Lecture 5 – Deterministic Trends, Nonparametric Trend Fitting, EMH, and Random-Walk Diagnostics

Jiajing Sun

School of Economics and Management
University of Chinese Academy of Sciences

Econometrics and Time Series Methods
Spring 2026

Textbook sequence for this lecture

This lecture follows the Chapter 8 logic of the textbook as closely as possible, while matching the course plan for Lecture 5.

1. **Deterministic trend models**
    - polynomial trends, trend-stationarity, and asymptotic scaling,
    - why unremoved deterministic trend can create spurious persistence.
2. **Nonparametric trend fitting**
    - moving averages, kernel smoothing, spline / HP filtering,
    - bandwidth, boundary bias, and generated-regressor issues.
3. **Efficient Market Hypothesis and random-walk testing**
    - weak-form EMH, random walk with drift, and returns as innovations,
    - autocorrelation, AR, and variance-ratio tests.
4. **R-integrated practice**
    - detrending, ADF / KPSS checks, and random-walk diagnostics.

Lecture goals

- Distinguish clearly between a **deterministic trend** and a **stochastic trend**.
- Understand when detrending is appropriate, and when differencing or returns are the right transformation instead.
- Learn the textbook's main nonparametric trend tools: moving averages, kernel smoothing, and spline / HP filtering.
- Connect the unit-root ideas from Lecture 4 to the weak-form EMH.
- Leave with a practical workflow for empirical finance: levels, trends, returns, unit-root tests, and predictability diagnostics.

Roadmap for the three-hour block

Block 1

Deterministic trend models and nonparametric trend fitting.

Block 2

Efficient Market Hypothesis, random-walk testing, and empirical interpretation.

Block 3

R block: detrending, unit-root testing, and random-walk diagnostics.

Lecture 5 map

Why Chapter 8 starts with deterministic trends

The textbook begins Chapter 8 by stressing that nonstationarity can arise for more than one reason.

- Some series have a **time-varying mean** driven by a smooth deterministic function of calendar time.
- Other series have a **stochastic trend**, typically induced by a unit root, so shocks accumulate permanently.
- If we do not separate these cases, we can misread autocorrelations, choose the wrong transformation, and misinterpret shocks.

Textbook message

Before testing for a stochastic trend, ask whether the apparent persistence could simply be the result of an unmodelled deterministic trend.

Deterministic trend versus stochastic trend

Trend-stationary model

$$Y_t = Q_p(t;\beta) + u_t, \qquad u_t \sim I(0).$$

- Mean changes with time in a deterministic way.
- Shocks are transitory.
- Stationarity is restored by **detrending**.

Difference-stationary model

$$Y_t = Y_{t-1} + u_t, \qquad \Delta Y_t = u_t.$$

- The level contains a stochastic trend.
- Shocks are permanent.
- Stationarity is restored by **differencing**.

Practical distinction

Trend-stationary series revert to a deterministic path; difference-stationary series do not.

Polynomial trend model

The textbook's benchmark deterministic-trend specification is

$$y_t = \beta_0 + \beta_1 t + \cdots + \beta_p t^p + u_t = Q_p(t;\beta) + u_t,$$

where $p \geq 0$ and the disturbance u_t is stationary, for example

$$A(L)u_t = B(L)\varepsilon_t.$$

- $Q_p(t;\beta)$ is a **global** or **strong** trend.
- u_t captures short-run stationary deviations around that path.
- Linear trend ($p = 1$) is the workhorse case; quadratic trend is often used when more curvature is needed.

Economic reading

A deterministic trend is suitable when the long-run path is smooth and shocks eventually die out.

Mean, variance, and the role of transitory shocks

Under the additive trend model,

$$\mathrm{E}(y_t) = \beta_0 + \beta_1 t + \cdots + \beta_p t^p, \qquad \mathrm{Var}(y_t) = \sigma_u^2.$$

So nonstationarity enters only through the mean. A simple extension also allows the scale to trend:

$$y_t = \alpha + \beta t + \sqrt{\omega + \gamma t}\, \varepsilon_t,$$

which implies

$$\mathrm{E}(y_t) = \alpha + \beta t, \qquad \mathrm{Var}(y_t) = (\omega + \gamma t)\sigma_\varepsilon^2.$$

- In the pure additive model, shocks are **temporary** deviations from a deterministic path.
- If variance also trends, inference becomes harder and convergence rates change.

Spurious persistence from an unremoved deterministic trend

For the additive model $y_t = Q_p(t;\beta) + u_t$ with stationary u_t,

$$\mathrm{Cov}(y_t, y_{t-s}) = \mathrm{Cov}(u_t, u_{t-s}).$$

If u_t is i.i.d., then the true autocovariance is zero for all $s > 0$ once the time-varying mean is removed. But in finite samples we often compute autocovariances around the **global sample mean**. For the simple case $y_t = t + u_t$,

$$\bar{y} = \frac{T+1}{2}, \qquad \widehat{\gamma}(k) \to \int_0^1 (u-0.5)^2 \, du > 0.$$

Key warning

A deterministic trend can make the sample ACF look very persistent even when the underlying disturbance is uncorrelated.

OLS estimation of a polynomial trend

Let X be the $T \times (p+1)$ design matrix

$$X = \begin{pmatrix} 1 & 1 & 1^2 & \cdots & 1^p \\ 1 & 2 & 2^2 & \cdots & 2^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & T & T^2 & \cdots & T^p \end{pmatrix}.$$

Then

$$\widehat{\beta} = (X'X)^{-1}X'y, \qquad \widehat{T}_t = x_t'\widehat{\beta}, \qquad \widehat{u}_t = y_t - \widehat{T}_t.$$

- Trend estimation is still least squares.
- But the regressors grow with t, so the matrix $X'X/T$ does *not* converge to a nondegenerate limit.
- Standard regression asymptotics must therefore be rescaled.

Scaling and asymptotic distribution

The textbook introduces a diagonal scaling matrix

$$\Delta = \mathrm{diag}(T, T^3, \ldots, T^{p+1}),$$

and shows that

$$\Delta^{-1/2} X' X \Delta^{-1/2} \to M,$$

for a positive definite matrix M. Under i.i.d. disturbances,

$$\Delta^{1/2}(\widehat{\beta} - \beta) \Longrightarrow N(0, \sigma^2 M^{-1}).$$

- The intercept is estimated at the usual $T^{1/2}$ rate.
- The slope and curvature terms are estimated faster because the regressors themselves trend strongly.
- For any fixed t, the fitted trend error is still of order $O_p(T^{-1/2})$.

Wald testing for trend terms

Because of the nonstandard scaling, tests on trend coefficients are naturally written in Wald form. For $H_0 : R\beta = r$, the usual statistic is

$$W = \widehat{\sigma}^{-2}(R\widehat{\beta} - r)' \left[R(X'X)^{-1}R' \right]^{-1} (R\widehat{\beta} - r),$$

which has an asymptotic χ_q^2 distribution under standard regularity conditions.

- A very common null is **no trend beyond the intercept**: $\beta_1 = \cdots = \beta_p = 0$.
- In practice this means: first ask whether a deterministic trend is statistically needed before moving on to differencing or unit-root testing.
- With serially correlated errors, use HAC or model-based long-run variance corrections.

Serial correlation, HAC, and trending regressors

When u_t is stationary but serially correlated,

$$\operatorname{Var}(\widehat{\beta}) = (X'X)^{-1}(X'\Gamma_T X)(X'X)^{-1},$$

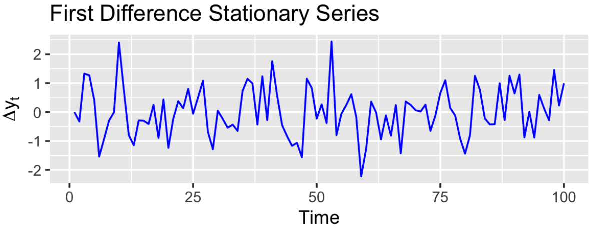
where Γ_T is the Toeplitz covariance matrix of the disturbance. The textbook emphasizes three points:

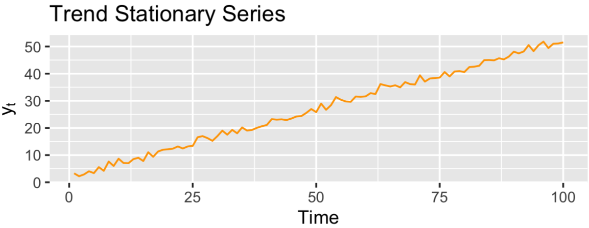
- OLS and GLS have the same asymptotic structure in this class of polynomial-trend models.
- Parametric ARMA corrections and nonparametric HAC estimators are both legitimate ways to estimate the long-run covariance.
- Weighted sums such as $\sum_{t=1}^{T} t^p u_t$ have variance of order T^{2p+1}, so trending regressors magnify long-run variance effects.

Empirical implication

Even when the trend is deterministic, serial correlation in the detrended disturbance still matters for inference.

Detrending is not the same as differencing


First Difference Stationary Series


Trend Stationary Series

Trend-stationary growth should be detrended; automatic differencing is the right response only when the nonstationarity is stochastic.

From global parametric trend to local smoothing

Polynomial trends are useful, but they can be too rigid.

- A global quadratic fit imposes one shape on the entire sample.
- Real economic or financial series often evolve in a way that is smoother locally than globally.
- This leads the textbook to **nonparametric trend fitting**: estimate a smooth function rather than commit to a fixed polynomial order.

Main idea

Replace “the mean is a line / parabola” with “the mean is a smooth function of time.”

Moving averages as linear filters

A very general smoother is the weighted moving average

$$\widehat{T}_t = \sum_{j=-n}^{n} w_j y_{t-j} = w(L) y_t, \qquad \sum_{j=-n}^{n} w_j = 1.$$

Examples include:

- two-sided equal weights,
- one-sided equal weights,
- exponentially weighted moving averages (EWMA).

Why practitioners like them

They smooth away high-frequency noise while keeping a transparent local interpretation. Seven-day COVID averages and Bollinger-band style moving averages are standard examples.

Nonparametric trend model

The textbook's flexible formulation is

$$y_t = g\left(\frac{t}{T^{\varkappa}}\right) + \sigma\left(\frac{t}{T^{\varkappa}}\right)\varepsilon_t, \qquad t = 1, \ldots, T,$$

where $g(\cdot)$ and $\sigma(\cdot)$ are smooth unknown functions and ε_t is stationary and weakly dependent.

- With $\varkappa = 1$, the trend lives on the compact interval $[0, 1]$ and is naturally interpreted as a **local** or **weak** trend.
- With $\varkappa < 1$, the effective domain expands and the asymptotics mix long-span and infill ideas.
- In this lecture we mostly follow the textbook's main case $\varkappa = 1$.

Kernel and Nadaraya–Watson trend estimation

A standard nonparametric estimator of $g(u)$ is the kernel smoother

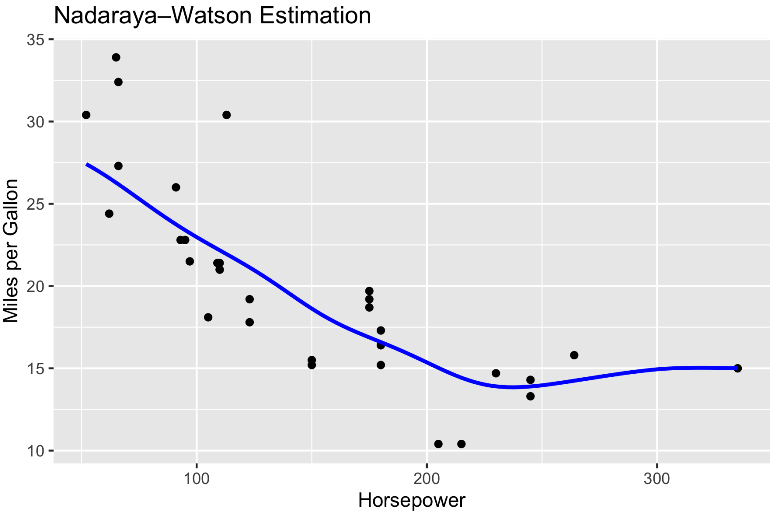
$$\widehat{g}(u) = \frac{\sum_{t=1}^{T} K\left(\frac{u-t/T}{h}\right) y_t}{\sum_{t=1}^{T} K\left(\frac{u-t/T}{h}\right)},$$

where h is a bandwidth and K is a kernel with

$$\int K(s)\,ds = 1, \qquad \int sK(s)\,ds = 0.$$

- Nearby observations receive more weight.
- The bandwidth controls the width of the local window.
- The estimator is a **linear smoother**: the fitted value is linear in the data.

A picture of smooth local fitting



Kernel smoothing does not force a single global line; it lets the fitted curve adapt to local curvature.

Asymptotic distribution of the kernel estimator

Under the textbook's smoothness and mixing assumptions, if $h \to 0$ and $Th^5 \to c > 0$, then

$$\sqrt{Th}\big(\widehat{g}(u) - g(u)\big) \Longrightarrow N\big(b(u), v(u)\big),$$

with

$$b(u) = c^{1/2} g''(u) \mu_2(K), \qquad v(u) = \sigma^2(u)\, \mathrm{lrvar}(\varepsilon)\, \|K\|_2^2.$$

- Bias depends on local curvature $g''(u)$.
- Variance depends on the long-run variance of the short-run disturbance.
- Dependence matters through $\mathrm{lrvar}(\varepsilon)$, not just $\mathrm{Var}(\varepsilon_t)$.

Bandwidth choice and the bias–variance trade-off

The optimal nonparametric rate is slower than the parametric regression rate.

$$\widehat{g}(u) \text{ converges at rate } T^{2/5}, \qquad \mathrm{MSE}(\widehat{g}(u)) = O(T^{-4/5}).$$

- Small h: low bias, high variance, wiggly fit.
- Large h: high bias, low variance, oversmoothed fit.
- The bandwidth is therefore the main tuning parameter in practice.

Textbook interpretation

The slower rate is the “price” of estimating an infinite-dimensional object $g(\cdot)$ instead of a finite vector of coefficients.

One-sided filters, EWMA, and boundary bias

Two-sided smoothers use future and past data. That is fine for ex-post trend extraction, but not for real-time forecasting.

- One-sided equal-weight filters and EWMA smoothers only use current and past observations.
- They are operationally natural for trading and monitoring.
- But the leading bias term is often worse at the boundary because the local window becomes asymmetric.

Textbook remedy

Either design one-sided weights with zero first moment, or use local linear methods, to reduce boundary bias back to order h^2.

Spline smoothing and the HP filter

The cubic smoothing spline solves

$$Q_\lambda(g) = \sum_{t=1}^{T} \big(y_t - g(t/T)\big)^2 + \lambda \int \{g''(u)\}^2 \, du.$$

In discrete form,

$$Q_\lambda(g) = (y-g)'(y-g) + \lambda g' D g, \qquad \widehat{g}_\lambda = (I + \lambda D)^{-1} y.$$

- As $\lambda \to 0$, the spline interpolates the data.
- As $\lambda \to \infty$, it approaches a least-squares line.
- The Hodrick–Prescott filter is the equally spaced-data macroeconomics version of this logic.

Generated regressors after detrending

Suppose we estimate a smooth trend and then work with residuals

$$\widehat{\varepsilon}_t = y_t - \widehat{g}(t/T).$$

The textbook emphasizes that these are **generated regressors**. For the empirical distribution function,

$$\sqrt{T}\big(\widehat{F}_T(e) - F(e)\big) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\big(1\{\varepsilon_t \le e\} - F(e)\big) + f(e)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t + o_p(1).$$

- The second term is the **price of detrending**.
- Ignoring the first-stage estimation step can distort downstream inference.

Robust trend fitting: median and Hampel filters

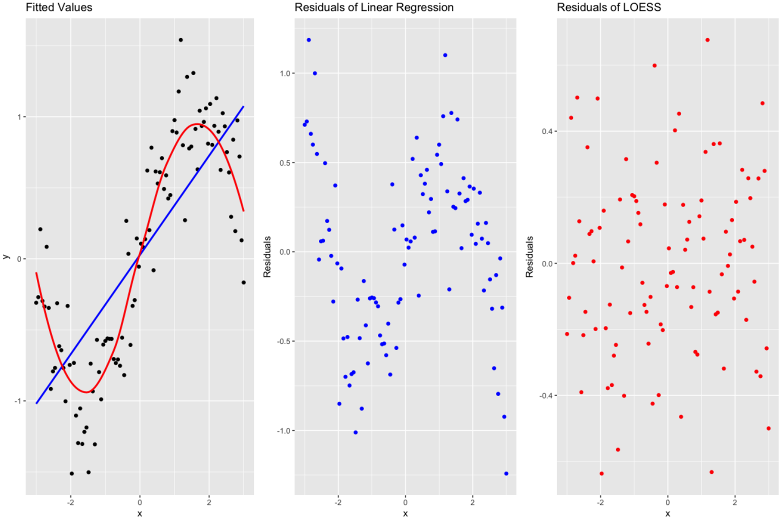
When outliers are important, mean smoothing can be fragile.

- A local median replaces the squared-loss criterion with absolute deviations.
- The Hampel filter compares each point with a local median and a local MAD scale estimate.
- Large outliers are pulled back toward the local center without smoothing away the entire local structure.

Use case

Robust filters are attractive when the series contains occasional extreme observations that are not representative of the underlying trend.

Linear fit versus flexible fit



Residual structure under a restrictive linear fit is evidence in favor of a flexible smoother.

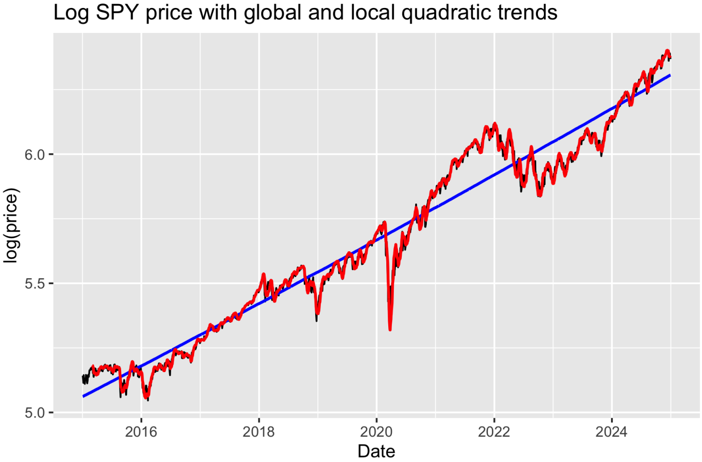
R illustration in the textbook: SPY global and local trends

The chapter's first applied block uses daily SPY prices from Yahoo Finance and compares:

1. a **global quadratic trend** fitted on the full sample,
2. a **rolling local quadratic trend** fitted on a moving window,
3. plots on both the **log-price** and **price** scales.

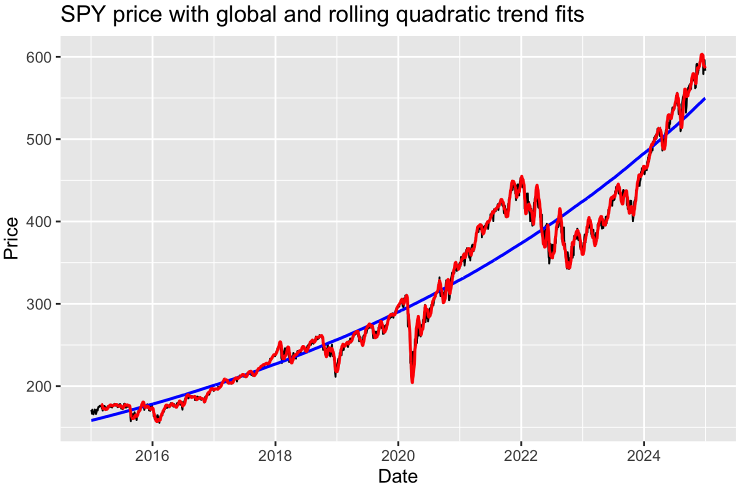
- This is a good example of the difference between a global deterministic approximation and a local adaptive fit.
- It also shows why the log scale is often the right place to think about long-run price growth.

SPY on the log scale: global versus local trend



On the log scale, the global trend is close to linear, while the rolling local fit tracks shorter-run departures from that path.

SPY on the price scale: global versus local trend



On the original price scale, the same difference is easier to see: the local fit reacts much more strongly to turning points and temporary swings.

What the SPY trend plots teach us

- A single global polynomial can summarize long-run growth, but it may miss sharp local turning points.
- A rolling local fit captures short-run movements better, but it is less stable and more sensitive to the window length.
- On the log scale, what looks nonlinear in levels often appears much closer to linear growth.

Takeaway

Trend fitting is not a purely mechanical preprocessing step. The chosen smoother changes the economic story you tell about the series.

Lecture 5 map

1. Deterministic trend models and nonparametric trend fitting
2. Efficient Market Hypothesis and random-walk testing
3. R block: detrending, unit-root testing, and random-walk diagnostics

Why EMH follows naturally from unit-root analysis

The textbook places the EMH after unit-root processes for a simple reason.

- In finance, prices are often modelled as random walks or random walks with drift.
- A random walk is exactly the canonical unit-root process.
- If prices are approximately $I(1)$, then returns or price changes are the natural stationary objects to analyze.

Bridge from Lecture 4

Lecture 4 asked whether a level series contains a unit root. Lecture 5 asks what that means economically in financial markets.

Samuelson, Fama, and the intuition of efficiency

The textbook summarizes the historical logic as follows.

- **Samuelson (1965):** properly anticipated prices should fluctuate randomly.
- **Fama (1970):** market prices fully reflect available information.
- The more rapidly information is incorporated into prices, the less predictable future price changes should be from currently available information.

Intuition

Random-looking price changes are not evidence of disorder; they are what we should expect when many traders rapidly exploit and thereby remove predictable profit opportunities.

Three forms of the Efficient Market Hypothesis

Form	**Information set**	**Empirical implication**
Weak	past prices and trading history	technical trading rules should not deliver persistent abnormal returns
Semi-strong	all public information	publicly available news should already be in prices
Strong	all public and private information	even insiders should not earn systematic excess returns

In this lecture we focus on the **weak form**, because it is the one most directly linked to time-series tests of random-walk behavior.

Weak-form EMH and the random-walk price model

The benchmark model is

$$P_t = P_{t-1} + \varepsilon_t,$$

or, with drift,

$$P_t = \mu + P_{t-1} + \varepsilon_t.$$

Iterating the second equation gives

$$P_t = P_0 + t\mu + \sum_{s=1}^{t} \varepsilon_s.$$

- The price level is nonstationary and behaves like an $I(1)$ process.
- The increment ΔP_t or log return is the economically relevant predictable / unpredictable object.
- Under weak-form EMH, past prices should not help forecast future price changes in a systematic way.

Prices, returns, and the martingale intuition

Under the random walk with drift,

$$\Delta P_t = \mu + \varepsilon_t.$$

In finance it is often more natural to work with log returns

$$r_t = \log P_t - \log P_{t-1}.$$

- A nonzero drift does *not* violate weak-form efficiency.
- What matters is whether the unpredictable component can be forecast from past information.
- This is why the textbook repeatedly moves from **price levels** to **returns** when discussing EMH.

Core distinction

Weak-form EMH is about the predictability of returns, not about whether price levels visibly trend upward over time.

Random-walk assumptions: rw1, rw2, and rw3

The textbook distinguishes three progressively weaker assumptions on the innovation ε_t.

- **rw1:** ε_t are i.i.d. with mean zero.
- **rw2:** ε_t are independent with mean zero, but not necessarily identically distributed.
- **rw3:** $\mathrm{E}(\varepsilon_t) = 0$ and $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$, but dependence is otherwise unrestricted.

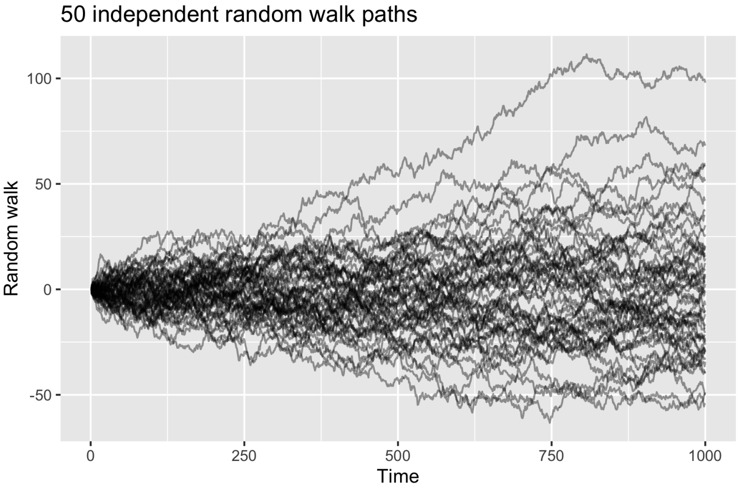
Hierarchy

If rw1 holds, then rw2 and rw3 automatically hold. The reverse is false.

Why this matters

Many classical tests are exact or asymptotic under rw1, while robust versions are needed once heteroskedasticity or more general dependence is allowed.

Fifty independent random walks



The widening fan of paths is exactly what we expect from a unit-root level process: the cross-sectional dispersion grows with time.

Autocorrelation tests for weak-form EMH

If past returns contain predictive information, we should see serial correlation.

$$\gamma(j) = \mathrm{Cov}(Y_t, Y_{t-j}), \qquad \rho(j) = \frac{\gamma(j)}{\gamma(0)}.$$

Under rw1 and mild conditions,

$$\sqrt{T}\big(\widehat{\rho}(1), \ldots, \widehat{\rho}(p)\big)' \Longrightarrow N(0, I_p).$$

- Individual sample autocorrelations can be compared with Bartlett bands of order $\pm z_{\alpha/2}/\sqrt{T}$.
- Significant autocorrelation means lagged returns help linearly predict future returns.

Portmanteau tests: Box–Pierce and Ljung–Box

The joint null of zero autocorrelation up to lag p can be tested by

$$Q_{BP} = T \sum_{j=1}^{p} \widehat{\rho}(j)^2,$$

or with the small-sample correction

$$Q_{LB} = T(T+2) \sum_{j=1}^{p} \frac{\widehat{\rho}(j)^2}{T-j}.$$

- Under rw1, these statistics are asymptotically χ^2_p.
- The Ljung–Box version is usually preferred in finite samples.
- Robust portmanteau variants are useful when returns are conditionally heteroskedastic.

Interpretation

Rejection says the return sequence is not behaving like i.i.d. noise at the chosen set of lags.

AR(p) predictability test

The textbook also tests EMH using an autoregression:

$$Y_t = \mu + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \varepsilon_t, \qquad \mathrm{E}(\varepsilon_t \mid Y_{t-1}, \ldots, Y_{t-p}) = 0.$$

Under weak-form EMH,

$$H_0 : \beta_1 = \cdots = \beta_p = 0.$$

A Wald statistic takes the form

$$W = T(\widehat{\beta} - 0)' \widehat{V}^{-1} (\widehat{\beta} - 0),$$

with $\widehat{V}$ estimated either classically (rw1) or with White-type robust covariance (rw2 / rw3).

Interpretation

This is a direct test of whether lagged returns have joint linear forecasting power.

Variance-ratio logic

The variance-ratio idea exploits a simple random-walk implication. If one-period returns are serially uncorrelated, then

$$\mathrm{Var}\left(\sum_{i=0}^{k-1} r_{t-i}\right) = k\mathrm{Var}(r_t).$$

So define

$$VR(k) = \frac{\mathrm{Var}(r_t + r_{t-1} + \cdots + r_{t-k+1})}{k\mathrm{Var}(r_t)}.$$

- Under the random-walk null, $VR(k) = 1$.
- $VR(k) < 1$ suggests mean reversion.
- $VR(k) > 1$ suggests momentum or positive serial correlation.

Robust variance-ratio tests under heteroskedasticity

The textbook moves beyond the i.i.d. benchmark because financial returns are rarely homoskedastic.

- Under rw2 and rw3, asymptotic variance formulas depend on higher-order moments such as λ_{ij} and averaged second moments.
- Robust automatic variance-ratio procedures estimate these quantities from the data and remain valid under conditional heteroskedasticity.
- In practice, the robust VR test is often more informative than the classical homoskedastic version.

Important nuance

Time-varying volatility does not by itself violate EMH. What matters is whether returns remain predictable after allowing for that heteroskedasticity.

What does rejection actually mean?

Not every statistical rejection is economically important.

- A huge sample can detect extremely small departures from the strict random-walk benchmark.
- A low p-value does not automatically imply a profitable trading strategy after costs, risk, and model uncertainty.
- The textbook therefore insists on reading **statistical evidence** and **economic magnitude** together.

Good empirical habit

Always report both significance and size: p-values, coefficient magnitudes, R^2, and the direction of the variance-ratio deviation.

Linking Lecture 4 and Lecture 5

The textbook's financial interpretation of Lectures 4 and 5 can be summarized in one table.

Object	**Typical empirical classification**	**Implication**
log price level	often close to $I(1)$	shocks shift the level permanently
log return	often close to $I(0)$	shocks are short-run innovations, suitable for predictability tests
trend-fitted level residual	depends on whether the original nonstationarity was deterministic or stochastic	detrending helps only if the trend was deterministic

Do not confuse the two questions

"Does price have a unit root?" and "Are returns predictable?" are related, but they are not the same test.

Empirical interpretation from the textbook: SPY returns

The chapter's main financial illustration uses daily SPY log returns and compares a long sample (2000–2024) with the recent 2024 subsample.

- Long sample: some serial dependence and mean reversion are statistically detectable.
- Recent sample: daily returns look much closer to serially uncorrelated noise.
- In both cases, the mean is tiny relative to volatility.

Textbook conclusion

Over long samples, weak-form EMH is not exact in a strict i.i.d. sense, but the economic predictability in daily returns is still modest.

Full sample versus 2024: textbook diagnostic table

Diagnostic	**2000–2024**	**2024 only**
Ljung–Box *p*-value (lag 20)	$< 2.2 \times 10^{-16}$	0.645
Robust Auto.Q *p*-value	0.0018	0.557
AR(5) robust Wald *p*-value	0.0288	0.6919
AR(5) R^2	0.84%	1.63%
Classical $VR(k) - 1$	negative for $k = 2, 5, 10, 20$	small at short k, negative at long k
Robust Auto.VR statistic	-5.04	0.59

- The long sample rejects the strict random-walk benchmark.
- The recent subsample is much more consistent with weak-form efficiency at the daily horizon.

Bottom line on weak-form efficiency

- Prices may still behave approximately like unit-root processes even when daily returns show small departures from i.i.d. behavior.
- A market can be "close to efficient" without satisfying every exact random-walk implication.
- The right empirical language is often:
  - *statistically detectable* deviations,
  - but *economically modest* predictability.

Practical interpretation

For many assets, model the **level** as persistent and the **return** as roughly stationary, then test carefully for any remaining structure in returns.

Lecture 5 map

1. Deterministic trend models and nonparametric trend fitting
2. Efficient Market Hypothesis and random-walk testing
3. R block: detrending, unit-root testing, and random-walk diagnostics

R workflow for Lecture 5

A coherent empirical workflow, following the textbook, is:

1. import prices and build log prices and log returns,
2. inspect the level series and ask whether deterministic trend fitting is sensible,
3. compare global and local smoothers,
4. run ADF / KPSS tests on prices and returns,
5. run ACF, portmanteau, AR, and variance-ratio diagnostics on returns,
6. interpret rejections economically, not just statistically.

Why this order?

Do not start with an EMH test on raw prices. First decide whether you should be working with levels, detrended levels, first differences, or returns.

R block 1: import data and construct log prices / returns

```
library(quantmod)
library(dplyr)

spy_xts <- getSymbols("SPY", src = "yahoo",
                      from = "2000-01-01",
                      to   = "2024-12-31",
                      auto.assign = FALSE)

spy_df <- data.frame(
  date  = index(spy_xts),
  price = as.numeric(Ad(spy_xts))
) %>%
  filter(!is.na(price)) %>%
  arrange(date) %>%
  mutate(
    log_price  = log(price),
    log_return = c(NA_real_, diff(log_price)),
    t          = row_number()
  )
```

From here, the level and return analyses should be kept conceptually separate.

R block 2: fit deterministic trends

```r
# Global quadratic trend on log prices
global_fit <- lm(log_price ~ t + I(t^2), data = spy_df)
spy_df$trend_global_log <- fitted(global_fit)

# Simple back-transformation with bias correction
bc_factor <- mean(exp(residuals(global_fit)))
spy_df$trend_global <- exp(spy_df$trend_global_log) * bc_factor

# Detrended residuals
spy_df$u_hat <- residuals(global_fit)
```

This is the textbook's parametric trend fit: smooth, global, and easy to interpret, but possibly too rigid locally.

R block 3: rolling local trend, LOESS, or HP-style smoothing

```r
# Rolling local quadratic trend (abbreviated idea)
n_window <- 40
trend_local_log <- rep(NA_real_, nrow(spy_df))

for (i in seq_len(nrow(spy_df))) {
  if (i < n_window) next
  win <- spy_df[(i - n_window + 1):i, ]
  fit_i <- lm(log_price ~ t + I(t^2), data = win)
  trend_local_log[i] <- predict(fit_i, newdata = spy_df[i, ])
}

# Alternative smoother:
# spy_df$trend_loess <- predict(loess(log_price ~ t, data = spy_df,
#                                     span = 0.08))
```

The key trade-off is adaptability versus stability. Window length or span plays the same role as a bandwidth.

R block 4: ADF and KPSS on prices and returns

```r
library(urca)

# Prices: ADF with drift, KPSS level + trend
adf_price  <- ur.df(spy_df$log_price, type = "drift",
                    lags = 1, selectlags = "AIC")
kpss_mu_p  <- ur.kpss(spy_df$log_price, type = "mu",  lags = "short")
kpss_tau_p <- ur.kpss(spy_df$log_price, type = "tau", lags = "short")

# Returns: ADF without drift, KPSS level
ret <- na.omit(spy_df$log_return)
adf_ret  <- ur.df(ret, type = "none", lags = 1, selectlags = "AIC")
kpss_ret <- ur.kpss(ret, type = "mu", lags = "short")
```

Textbook pattern: log prices often look close to $I(1)$, while log returns are much closer to $I(0)$.

R block 5: ACF and Ljung–Box diagnostics

```
# ACF plot for returns
acf(ret, main = "ACF of SPY daily log returns")

# Ljung-Box test up to lag 20
Box.test(ret, lag = 20, type = "Ljung-Box")

# If heteroskedasticity is a concern, add a robust portmanteau test
# from a package implementing Auto.Q or related procedures.
```

These diagnostics ask whether serial correlation remains once we have moved from prices to returns.

R block 6: AR predictability and robust Wald tests

```
library(sandwich)
library(lmtest)

ret_df <- data.frame(
  ret = ret,
  L1 = dplyr::lag(ret, 1), L2 = dplyr::lag(ret, 2),
  L3 = dplyr::lag(ret, 3), L4 = dplyr::lag(ret, 4),
  L5 = dplyr::lag(ret, 5)
) %>% na.omit()

fit_ar5 <- lm(ret ~ L1 + L2 + L3 + L4 + L5, data = ret_df)
coeftest(fit_ar5, vcov. = vcovHC(fit_ar5, type = "HC0"))
waldtest(fit_ar5, . ~ 1, vcov = vcovHC(fit_ar5, type = "HC0"))
```

The joint null is that lagged returns have no linear forecasting power.

R block 7: variance-ratio diagnostics

```
library(vrtest)

# Classical variance-ratio diagnostics
kvec <- c(2, 5, 10, 20)
vr_out <- VR.minus.1(ret, kvec = kvec)
vr_out$VR.kvec

# Robust automatic VR test
autoVR <- Auto.VR(ret)
autoVR
```

Read the sign economically: $VR(k) - 1 < 0$ indicates mean reversion; $VR(k) - 1 > 0$ indicates momentum.

A practical workflow: detrend, difference, or use returns?

Observed pattern	First transformation to try	What to test next
smooth macro growth around a path	polynomial trend or local smoother	check residual stationarity and HAC inference
price level that wanders with permanent shocks	log difference / return	ADF / KPSS on transformed series
daily returns with visible dependence	keep returns in levels	ACF, Ljung–Box, AR, VR, volatility modelling
series with outliers	robust local smoother	re-check diagnostics after filtering

This is the operational version of the textbook's deterministic-trend / stochastic-trend distinction.

Common pitfalls in empirical work

- Treating any upward-sloping series as evidence of a unit root without considering deterministic trend.
- Detrending mechanically and then forgetting that the residual series contains first-stage estimation error.
- Testing EMH on price levels instead of on returns or price changes.
- Reporting statistical rejection without asking whether the effect is economically meaningful.
- Ignoring heteroskedasticity when using ACF, AR, or variance-ratio tests on financial returns.

Rule of thumb

First decide what object you are modelling. Only then decide what test or estimator is appropriate.

Lecture 5 summary

1. Deterministic trends and stochastic trends are fundamentally different sources of nonstationarity.
2. Polynomial trend models are easy to estimate, but their asymptotics require nonstandard scaling because the regressors trend with time.
3. Nonparametric smoothers add flexibility, at the cost of slower convergence and bandwidth sensitivity.
4. Weak-form EMH links naturally to random-walk pricing: prices may be close to $I(1)$ while returns behave roughly like $I(0)$ innovations.
5. ACF, Ljung–Box, AR, and variance-ratio tests answer complementary predictability questions.
6. In practice, combine trend modelling, unit-root testing, and return diagnostics rather than relying on any single test.