Lecture 4 – Unit-Root Processes, Difference Stationarity, Dickey–Fuller / ADF, and KPSS

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 sequence as closely as possible, while matching the course plan for Lecture 4.

1. **Chapter 8:** *Unit Root Process*
   - unit roots, random walks, and nonstationarity,
   - factorizing the autoregressive polynomial,
   - integrated processes and difference stationarity.
2. **Chapter 8:** *Dickey–Fuller Test*
   - testing a unit root in the AR(1) model,
   - nonstandard critical values,
   - deterministic versus stochastic trend.
3. **Chapter 8:** *Augmented Dickey–Fuller Test*
   - lagged differences, serial correlation, and specification,
   - random walk with drift illustrations,
   - interpretation of the test statistics.
4. **Chapter 8:** *KPSS Test*
   - reversing the null and alternative,
   - residual partial sums and long-run variance,
   - why combining unit-root and stationarity tests is useful.

Lecture goals

- Understand why a unit root is not just "high persistence," but a boundary case with fundamentally different asymptotic behavior.
- Distinguish clearly between trend-stationary and difference-stationary representations.
- See how the Dickey–Fuller and Augmented Dickey–Fuller tests are built directly from the AR model.
- Learn the logic of the KPSS test and why it complements DF / ADF rather than replacing them.
- Leave the lecture with a practical decision rule for classifying a series as approximately $I(0)$, $I(1)$, or ambiguous.

Roadmap for the three-hour block

Block 1

Unit-root processes, random walks, and difference stationarity.

Block 2

Dickey–Fuller and Augmented Dickey–Fuller tests.

Block 3

KPSS test and the logic of combining stationarity and unit-root tests.

Lecture 4 map

Why Chapter 8 turns to unit roots

The textbook opens the unit-root part of Chapter 8 with a simple observation: many macroeconomic and financial series display **long-term memory characteristics** and a persistence that is hard to reconcile with weak stationarity.

- Output levels, price indices, exchange rates, interest rates, and asset prices often drift over long horizons.
- In such series, random shocks can have **lasting** rather than transitory effects.
- This is exactly the setting in which a **unit root** becomes the natural econometric object.

Working definition

A unit root exists when the characteristic equation of the autoregressive part has a root with modulus equal to one.

The textbook's starting example: AR(1)

Consider the AR(1) model

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2).$$

The lag polynomial is

$$1 - \phi L,$$

and the characteristic equation is

$$1 - \phi z = 0.$$

- The root is $z = 1/\phi$.
- A **unit root** occurs when $\phi = 1$, so that $z = 1$.
- This special case is the **random walk**: $Y_t = Y_{t-1} + \varepsilon_t$.

Boundary interpretation

The unit-root case is the boundary between stationary autoregression ($|\phi| < 1$) and explosive behavior ($|\phi| > 1$).

Why $\phi = 1$ breaks weak stationarity

For the autoregressive process

$$Y_t = \theta Y_{t-1} + \varepsilon_t,$$

stationarity requires finite, time-invariant mean and variance. If $\theta = 1$, then

$$Y_t = Y_{t-1} + \varepsilon_t, \qquad \mathrm{Var}(Y_t) = \mathrm{Var}(Y_{t-1}) + \sigma^2.$$

Iterating the variance recursion gives

$$\mathrm{Var}(Y_t) = \mathrm{Var}(Y_0) + t\sigma^2.$$

- The unconditional variance grows linearly with t.
- There is no finite time-invariant variance unless $\sigma^2 = 0$.
- So the process cannot be weakly stationary.

Textbook conclusion

An AR(1) process is stationary only when the autoregressive root lies outside the unit circle, equivalently when $|\theta| < 1$.

Stationarity of ARMA models in root language

The textbook extends the same logic to the general ARMA(p, q) model

$$\theta(L)Y_t = \alpha(L)\varepsilon_t.$$

- The process is stationary if and only if every root $z_1, \ldots, z_p$ of $\theta(z) = 0$ satisfies $|z_j| > 1$.
- Equivalently, all roots of the autoregressive polynomial lie **outside** the unit circle.
- If one root is exactly 1, stationarity fails, but differencing may restore stationarity.

Why this matters

The root language is the clean bridge between the stationary ARMA world of Lecture 3 and the integrated-process world of Lecture 4.

A single unit root leads to differencing

Suppose one autoregressive root is exactly 1 while all other roots remain outside the unit circle. Then the textbook factorizes the AR polynomial as

$$\theta(L) = (1 - L)\theta^*(L).$$

Hence

$$\theta^*(L)(1 - L)Y_t = \theta^*(L)\Delta Y_t = \alpha(L)\varepsilon_t,$$

where $\Delta = 1 - L$ is the first-difference operator.

- If Y_t has a unit root, then ΔY_t can be stationary.
- The unit-root series is nonstationary in levels but stationary in first differences.
- This is the essence of **difference stationarity**.

Key phrase

A series that becomes stationary after first differencing is called **first-order integrated**, or $I(1)$.

Worked example from the textbook: ARMA(2,1) with a unit root

The textbook uses

$$Y_t = 1.2Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t - 0.5\varepsilon_{t-1}.$$

In lag notation,

$$(1 - 1.2L + 0.2L^2)Y_t = (1 - 0.5L)\varepsilon_t.$$

Factor the AR part:

$$(1 - 0.2L)(1 - L)Y_t = (1 - 0.5L)\varepsilon_t.$$

So $z = 1$ is a root of the characteristic equation, and after differencing,

$$\Delta Y_t = 0.2\Delta Y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1},$$

which is a stationary ARMA(1,1) process.

Interpretation

The original series is nonstationary because of the unit root, but its first difference inherits an ordinary stationary ARMA representation.

Integrated processes and ARIMA language

- $I(0)$: stationary in levels.
- $I(1)$: nonstationary in levels, stationary after one difference.
- $I(2)$: stationary only after two differences.

If ΔY_t follows a stationary ARMA model, then Y_t follows an

$$\text{ARIMA}(p, 1, q)$$

model. The textbook stresses two practical points:

- Many economic and financial *levels* or *log-levels* behave as $I(1)$ processes.
- Some series require second differencing, in which case the characteristic equation has two unit roots and the series is $I(2)$.

Operational meaning

Integration order tells you how many differences are required before the stationary tools from Lecture 3 become appropriate again.

Trend-stationary versus difference-stationary

The textbook's broader Chapter 8 logic is that there are two very different reasons a series can be nonstationary.

Trend-stationary

$$Y_t = \beta_0 + \beta_1 t + u_t, \qquad u_t \sim I(0).$$

- Mean varies deterministically with time.
- Shocks are transitory.
- Stationarity is restored by **detrending**.

Difference-stationary

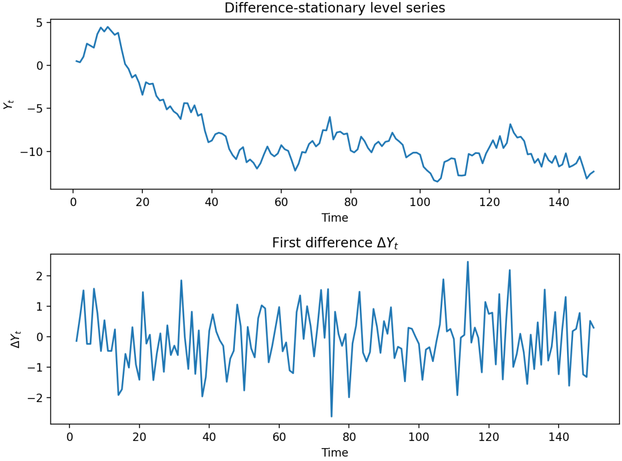
$$Y_t = Y_{t-1} + u_t, \qquad \Delta Y_t = u_t.$$

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

Main distinction

A trend-stationary series reverts to a deterministic path; a difference-stationary series does not.

Difference stationarity: level versus first difference



The level series wanders, while the first difference fluctuates around a stable mean and variance. That is exactly the content of the statement "Y_t is $I(1)$ but ΔY_t is $I(0)$."

Random walk and random walk with drift

The basic random walk is

$$Y_t = Y_{t-1} + \varepsilon_t.$$

A very important extension is the random walk with drift:

$$Y_t = \delta + Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.(0, \sigma^2).$$

Starting from Y_0, repeated substitution gives

$$Y_t = Y_0 + \delta t + \sum_{i=1}^{t} \varepsilon_i.$$

- The term δt is a deterministic component.
- The cumulative shock term $\sum_{i=1}^{t} \varepsilon_i$ is the stochastic trend.

Important subtlety

A random walk with drift contains **both** deterministic and stochastic trend components.

Moments of the random walk with drift

From

$$Y_t = Y_0 + \delta t + \sum_{i=1}^{t} \varepsilon_i,$$

we obtain

$$\mathrm{E}(Y_t) = Y_0 + \delta t, \qquad \mathrm{Var}(Y_t) = t\sigma^2.$$

- The mean drifts linearly over time.
- The variance also grows linearly over time.
- Hence nonstationarity comes from both the mean and the variance.

Connection to the textbook discussion

This is why a random walk with drift may look like a deterministic trend series in finite samples even though its stochastic structure is fundamentally different.

Transitory versus permanent shocks

The textbook emphasizes the contrast between $I(0)$ and $I(1)$ processes.

- In an $I(0)$ process, a shock changes the series now, but the effect eventually dies out. This is mean reversion.
- In an $I(1)$ process, a shock permanently shifts the future path of the level series.
- This is why the textbook says an $I(1)$ process has “infinite memory” of past shocks.

Forecasting implication

For a stationary AR process, long-horizon forecasts move back toward the mean. For a random walk, long-horizon forecasts inherit the current level and every shock that has accumulated into it.

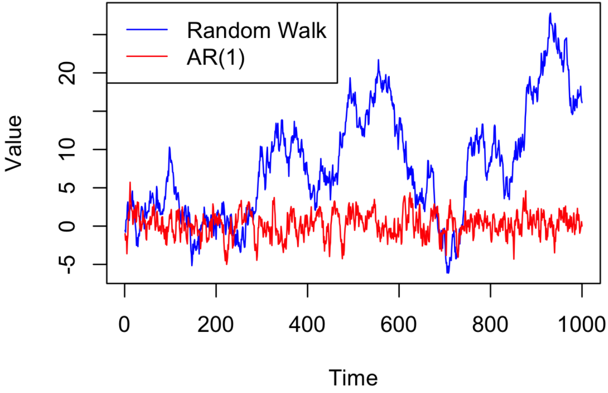
A compact comparison of $I(0)$ and $I(1)$ behavior

Feature	$I(0)$	$I(1)$
Mean behavior	fluctuates around a fixed mean	no fixed mean in levels
Variance	finite and time-invariant	grows with time in the level series
Effect of shocks	temporary	permanent
Memory	limited	effectively infinite in levels
ACF	decays relatively quickly	decays very slowly
How to restore stationarity	usually none or detrend	difference the series

Practical reading

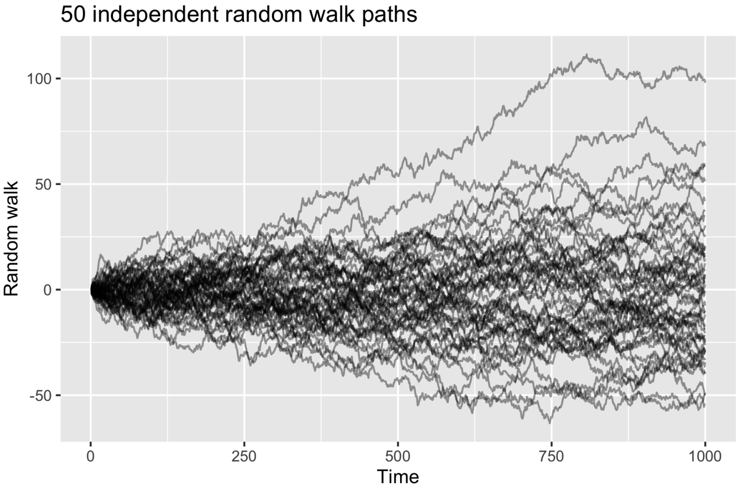
The difference between $I(0)$ and $I(1)$ is not cosmetic. It changes modelling, inference, and the economic interpretation of shocks.

Simulated AR(1) and random walk



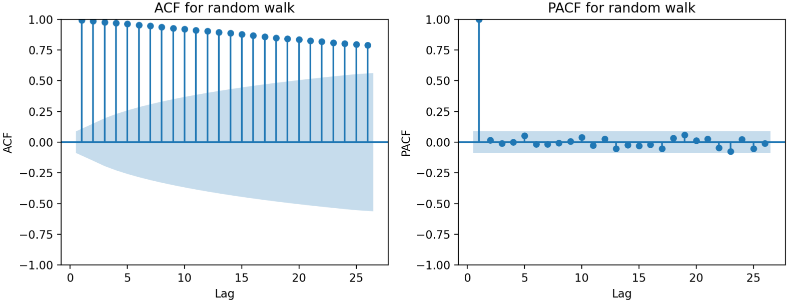
Stationary AR(1) dynamics revert; random-walk dynamics wander.

Many random walks fan out over time



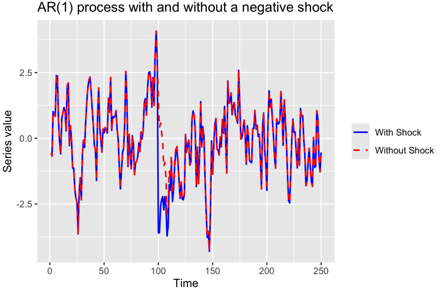
The textbook's random-walk intuition is visible here: the cross-sectional spread of paths widens as t grows because the level variance increases with time.

Random-walk series and its ACF / PACF

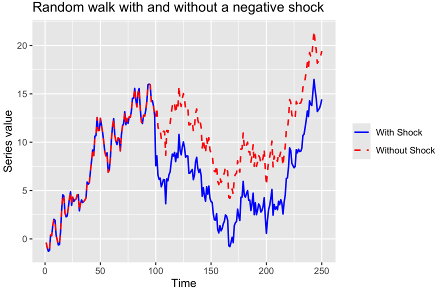


The ACF stays close to one for many lags. This is the visual signature of the permanent effect of shocks in the level series.

Shock response: AR(1) versus random walk



AR(1): shock effect dies out



Random walk: shock changes the level path permanently

This is one of the cleanest textbook contrasts: stationary dynamics absorb shocks; unit-root dynamics accumulate them.

Why stochastic trends create econometric problems

The textbook lists three immediate consequences of a stochastic trend.

1. In the model $Y_t = Y_{t-1} + \varepsilon_t$, the OLS estimator is biased toward zero even though the true autoregressive coefficient is one.
2. If a regressor has a stochastic trend, the usual large-sample t-distribution for OLS coefficients no longer applies.
3. Two unrelated $I(1)$ series can look strongly related in levels, producing **spurious regression**.

Bottom line

Visual persistence is not enough. Once stochastic trends are plausible, we need formal unit-root and stationarity tests.

Why ACF evidence is suggestive but not decisive

A slow sample ACF decay often points toward a unit root, but the textbook's logic is careful here.

- A deterministic trend can also create large sample autocorrelations if the time-varying mean is not removed correctly.
- Near-unit-root stationary processes can look very persistent in finite samples.
- Small samples make it difficult to distinguish "very persistent" from "truly integrated."

Hence the next step

The Dickey–Fuller family asks a sharper question: does the autoregressive root actually equal one?

Lecture 4 map

The Dickey–Fuller idea starts from AR(1)

The textbook introduces the DF test through the AR(1) model

$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t.$$

We want to test

$$H_0 : \theta = 1 \qquad \text{against} \qquad H_1 : \theta < 1.$$

The formal DF statistic is the usual t-ratio for testing $\theta = 1$:

$$DF = \frac{\widehat{\theta} - 1}{se(\widehat{\theta})}.$$

But there is a twist

Under the null, the process is nonstationary, so the statistic does *not* follow the ordinary t-distribution.

Equivalent difference form of the DF regression

Subtract Y_{t-1} from both sides of the AR(1) model:

$$\Delta Y_t = \delta + (\theta - 1) Y_{t-1} + \varepsilon_t.$$

Define

$$\pi = \theta - 1.$$

Then the regression becomes

$$\Delta Y_t = \delta + \pi Y_{t-1} + \varepsilon_t,$$

with hypotheses

$$H_0 : \pi = 0 \qquad \text{versus} \qquad H_1 : \pi < 0.$$

Why this form is convenient

Testing whether the autoregressive root equals one becomes testing whether the coefficient on Y_{t-1} in the differenced regression is zero.

The three common DF specifications

In practice, the deterministic part of the test equation matters.

1. **No constant, no trend**

$$\Delta Y_t = \pi Y_{t-1} + \varepsilon_t.$$

   Appropriate when the series fluctuates around zero with no visible drift.

2. **Constant only (drift)**

$$\Delta Y_t = \delta + \pi Y_{t-1} + \varepsilon_t.$$

   Appropriate when the series may have nonzero mean or drift but no deterministic trend.

3. **Constant plus linear trend**

$$\Delta Y_t = \delta + \gamma t + \pi Y_{t-1} + \varepsilon_t.$$

   Appropriate when a deterministic linear trend is plausible under the alternative.

How the deterministic specification changes the null and alternative

The null is always a unit root, but the alternative depends on what deterministic terms are included.

- Without constant and trend: alternative is stationarity around zero.
- With constant only: alternative is level stationarity around a nonzero mean.
- With constant and trend: alternative is **trend stationarity**.

Econometric lesson

You are not only testing “unit root or not.” You are also choosing *what kind of stationary alternative* you are willing to compare it to.

Why the DF statistic is nonstandard

Dickey and Fuller showed that under the null hypothesis $\theta = 1$, the t-ratio is not t-distributed, not even asymptotically.

- Under H_0, Y_t is nonstationary, so the usual central limit arguments for OLS change.
- The limiting distribution is skewed to the right relative to a standard t.
- Critical values therefore depend on the deterministic specification of the regression.

Practical implication

Never compare a DF or ADF test statistic to a standard normal or Student-t critical value. Use the correct Dickey–Fuller critical values for the chosen regression.

Approximate Dickey–Fuller critical values

	Without trend		With trend	
Sample size	1%	5%	1%	5%
$T = 25$	-3.75	-3.00	-4.38	-3.60
$T = 50$	-3.58	-2.93	-4.15	-3.50
$T = 100$	-3.51	-2.89	-4.40	-3.45
$T = 250$	-3.46	-2.88	-3.99	-3.43
$T = 500$	-3.44	-2.87	-3.98	-3.42
$T = \infty$	-3.43	-2.86	-3.96	-3.41

These are the textbook's benchmark values adapted from Fuller. The trend case requires more negative critical values because the alternative is broader.

How to read the DF decision rule

Because the alternative is $\pi < 0$, the rejection region is in the **left tail**.

- If the test statistic is *more negative* than the critical value, reject the unit-root null.
- If it is not negative enough, fail to reject the null.

Interpretation

- Rejecting H_0 supports an $I(0)$ or trend-stationary description, depending on the specification.
- Failing to reject H_0 means the data are consistent with a unit root; it does *not* prove the root is exactly one.

Power warning

DF and ADF tests often have low power against highly persistent but stationary alternatives.

Deterministic trend versus stochastic trend

The textbook places this comparison directly next to the DF discussion because it determines which version of the test makes sense.

Deterministic trend

$$Y_t = \delta + \gamma t + \alpha(L)\varepsilon_t.$$

Then

$$\mathrm{E}(Y_t) = \delta + \gamma t.$$

- Mean depends on time.
- Variance can still be time-invariant.
- Stationarity can hold around the trend.

Stochastic trend

$$Y_t = Y_{t-1} + \varepsilon_t.$$

Then

$$\mathrm{Var}(Y_t) = t\sigma^2.$$

- Variance depends on time.
- Shocks accumulate into the level.
- Differencing, not detrending, restores stationarity.

Random walk with drift revisited

The process

$$Y_t = \delta + Y_{t-1} + \varepsilon_t$$

expands to

$$Y_t = Y_0 + \delta t + \sum_{i=1}^{t} \varepsilon_i.$$

- The drift δt generates a deterministic trend in the mean.
- The cumulative innovations generate a stochastic trend in the level.
- In finite samples, a large drift can visually dominate the stochastic component.

Why this matters for testing

If a trend is omitted when the alternative is trend-stationary, DF / ADF inference can be badly distorted.

Typical Dickey–Fuller pitfalls

- **Wrong deterministic specification:** using a no-trend regression for a series that clearly has a deterministic trend.
- **Over-interpreting nonrejection:** failure to reject a unit root is often weak evidence in short samples.
- **Ignoring structural breaks:** a broken trend can mimic a unit root.
- **Treating levels and growth rates the same way:** price levels and returns often belong in different DF specifications.

Empirical discipline

Always inspect the series, think about the economics, and match the deterministic terms in the test regression to the plausible data-generating mechanism.

Lecture 4 map

Why augment the Dickey–Fuller regression?

The basic DF regression is derived for a first-order autoregressive environment. The textbook extends it by adding lagged differences so that the error term becomes approximately white noise.

- If serial correlation remains in the residuals, the plain DF regression is misspecified.
- Lagged differences absorb higher-order autoregressive dynamics.
- The resulting test is the **Augmented Dickey–Fuller (ADF)** test.

Same null, richer short-run dynamics

ADF still tests for a unit root, but it controls for extra serial dependence in the differenced equation.

AR(2) algebra gives the ADF(1) regression

Start from the AR(2) model in the textbook:

$$Y_t = \delta + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \varepsilon_t.$$

Rewrite it as

$$\Delta Y_t = \delta + (\theta_1 + \theta_2 - 1) Y_{t-1} - \theta_2 \Delta Y_{t-1} + \varepsilon_t.$$

The characteristic equation is

$$1 - \theta_1 z - \theta_2 z^2 = 0.$$

If $z = 1$ is a unit root, then

$$1 - \theta_1 - \theta_2 = 0.$$

Testing implication

Testing for a unit root is equivalent to testing whether the coefficient on Y_{t-1} in the transformed regression is zero.

The general ADF regression

For the AR(p) model

$$Y_t = \delta + \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p} + \varepsilon_t,$$

the textbook gives the equivalent ADF representation

$$\Delta Y_t = \delta + \pi Y_{t-1} + c_1 \Delta Y_{t-1} + \cdots + c_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,$$

where

$$\pi = \theta_1 + \cdots + \theta_p - 1.$$

- Null: $H_0 : \pi = 0$.
- Alternative: $H_1 : \pi < 0$.
- A trend version adds γt on the right-hand side.

Deterministic terms in ADF: none, drift, trend

The ADF logic mirrors the DF logic, but now with lagged differences included.

$$\text{ADF(none)}: \quad \Delta Y_t = \pi Y_{t-1} + \sum_{j=1}^{k} c_j \Delta Y_{t-j} + \varepsilon_t,$$

$$\text{ADF(drift)}: \quad \Delta Y_t = \delta + \pi Y_{t-1} + \sum_{j=1}^{k} c_j \Delta Y_{t-j} + \varepsilon_t,$$

$$\text{ADF(trend)}: \quad \Delta Y_t = \delta + \gamma t + \pi Y_{t-1} + \sum_{j=1}^{k} c_j \Delta Y_{t-j} + \varepsilon_t.$$

- The critical values remain nonstandard.
- The appropriate version depends on whether the alternative is stationarity around zero, around a constant mean, or around a deterministic trend.

Choosing the lag length in ADF

The textbook emphasizes why lagged differences are included; empirically, we also need to decide how many of them to include.

- Too few lags leave serial correlation in the residuals and can invalidate the test.
- Too many lags waste degrees of freedom and reduce power.
- In practice, information criteria, sequential testing, or residual diagnostics are commonly used.

Interpretation

The augmentation part of ADF is about cleaning the short-run dynamics so that the long-run question – does the root equal one? – is tested in a credible regression.

How software often reports ADF output

Many implementations report more than one statistic.

- A τ **statistic**: the Dickey–Fuller type t-statistic for the coefficient on Y_{t-1}.
- Sometimes a ϕ **statistic**: a joint test involving the deterministic terms, especially in the drift or trend specification.

Reading the output

- The τ statistic addresses the unit-root null directly.
- The ϕ statistic helps assess whether a drift or trend term is relevant under the alternative specification.

Do not mix them up

A significant trend-related ϕ statistic does not by itself overturn a nonrejected unit-root null.

The textbook's simulation design for ADF with drift

The simulation uses random walks with drift,

$$Y_t = Y_{t-1} + \delta + \varepsilon_t,$$

for three values of the drift parameter:

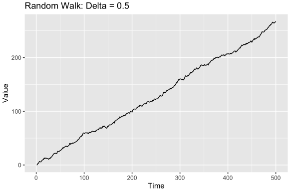
$$\delta \in \{0.5,\ 0.1,\ 0.05\}.$$

- The ADF regression is run with `type = "drift"`.
- The reported statistics are τ_2 and ϕ_1.
- The point of the exercise is to show how deterministic drift can coexist with a unit root.

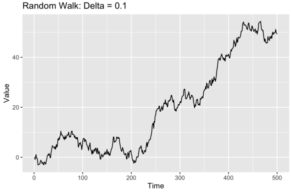
Question behind the figure

If the process truly has a unit root, can a strong deterministic drift make it look deceptively trend-like in finite samples?

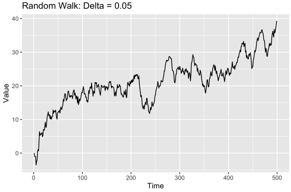
ADF results for different drifts



(a) $\delta = 0.5$
$\tau_2 = 0.4965$, $\phi_1 = 64.55$
τ_2 critical values: -3.44, -2.87, -2.57
ϕ_1 critical values: 6.47, 4.61, 3.79



(b) $\delta = 0.1$
$\tau_2 = -0.0598$, $\phi_1 = 2.46$
τ_2 critical values: -3.44, -2.87, -2.57
ϕ_1 critical values: 6.47, 4.61, 3.79



(c) $\delta = 0.05$
$\tau_2 = -2.0785$, $\phi_1 = 3.75$
τ_2 critical values: -3.44, -2.87, -2.57
ϕ_1 critical values: 6.47, 4.61, 3.79

For all three cases, the τ_2 statistic fails to reject the unit-root null, even though the realized series can look strongly trending when δ is large.

How to interpret the drift simulation

- For every drift value, τ_2 remains above the relevant negative critical values, so the unit-root null is not rejected.
- When $\delta = 0.5$, the ϕ_1 statistic is extremely large, indicating a strong deterministic component in the observed path.
- As δ shrinks, the trend component becomes visually less dominant, but the nonrejection of the unit root remains.

Textbook lesson

A series can contain both a deterministic trend and a stochastic trend. A strong deterministic trend may dominate the picture without eliminating the unit root.

Beveridge–Nelson decomposition: the conceptual message

The textbook briefly uses the Beveridge–Nelson idea to formalize the difference between permanent and transitory movements.

- A nonstationary series can often be decomposed into a **stochastic trend** component and a **stationary** component.
- The stochastic trend carries the permanent effects of shocks.
- The stationary component carries cyclical or short-run fluctuations.

Why it belongs here

DF / ADF tests are fundamentally about whether the permanent component is present in the level series.

What ADF can and cannot tell you

- ADF can tell you whether the data provide enough evidence against the unit-root null in a specified regression.
- It cannot, by itself, settle every borderline case.
- Near-unit-root processes, short samples, volatility changes, and structural breaks can all complicate interpretation.

Hence the need for a complement

The KPSS test asks the reverse question: instead of taking the unit root as the null, it takes stationarity as the null.

Lecture 4 map

KPSS reverses the hypotheses

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test flips the logic of the Dickey–Fuller family.

DF / ADF

$$H_0: \ \text{unit root}$$
$$H_1: \ \text{(trend-)stationary}$$

KPSS

$$H_0: \ \text{(trend-)stationary}$$
$$H_1: \ \text{stochastic trend}$$

Why this is valuable

One test's weak nonrejection becomes less informative in isolation. Two tests with opposite nulls can provide a much sharper overall reading.

The KPSS model in the textbook

KPSS starts from the decomposition

$$y_t = \alpha + \beta t + r_t + \varepsilon_t, \qquad r_t = r_{t-1} + u_t,$$

where u_t is i.i.d. with variance σ_u^2 and ε_t is stationary and weakly dependent.

- r_t is the stochastic-trend component.
- If $\sigma_u^2 = 0$, then r_t is constant and the series is stationary around the deterministic part.

Hence the KPSS hypotheses are

$$H_0 : \sigma_u^2 = 0, \qquad H_1 : \sigma_u^2 > 0.$$

How the KPSS statistic is constructed

Under the null model, regress y_t on the chosen deterministic terms and collect the residuals

$$\widehat{\varepsilon}_t = y_t - \widehat{\alpha} - \widehat{\beta} t.$$

Form the partial sums

$$S_t = \sum_{s=1}^{t} \widehat{\varepsilon}_s, \qquad t = 1, \ldots, T.$$

Let $\widehat{\omega}^2$ be a consistent estimate of the long-run variance of $\widehat{\varepsilon}_t$. Then

$$\mathrm{KPSS} = \frac{1}{T^2} \sum_{t=1}^{T} \left(\frac{S_t}{\widehat{\omega}} \right)^2 .$$

Intuition

If the residuals are stationary, their cumulative sums should not wander too far. If they contain a stochastic trend, the partial sums grow too much and the statistic becomes large.

Level-stationary and trend-stationary KPSS versions

Just as in DF / ADF, the deterministic specification matters.

- **Level-stationary null:** regress on a constant only.
- **Trend-stationary null:** regress on a constant and linear trend.

Interpretation

- The level version asks whether the series is stationary around a constant mean.
- The trend version asks whether the series is stationary around a deterministic linear trend.

Symmetry with ADF

Matching the deterministic component to the economic context is just as important in KPSS as it is in DF / ADF.

Why long-run variance estimation matters in KPSS

The statistic uses a long-run variance estimate $\widehat{\omega}^2$, often based on a Newey–West type HAC estimator.

- If the bandwidth is too small, short-run dependence may be underestimated.
- If the bandwidth is too large, finite-sample performance may deteriorate.
- Different software defaults can change the test outcome in borderline cases.

Practical reading

KPSS is not just "a number." It is a residual-based partial-sum statistic normalized by an estimated long-run variance.

Local-to-unity alternatives

The textbook notes that unit-root and stationarity tests are naturally studied under *local-to-unity* alternatives,

$$y_t = \phi_T y_{t-1} + u_t, \qquad \phi_T = 1 - \frac{c}{T}, \quad c > 0.$$

- Here the autoregressive root approaches one at rate $1/T$.
- This captures the empirically important distinction between clearly stationary and nearly integrated behavior.
- It also explains why finite-sample discrimination between "very persistent" and "unit root" is genuinely hard.

Implication

Borderline outcomes from ADF and KPSS are not surprising; the tests are working near an asymptotic boundary.

Why combine ADF and KPSS?

Each test has a one-sided null that can be weakly informative on its own.

ADF outcome	**KPSS outcome**	**Interpretation**
Reject unit root	Do not reject stationarity	Strong evidence for $I(0)$
Fail to reject unit root	Reject stationarity	Strong evidence for $I(1)$
Reject unit root	Reject stationarity	Conflicting evidence: possible break, misspecified trend, heteroskedasticity, or near-unit-root behavior
Fail to reject unit root	Do not reject stationarity	Low power / small sample / borderline persistence; investigate further

A practical decision rule for combining the tests

1. Decide whether the deterministic specification should be *none*, *drift*, or *trend*.
2. Run the matching ADF regression and the matching KPSS test.
3. Look for convergence of evidence rather than a single decisive p-value.
4. If the evidence conflicts, inspect plots, residual dependence, possible breaks, and the first-differenced series.

Good empirical habit

The combined use of ADF and KPSS turns the question from "did one test reject?" into "which description of the series is jointly most plausible?"

What to do when the tests conflict

Conflicting outcomes are common in finite samples. A good response is diagnostic rather than mechanical.

- Reconsider the deterministic terms: should a time trend be included?
- Check whether the differenced series looks plausibly $I(0)$.
- Examine residual serial correlation and the lag length in ADF.
- Think about structural breaks, regime changes, and outliers.
- Remember that near-unit-root processes can fool both visual inspection and formal tests.

Interpretive principle

Ambiguous evidence usually means the data are close to the boundary between $I(0)$ and $I(1)$ or the deterministic component has been misspecified.

A simple empirical workflow after this lecture

1. Plot the level series and ask whether a deterministic trend is plausible.
2. Plot the first difference and ask whether stationarity now looks plausible.
3. Use ADF to test the unit-root null under a sensible deterministic specification.
4. Use KPSS to test the complementary stationarity null under the same logic.
5. Model the series in levels if evidence supports $I(0)$ / trend stationarity; difference it if evidence supports $I(1)$.

Why this matters

The decision between levels, detrended levels, and differences is one of the most consequential modelling choices in time-series econometrics.

Lecture 4 map

Lecture 4 takeaways

- A unit root is a root of the AR polynomial at one; in levels, it creates nonstationary variance and permanent shock effects.
- Difference-stationary series are made stationary by differencing, not by subtracting a deterministic trend.
- The Dickey–Fuller test asks whether the autoregressive root equals one in an AR(1)-type environment.
- The Augmented Dickey–Fuller test adds lagged differences to handle higher-order short-run dynamics.
- The KPSS test reverses the null and asks whether the series is stationary around a constant or a trend.
- Combining ADF and KPSS usually gives a more informative classification than relying on either test alone.

From theory to modelling choices

After today's theory block, the modelling logic should be much clearer.

- If the series is plausibly $I(0)$, the stationary ARMA tools from Lecture 3 are appropriate in levels.
- If the series is trend-stationary, detrend first and then model the stationary remainder.
- If the series is $I(1)$, difference it and move to ARIMA-type modelling or, in multivariate settings, cointegration analysis.

Conceptual bridge

Lecture 3 taught us how to work *inside* the stationary world. Lecture 4 tells us how to decide whether the data even belong there.

Where Chapter 8 goes after this core theory block

The remainder of the chapter pushes these ideas into applications.

- Empirical unit-root and stationarity testing on financial data,
- interpreting prices as approximately $I(1)$ and returns as approximately $I(0)$,
- linking random-walk behavior to the weak-form Efficient Market Hypothesis.

For us

This lecture focused on the core theory: unit roots, differencing, DF / ADF, KPSS, and the logic of combining the tests.