+1}/P_t - 1$ the simple return
- Consequence of the fair game: $\mathrm{corr}(\varepsilon_{t+1}, g(\mathcal{I}_t)) = 0$ for any function g of past information
  - where $\varepsilon_{t+1} = r_{t+1} - \mu_t$ is the unexpected part of the return
- If $\mu_t = \mu$ is constant, all autocorrelations $\rho_k = \mathrm{corr}(r_t, r_{t-k})$, $k \geq 1$, are zero
  - this is the hypothesis tested in section 2
- The converse fails: zero autocorrelations do not rule out a nonlinearly predictable mean
  - the MDH (Martingale Difference Hypothesis), next topic

Discounting and the risk-neutral measure

- Why: the martingale above uses the actual probabilities and the required return m.
  - pricing theory uses, equivalently, the risk-free rate and risk-adjusted probabilities.
- A risk-neutral measure uses pricing probabilities, in an arbitrage-free model.
  - under which discounted traded prices, including dividends, are martingales.
- Example: a stock without dividends costs 100 today.
  - in one year it is worth 120 or 80.
  - the risk-free rate is 5%.
  - The pricing probability $q = 0.625$ gives $q\,120 + (1-q)80 = 105$, and $105/1.05 = 100$, today's price.
- This q need not equal the actual probability of an increase.
  - pricing probabilities are not forecasts of frequencies.

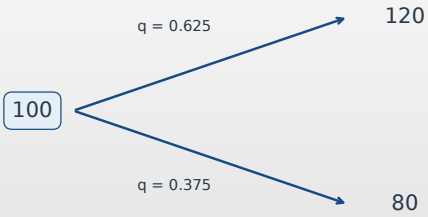
Probabilities and conditional means

Martingale under actual probabilities



Next-value mean = 100
Expected increment = 0

Risk-neutral pricing example



Next-value mean = 105
Discounted: 105 / 1.05 = 100

Left: the actual conditional mean stays at 100. Right: pricing probabilities give 105, whose discounted value is 100.

Fundamental value, rational bubbles and fads

- With a constant required simple return m, the fair game is $\mathbb{E}[P_{t+1} + D_{t+1} \mid \mathcal{I}_t] = (1+m)\,P_t$
  - D_{t+1}: the cash flow (dividend) paid in $t+1$
  - Fundamental value $P_t^* = \mathbb{E}\big[\sum_{j\geq 1} D_{t+j}/(1+m)^j \mid \mathcal{I}_t\big]$; pricing error $B_t = P_t - P_t^*$
  - The fundamental value P_t^* satisfies the same fair-game equation as the price
  - Subtracting the two equations, the gap B_t satisfies $\mathbb{E}[B_{t+1} \mid \mathcal{I}_t] = (1+m)\,B_t$
- A fair game ties the price to the fundamental value only up to a **rational bubble**
  - A gap $B_t > 0$ sustained by the expectation of resale
  - A gap $B_t \neq 0$ is admissible only if it is expected to grow at the required rate m (Chapter 17)
  - A mean-reverting gap (fad) makes ε_{t+1} predictable and violates the fair game, although short-horizon tests have little power against it Summers (1986)
- The restriction is on the conditional mean only
  - The variance is not restricted
  - volatility may be predictable (Chapter 1)

Rational bubble: a numerical example

- A share pays 10 RON (Romanian leu) at the end of every year, forever.
  - The first dividend is paid one year from today.
  - The constant required return is $m = 10\%$; fundamental value is $P_t^* = 10/0.10 = 100$ RON.
- Today $P_t = 120$ RON, so $B_t = 20$ RON.
  - Without a crash, $B_{t+1} = 1.1 \times 20 = 22$ RON.
  - After next year's dividend, fundamental value remains 100 RON; the price is therefore 122 RON.
- The total simple return is $(122 + 10 - 120)/120 = 10\%$.

Rational bubble: what happens if it can burst?

- Initial bubble $B_t = 20$ RON and required return $m = 10\%$.
  - Survival probability $\pi = 0.8$; crash probability $1 - \pi = 0.2$.
  - A crash makes the bubble zero; investors know the probabilities.
- The fair game for the bubble, $\mathbb{E}[B_{t+1} \mid \mathcal{I}_t] = (1+m)B_t$, requires an expected bubble of $(1+m)B_t = 22$ RON.
  - Conditional on survival: $B_{t+1} = 22/0.8 = 27.5$ RON.
  - Conditional growth is $27.5/20 - 1 = 37.5\%$.
- Crash mechanism: Blanchard & Watson (1982), Bubbles, Rational Expectations and Financial Markets.

- The purchase price is 120 RON.
  - the dividend at year-end is 10 RON.
- Survival: the ex-dividend price is 127.5 RON.
  - Simple return: $(127.5 + 10 - 120)/120 \approx 14.6\%$.
- Crash: the price falls to fundamental value, 100 RON.
  - Simple return: $(100 + 10 - 120)/120 \approx -8.3\%$.
- Expected return: $0.8 \times (127.5 + 10 - 120)/120 + 0.2 \times (100 + 10 - 120)/120 = 10\%$.
  - The expected return satisfies the fair-game condition; either realised return can differ from it.

Rational bubble and the no-bubble condition

- The fair game restricts only the expected return over the next period.
  - it does not pin down the price level.
- **No-bubble condition**: $\lim_{k\to\infty} \dfrac{\mathbb{E}[P_{t+k} \mid \mathcal{I}_t]}{(1+m)^k} = 0$
  - also called the transversality condition
  - k: the resale horizon; the discounted expected resale price vanishes
- Under this condition the price equals the fundamental value P_t^*.
  - it is supported entirely by dividends.
- A rational bubble violates the condition.
  - from $\mathbb{E}[B_{t+k} \mid \mathcal{I}_t] = (1+m)^k B_t$, the discounted value of the bubble stays $B_t \neq 0$ at every horizon.
- In the numerical example, the 20 RON bubble is still worth 20 RON in today's money, however distant the resale.
- Tirole (1985): in overlapping-generations models, a bubble can exist only if:
  - the interest rate is below the growth rate of the economy.

The Grossman–Stiglitz Paradox

- Gathering information is costly (analysts, data, computing)
- If prices reflected **all** information, nobody would be paid for collecting it Grossman & Stiglitz (1980)
  - Then nobody would collect it, and prices could not reflect it
- Resolution: an **equilibrium degree of inefficiency**
  - just large enough to reward informed traders for their costs
- Markets with cheap information and low costs (large, liquid) should be closer to efficiency
  - a prediction for our comparison of developed, emerging and crypto markets
- Efficiency does not rule out bubbles either
  - it says they are hard to time (Chapter 17)

Pure arbitrage and relative-value trading

- Pure arbitrage is a strategy with:
  - no positive initial outlay.
  - no possible loss.
  - a positive gain in at least one state.
- Relative-value trading is risky when price convergence is uncertain.
- Example: a price gap of 20 points may widen to 30 before it closes.
  - the position then needs extra collateral (margin) and may be liquidated.
- The link with efficiency: arbitrageurs with limited capital cannot eliminate deviations from fundamental value Shleifer & Vishny (1997).
  - when convergence is uncertain.

Autocorrelation and Its Robust Standard Error

- Sample autocorrelation: $\hat{\rho}_k = \sum_{t=k+1}^{T} e_t e_{t-k} / \sum_{t=1}^{T} e_t^2$, with $e_t = r_t - \bar{r}$
  - T: number of daily returns; $\bar{r}$: their mean; k: the lag; $\hat{\rho}_k > 0$: persistence, $\hat{\rho}_k < 0$: reversal
- Under RW1 (i.i.d.): $\mathrm{SE}(\hat{\rho}_k) \approx 1/\sqrt{T}$; 95% band $\pm 1.96/\sqrt{T}$
- Unpredictable mean with volatility clustering (finite fourth moments): $\mathrm{SE}(\hat{\rho}_k) \approx \sqrt{\hat{\tau}_k}$
  - $\hat{\tau}_k = \sum_t e_t^2 e_{t-k}^2 / \left(\sum_t e_t^2\right)^2$ (robust variance under H_0; for test bands centred at 0)
  - When large moves cluster, $\hat{\tau}_k > 1/T$: the i.i.d. band is too narrow and classical tests reject too often; the robust statistics $\tilde{Q}$ (Ljung–Box) and z^* (variance ratio) use $\hat{\tau}_k$ instead of $1/T$

Intervals for ρ_k: the HAC Standard Error

- $\sqrt{\hat{\tau}_k}$ is derived under H_0: valid for a test band centred at 0, not for an interval around $\hat{\rho}_k \neq 0$.
- Around the estimate we use the HAC variance of the estimator Newey & West (1987):
  - $\hat{\psi}_t = (e_t e_{t-k} - \hat{\rho}_k e_t^2)/\hat{\gamma}_0$, with $\hat{\gamma}_0 = \frac{1}{T}\sum_t e_t^2$: the contribution of day t to $\hat{\rho}_k - \rho_k$.
  - $\widehat{\mathrm{Var}}(\hat{\rho}_k) = \frac{1}{T}\left[\hat{c}_0 + 2\sum_{j=1}^{\ell}\left(1 - \frac{j}{\ell+1}\right)\hat{c}_j\right]$, with $\hat{c}_j$ the lag-j autocovariance of $\hat{\psi}_t$.
  - bandwidth $\ell = \lfloor 4(T/100)^{2/9} \rfloor$: 9–10 lags for the full samples, 4–5 for one calendar year, 6 for the ETF (Exchange-Traded Fund) windows.
- Use: intervals around $\hat{\rho}_k$ and comparisons between periods (crypto years, before/after the ETFs).
- Check: a moving-block bootstrap (blocks of $T^{1/3}$ days) gives similar values.
  - S&P 500: HAC 0.030, block bootstrap 0.031, $\mathrm{SE}_0 = 0.027$.
- The HAC SE is smaller than SE_0 for 14 of the 16 series (BET: 0.023 against 0.028).

Joint tests and reference distributions

- A portmanteau test combines several lag correlations into one statistic.
  - instead of selecting the most striking lag.
- A Wald test similarly tests coefficient restrictions jointly
  - five equal weekday means impose four restrictions, $\beta_1 = \beta_2$, $\beta_1 = \beta_3$, $\beta_1 = \beta_4$, $\beta_1 = \beta_5$, so the reference is χ^2_4, with 5% critical value 9.49.
- Five restrictions, all $\beta_d = 0$, would test a different null: every weekday mean equal to zero.

Portmanteau Tests: Ljung–Box and Its Robust Version

- Ljung & Box (1978): $Q(m) = T(T+2)\sum_{k=1}^{m} \hat{\rho}_k^2/(T-k) \sim \chi_m^2$ under i.i.d. returns
  - Tests jointly H_0: $\rho_1 = \cdots = \rho_m = 0$
  - m: the number of lags
- Robust version Lobato, Nankervis & Savin (2001): $\tilde{Q}(m) = \sum_{k=1}^{m} \hat{\rho}_k^2/\hat{\tau}_k \to \chi_m^2$ under RW3
  - if the $\hat{\rho}_k$ at different lags are asymptotically uncorrelated
  - the condition, $\mathbb{E}[\varepsilon_t^2 \varepsilon_{t-j}\varepsilon_{t-k}] = 0$ for $j \neq k$, holds for a GARCH with symmetric innovations
- Three markets at $m = 10$ (5% critical value 18.31)
  - S&P 500: $Q = 97.1$ but $\tilde{Q} = 20.0$ ($p = 0.029$)
  - BET: $Q = 108.0$, $\tilde{Q} = 24.4$ ($p = 0.007$)
  - Bitcoin: $Q = 20.2$ ($p = 0.028$) but $\tilde{Q} = 11.9$ ($p = 0.294$)
- The robust test keeps the rejections for S&P 500 and BET, but not for Bitcoin
  - $\tilde{Q}$ requires a finite fourth moment, which Chapter 1 questions

The Runs Test

- The runs test Wald & Wolfowitz (1940) uses only the **signs** of returns (zero-return days dropped), so heavy tails do not affect it
- A **run** is a maximal block of equal signs: ++|-|+++ has 3 runs
- With n_1 positive days, n_2 negative days and $n = n_1 + n_2$, if the signs are independent:
  - $\mathbb{E}[R] = \dfrac{2n_1n_2}{n} + 1, \quad \mathrm{Var}[R] = \dfrac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}, \quad z = \dfrac{R - \mathbb{E}[R]}{\sqrt{\mathrm{Var}[R]}} \approx N(0,1)$
  - R: the observed number of runs; $|z| > 1.96$ rejects independence at 5%
  - the test assumes a constant probability of a positive day
- Too **few** runs ($z < 0$) mean signs that persist
  - too **many** ($z > 0$) mean signs that alternate
- S&P 500: $z = 4.34$, signs alternate, as $\hat{\rho}_1 < 0$
- BET: $z = -6.05$, signs persist, as $\hat{\rho}_1 > 0$
- Bitcoin: $z = 2.68$ rejects independent signs although $\hat{\rho}_1$ is not significant
  - signs and magnitudes carry different information
  - all markets: Appendix

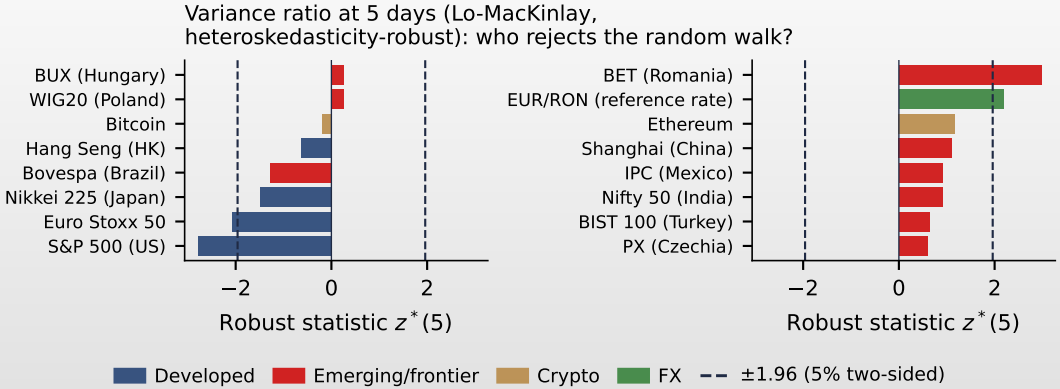
The Lo–MacKinlay Test

- Lo & MacKinlay (1988) estimate $\mathrm{VR}(q)$ from overlapping sums of q daily returns and correct the small-sample bias
  - $\hat{\sigma}_a^2 = \frac{1}{T-1}\sum_t (r_t - \hat{\mu})^2, \quad \hat{\sigma}_c^2(q) = \frac{1}{m}\sum_{t=q}^{T}\left(\sum_{j=0}^{q-1} r_{t-j} - q\hat{\mu}\right)^2$
  - $m = q(T-q+1)(1-q/T); \quad \widehat{\mathrm{VR}}(q) = \hat{\sigma}_c^2(q)/\hat{\sigma}_a^2$
  - $\hat{\mu}$: mean daily return; $\hat{\sigma}_a^2$: variance of one-day returns; $\hat{\sigma}_c^2(q)$: variance of the overlapping q-day sums divided by q; m: a divisor that removes the small-sample bias
- Test statistic under RW1 (homoskedastic returns, i.e. with constant variance)
  - $z(q) = \dfrac{\widehat{\mathrm{VR}}(q) - 1}{\sqrt{2(2q-1)(q-1)/(3qT)}} \approx N(0,1)$
  - the denominator is the standard error of $\widehat{\mathrm{VR}}(q)$ under RW1; $|z| > 1.96$ rejects $\mathrm{VR}(q) = 1$ at 5%
- This z rejects too often under volatility clustering (next slide)

The Heteroskedasticity-Robust Statistic z^*

- Under RW3 plus the conditions of Lo & MacKinlay (1988), the $\hat{\rho}_k$ are asymptotically uncorrelated with variances $\hat{\delta}_k / T$
  - finite moments, weak dependence (mixing), $\mathbb{E}[\varepsilon_t^2 \varepsilon_{t-j} \varepsilon_{t-k}] = 0$ for $j \neq k$
  - $\hat{\delta}_k = T\hat{\tau}_k$, with $\hat{\tau}_k$ of the autocorrelation slide: the robust variance of $\sqrt{T}\hat{\rho}_k$, equal to 1 without volatility clustering
- Asymptotic variance of $\widehat{\mathrm{VR}}(q)$ and the test statistic
  - $\hat{\theta}(q) = \sum_{k=1}^{q-1} \left(\frac{2(q-k)}{q} \right)^2 \hat{\delta}_k, \qquad z^*(q) = \frac{\sqrt{T}\left(\widehat{\mathrm{VR}}(q) - 1\right)}{\sqrt{\hat{\theta}(q)}} \approx N(0,1)$
  - $2(q-k)/q$: the weight of $\hat{\rho}_k$ in $\widehat{\mathrm{VR}}(q)$; $|z^*| > 1.96$ rejects $\mathrm{VR}(q) = 1$ at 5%
- RW3 alone is not enough
  - the conditions hold for GARCH with symmetric innovations
  - but can fail with a leverage effect
  - Without the condition $\mathbb{E}[\varepsilon_t^2 \varepsilon_{t-j} \varepsilon_{t-k}] = 0$, $\hat{\theta}(q)$ needs the full covariance matrix of the $\hat{\rho}_k$
- At $q = 5$, the S&P 500 has $z^*(5) = -2.78$
  - the reversal stays significant with the robust standard error
  - the comparison of z with z^* at $q = 2$: section 2

The Variance Ratio at 5 Days on 16 Series



MFM_ch2_variance_ratio

The 5-Day Variance Ratio and the Chow–Denning Test

- Each bar on the previous chart is the robust statistic $z^*(5)$ of one of the 16 series.
  - the dashed lines mark ± 1.96.
- At 5 days, $|z^*(5)| > 1.96$ for:
  - the BET (2.98), the S&P 500 (-2.78).
  - EUR/RON (2.19) and the Euro Stoxx 50 (-2.08).
- The joint Chow & Denning (1993) test rejects RW3 for 5 of the 16 series.
  - critical value 2.49 (section 2).
  - BET, S&P 500, EUR/RON.
  - IPC (at 2 days, $z^* = 3.72$) and Ethereum (at 20 days, $z^* = 2.70$).
- The Euro Stoxx 50 rejects at 5 days but not in the joint test: its largest $|z^*|$ is $2.24 < 2.49$.
- No ranking "developed = efficient": the S&P 500 shows reversal, the BET and EUR/RON persistence.

MFM_ch2_variance_ratio

Beyond Linear Predictability: the Martingale Difference Hypothesis

- **MDH** (Martingale Difference Hypothesis): $\mathbb{E}[r_t \mid \mathcal{I}_{t-1}] = \mu$
  - no function of the past predicts the mean, linear or not
- RW3 (all $\rho_k = 0$) is weaker. Counterexample: $r_t = \varepsilon_t + \theta(\varepsilon_{t-1}^2 - \sigma^2)$, ε_t i.i.d. symmetric
  - $\sigma^2 = \mathrm{Var}(\varepsilon_t)$, so the extra term has mean 0; $\theta \neq 0$: how strongly yesterday's squared shock moves today's mean
  - all $\rho_k = 0$, yet $\mathbb{E}[r_t \mid \varepsilon_{t-1}] = \theta(\varepsilon_{t-1}^2 - \sigma^2)$ is predictable: autocorrelation tests cannot see it
  - The MDH tests on the next slide look for exactly this kind of dependence (results: Appendix)
- **BDS** (Brock–Dechert–Scheinkman) Brock, Dechert, Scheinkman & LeBaron (1996) compares the frequency of nearby return sequences
  - with that implied by i.i.d. observations
  - it tests i.i.d., not the MDH
  - GARCH alone makes it reject
  - A BDS rejection is evidence against RW1, not against weak-form efficiency

What nonlinear tests examine

- Why: the counterexample on the previous slide shows that zero autocorrelations do not rule out a predictable mean
- The generalised spectral test Escanciano & Velasco (2006) measures the covariance between r_t and nonlinear functions of past returns
  - $e^{\mathrm{i}xr_{t-j}}$, at all lags j
  - it detects the counterexample $r_t = \varepsilon_t + \theta(\varepsilon_{t-1}^2 - \sigma^2)$, which autocorrelations miss
- The automatic portmanteau test Escanciano & Lobato (2009) is a heteroskedasticity-robust Ljung–Box
  - with the number of lags chosen from the data by a penalty
  - it tests autocorrelations only, so it does not detect the nonlinear counterexample

HAC standard errors and bandwidth

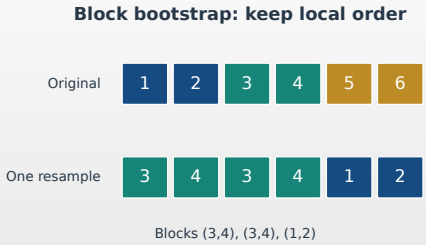
OLS: Ordinary Least Squares.

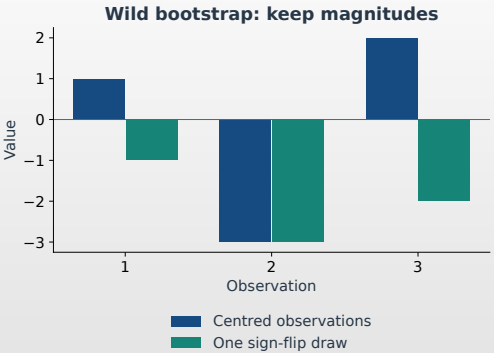
- Why: the residuals are autocorrelated and heteroskedastic
  - in predictive regressions and on overlapping rolling windows
  - and classical OLS standard errors are then not valid
  - in the dividend–price regression they are too small (OLS $t = 1.16$, Newey–West $t = 0.91$)
  - HAC errors can also be smaller than OLS ones (the calendar-effects table)
- HAC standard errors estimate the long-run variance
  - the variance plus twice the sum of the autocovariances
- Newey & West (1987) weight the autocovariance at lag j by $1 - j/(\ell + 1)$, so that the estimate is never negative
- The bandwidth ℓ is the number of lags included
  - with $\ell = 5$ the weights of lags 1–5 are $5/6, 4/6, \ldots, 1/6$

Bootstrap: constructing repeated samples

- Why: neither the Stambaugh bias nor the DFA null band has a known exact distribution
  - the bootstrap approximates it
- The bootstrap generates many artificial samples under the null and recomputes the statistic in each
- The i.i.d. bootstrap draws observations with replacement
  - and destroys all dependence, volatility clustering included
- The block bootstrap Künsch (1989) draws consecutive blocks and keeps the dependence within each block
- The wild bootstrap multiplies each demeaned return by a random sign
  - it keeps each day's volatility but destroys serial correlation: it reproduces an RW3-type null with symmetric signs (used for DFA)

What each bootstrap preserves





Blocks preserve local ordering; sign flips preserve absolute magnitudes. They impose different assumptions on the artificial samples.

Predictive Regressions and the Stambaugh Bias (1/2)

- Back to the regression of section 3: why can *dp* look like a good predictor even when $\beta = 0$?
- Model: $r_{t+1} = \alpha + \beta x_t + u_{t+1}, \quad x_{t+1} = \theta + \rho x_t + v_{t+1}$, with $\mathrm{Cov}(u, v) = \sigma_{uv}$
  - r: excess return; x: log dividend–price ratio $dp_t = \ln D_t - \ln P_t$, persistent ($\rho \approx 1$)
  - D_t: dividends paid over the last year; β: the predictive slope, 0 if returns are unpredictable; α, θ: intercepts
  - ρ: the persistence of x, the AR(1) coefficient of *dp* (not the return autocorrelation ρ_k, nor the ρ of the Campbell–Shiller identity); u, v: the shocks of the two equations; σ_{uv}: their covariance; $\sigma_v^2 = \mathrm{Var}(v)$; T: number of periods

Predictive Regressions and the Stambaugh Bias (2/2)

- x_t is predetermined, not strictly exogenous: OLS is consistent but biased Stambaugh (1999)
  - Project u on v: $u = (\sigma_{uv}/\sigma_v^2)\, v + e$, so $\hat{\beta} - \beta = (\sigma_{uv}/\sigma_v^2)(\hat{\rho} - \rho) + \text{noise}$; e: the part of u unrelated to v
  - Both regressions use the same months and the same x_t, so the estimation errors $\hat{\beta} - \beta$ and $\hat{\rho} - \rho$ are linked through the correlation of the shocks u and v; the noise term has mean zero when e is independent of the whole path of x, e.g. (u, v) i.i.d. with a bivariate Normal distribution
  - $\hat{\rho}$ is biased downward in small samples Kendall (1954), and this error passes into $\hat{\beta}$
  - With $\mathbb{E}[\hat{\rho} - \rho] \approx -(1 + 3\rho)/T$: $\quad \mathbb{E}[\hat{\beta} - \beta] \approx -\dfrac{\sigma_{uv}}{\sigma_v^2}\,\dfrac{1 + 3\rho}{T}$
- Price up $\Rightarrow$ *dp* down: $\sigma_{uv} < 0$, so the bias is **upward**: predictability looks stronger than it is
- Near a unit root the t-statistic is not $N(0, 1)$ either
  - valid inference by a Bonferroni Q-test over a confidence interval for ρ Campbell & Yogo (2006)

Bonferroni Q: allowing for uncertain persistence

- Campbell–Yogo inference allows for uncertainty about predictor persistence. Campbell & Yogo (2006)
  - instead of treating its estimate as known.
- It combines a confidence set for persistence with tests of the predictive slope over that set.
  - allocating the error probability across the steps.
- For the S&P 500, $\hat{\rho} = 0.9988$ and $T(1 - \hat{\rho}) = 1.4$.
  - a unit-root test cannot reject $\rho = 1$.
  - there the t-statistic of $\hat{\beta}$ departs most from $N(0, 1)$.
- This Q-test concerns predictive regressions
  - it is not the Ljung–Box Q statistic for autocorrelations.

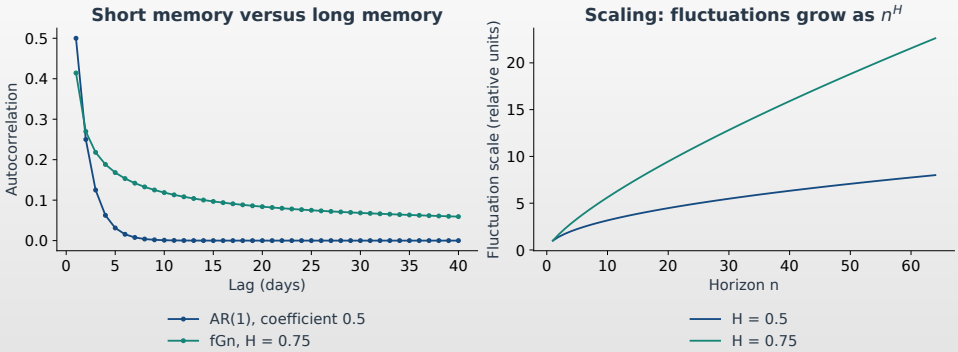
Worked Example: the Stambaugh Bias for the S&P 500

- Monthly log excess return regressed on the lagged dp_t, 1934–2026, $T = 1\,111$ months
  - data construction: Appendix
- OLS: $\hat{\beta} = 0.267$, Newey–West $t = 0.91$: not significant even before any correction
  - in % per month per unit of dp
- Inputs: $\hat{\rho} = 0.9988$, $\text{corr}(\hat{u}, \hat{v}) = -0.97$, $\hat{\phi} = \hat{\sigma}_{uv}/\hat{\sigma}_v^2 = -0.925$ (as fractions, not in %)
- Substitution: bias $= -100\,\hat{\phi}\,(1 + 3\hat{\rho})/T = 0.333$ (in %, like $\hat{\beta}$)
  - corrected slope $= 0.267 - 0.333 = -0.066$: the estimated first-order bias correction exceeds the estimated slope
  - the correction is itself estimated: formal inference by Campbell & Yogo (2006) or a bootstrap
  - Kendall's formula is first order in $1/T$ and least accurate when $T(1 - \hat{\rho})$ is small, here 1.4: read -0.066 as indicative of the sign and size, not as a precise slope
- Simulation: the nominal 5% t-test rejects $\beta = 0$ in 26.5% of samples
  - with the same ρ and $\text{corr}(u, v)$, $\beta = 0$ and a 30-year sample ($T = 360$ months)
  - more than five times too often

Comparing Forecasts: Diebold–Mariano and Clark–West

- Section 3 defined R^2_{OS} and applied the Clark–West test to the S&P 500
  - here we show why the test has this form
- Most published predictors have $R^2_{OS} < 0$ Welch & Goyal (2008)
  - the update Goyal, Welch & Zafirov (2024) confirms it
- The Diebold & Mariano (1995) test compares the mean difference between the squared errors of the two forecasts
- The *dp* model contains the historical mean (nested models)
  - under H_0 it estimates a zero β
  - the noise of that estimate inflates its mean squared prediction error (MSPE)
- So the Diebold–Mariano test rejects less often than its nominal level
- Clark & West (2007) add this noise back: $f_t = (r_t - \bar{r}_t)^2 - \left[(r_t - \hat{r}_t)^2 - (\bar{r}_t - \hat{r}_t)^2\right]$
  - $(\bar{r}_t - \hat{r}_t)^2$: the part of the model's error due to estimating β
  - the t test on the mean of f_t is one-sided: only a positive gain is evidence for the predictor

Dependence across lags and horizons

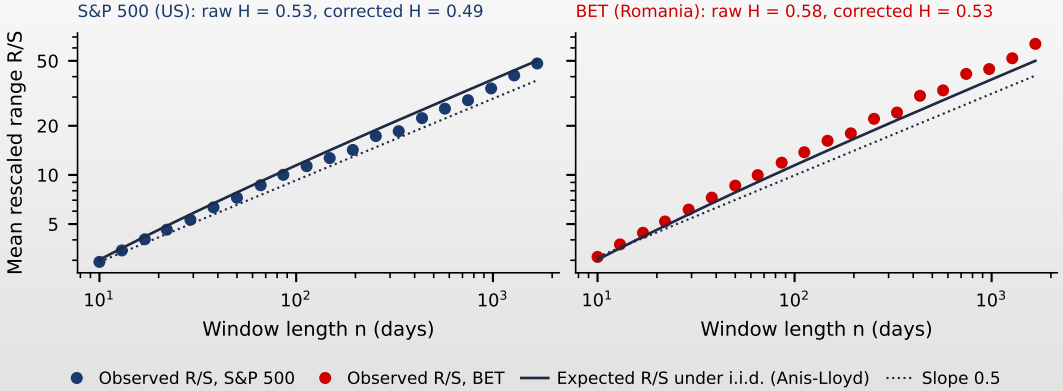


AR(1) correlations fade geometrically; the fGn example fades slowly. The right panel compares fluctuation growth across horizons.

The Rescaled Range (R/S) Statistic

- Using the rescaled range, Hurst (1951) found that years of high Nile floods followed other high-flood years
  - more often than chance allows
  - history: Appendix
- Construction for a window of n returns with mean $\bar{r}_n$ and standard deviation s_n
  - Cumulative deviations $Y_k = \sum_{i=1}^{k}(r_i - \bar{r}_n)$, $k = 1, \ldots, n$
  - Range $R_n = \max_k Y_k - \min_k Y_k$; rescaled range R_n/s_n
- Scaling law: $\mathbb{E}[R_n/s_n] \approx c\, n^H$, where H is the **Hurst exponent**
  - c: a constant; taking logs, $\ln(R/S) \approx \ln c + H \ln n$, so H is a slope
  - Estimate: slope of $\ln(R/S)$ on $\ln n$, averaging R/S over non-overlapping windows
- $H = 0.5$: no long memory (short-run dependence allowed)
  - $H > 0.5$: persistence, long memory with $\rho_k \sim k^{2H-2}$
  - $H < 0.5$: anti-persistence

R/S Analysis: S&P 500 and BET (1/2)



MFM_ch2_hurst

R/S Analysis: S&P 500 and BET (2/2)

- On the previous chart, the dots show the average R/S for each window length n.
- The solid line is the Anis & Lloyd (1976) expectation for i.i.d. returns from the Normal distribution.
  - formula: Appendix.
- The Anis–Lloyd correction removes the bias of the raw slope.
  - The raw slope exceeds 0.5 in small samples even for i.i.d. returns.
  - $H_{AL} = 0.5+$ the slope of the regression of $\ln(R/S) - \ln \mathbb{E}[R_n/s_n]$ on $\ln n$.
- For the S&P 500, the raw exponent is 0.530 and the corrected one 0.485.
  - the apparent "persistence" came from small-sample bias.
- For the BET, the raw exponent is 0.578 and the corrected one 0.533.
  - persistence remains after the correction.
- Open question: is BET's $H > 0.5$ long memory, or just its lag-1 autocorrelation $\hat{\rho}_1 = 0.109$?
- The classical R/S statistic cannot tell the two apart.

MFM_ch2_hurst

Lo's Modified R/S Test

- Lo (1991): short-run autocorrelation inflates R/S and mimics long memory
- Modification: s_n is replaced by a long-run (HAC) standard deviation $\hat{\sigma}_\ell$ with ℓ lags
  - $\hat{\sigma}_\ell^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{\ell}\left(1 - \frac{j}{\ell+1}\right)\hat{\gamma}_j$, $\hat{\gamma}_j$: autocovariance at lag j ($\hat{\gamma}_0$: the variance), with declining weights
  - $V = \dfrac{\max_k Y_k - \min_k Y_k}{\hat{\sigma}_\ell \sqrt{T}}$; Y_k: cumulative deviations over the full sample of T returns; bandwidth ℓ chosen as in Andrews (1991)
- Under short memory, V lies in $[0.809, 1.862]$ with probability 95%
- V is inside the interval: short memory is not rejected
  - for the S&P 500 ($V = 1.560$), Bitcoin (1.547), PX (1.749) and EUR/RON (1.429)
- For the BET ($V = 1.961$, with $\ell = 7$ lags) and Ethereum ($V = 1.980$), V is just above 1.862
- This is weak evidence of long memory

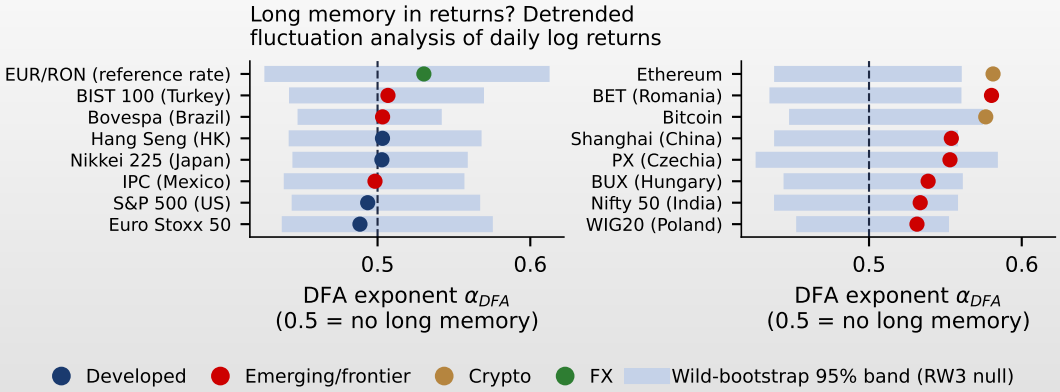
Detrended Fluctuation Analysis (DFA) (1/2)

- Peng et al. (1994): DFA measures how the fluctuation of cumulative returns grows with the window length.
- Removing a linear trend from the profile removes a constant mean of returns in each box.
- If the mean of returns itself trends, higher-order DFA is needed.
- The algorithm
  - The profile $y_k = \sum_{t \leq k}(r_t - \bar{r})$ is the running sum of deviations from the mean, over the full sample.
  - The profile is split into boxes of length n.
  - In each box a linear trend is fitted and the residuals are kept.
  - $F(n)$ is the root mean square of all residuals; the calculation is repeated for many n.
- $F(n) \propto n^{\alpha_{DFA}}$: the DFA exponent α_{DFA} plays the role of H ($\alpha_{DFA} = 0.5$: no long memory).

Detrended Fluctuation Analysis (DFA) (2/2)

- The null band comes from a **wild bootstrap** (199 replications).
  - each residual $e_t = r_t - \bar{r}$ is multiplied by a random sign $\eta_t = \pm 1$, with probability $1/2$.
- The sign flip keeps the volatility $|e_t|$, allowed under RW3, and destroys only serial correlation.
  - a shuffle would destroy both.
- The implicit null is stricter than RW3: signs symmetric and independent, given volatility.
- The band is therefore a benchmark, not an exact test of RW3 and not a confidence interval for α_{DFA}.

DFA Exponent Across Markets (1/2)



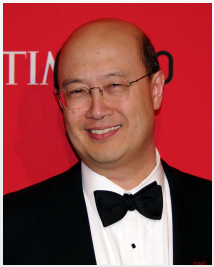
MFM_ch2_hurst

DFA Exponent Across Markets (2/2)

- Each dot on the previous chart is the exponent α_{DFA} of one series.
  - the bar is the 95% wild-bootstrap null band.
- S&P 500 (0.494), Euro Stoxx 50 (0.488) and Nikkei (0.503) are close to 0.5, inside the null band.
- Three of the 16 series exceed the null band.
  - Ethereum (0.581).
  - BET (0.580, against an upper limit of 0.561).
  - and, marginally, Bitcoin (0.576 against 0.575).
- PX (0.553) and Shanghai (0.554) exceed only a null band built by shuffling.
  - Shuffling also destroys volatility clustering.
  - their apparent memory comes from volatility clustering.
- The local Whittle estimator Robinson (1995) estimates $d = H - 0.5$ from the lowest frequencies.
  - for the BET, $\hat{d} = 0.118$, 95% interval $[0.062, 0.174]$ (Appendix).
- The interval assumes conditionally homoskedastic innovations.
  - volatility clustering violates this assumption.
  - so it is not decisive evidence.
- Conclusion: the evidence of long memory in returns is weak and depends on the method.
  - memory in volatility (Chapter 1) is strong.

MFM_ch2_hurst

Adaptive Markets Hypothesis (AMH) (1/2)



Andrew W. Lo, MIT

Photo: David Shankbone (2012); CC BY 3.0; Wikimedia Commons

- Full-sample tests give weak and unstable evidence
  - one possible reason: efficiency varies over time
- Andrew Lo proposed the Adaptive Markets Hypothesis (AMH) in Lo (2004) and developed it in the book Lo (2017)
- Markets are an ecology of competing species that adapt by trial and error
  - traders, funds, algorithms
  - not by perfect optimisation
  - efficiency is a matter of degree and depends on the number of competitors, the profit opportunities and the environment
- Anomalies appear, weaken and may reappear
- Time-varying statistics alone do not identify the AMH
  - since an Efficient Market Hypothesis (EMH) with time-varying risk premia also allows them
  - comparison table: Appendix

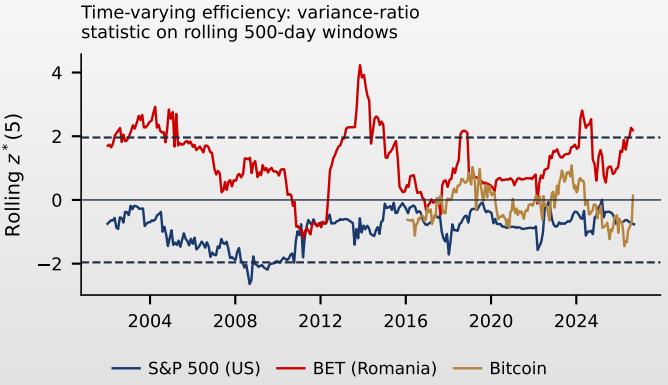
Adaptive Markets Hypothesis (AMH) (2/2)

- Lim & Brooks (2011), a survey: stock-market efficiency varies over time, especially around crises
- Urquhart & McGroarty (2016): serial dependence appears and disappears across subperiods with market conditions
  - on four major indices (US, UK, Japan, euro area)

Measuring Time-Varying Efficiency

- Compute a statistic on a **rolling window**
  - 500 trading days (about two years)
  - moved forward by 21 days (one month)
  - Statistics: $z^*(5)$ of the variance-ratio test and the DFA exponent
- The DFA band is computed as before (random sign flips), but separately in each window
  - so it follows the window's volatility
- For comparison, a fixed null band is $[0.380, 0.615]$
  - simulated from i.i.d. Student-t returns with 4 degrees of freedom
  - heavy tails, no volatility clustering
- The fixed band ignores volatility clustering, allowed under RW3
  - volatility clustering raises α_{DFA}
  - so the fixed band flags too many windows
- Two cautions for reading the results
  - Overlapping windows are strongly correlated: runs of rejections are not independent evidence
  - Even under the null about 5% of windows reject; a share near 5% is a description, not a calibrated test of stability

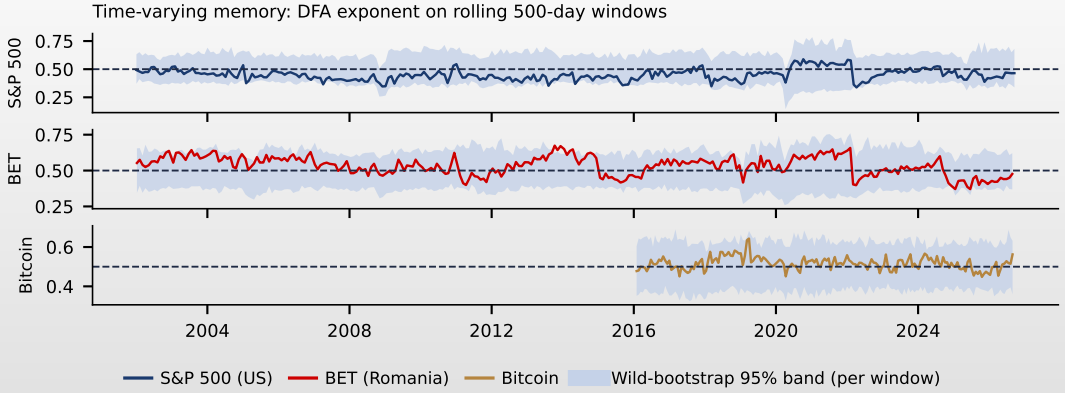
Rolling Variance Ratio: S&P 500, BET, Bitcoin



- $z^*(5)$ on rolling 500-day windows, dated by the last day
  - dashed lines ± 1.96
- Share of windows with $|z^*(5)| > 1.96$: S&P 500 6.7%, BET 22.4%, Bitcoin 0.0%
- S&P 500: negative in almost every window, beyond -1.96 only in windows ending in 2008–2010
- BET: episodes (windows ending in 2002–2005, 2013–2014, 2018 and after 2023)
- The upgrade by the index provider FTSE Russell (FTSE: Financial Times Stock Exchange) of 21 Sep 2020 (Appendix) shows no break
  - BET was near zero from 2016 and rises again after 2023

MFM_ch2_rolling_efficiency

Rolling DFA Exponent: S&P 500, BET, Bitcoin (1/2)



MFM_ch2_rolling_efficiency

Rolling DFA Exponent: S&P 500, BET, Bitcoin (2/2)

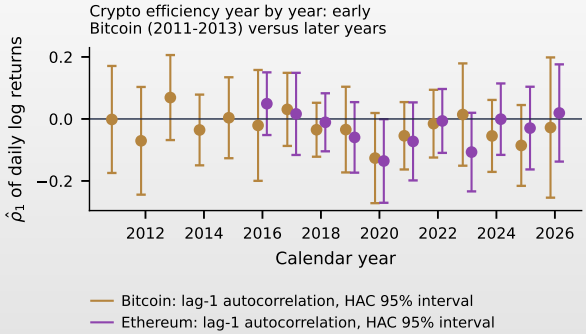
- α_{DFA} on rolling 500-day windows.
  - shaded: the wild-bootstrap 95% null band computed in each window.
- Outside their own band: S&P 500 1.7%, BET 3.1%, Bitcoin 0.0% of windows.
- With a fixed Student-t_4 band: 6.7%, 11.2%, 1.1%.
  - the fixed band ignores volatility clustering.
  - the extra DFA exceedances of the BET came from volatility clustering.
- The BET episodes in the variance-ratio chart remain.
  - they concern autocorrelation over a few days (short memory).
  - the DFA exponent does not target short memory.
- Low power: the BET band averages $[0.36, 0.64]$.
  - 500 observations cannot pin down a scaling exponent.

MFM_ch2_rolling_efficiency

The AMH, Tested

- Kim, Shamsuddin & Lim (2011) show that predictability varies with market conditions (crises, bubbles).
  - rather than along a declining trend.
  - over a century of US data.
  - and make the AMH testable by linking it to state variables.
- Our version: regress the rolling $z^*(5)$ on the window's annualised volatility (%) and a crisis indicator
  - The crisis indicator is 1 if the window overlaps Sep 2008–Jun 2009 or Feb–Jun 2020.
  - Newey–West errors with 24 lags, because consecutive windows share 95% of their days.
  - Volatility is the control variable: windows with similar volatility are compared.
- S&P 500: the crisis coefficient is -0.71 ($t = -2.53$).
  - the volatility coefficient 0.008 ($t = 0.39$).
  - $R^2 = 0.21$.
- BET: the crisis coefficient is -0.82 ($t = -2.77$).
  - the volatility coefficient -0.008 ($t = -0.36$).
  - $R^2 = 0.15$.
- In crisis windows $z^*(5)$ is lower, i.e. closer to reversal, even at equal volatility.
- Type of evidence: association.
  - conditional on volatility, not a causal effect.
- State dependence is consistent with the AMH but does not identify it.
  - a change in $z^*(5)$ can come from $\widehat{\mathrm{VR}}(5) - 1$ or from its standard error.

Crypto Efficiency Year by Year (1/2)



- The AMH predicts that a new market becomes more efficient over time
  - $\hat{\rho}_1 \neq 0$ in Bitcoin's early years, values near zero later
- The chart: $\hat{\rho}_1$ of daily returns in each calendar year, with the HAC 95% interval, for comparing years
- Bitcoin 2011–2013: the intervals are wide and include zero
  - the data do not show the early inefficiency the AMH predicts

MFM_ch2_crypto_efficiency

Crypto Efficiency Year by Year (2/2)

- Only one of the 27 yearly autocorrelations is significant at 5%: Ethereum in 2023, $\hat{\rho}_1 = -0.107$
  - 16 for Bitcoin, 11 for Ethereum
  - by chance alone we would expect about 1.35
  - the test uses SE_0 (derived under H_0), the chart's bars the HAC SE: the bar of Ethereum 2020 excludes 0, that of 2023 does not
- Urquhart (2016) rejects efficiency in the early years (Appendix)
  - with other tests, another price source and classical standard errors
- Power: one year of 365 returns gives a robust SE of about 0.05–0.13, so small effects are invisible
  - the robust SE requires a finite fourth moment, which Chapter 1 questions

Event Studies: the Prediction Error

- Prediction-error variance MacKinlay (1997): $\mathrm{Var}(AR_\tau) = \sigma_\varepsilon^2\left[1 + \frac{1}{T_0} + \frac{(r_{m,\tau} - \hat{\mu}_m)^2}{(T_0 - 1)\hat{\sigma}_m^2}\right]$, T_0: estimation days
  - σ_ε^2: the residual variance of the market model; $r_{m,\tau}$: the market return on event day τ.
  - $\hat{\mu}_m$, $\hat{\sigma}_m^2$: mean and variance of the market return in the estimation window.
  - The bracket grows when the event-day market move is unusual.
- For the bank tax (section 4), the extreme BET move makes the bracket 2.12.
  - hence $t = -6.85$ instead of -10.0 for TLV.
- With the same correction, BRD has $t = -4.03$.
- Boehmer, Musumeci & Poulsen (1991) standardise each AR by its own prediction error.
  - estimate the variance from the firms' standardised values on the event day, not from calm days.
  - the test stays valid if the event raises the variance.

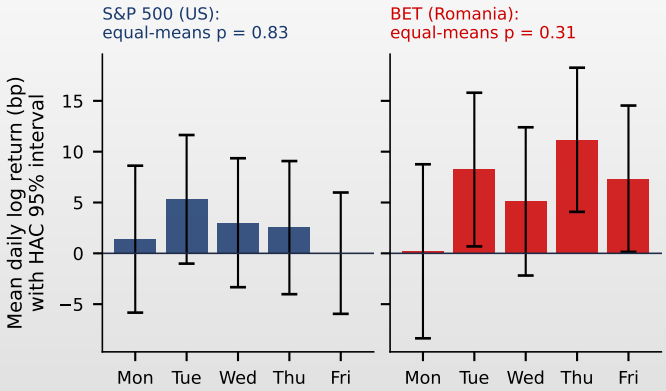
Event Studies: Correlation Across Firms

- The question: what happens to the test of $\overline{AR}_0$ when all firms have the event on the same day?
- Let N be the number of firms and $\bar{\varrho}$ the average correlation of their abnormal returns.
  - estimated in the estimation window.
  - for TLV and BRD, $\bar{\varrho} = 0.153$.
- The test that assumes independence ignores the covariances.
  - with $\bar{\varrho} > 0$ the true variance of the mean is $1 + (N-1)\bar{\varrho}$ times larger.
- For TLV and BRD, the t computed under independence is -7.55.
  - divided by $\sqrt{1+\bar{\varrho}} = 1.074$ it becomes -7.03.
- That is the t of the equal-weighted portfolio, which allows for the correlation automatically.
- If the variance is estimated cross-sectionally, as in Boehmer, Musumeci & Poulsen (1991), it further understates the true variance.
  - by the factor $1 - \bar{\varrho}$.
  - this is why Kolari & Pynnönen (2010) multiply the statistic by $\sqrt{(1-\bar{\varrho})/(1+(N-1)\bar{\varrho})}$.
- With ten firms and the same correlation, the independence-based test understates the variance.
  - By a factor of $1 + 9\bar{\varrho} = 2.38$.
  - the cross-sectional test by $(1+9\bar{\varrho})/(1-\bar{\varrho}) = 2.81$.

Anomalies and Calendar Effects

- The question: which patterns survive the correction for multiple testing, and what explains them?
- An **anomaly** is a return pattern the asset-pricing model does not explain.
- Three possible explanations.
  - An omitted risk factor.
  - Mispricing (a behavioural bias).
  - Data snooping.
- Families:
  - calendar (section 5): TOM Lakonishok & Smidt (1988), first half of the month Ariel (1987).
  - time-series (time-series momentum).
  - cross-sectional (size, value, momentum: **Chapter 3**).
- Returns on published anomalies fall McLean & Pontiff (2016):
  - by 26% after the end of the sample studied.
  - by 58% after publication.
- Day of week: $r_t = \sum_{d=1}^{5} \beta_d D_{d,t} + u_t$.
  - $D_{d,t} = 1$ if day t is weekday d (Monday, ..., Friday), else 0.
  - β_d is the mean return on weekday d.
- The Wald test of $\beta_1 = \cdots = \beta_5$ (χ^2_4) uses HAC errors Newey & West (1987) with 5 lags.
  - because u_t is heteroskedastic and autocorrelated.

Day-of-the-Week Effect: S&P 500 and BET



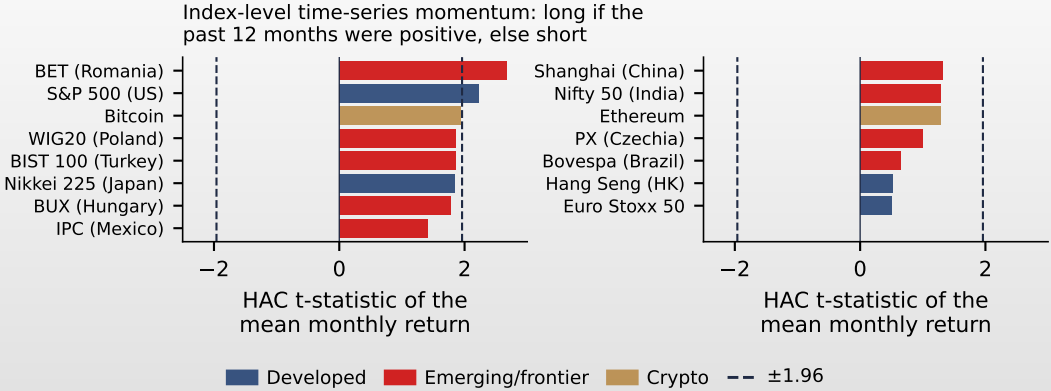
- Mean daily log return by weekday, in basis points (bp), 2000–2026, HAC 95% intervals
  - 1 bp = 0.01%
- S&P 500: we find no Monday effect since 2000, unlike French (1980)
  - from 0.0 bp (Friday) to 5.3 bp (Tuesday)
  - equal-means $p = 0.83$
- BET: Thursday 11.2 bp ($t = 3.09$), yet the joint test gives $p = 0.31$
- Two different nulls:
  - a day's mean equal to zero (mixes the day effect with the positive average return)
  - all five means equal (the day effect itself)

MFM_ch2_calendar_anomalies

Time-Series Momentum on Indices (1/3)

- Time-series momentum (TSMOM) Moskowitz, Ooi & Pedersen (2012): signal $s_t = \text{sign}\left(\sum_{j=1}^{12} r_{t-j}\right)$
  - $+1$ (long) if the past 12 monthly log returns sum to a gain, -1 (short) otherwise
  - The strategy return is $s_t R_t$, R_t the index return of month t
- On the next chart, each bar is the HAC t (6 lags) of the strategy's mean monthly return in one market.

Time-Series Momentum on Indices (2/3)



MFM_ch2_tsmom

Time-Series Momentum on Indices (3/3)

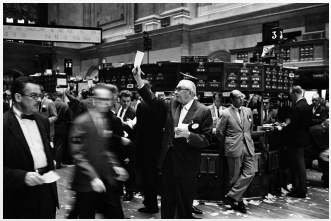
- Taken alone, the strategy has a mean return significant at 5%:
  - for the BET (1.40% a month, $t = 2.68$).
  - the S&P 500 ($t = 2.22$).
- For Bitcoin, $t = 1.94$, below 1.96.
- But we tested 15 markets.
  - the first Holm threshold is $0.05/15 = 0.0033$.
  - the smallest p-value, the BET's, is 0.007.
- After the multiple-testing correction, no market remains significant.
- The TSMOM Sharpe ratio is below buy-and-hold: 0.65 vs 0.75 for the BET, 0.48 vs 0.52 for the S&P 500.
- Type of evidence: comparison of procedures.
- Part of the TSMOM return may be the equity premium, earned in the months spent long.
  - the mean stock return over the risk-free rate.
- The Sharpe ratio is annualised (full table: Appendix).
  - $\sqrt{12}\times$ monthly mean over monthly SD.
  - on price indices without dividends or a risk-free rate.

MFM_ch2_tsmom

Behavioural concepts

- Prospect theory: gains and losses are judged relative to a reference point.
  - losses weigh more than equal gains (loss aversion).
  - small probabilities are overweighted.
- Disposition effect: the tendency to sell winners and keep losers.
- Anchoring: excess weight on an initial number, such as the purchase price.
- Overconfidence: overstating the precision of one's beliefs, which leads to too much trading.
- Limited attention: relevant information left unprocessed.
- Herding: following others' actions without an independent assessment.
- Underreaction and overreaction: a price adjusts too little or too much to news at first.
  - the second is followed by reversal.

Psychology in Prices, from the Floor to the Feed



New York Stock Exchange trading floor, 1963

Photo: Thomas J. O'Halloran (1963); public domain; Wikimedia Commons

- **Excess volatility** Shiller (1981): prices move far more than the dividends that should justify them
- **Overreaction**: De Bondt & Thaler (1985) use it to explain 3–5-year return reversals
  - portraits: Appendix
- **Prospect theory** Kahneman & Tversky (1979) explains the disposition effect
- **Herding** in an illiquid market (BVB): herding produces log-periodic bubbles Pele, Mazurencu-Marinescu & Nijkamp (2013)
  - log-periodic bubble: prices rising super-exponentially, with ever faster oscillations before the crash
  - today its channels are order books, algorithms and social media (Chapter 10)
- Biases alone do not break efficiency.
  - rational arbitrageurs could trade against them.
  - Can they?

Limits to Arbitrage



Andrei Shleifer, Harvard

Photo: Manuel Castells (2018); CC BY-SA 2.0; Wikimedia Commons

- Shleifer & Vishny (1997): real arbitrage is risky and needs capital
- **Noise-trader risk** De Long, Shleifer, Summers & Waldmann (1990): mispricing can get worse before it corrects
  - Noise traders: investors who trade on signals without information content (sentiment, rumours)
- Funding constraints: managers lose clients after losses and must close positions at the worst time
- Costs: short-selling limits, margins, trading costs, no perfect substitutes
- Anomalies can persist where arbitrage is hard
  - in small, illiquid markets such as the Romanian one (BET) we find the strongest persistence among the stock indices
- The AMH reconciles both views: rational and biased behaviour coexist and evolve

Can Machine Learning Beat Efficient Markets?

- ML (machine learning) learns a prediction rule from data (Chapter 13)
  - captures nonlinear patterns that autocorrelation tests miss
- Gu, Kelly & Xiu (2020): trees and neural networks give large economic gains over linear regressions
  - in the US cross-section
  - yet their monthly out-of-sample R^2 is below 1%
- Large language models read news and filings: Tak & Pele (2025) forecast stock trends with LLaMA

Validation, leakage and shrinkage

- Flexible models, such as trees and neural networks, can also fit accidental sample patterns (overfitting).
- Walk-forward validation: the model is estimated only on data before the test period.
- Leakage: a forecast uses information unavailable at its date.
  - such as an accounting figure revised later.
- Ridge: $\hat{g} = \arg\min_g \sum_t (r_t - x_t' g)^2 + \lambda \sum_j g_j^2 = (X'X + \lambda I)^{-1} X' r$.
  - $\lambda > 0$ pulls the coefficients towards zero (shrinkage).
  - reduces estimation variance.
- With $N(0, \sigma^2)$ errors and a prior $g \sim N(0, \tau^2 I)$, ridge with $\lambda = \sigma^2/\tau^2$ is the posterior mean of the coefficients.
  - Bayesian learning.
- Cross-validation: leave out one block of data at a time (e.g. a year).
  - estimate on the rest.
  - pick the λ with the smallest error on the blocks left out.

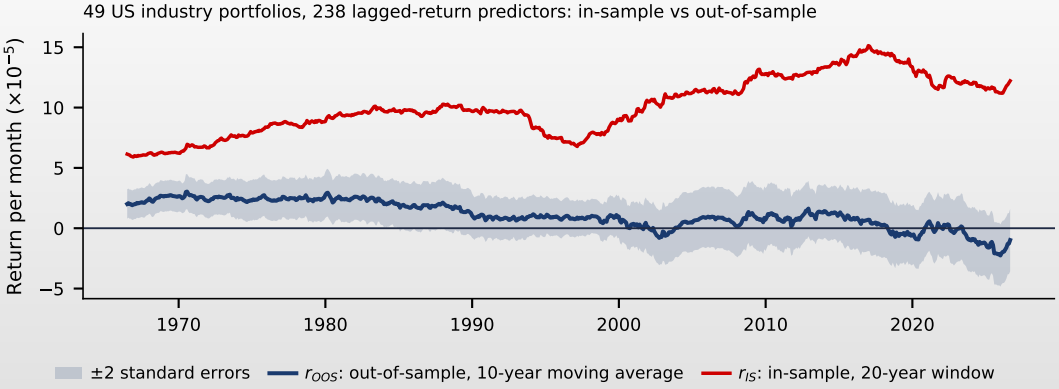
Case Study: Rational Learning with Many Predictors (1/2)

- Martin & Nagel (2022): investors price N assets whose cash flows depend on J characteristics with **unknown** coefficients
  - learn them optimally (Bayes) from past data
  - research question: Appendix
  - The posterior mean of the coefficients is the ridge estimator of the previous slide.
  - Investors are risk-neutral, so predictability cannot be a risk premium.
- **In sample**: the test of no predictability rejects with probability tending to 1 (Martin & Nagel (2022), Prop. 4).
  - if the number of predictors per asset, J/N, tends to a constant $\psi > 0$.
- The mechanism: the price embeds the shrunk coefficients, so it underreacts to X.
  - ex post, X predicts returns even though investors are rational.
- A rejection is not evidence of irrationality (formal results: Appendix).
  - the rational-expectations null is the wrong benchmark when investors must learn the parameters.

Case Study: Rational Learning with Many Predictors (2/2)

- r_{OOS}: the return of a portfolio built from forecasts.
  - in a period not used to estimate its coefficients.
- **Out of sample**: mean return zero (Martin & Nagel (2022), Prop. 6).
  - For a portfolio built with coefficients from other periods.
  - $\mathbb{E}[r_{OOS}] = 0$.
  - model and empirical design: Appendix.
- So out-of-sample tests keep their economic meaning: in this model, they test efficiency.
- For US stocks, the in-sample R^2 is about 1%.
- The out-of-sample return r_{OOS} is much lower than the in-sample one (Martin & Nagel (2022), Fig. 8).
  - after 2014, within two standard errors of zero.

Replication on Course Data: In Sample vs Out of Sample (1/2)



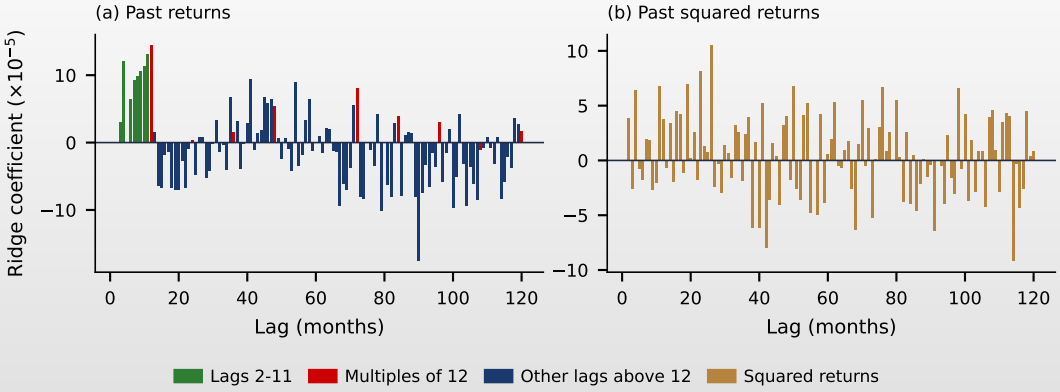
MFM_ch2_big_data

Replication on Course Data: In Sample vs Out of Sample (2/2)

- Data: 40–49 industry portfolios (Kenneth French Data Library), July 1926 – August 2026.
  - 49 in recent years.
  - 238 predictors (lagged returns and their squares, lags 2–120).
  - rolling 20-year OLS, as in Martin & Nagel (2022), Sec. 7.
- The portfolio: each month, every industry gets a weight proportional to its forecast return.
- Vertical axis: the portfolio's monthly return, in units of 10^{-5}.
  - one series is r_{IS} (coefficients from the same window).
  - the other the 10-year moving average of r_{OOS} (coefficients from the previous window), with a descriptive ± 2 standard-error band.
- r_{IS} is between 5.90 and 15.14, positive by construction.
- r_{OOS} has a mean of 1.05×10^{-5} over 842 months (from July 1956), about 11% of r_{IS}.
- In the first 30 out-of-sample years (to June 1986), the mean of r_{OOS} is 4.16 standard errors above zero, afterwards only 0.62.
- With $J/N \approx 5$, the in-sample fit is inflated: an R^2 of 5.4% against about 1% in Martin & Nagel (2022) (Appendix).
- Type of evidence: comparison of procedures.

MFM_ch2_big_data

Replication on Course Data: Ridge Coefficients by Lag (1/2)



MFM_ch2_big_data

Replication on Course Data: Ridge Coefficients by Lag (2/2)

- Full-sample ridge, January 1971 – August 2026.
  - the penalty λ is chosen by leave-one-year-out cross-validation (56 years).
- The previous chart: the ridge coefficient of each standardised predictor, by lag in months.
  - compare Martin & Nagel (2022), Fig. 7.
- For lags 2–11, 8 of 10 coefficients are positive, matching **momentum** Jegadeesh & Titman (1993).
- For multiples of 12, 9 of 10 coefficients are positive, matching **seasonality** Heston & Sadka (2008).
- For the other lags above 12, 67% of the coefficients are negative, matching long-run **reversal** De Bondt & Thaler (1985).
  - the coefficients on squared returns have no dominant sign.
- R^2 on the years left out, $1 - \sum(r - \hat{r})^2 / \sum r^2$: ridge 0.013%, OLS −2.26%.
- Ridge recovers the known anomalies (limits and extensions: Appendix).
  - yet they predict almost nothing in the years left out.

MFM_ch2_big_data

ETFs, placebo and price discovery

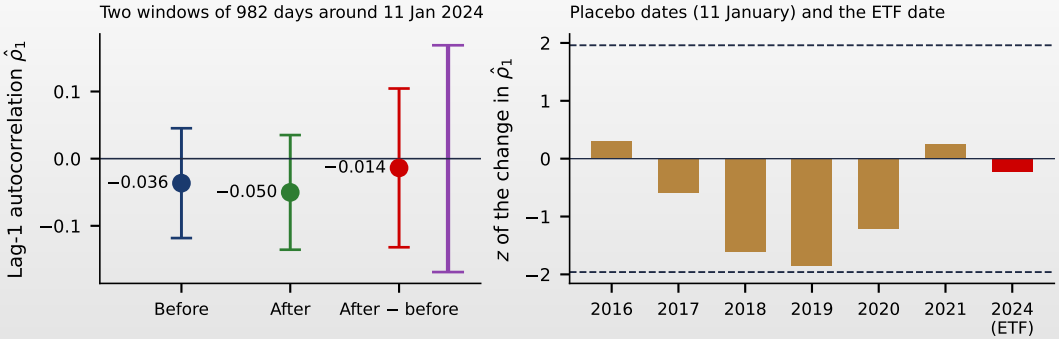
ETF: Exchange-Traded Fund.

- A spot ETF holds the underlying asset (here Bitcoin).
  - not futures contracts like the Bitcoin ETFs launched in 2021.
  - arbitrage between the fund and the spot market becomes direct.
- A placebo repeats the analysis at a date without the event.
  - a similar apparent change in 2019 weakens a claim that the 2024 launch caused the change.
- Price discovery is the incorporation of information into prices.
  - if exchange A reacts to news first and B follows, A may contribute more to price discovery.

Estimation and Power (1/3)

- The test: two non-overlapping windows of 982 days, before and after 11 January 2024.
- $z = \Delta\hat{\rho}_1 / \sqrt{s^2_{\text{pre}} + s^2_{\text{post}}}$, with $\Delta\hat{\rho}_1 = \hat{\rho}_{1,\text{post}} - \hat{\rho}_{1,\text{pre}}$ and the HAC standard errors s of the two estimates (valid also when $\rho_1 \neq 0$).
- **Minimum detectable effect (MDE)**: the smallest true change detected with 80% power at the 5% level.
- $\text{MDE} = (z_{0.975} + z_{0.80}) \times \text{SE}(\Delta\hat{\rho}_1) \approx 2.8 \times \text{SE}(\Delta\hat{\rho}_1)$
  - z_p: the p-quantile of the standard Normal distribution; $z_{0.975} = 1.96$ and $z_{0.80} = 0.84$
- Placebo dates: 11 January of each year 2016–2021, with windows of 982 returns.

Mini-Case: Estimate, Placebo and MDE



MFM_ch2_ai_discovery

Estimation and Power (2/3)

- The previous chart, left: $\hat{\rho}_1$ before and after the launch, with HAC 95% intervals.
  - the third point is the difference, with the $\pm$MDE bar.
- Right: the z statistic for the ETF date, among those of the six placebo dates.
- Before (May 2021 – Jan 2024): $\hat{\rho}_1 = -0.036$ ($s = 0.042$), VR(5) $= 1.000$ ($z^* = 0.00$).
- After (Jan 2024 – Sep 2026): $\hat{\rho}_1 = -0.050$ ($s = 0.044$), VR(5) $= 0.939$ ($z^* = -0.64$).
- $\Delta\hat{\rho}_1 = -0.014$, $z = -0.23$: an unchanged ρ_1 is not rejected.
  - The HAC standard errors s require a finite fourth moment, which Chapter 1 questions.
- None of the six placebo dates gives a significant change in ρ_1 at 5%.
  - the largest statistic is $|z| = 1.85$.
- Type of evidence: description.
  - A before/after comparison without a control group (difference-in-differences).

Estimation and Power (3/3)

- Here $\mathrm{SE}(\Delta\hat{\rho}_1) = \sqrt{0.042^2 + 0.044^2} \approx 0.060$, so the MDE is about 0.17.
- "Not significant" is not "no effect": the MDE is about 12 times the estimated change.
- Before 2024, $\hat{\rho}_1 = -0.036$.
  - even a full move to $\rho_1 = 0$ would be a change almost five times smaller than the MDE.
  - this design would detect it with probability well below 80%.
- A better outcome variable: intraday price discovery across venues, not daily ρ_1.

MFM_ch2_ai_discovery

Difference-in-differences: example and assumptions

- The before/after test cannot separate the effect of the ETFs from other changes in 2024.
- A causal conclusion needs a control group, for example Ethereum before its own ETFs (July 2024).
- Difference-in-differences (DiD) compares two before–after changes.
  - It subtracts the control group's change from the treated group's change.
- Parallel trends:
  - without treatment, the treated and control groups would have changed in the same way.
- Similar pre-treatment trends support this assumption but cannot prove it.
- No anticipation: the treated group does not react before treatment.
- No spillover: treatment does not affect the control group.
  - for example, a Bitcoin ETF must not change Ethereum trading.

Appendix: Pioneers, Bachelier and Samuelson



Louis Bachelier (1870–1946)

Photo: unknown author; public domain; Wikimedia Commons



Paul A. Samuelson (1915–2009)

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

- Bachelier (1900): prices of French government bonds move like a random walk
  - "The mathematical expectation of the speculator is zero"
- Samuelson (1965): if prices are **properly anticipated**, their changes are unpredictable
- Both results are the ancestors of the Efficient Market Hypothesis

Back

Appendix: Eugene Fama



Eugene F. Fama, Nobel Prize 2013

Photo: Bengt Nyman (2013); CC BY 2.0; Wikimedia Commons

- Fama (1970) gave the definition used in this chapter and the three forms of efficiency
- Nobel Prize 2013, shared with Robert Shiller (a sceptic of efficiency) and Lars Peter Hansen
- Update and survey: Fama (1991)
  - an accessible defence of the EMH against its critics: Malkiel (2003)

Back

Appendix: Summary Table, Weak-Form Tests 2000–2026

Market	T	$\hat{\rho}_1$	SE_0	$\mathrm{SE}_{\mathrm{HAC}}$	$\widehat{\mathrm{VR}}(5)$	$z^*(5)$	CD p	Runs z
S&P 500 (US)	6 717	−0.099	0.027	0.030	0.829	−2.78	0.001	4.34
Euro Stoxx 50	6 836	−0.029	0.019	0.016	0.910	−2.08	0.097	3.41
Nikkei 225 (Japan)	6 545	−0.038	0.024	0.018	0.923	−1.48	0.317	1.97
Hang Seng (HK)	6 583	−0.004	0.024	0.017	0.967	−0.64	0.872	−0.17
BET (Romania)	6 670	0.109	0.028	0.023	1.173	2.98	0.001	−6.05
WIG20 (Poland)	6 686	0.018	0.018	0.016	1.010	0.25	0.775	2.25
BUX (Hungary)	6 666	0.037	0.024	0.021	1.013	0.26	0.427	−0.42
PX (Czechia)	6 697	0.051	0.032	0.020	1.041	0.60	0.354	−2.53
BIST 100 (Turkey)	6 692	0.005	0.024	0.018	1.033	0.64	0.949	−0.63
Bovespa (Brazil)	6 622	−0.028	0.025	0.027	0.928	−1.29	0.521	0.88
IPC (Mexico)	6 735	0.071	0.019	0.016	1.038	0.92	0.001	−3.27
Nifty 50 (India)	6 613	0.048	0.024	0.019	1.048	0.91	0.178	−5.46
Shanghai (China)	6 474	0.021	0.019	0.018	1.046	1.11	0.409	1.02
Bitcoin	4 384	−0.022	0.026	0.021	0.991	−0.19	0.835	2.68
Ethereum	3 913	−0.012	0.027	0.023	1.067	1.17	0.027	3.91
EUR/RON (reference rate)	5 352	0.170	0.036	0.033	1.157	2.19	0.000	−1.66

- T: number of daily returns
  - SE_0: robust SE of $\hat{\rho}_1$ under H_0, for the test; $\mathrm{SE}_{\mathrm{HAC}}$: HAC SE of the estimator, for intervals
  - CD: Chow–Denning over $q = 2, 5, 10, 20$
  - EUR/RON from July 2005, crypto on 365 days per year

Back

Appendix: Runs Test on All Markets, Other VR Tests

- Runs z (table on the previous slide):
  - Nifty 50 -5.46, IPC -3.27 (persistence in emerging markets)
  - Ethereum 3.91, significant although $\hat{\rho}_1$ is not
- Choi (1999): an automatic VR test that chooses the horizon from the data
  - spectral estimate at frequency zero
- Wild-bootstrap versions for small samples: Kim (2006)
  - applications to Asian markets: Kim & Shamsuddin (2008)

Back

Appendix: Automatic Portmanteau Test of the MDH

- What we derive: a portmanteau test valid under the MDH with conditional heteroskedasticity Escanciano & Lobato (2009).
  - the number of lags chosen from the data.
- $AQ = T\sum_{j\le\tilde{p}}\tilde{\rho}_j^2$, with $\tilde{\rho}_j = \hat{\gamma}_j/\sqrt{\hat{\kappa}_j}$, $\hat{\gamma}_j = \frac{1}{T-j}\sum_t e_t e_{t-j}$ and $\hat{\kappa}_j = \frac{1}{T-j}\sum_t e_t^2 e_{t-j}^2$
  - $e_t = r_t - \bar{r}$; $\hat{\gamma}_j$: the autocovariance at lag j; $\hat{\kappa}_j$: its robust variance
  - $\tilde{p} \le 10$ maximises $T\sum_{j\le p}\tilde{\rho}_j^2$ minus a penalty: BIC ($p \ln T$), but AIC ($2p$) when $\max_j \sqrt{T}|\tilde{\rho}_j|$ exceeds $\sqrt{2.4 \ln T}$
- Under the MDH, $\tilde{p} \to 1$ with probability 1, hence the χ_1^2 limit.
  - a large AQ rejects the MDH.
- AQ: S&P 500 13.23 ($p = 0.0003$), BET 14.92 ($p = 0.0001$), Bitcoin 0.75 ($p = 0.385$).
  - the MDH is rejected at 5% for 5 of the 16 series.

Back

Appendix: Predictive Regression, Data and Full Output

FRED: Federal Reserve Economic Data.

- Monthly log excess return regressed on the lagged dp_t
  - dividends included, minus the 3-month T-bill rate
  - Data: Shiller (2026) (S&P Composite, to June 2026), extended with monthly averages of our daily S&P 500 closes (scaled over the last 12 common months) and the dividends of the SPDR (Standard & Poor's Depositary Receipts) S&P 500 ETF (SPY); 3-month US Treasury bill yield (FRED, series TB3MS; annual %), monthly log rate $\ln(1 + \text{TB3MS}/1200)$ lagged one month
- OLS: $\hat{\beta} = 0.267$, Newey–West SE 0.293 (12 lags), $t = 0.91$, OLS $t = 1.16$, $R^2 = 0.12\%$
- Corrected slope / unadjusted HAC SE $= -0.22$: only illustrative
  - it ignores the uncertainty of the correction
- Economic value Campbell & Thompson (2008): proportional rise in the expected return of the market timer
  - number: slide The Dividend–Price Ratio: Results for the S&P 500
  - $\dfrac{R^2}{1-R^2} \cdot \dfrac{1+S^2}{S^2}$, S the monthly Sharpe ratio
- The Diebold & Mariano (1995) test gives $t = -0.57$
  - but it favours the small model when the models are nested
  - Clark & West (2007) correct for this and give $t = 0.42$

Back

Appendix: The Speed of Price Adjustment

CME: Chicago Mercantile Exchange.

- At high frequency, futures prices react to scheduled macro news within microseconds or less
  - Tak, Schlamp, Osterrieder & Pele (2026) measure this race at Eurex and CME
- Crypto events: Pele, Lin, Wang & Gaman (2025) study market responses to Ethereum development milestones
- Speed of adjustment is itself a measure of efficiency
  - it depends on technology (Chapter 10, market microstructure)

Back

Appendix: Harold Hurst and the Nile



Harold Edwin Hurst (1880–1978)

Photo: Elliott & Fry (1953); public domain; Wikimedia Commons



Nilometer on Elephantine Island, Aswan

Photo: Olaf Tausch (2011); CC BY 3.0; Wikimedia Commons

- Hurst (1951) designed Nile reservoirs
- He found that the range of cumulative Nile flows grows faster than $\sqrt{n}$ with the horizon
  - years of high floods follow other years of high floods, hence the **rescaled range** statistic (R/S)

Back

Appendix: The Anis–Lloyd Expectation

- Anis & Lloyd (1976): exact $\mathbb{E}[R_n/s_n]$ under i.i.d. returns from the Normal distribution
  - $\mathbb{E}[R_n/s_n] = \dfrac{\Gamma((n-1)/2)}{\sqrt{\pi}\,\Gamma(n/2)} \sum_{i=1}^{n-1} \sqrt{\dfrac{n-i}{i}}$
  - Γ: the gamma function
- This $\mathbb{E}[R_n/s_n]$ is the reference subtracted in H_{AL} (R/S Analysis slide)
- Uncorrected estimates overstate memory by about 0.04–0.05 with daily data

Back

Appendix: Long Memory with Inference, the Local Whittle Estimator

- $d = H - 0.5$, with spectral density $f(\lambda) \sim G\lambda^{-2d}$ as $\lambda \to 0$
  - $G > 0$ a constant
  - $d > 0$: the density explodes at low frequencies, long memory
  - fit only the first m Fourier frequencies $\lambda_j = 2\pi j/T$ (m: bandwidth)
  - $\hat{d} = \arg\min_d \left[\ln\left(\frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2d} I(\lambda_j)\right) - \frac{2d}{m}\sum_{j=1}^{m} \ln \lambda_j \right]$, I: the periodogram (sample spectral density)
  - Robinson (1995): for $|d| < 1/2$, $m \to \infty$, $m/T \to 0$, a smooth f near 0 and a linear process with conditionally homoskedastic martingale-difference innovations, $\sqrt{m}(\hat{d} - d) \to N(0, 1/4)$ without Gaussianity: a confidence interval instead of a band
- Estimates with $m = T^{0.65}$ and the 95% CI $\hat{d} \pm 1.96/(2\sqrt{m})$
  - S&P 500 -0.013 $[-0.069, 0.043]$; Bitcoin 0.063 $[-0.002, 0.127]$
  - BET 0.118 $[0.062, 0.174]$; Ethereum 0.112 $[0.045, 0.179]$
- Bandwidth trade-off: a large m reaches frequencies where short-run autocorrelation biases $\hat{d}$ upward
  - BET $\hat{\rho}_1 = 0.109$
  - with $m = T^{0.5}$ BET gives $\hat{d} = 0.135$ (SE 0.056): a sensitivity check supporting, not proving, that its $d > 0$ is not only short memory
  - GARCH returns also strain the conditions above

Back

Appendix: EMH versus AMH

	EMH	AMH
Investors	Rational, optimising	Bounded rationality, heuristics
Efficiency	An equilibrium property, tested as a restriction	A matter of degree that changes with conditions
Risk–return	Equilibrium risk premia, possibly time-varying	Changes with the population of investors
Anomalies	Omitted risk factors or data mining	Transient profit opportunities
Empirical tool	Tests of an equilibrium restriction	Tests of time and state variation (rolling windows)

Back

Appendix: Is Bitcoin Efficient? The Literature

- Urquhart (2016): Bitcoin inefficient over 2010–2016
  - some tests indicate efficiency in the later subperiod
  - His battery (Ljung–Box, runs, the rank test of Bartels (1982), automatic VR, BDS, R/S) includes BDS, which rejects under volatility clustering alone: not a test of RW3
- Nadarajah & Chu (2017): with a power transformation of returns, Bitcoin is weak-form efficient
- Kristoufek (2018): Bitcoin markets mostly inefficient in 2010–2017, with efficiency changing over time
- Tran & Leirvik (2020): time-varying efficiency across many crypto-assets, increasing over time
- Pele et al. (2023): crypto-assets have become a distinct asset class
  - their returns share the stylised facts of other markets
- Our own estimates, calendar year by calendar year
  - slide Crypto Efficiency Year by Year
  - Bitcoin, lag-1 autocorrelation $\hat{\rho}_1$: -0.002 in 2011, -0.071 in 2012, 0.069 in 2013, all close to zero
  - Robust variance-ratio statistic $z^*(5)$: at most 1.71 in absolute value over all crypto-asset years, below the critical value 1.96
  - In 2011–2013 the robust standard error of $\hat{\rho}_1$ is 0.09–0.11: the yearly non-rejection has little power and does not by itself contradict Urquhart (2016)

Back

Appendix: Romania, Frontier then Emerging

BVB: Bursa de Valori București — the Bucharest Stock Exchange.

- FTSE Russell (2019): Romania reclassified from Frontier to Secondary Emerging
  - announced in September 2019
  - effective 21 September 2020
- Expected channels: more foreign investors, more analysts, more liquidity ⇒ faster price discovery
- Earlier BVB evidence:
  - crashes and log-periodic bubbles Pele & Mazurencu-Marinescu (2012)
  - herding in an illiquid market Pele, Mazurencu-Marinescu & Nijkamp (2013)
- Did efficiency improve? Seminar C2 compares five years before and after
  - with WIG20 and BUX as a comparison group
  - with a block bootstrap Künsch (1989)
  - A before/after comparison is descriptive unless confounders are controlled

Back

Appendix: Reality Check, SPA and StepM explained

SPA: Superior Predictive Ability (test).

- Reality Check asks whether the best of all tried rules beats a benchmark. White (2000)
  - using a bootstrap distribution of their maximum performance.
- SPA standardises performance by its uncertainty (studentisation) and reduces distortion from poor rules. Hansen (2005)
- StepM tests the strongest statistic first, then the remaining family using jointly resampled critical values
  - it controls FWER under its assumptions. Romano & Wolf (2005)
- Example: the same months are drawn for all 15 calendar tests, so common market shocks are kept
  - separate resampling would destroy that dependence.

Back

Appendix: StepM on the 15 Calendar Tests

- Bonferroni and Holm ignore the correlation between tests (same markets, same days)
  - Reality Check, SPA and StepM allow for it (explained on the previous slide)
- Our 15 calendar tests: StepM with a bootstrap of blocks of 3 calendar months
  - common to all markets
  - 999 replications
  - Critical p at the first step 0.0016 (Bonferroni: 0.0033); 2 rejections, BET and BUX TOM, as with Holm
  - Why stricter than Bonferroni? At nominal 5%, the bootstrap rejection rates of the HAC tests are 6.0% (day of week), 6.4% (TOM), 6.4% (January)
- Chordia, Goyal & Saretto (2020): thresholds $|t| > 3.8$ (time series) and $|t| > 3.4$ (cross-section)
  - from over two million strategies generated at random from real accounting data
  - without them, about 45% of rejections are false
- Lowering false discoveries raises missed ones Harvey & Liu (2020)

Back

Appendix: Time-Series Momentum, All Markets

Market	Mean (%/month)	t_{HAC}	Sharpe TSMOM	Sharpe buy-and-hold	$\hat{\rho}_1$ monthly
S&P 500 (US)	0.61	2.22	0.48	0.52	0.03
Euro Stoxx 50	0.16	0.51	0.11	0.16	0.05
Nikkei 225 (Japan)	0.61	1.85	0.39	0.42	0.09
Hang Seng (HK)	0.18	0.53	0.10	0.19	0.04
BET (Romania)	1.40	2.68	0.65	0.75	0.19
WIG20 (Poland)	0.67	1.86	0.36	0.26	0.02
BUX (Hungary)	0.72	1.79	0.40	0.64	0.10
PX (Czechia)	0.38	0.99	0.24	0.45	0.10
BIST 100 (Turkey)	1.03	1.85	0.37	0.75	−0.08
Bovespa (Brazil)	0.28	0.65	0.14	0.51	0.11
IPC (Mexico)	0.44	1.41	0.32	0.63	0.03
Nifty 50 (India)	0.50	1.29	0.28	0.64	0.02
Shanghai (China)	0.59	1.32	0.30	0.23	0.08
Bitcoin	4.47	1.94	0.72	1.08	0.13
Ethereum	6.50	1.28	0.57	0.90	0.16

- 307 months for equities, fewer for crypto
  - Sharpe ratio $\sqrt{12}\,\bar{x}/s_x$ of price returns, no risk-free rate subtracted
  - $\bar{x}$, s_x: mean and SD of the monthly strategy returns
  - Moskowitz, Ooi & Pedersen (2012) use futures excess returns scaled to 40% ex-ante volatility

Back

Appendix: Robert Shiller and Richard Thaler



Robert J. Shiller, Nobel Prize 2013

Photo: Bengt Nyman (2013); CC BY 2.0; Wikimedia Commons



Richard H. Thaler, Nobel Prize 2017

Photo: Bengt Nyman (2017); CC BY 2.0; Wikimedia Commons

- Excess volatility Shiller (1981) and overreaction De Bondt & Thaler (1985): Psychology in Prices slide

Back

Appendix: Case Study, Research Question

- Martin & Nagel (2022): what does a rejected no-predictability test mean?
  - when the number J of characteristics is comparable to the number N of assets
  - the coefficients must be learned from past data: a high-dimensional prediction problem, as in machine learning
- Semi-strong efficiency: prices "fully reflect" public information Fama (1970), with two readings
  - investors know the parameters of the cash-flow model (rational expectations: the null of most efficiency tests)
  - investors update their beliefs optimally (Bayes) with the data available in real time
- Why it matters: hundreds of in-sample return predictors, the "factor zoo" Harvey, Liu & Zhu (2016)
  - text, satellite images and social media keep increasing J
  - does every significant predictor need a risk-based or a behavioural explanation?
- Link to this chapter: out-of-sample predictability Campbell & Thompson (2008) and the decay of anomalies after publication McLean & Pontiff (2016)

Back

Appendix: Case Study, Model and Method (1/2)

- **Model** (Martin & Nagel (2022), Sec. 2)
  - cash-flow growth $\Delta y_t = Xg + e_t$, with $e_t \sim N(0, I)$ and a zero interest rate
  - Δy_t: the N cash-flow growth rates; X: the $N \times J$ matrix of characteristics; g: the J unknown coefficients; I: identity matrix
- **Learning** (Prop. 1–2)
  - prior $g \sim N(0, (\theta/J)\, I)$, the true distribution from which g is drawn; θ: signal strength
  - schematically, one period: the price is $X\hat{g}_t$, with $\hat{g}_t = (tX'X + (J/\theta)I)^{-1} X' \sum_{s \le t} \Delta y_s$, so $\mathbb{E}[\hat{g}_t \mid g] = Ag$, with A "smaller" than I
  - $r_{t+1} = X(I - A)g$ (underreaction from shrinkage) $-X(\hat{g}_t - Ag)$ (estimation noise) $+e_{t+1}$ (unpredictable shock)
- **Asymptotics and simulations** (Sec. 4)
  - $N, J \to \infty$ with $J/N \to \psi > 0$
  - simulations with $N = 1000$, entries of X drawn from the standard Normal distribution and $\theta = 1$

Back

Appendix: Case Study, Model and Method (2/2)

- **CRSP**: Center for Research in Security Prices.
- **NYSE**: New York Stock Exchange.
- **Tests**
  - in sample: a cross-sectional OLS of r_{t+1} on X (requires $J < N$) with a joint χ^2 test
  - out of sample: a portfolio with weights $\frac{1}{N} X h_{s+1}$, from coefficients estimated in other periods
  - $r_{\text{OOS},t+1} = \frac{1}{N} h'_{s+1} X' r_{t+1}$, $s \neq t$ (eq. 19); r_{IS}: the same with $s = t$
  - h_{s+1}: the estimated coefficient vector; r_{t+1}: the N returns
- **Empirical design** (Sec. 7)
  - US stocks in CRSP, without those below the 20th NYSE size percentile or priced below \$1
  - predictors: simple and squared monthly returns at lags $t-2$ to $t-120$ (238 predictors), demeaned and standardised across stocks every month
  - full-sample ridge, 1971 – June 2019, penalty chosen by leave-one-year-out cross-validation
  - rolling 20-year OLS gives r_{OOS} from January 1956

Back

Appendix: Case Study, Formal Results (1/2)

- In sample: the probability of no rejection is of the order $e^{-0.015N}$ (Fig. 2)
  - Prop. 4 of Martin & Nagel (2022), main slide
  - for $\psi > 0.4$ ($\theta = 1$, one period of learning)
- Simulations: a test of nominal size 5% rejects with probability above 90% once $J > N/2$ (Fig. 3b)
  - $N = 1000$, four periods of learning
  - not a size distortion: the in-sample predictable component really is in the prices
- Single-predictor tests with k-FWER control (Fig. 4)
  - probability of k or more false rejections at most 5%
  - for $k = 20$ and $J/N = 0.05$, the probability of at least k rejections is 25% instead of 5%
  - about 80% for J/N close to 1
  - multiple-testing corrections do not help: the problem is the null hypothesis, not the test

Back

Appendix: Case Study, Formal Results (2/2)

- Out of sample: $\mathbb{E}[r_{\text{OOS},t+1}] = 0$ holds forward and backward in time (Prop. 6)
  - positive only if investors shrink more than the data warrant (Fig. 6)
- US stocks: the ridge coefficients recover three patterns (Fig. 7)
  - momentum
  - long-run reversal
  - momentum seasonality
  - $r_{\text{IS}} \approx 0.006\%$ per month (Fig. 8)

Appendix: Case Study, In Sample vs Out of Sample in Detail

- $r_{\text{IS}} \geq 0$ in every 20-year window by construction: $r_{\text{IS}} = b' G^{-1} b / 240$
  - $G = \sum_s X_s' X_s / N_s$ and $b = \sum_s X_s' r_s / N_s$ over the 240 months s of the window
  - X_s: the $N_s \times 238$ predictor matrix of month s; r_s: the N_s portfolio returns; $G^{-1} b$: the pooled OLS coefficients
  - mean in-sample R^2 of 5.4%, against about 1% for individual stocks in Martin & Nagel (2022): 238 predictors against only 40–49 portfolios a month, pooled over 240 months, inflate the in-sample fit
- Out of sample (from July 1956):
  - until June 1986: mean $r_{\text{OOS}} = 1.98 \times 10^{-5}$, 4.16 standard errors above 0
  - afterwards 0.36×10^{-5}, only 0.62 standard errors
  - the 10-year average is last above two standard errors in the window ending in December 1987; in the windows ending in 2014–2019 its largest ratio to the standard error is 1.25, as for stocks
- Interpretation suggested by Martin & Nagel (2022): the gap is consistent with investor learning
  - but does not establish it or rule out risk premia
  - early out-of-sample gains are also consistent with excess shrinkage before the computer era (Fig. 6)

Back

Appendix: Case Study, Replication, Limitations and Extensions

- **Replicate with the course data** (stock-level CRSP files are not part of them)
  - the 49 value-weighted industry portfolios of the main replication slides; same predictors, standardisation and windows
  - the size and price filters do not apply to portfolios
- **Limits stated by the authors**
  - lagged returns are equilibrium outcomes, not the exogenous X of the model: the application illustrates the wedge, it does not test the model
  - a t-test on the monthly r_{OOS} is not a valid test of $\mathbb{E}[r_{OOS}] = 0$: an expectation that also averages over the prior distribution of g
  - the backward result may rely on i.i.d. cash-flow growth, homogeneous investors and constant characteristics
- **Open extensions**
  - time series: the dividend–price regression of section 3 has a single predictor ($J = 1$), yet an in-sample R^2 of 0.12% and $R^2_{OS} = -0.51\%$: what explains the gap, if not the number of predictors?
  - nonlinear models (trees, neural networks Gu, Kelly & Xiu (2020)) add functional flexibility without necessarily changing J: does the wedge grow?
  - AMH: is early out-of-sample predictability the excess shrinkage of investors with less computing power?

Back

Appendix: The Discovery Loop, Reusable Prompts (1/2)

- **Literature**
  - `List peer-reviewed studies since 2020 on how ETF introductions change the efficiency or price discovery of the underlying. For each: DOI, design (event study, difference-in-differences (DiD), structural break), efficiency measure. Flag any entry you are unsure exists.`
  - Resolve every DOI (Digital Object Identifier) on Crossref; the title and authors must match
- **Hypothesis**
  - `Give three mechanisms by which spot ETFs could change Bitcoin's serial dependence (arbitrage, investor mix, trading hours). For each, the sign of the change in rho1 and VR(q), and a variable that separates it from the others.`
- **Data and code**
  - `Read the daily Bitcoin closing prices from the course data; log returns; rho1 with a HAC (Newey-West) standard error of the estimate and the Lo-MacKinlay VR(5) with z*, in two windows of equal length around 11 January 2024.`

Back

Appendix: The Discovery Loop, Reusable Prompts (2/2)

- **Robustness**
  - `Write a pre-registered robustness plan: window lengths, weekday-only closes, VR(q) for q = 2 to 20 with Chow-Denning, placebo dates, and Ethereum as a control until its own spot ETFs (July 2024).`
- **Critique**
  - `Act as a hostile referee: if Bitcoin was already efficient before 2024, what could this test ever detect? Compute the minimum detectable change in rho1 and name a better outcome variable.`
- **Report**
  - Estimates, robust standard errors, the placebo distribution, every test run (with a multiple-testing correction)
  - Prompts that shaped a decision go into `AI_USE.md`

Back

Appendix: What the Researcher Must Check

- **i.i.d. standard errors**: $1/\sqrt{T}$ rejects too often under volatility clustering
  - test: robust SE under H_0 and z^*, as in the weak-form tests of this chapter; HAC SE for intervals and before/after comparisons
- **Unequal windows**: different power before and after
  - test: equal lengths, fixed in advance
- **Event date chosen from the data**
  - test: the launch date (first trading day) only, plus a placebo distribution at dates without an event
- **Confounding**: before/after is not causal
  - test: a control asset (Ethereum before July 2024) in a difference-in-differences
  - causal only under the conditions of the difference-in-differences slide; the common post-period (Jan–Jul 2024) is short
- **Many statistics**: ρ_1, VR(2–20), runs
  - test: Chow–Denning or Holm, not the smallest single p-value
- **Low power**: report the MDE of the Estimation and Power slide

Back

Appendix: Further Study on Quantinar

Advanced courses on quantinar.com

- Efficiency of cryptocurrency markets - A GMM-based analysis
- Generalized Method of Moments
- Market Responses to ETH Development Milestones

Code, data and textbooks

- All charts and tables: github.com/danpele/MFM/tree/main/Quantlets/Ch_02
- Textbooks: Campbell, Lo & MacKinlay (1997), Ch. 2; Franke, Härdle & Hafner (2019)

Back

Acronyms used in this material (1/3)

- **AI**: Artificial Intelligence
- **AIC**: Akaike Information Criterion
- **AMH**: Adaptive Markets Hypothesis
- **AR(p)**: AutoRegressive model of order p
- **AR**: Abnormal Return, in event studies.
- **BDS**: Brock–Dechert–Scheinkman (test)
- **BET**: Bucharest Exchange Trading (index)
- **BIC**: Bayesian Information Criterion
- **BIST**: Borsa İstanbul — *the Istanbul Stock Exchange (BIST 100 index)*
- **BNR**: Banca Națională a României — *the National Bank of Romania*
- **BUX**: Budapesti Értéktőzsde index — *the Budapest Stock Exchange index*
- **BVB**: Bursa de Valori București — *the Bucharest Stock Exchange*
- **CAPM**: Capital Asset Pricing Model
- **CAR**: Cumulative Abnormal Return
- **CD**: Chow–Denning (multiple variance-ratio test)
- **CI**: Confidence Interval
- **CME**: Chicago Mercantile Exchange
- **CRSP**: Center for Research in Security Prices
- **DFA**: Detrended Fluctuation Analysis
- **DiD**: Difference-in-Differences

Acronyms used in this material (2/3)

- **EMH**: Efficient Market Hypothesis
- **EOD**: End of Day
- **EODHD**: EOD Historical Data (market data provider)
- **ETF**: Exchange-Traded Fund
- **fGn**: fractional Gaussian noise
- **FRED**: Federal Reserve Economic Data
- **FTSE**: Financial Times Stock Exchange (FTSE Russell, index provider)
- **FWER**: Family-Wise Error Rate
- **FX**: Foreign eXchange
- **GARCH**: Generalized AutoRegressive Conditional Heteroskedasticity
- **GMM**: Generalized Method of Moments
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors)
- **HK**: Hong Kong
- **IPC**: Índice de Precios y Cotizaciones — *the Mexican Stock Exchange index*
- **MDE**: Minimum Detectable Effect
- **MDH**: Martingale Difference Hypothesis
- **MIT**: Massachusetts Institute of Technology
- **ML**: Machine Learning
- **MSPE**: Mean Squared Prediction Error
- **NBER**: National Bureau of Economic Research

Acronyms used in this material (3/3)

- **NW**: Newey–West (standard errors)
- **NYSE**: New York Stock Exchange
- **OLS**: Ordinary Least Squares
- **OOS**: Out-Of-Sample
- **OS**: Out-of-Sample (subscript of R^2_{OS})
- **PX**: Prague Stock Exchange index (index Burzy cenných papírů Praha)
- **RW1**: Random Walk 1 (i.i.d. increments)
- **RW2**: Random Walk 2 (independent increments)
- **RW3**: Random Walk 3 (uncorrelated increments)
- **SD**: Standard Deviation
- **SE**: Standard Error
- **SPA**: Superior Predictive Ability (test)
- **SPDR**: Standard & Poor's Depositary Receipts
- **TOM**: Turn Of the Month
- **TSMOM**: Time-Series MOMentum
- **UK**: United Kingdom
- **US**: United States
- **UTC**: Coordinated Universal Time (Temps Universel Coordonné)
- **VR**: Variance Ratio
- **WIG20**: Warszawski Indeks Giełdowy 20 — *the Warsaw Stock Exchange index of the 20 largest companies*

References (1/7)

- Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3), 817–858.
- Anis, A.A., Lloyd, E.H. (1976). The expected value of the adjusted rescaled Hurst range of independent normal summands. *Biometrika*, 63(1), 111–116.
- Ariel, R.A. (1987). A monthly effect in stock returns. *Journal of Financial Economics*, 18(1), 161–174.
- Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure*, 17, 21–86.
- Bartels, R. (1982). The rank version of von Neumann's ratio test for randomness. *Journal of the American Statistical Association*, 77(377), 40–46.
- Blanchard, O.J., Watson, M.W. (1982). Bubbles, rational expectations and financial markets. In P. Wachtel (ed.), *Crises in the Economic and International Financial System*, Lexington Books, 295–315.
- Boehmer, E., Musumeci, J., Poulsen, A.B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, 30(2), 253–272.
- Brock, W.A., Dechert, W.D., Scheinkman, J.A., LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3), 197–235.
- Campbell, J.Y., Lo, A.W., MacKinlay, A.C. (1997). The Econometrics of Financial Markets. Princeton University Press.
- Campbell, J.Y., Shiller, R.J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3), 195–228.
- Campbell, J.Y., Thompson, S.B. (2008). Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies*, 21(4), 1509–1531.
- Campbell, J.Y., Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1), 27–60.
- Choi, I. (1999). Testing the random walk hypothesis for real exchange rates. *Journal of Applied Econometrics*, 14(3), 293–308.
- Chordia, T., Goyal, A., Saretto, A. (2020). Anomalies and false rejections. *Review of Financial Studies*, 33(5), 2134–2179.

References (2/7)

- Chow, K.V., Denning, K.C. (1993). A simple multiple variance ratio test. *Journal of Econometrics*, 58(3), 385–401.
- Clark, T.E., West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291–311.
- Cochrane, J.H. (2008). The dog that did not bark: a defense of return predictability. *Review of Financial Studies*, 21(4), 1533–1575.
- De Bondt, W.F.M., Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, 40(3), 793–805.
- De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4), 703–738.
- Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Escanciano, J.C., Lobato, I.N. (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics*, 151(2), 140–149.
- Escanciano, J.C., Velasco, C. (2006). Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics*, 134(1), 151–185.
- Fama, E.F. (1970). Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, 25(2), 383–417.
- Fama, E.F. (1991). Efficient capital markets: II. *Journal of Finance*, 46(5), 1575–1617.
- Fama, E.F., Fisher, L., Jensen, M.C., Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, 10(1), 1–21.
- Fama, E.F., French, K.R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1), 3–25.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). Statistics of Financial Markets: An Introduction, 5th ed. Springer.
- French, K.R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics*, 8(1), 55–69.
- French, K.R. (2026). Kenneth R. French Data Library. Tuck School of Business, Dartmouth College.

References (3/7)

- FTSE Russell (2019). FTSE Russell announces results of country classification review (press release, 26 September 2019).
- Goyal, A., Welch, I., Zafirov, A. (2024). A comprehensive 2022 look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 37(11), 3490–3557.
- Grossman, S.J., Stiglitz, J.E. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, 70(3), 393–408.
- Gu, S., Kelly, B., Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273.
- Hansen, P.R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4), 365–380.
- Harvey, C.R., Liu, Y. (2020). False (and missed) discoveries in financial economics. *Journal of Finance*, 75(5), 2503–2553.
- Harvey, C.R., Liu, Y., Zhu, H. (2016). . . . and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5–68.
- Heston, S.L., Sadka, R. (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics*, 87(2), 418–445.
- Hurst, H.E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116(1), 770–799.
- Jegadeesh, N., Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, 48(1), 65–91.
- Kahneman, D., Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2), 263–291.
- Kendall, M.G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika*, 41(3–4), 403–404.
- Kia, K., Liu, B., Li, Q., Song, V., Xu, K. (2026). Price discovery in Bitcoin ETF market. *Financial Review*, 61(2), 435–449.

References (4/7)

- Kim, J.H. (2006). Wild bootstrapping variance ratio tests. *Economics Letters*, 92(1), 38–43.
- Kim, J.H., Shamsuddin, A. (2008). Are Asian stock markets efficient? Evidence from new multiple variance ratio tests. *Journal of Empirical Finance*, 15(3), 518–532.
- Kim, J.H., Shamsuddin, A., Lim, K.-P. (2011). Stock return predictability and the adaptive markets hypothesis: evidence from century-long U.S. data. *Journal of Empirical Finance*, 18(5), 868–879.
- Kolari, J.W., Pynnönen, S. (2010). Event study testing with cross-sectional correlation of abnormal returns. *Review of Financial Studies*, 23(11), 3996–4025.
- Kristoufek, L. (2018). On Bitcoin markets (in)efficiency and its evolution. *Physica A*, 503, 257–262.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3), 1217–1241.
- Lakonishok, J., Smidt, S. (1988). Are seasonal anomalies real? A ninety-year perspective. *Review of Financial Studies*, 1(4), 403–425.
- Lim, K.-P., Brooks, R. (2011). The evolution of stock market efficiency over time: a survey of the empirical literature. *Journal of Economic Surveys*, 25(1), 69–108.
- Liu, Y., Tsyvinski, A. (2021). Risks and returns of cryptocurrency. *Review of Financial Studies*, 34(6), 2689–2727; NBER Working Paper 24877.
- Ljung, G.M., Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.
- Lo, A.W. (1991). Long-term memory in stock market prices. *Econometrica*, 59(5), 1279–1313.
- Lo, A.W. (2004). The adaptive markets hypothesis. *Journal of Portfolio Management*, 30(5), 15–29.
- Lo, A.W. (2017). Adaptive Markets: Financial Evolution at the Speed of Thought. Princeton University Press.
- Lo, A.W., MacKinlay, A.C. (1988). Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies*, 1(1), 41–66.

References (5/7)

- Lobato, I., Nankervis, J.C., Savin, N.E. (2001). Testing for autocorrelation using a modified Box–Pierce Q test. *International Economic Review*, 42(1), 187–205.
- MacKinlay, A.C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35(1), 13–39.
- Malkiel, B.G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1), 59–82.
- Martin, I.W.R., Nagel, S. (2022). Market efficiency in the age of big data. *Journal of Financial Economics*, 145(1), 154–177.
- McLean, R.D., Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, 71(1), 5–32.
- Moskowitz, T.J., Ooi, Y.H., Pedersen, L.H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2), 228–250.
- Nadarajah, S., Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters*, 150, 6–9.
- Newey, W.K., West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Pele, D.T., Lin, M.-B., Wang, R., Gaman, S. (2025). Market responses to ETH development milestones. IDA Institute Digital Assets, Bucharest University of Economic Studies.
- Pele, D.T., Mazurencu-Marinescu, M. (2012). Modelling stock market crashes: the case of Bucharest Stock Exchange. *Procedia – Social and Behavioral Sciences*, 58, 533–542.
- Pele, D.T., Mazurencu-Marinescu, M., Nijkamp, P. (2013). Herding behaviour, bubbles and log periodic power laws in illiquid stock markets: a case study on the Bucharest Stock Exchange. SSRN Working Paper.
- 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.
- Peng, C.-K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E., Goldberger, A.L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2), 1685–1689.

References (6/7)

- Roberts, H.V. (1959). Stock-market "patterns" and financial analysis: methodological suggestions. *Journal of Finance*, 14(1), 1–10.
- Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23(5), 1630–1661.
- Romano, J.P., Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4), 1237–1282.
- Rozeff, M.S., Kinney, W.R. (1976). Capital market seasonality: the case of stock returns. *Journal of Financial Economics*, 3(4), 379–402.
- Samuelson, P.A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6(2), 41–49; reprinted 2015, World Scientific.
- Shiller, R.J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71(3), 421–436.
- Shiller, R.J. (2026). U.S. stock markets 1871–present and CAPE ratio (monthly data, accessed September 2026).
- Shleifer, A., Vishny, R.W. (1997). The limits of arbitrage. *Journal of Finance*, 52(1), 35–55.
- Stambaugh, R.F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3), 375–421.
- Sullivan, R., Timmermann, A., White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, 54(5), 1647–1691.
- Summers, L.H. (1986). Does the stock market rationally reflect fundamental values? *Journal of Finance*, 41(3), 591–601.
- Tak, R., Pele, D.T. (2025). LLM-driven stock prediction: capturing market trends with LLaMA. *Proceedings of the International Conference on Business Excellence*, 19(1), 529–540.
- Tak, R., Schlamp, S., Osterrieder, J., Pele, D.T. (2026). Racing the news: nanosecond market efficiency at Eurex and CME. SSRN Working Paper.

References (7/7)

- Tirole, J. (1985). Asset bubbles and overlapping generations. *Econometrica*, 53(6), 1499–1528.
- Tran, V.L., Leirvik, T. (2020). Efficiency in the markets of crypto-currencies. *Finance Research Letters*, 35, 101382.
- Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*, 148, 80–82.
- Urquhart, A., McGroarty, F. (2016). Are stock markets really efficient? Evidence of the adaptive market hypothesis. *International Review of Financial Analysis*, 47, 39–49.
- Wald, A., Wolfowitz, J. (1940). On a test whether two samples are from the same population. *Annals of Mathematical Statistics*, 11(2), 147–162.
- Welch, I., Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455–1508.
- White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097–1126.