Lecture 8 — Nonstationary VAR Systems, Cointegration, and Vector Error-Correction Models

Chapter 3 continuation: long-run equilibrium, common stochastic trends, Johansen testing, and an R-integrated workflow

Jiajing Sun

School of Economics and Management, University of Chinese Academy of Sciences

Econometrics and Time Series Methods
Spring 2026

Why this lecture comes right after Lectures 6 and 7

Lectures 6 and 7 developed the **stationary** VAR toolkit:

- reduced-form VARs,
- stability and companion-form logic,
- estimation and lag selection,
- VMA representations and impulse responses.

But many macroeconomic variables are not covariance-stationary in levels. Output, consumption, income, prices, money, and exchange rates are often well approximated by **integrated** processes.

The new question

If several variables are individually nonstationary, can they still obey a stable long-run equilibrium relation?

Answer

Yes. That is exactly the point of **cointegration**. And once cointegration is present, the correct dynamic model is not a stationary VAR in levels and not a pure VAR in differences, but a **vector error-correction model (VECM)**.

Textbook logic for today's three-hour lecture

I follow the Chapter 3 sequence in the same spirit as the textbook:

1. nonstationary VAR systems and common stochastic trends;
2. the definition of cointegration and the Granger representation idea;
3. residual-based and system-based tests for cointegration;
4. vector error-correction form and economic interpretation;
5. an R workflow using consumption and income data.

Boundary with earlier lectures

Lecture 4 introduced unit roots in the *univariate* setting. Today we move to the *multivariate* case and ask whether nonstationary series wander independently or are tied together by equilibrium forces.

Learning goals

By the end of Lecture 8, students should be able to:

1. distinguish a nonstationary VAR system from a stationary one;
2. define cointegration in terms of stationary linear combinations;
3. understand why the rank of the long-run impact matrix is central;
4. derive the VECM from a VAR and interpret α and β economically;
5. explain the logic of the Engle-Granger and Johansen procedures;
6. implement a basic cointegration workflow in R, including lag selection, rank testing, VECM estimation, and impulse-response analysis.

Practical plan for the three contact hours

Hour 1

Why nonstationary multivariate systems are different; random walks in several dimensions; cointegration as a stationary linear combination; Granger representation theorem.

Hour 2

Residual-based and system-based cointegration testing; superconsistency; vector error-correction form; interpretation of the loading and cointegrating matrices.

Hour 3

R block: FRED consumption and income example; unit-root checks; Engle-Granger and Johansen tests; VECM estimation; practical interpretation and IRFs.

Why a stationary VAR in levels can be misleading

Suppose we estimate

$$y_t = c + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \varepsilon_t$$

on variables whose levels are actually $I(1)$.

- The usual stationary asymptotics from Lectures 6 and 7 no longer apply automatically.
- Regressions among unrelated trending variables can look highly significant.
- Differencing all variables may restore stationarity, but it can also destroy information about long-run equilibrium.

Central econometric problem

How do we model short-run changes without losing the long-run relation among the levels?

Vector $I(1)$ behavior and common stochastic trends

A vector process y_t is often called $I(1)$ when each component needs one difference to become stationary:

$$\Delta y_t = y_t - y_{t-1}$$

is $I(0)$, while y_t itself is not.

Two possibilities

1. Each component may drift with its own stochastic trend, so the variables can separate indefinitely.
2. Or some linear combinations may cancel the common trends, producing stationary equilibrium errors.

Name of the second case

That second case is **cointegration**.

Multivariate random walk as the benchmark nonstationary system

A random walk in $\mathbb{R}^n$ is

$$y_t = y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d. } (0, \Omega_\varepsilon).$$

By recursive substitution,

$$y_t = \sum_{s=1}^{t} \varepsilon_s + y_0.$$

Hence

$$\mathrm{E}(y_t \mid y_0) = y_0, \qquad \mathrm{E}\big[(y_t - y_0)(y_t - y_0)' \mid y_0\big] = t\,\Omega_\varepsilon.$$

Meaning

The variance grows linearly with time. Shocks do not die out; they accumulate permanently.

Dimensionality and the “drunken man” metaphor

The textbook uses a helpful geometric metaphor.

- In one dimension, a random walker on a line repeatedly crosses previously visited points.
- In two dimensions, returns still occur, but less frequently.
- In three or more dimensions, the path becomes much more dispersed, and revisiting a particular point becomes far less likely.

Econometric analogy

As the dimension of the system increases, finding stationary combinations becomes a harder statistical problem. Cointegration asks whether the apparently wandering coordinates are actually linked by lower-dimensional equilibrium restrictions.

Spurious regression remains a danger in multivariate work

If two independent random walks are regressed on each other in levels, the estimated slope can look significant even though there is no genuine equilibrium link.

Lesson

Nonstationary comovement is not enough. We need to know whether the *difference between the series after an economically meaningful linear combination* is stationary.

Start from a VAR(1) in levels

To formalize the idea of common stochastic trends, the chapter starts from

$$\mathrm{y}_t = \Phi \mathrm{y}_{t-1} + \varepsilon_t.$$

This is not restrictive, because any VAR(p) can be written as a VAR(1) in companion form.

Subtract y_{t-1} from both sides:

$$\Delta \mathrm{y}_t = (\Phi - I_n)\mathrm{y}_{t-1} + \varepsilon_t.$$

Define

$$\Pi = \Phi - I_n.$$

Then

$$\Delta \mathrm{y}_t = \Pi \mathrm{y}_{t-1} + \varepsilon_t.$$

Key object

The matrix Π carries the long-run information.

Eigenvalue view: why some directions are stationary and others are not

If Φ is symmetric, write

$$\Phi = Q\Lambda Q',$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and Q is orthonormal. Premultiplying by Q' gives

$$\mathrm{x}_t = \Lambda \mathrm{x}_{t-1} + \varepsilon_t^*, \qquad \mathrm{x}_t = Q'\mathrm{y}_t.$$

- The component x_{jt} is stationary if and only if $|\lambda_j| < 1$.
- If $\lambda_j = 1$, that direction contains a unit root.
- Therefore, some linear combinations of the original variables can be stationary even when the levels themselves are not.

Core intuition

Cointegration is about separating the stationary directions from the unit-root directions in a multivariate system.

The rank of Π determines the economic case

In

$$\Delta y_t = \Pi y_{t-1} + \varepsilon_t,$$

three cases matter:

1. $\text{rank}(\Pi) = 0$: then $\Pi = 0$ and the system behaves like a pure multivariate random walk. There is **no cointegration**.
2. $\text{rank}(\Pi) = n$: then the level vector is already stationary. There is **no unit root problem**.
3. $0 < \text{rank}(\Pi) = r < n$: then the system is nonstationary, but there exist r stationary long-run relations. This is the **cointegrated case**.

Equivalent factorization

When $0 < r < n$, we can write

$$\Pi = \alpha\beta',$$

with α and β both $(n \times r)$ and full column rank.

Definition of cointegration

Definition

Suppose $y_t \in \mathbb{R}^n$ is a vector of unit-root, or $I(1)$, processes. If there exists a nonzero vector $a \in \mathbb{R}^n$ such that

$$a' y_t$$

is stationary, or $I(0)$, then the components of y_t are said to be **cointegrated**. The vector a is a **cointegrating vector**.

- Individual levels drift.
- But certain economically meaningful spreads, gaps, or ratios do not drift without bound.

Economic language

Cointegration means long-run equilibrium with short-run deviations.

Bivariate normalization

In the simplest two-variable case, take

$$\mathrm{y}_t = (y_{1t}, y_{2t})', \qquad a = (1, -\beta)'.$$

Then cointegration means

$$u_t = y_{1t} - \beta y_{2t}$$

is stationary, so the long-run relation can be written as

$$y_{1t} = \beta y_{2t} + u_t.$$

Interpretation

- y_{1t} and y_{2t} may each be $I(1)$.
- Their spread u_t is $I(0)$.
- Therefore the two series may drift, but not drift apart indefinitely.

OLS cointegrating regression and superconsistency

Regressing y_{1t} on y_{2t} gives

$$\widehat{\beta} = \left(\sum_{t=1}^{T} y_{2t}^2 \right)^{-1} \left(\sum_{t=1}^{T} y_{2t} y_{1t} \right).$$

Even if y_{2t} is endogenous relative to the stationary error, the estimator is not merely consistent; it is **superconsistent**.

If $y_{2t} = y_{2,t-1} + \varepsilon_t$ and u_t is stationary, the textbook gives

$$\frac{1}{T^2} \sum_{t=1}^{T} y_{2t}^2 \Longrightarrow \int B_\varepsilon(s)^2 \, ds, \qquad \frac{1}{T} \sum_{t=1}^{T} y_{2t} u_t \Longrightarrow \int B_\varepsilon(s) \, dB_u(s),$$

so that

$$T(\widehat{\beta} - \beta) \Longrightarrow \frac{\int B_\varepsilon(s) \, dB_u(s)}{\int B_\varepsilon(s)^2 \, ds}.$$

Important warning

Superconsistency does *not* imply ordinary t-tests are valid. The limit law is nonstandard.

Why cointegration is stronger than correlation

Correlation only says that variables move together contemporaneously.

Cointegration says something stronger:

- the variables share a common stochastic trend,
- deviations from a particular linear relation are mean-reverting,
- and there is a genuine distinction between short-run disequilibrium and long-run equilibrium.

Examples from economics and finance

- consumption and income,
- money demand and income / interest rates,
- purchasing power parity variables,
- pairs trading based on linked asset prices.

Granger representation theorem: why the theory matters

A central theorem says that if a VAR has unit roots and reduced-rank long-run matrix, then the process admits a representation with:

1. a nonstationary common-trend component,
2. a stationary transitory component,
3. and a finite-dimensional error-correction form.

Assumptions in words

- characteristic roots are either outside the unit circle or equal to one;
- the long-run matrix has reduced rank $r < n$;
- the cointegrating restrictions are non-redundant.

Granger representation theorem: a useful formula

The theorem in the textbook writes the solution as

$$y_t = C\sum_{s=1}^{t}\varepsilon_s + (I_n - C)\sum_{i=0}^{\infty}(I_n + \alpha\beta')^i\varepsilon_{t-i} + Cy_0,$$

where

$$C = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp.$$

- The term

  $$C\sum_{s=1}^{t}\varepsilon_s$$

  is the **nonstationary common-trend part**.
- The infinite sum involving $(I_n + \alpha\beta')^i$ is the **stationary transitory part**.

Practical implication

Cointegration is exactly what justifies an **error-correction** model.

How many common stochastic trends are there?

If y_t has dimension n and cointegrating rank r, then:

$$\text{number of cointegrating relations} = r, \qquad \text{number of common stochastic trends} = n - r.$$

Interpretation

- $\beta' y_t$ is stationary: these are the long-run equilibria.
- $\beta'_\perp y_t$ accumulates permanent shocks: these are the common trends.

Economic reading

A high-dimensional macro system can often be driven by a smaller number of permanent shocks plus a set of equilibrium-correction forces.

Residual-based cointegration testing: Engle-Granger logic

To test the null of no cointegration, estimate the static regression

$$y_{1t} = \delta + \pi' \mathrm{y}_{2t} + \varepsilon_t,$$

where y_{1t} is one variable and y_{2t} contains the others.

Under the null of no cointegration:

- the regression is spurious in levels,
- the residual $\widehat{\varepsilon}_t$ still contains a stochastic trend.

Under the alternative:

- the residual is stationary,
- because the estimated linear combination approximates a cointegrating relation.

ADF on the residuals

Let x_t denote the OLS residual from the cointegrating regression. Then test

$$\Delta x_t = (\theta - 1)x_{t-1} + c_1 \Delta x_{t-1} + \cdots + c_p \Delta x_{t-p} + \text{error}.$$

- This looks like an ADF regression.
- But the critical values are **not** the ordinary Dickey-Fuller ones, because the residual is generated by first-stage estimation.
- The asymptotic distribution also depends on the dimension of the system and on deterministic terms.

Key message

Residual-based unit-root testing is valid only with **cointegration-specific critical values**.

A small textbook table of residual-ADF critical values

For the regression

$$y_{1t} = \delta + \pi' \mathbf{y}_{2t} + \varepsilon_t$$

with no drift in the regressors, the textbook reports the following asymptotic critical values for the ADF statistic on the residuals:

Number of regressors	1%	2.5%	5%	10%
1	-3.96	-3.64	-3.37	-3.07
2	-4.31	-4.02	-3.77	-3.45
3	-4.73	-4.37	-4.11	-3.83
4	-5.07	-4.71	-4.45	-4.16
5	-5.28	-4.98	-4.71	-4.43

Interpretation

As the number of regressors grows, the critical values become more negative. So using ordinary ADF cutoffs would be wrong.

Limitations of the Engle-Granger approach

Residual-based testing is simple and intuitive, but it has important limitations:

- it is essentially a single-equation method;
- it depends on normalization, that is, on which variable is placed on the left-hand side;
- it is not ideal when there may be several cointegrating relations;
- it separates long-run estimation from short-run dynamics.

Why this matters

In a larger system, we typically prefer a **system method** that estimates the cointegration rank and the adjustment structure jointly. That leads to the Johansen framework.

Phillips triangular system

A useful theoretical device is the triangular system

$$\mathrm{y}_t = \begin{bmatrix} \mathrm{y}_{1t} \\ \mathrm{y}_{2t} \end{bmatrix}, \qquad \mathrm{y}_{1t} = B\mathrm{y}_{2t} + \mathrm{u}_{1t}, \qquad \mathrm{y}_{2t} = \mathrm{y}_{2,t-1} + \mathrm{u}_{2t},$$

where $\mathrm{u}_t = (\mathrm{u}'_{1t}, \mathrm{u}'_{2t})'$ is stationary and weakly dependent.

Meaning

- y_{2t} carries the stochastic trends.
- $\mathrm{y}_{1t} - B\mathrm{y}_{2t}$ is stationary.
- The matrix B contains the long-run equilibrium coefficients.

Error-correction form of the triangular system

The same system can be written in ECM form:

$$\Delta \mathrm{y}_t = -\begin{bmatrix} I_{n_1} & B \\ 0 & I_{n_2} \end{bmatrix} \Pi \mathrm{y}_{t-1} + \mathrm{v}_t, \qquad \Pi = (I, -B).$$

Economic reading

The term $\Pi \mathrm{y}_{t-1}$ is the lagged disequilibrium. If the variables drift away from their long-run relation, the minus sign implies that subsequent changes push them back toward equilibrium.

This is the main idea of error correction

Changes today depend on yesterday's equilibrium error.

Conditional mean representation in the triangular system

Under Gaussian shocks, the textbook writes the conditional mean of y_{1t} given y_{2t} as

$$y_{1t} = By_{2,t-1} + C\Delta y_{2,t} + v_{1\cdot 2,t}, \qquad C = \Omega_{12}\Omega_{22}^{-1}.$$

- $By_{2,t-1}$ is the long-run component.
- $C\Delta y_{2,t}$ captures short-run comovement in changes.
- $v_{1\cdot 2,t}$ is the orthogonalized innovation.

Interpretation

The model separates **equilibrium levels** from **short-run adjustment in differences**.

Why the cointegrating estimator is superconsistent but nonstandard

The triangular-system asymptotics produce

$$T(\widehat{B} - B) \Longrightarrow \frac{\int_0^1 dS_{1\cdot 2} S_2'}{\left(\int_0^1 S_2 S_2'\right)^{-1}},$$

a mixed normal limit rather than the standard stationary Gaussian limit.

- This is one reason the long-run coefficients can be estimated very precisely.
- But standard stationary inference needs modification.
- The special asymptotics are driven by the integrated regressors.

From VAR(p) to VECM

Start with the level VAR

$$y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \varepsilon_t.$$

It can be rewritten as

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + \Pi y_{t-p} + \varepsilon_t,$$

where the textbook uses

$$\Gamma_i = \Phi_1 + \cdots + \Phi_i - I_n, \qquad \Pi = \Phi_1 + \cdots + \Phi_p - I_n.$$

Equivalent common notation

Many texts write the disequilibrium term as Πy_{t-1} instead of Πy_{t-p} after re-indexing. The econometric idea is the same.

Why the VECM is the right compromise

A VECM combines both pieces of information:

$$\Delta y_t = \underbrace{\Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1}}_{\text{short-run dynamics}} + \underbrace{\Pi y_{t-1}}_{\text{long-run disequilibrium}} + \varepsilon_t .$$

- Differenced terms remove unit roots from the short-run dynamics.
- The level term keeps the long-run equilibrium restriction alive.

What would go wrong otherwise?

A VAR in differences throws away the long-run equilibrium. A VAR in levels ignores the nonstationary nature of the data. The VECM is the natural middle ground.

Factorizing the long-run matrix

When the rank is r with $0 < r < n$, we write

$$\Pi = \alpha\beta',$$

where:

- β is the $(n \times r)$ matrix of **cointegrating vectors**;
- α is the $(n \times r)$ matrix of **loading coefficients**.

Then the VECM becomes

$$\Delta \mathrm{y}_t = \Gamma_1 \Delta \mathrm{y}_{t-1} + \cdots + \Gamma_{p-1} \Delta \mathrm{y}_{t-p+1} + \alpha\beta' \mathrm{y}_{t-1} + \varepsilon_t.$$

Separation of roles

$\beta' \mathrm{y}_{t-1}$ measures yesterday's deviation from equilibrium; α tells us which variables respond to that deviation and how strongly.

Economic meaning of the cointegrating vectors β

Each column of β defines one stationary long-run relation:

$$\beta_j' \mathrm{y}_t \sim I(0), \qquad j = 1, \ldots, r.$$

Examples:

- consumption relative to disposable income,
- real money balances relative to income and interest rates,
- price differentials under purchasing power parity.

Normalization issue

Only the *span* of β is identified. Multiplying β by a nonsingular $r \times r$ matrix and adjusting α accordingly leaves $\Pi = \alpha\beta'$ unchanged.

Economic meaning of the loading matrix α

The rows of α show which equations adjust to restore equilibrium.

For a scalar cointegrating relation ($r = 1$), equation i looks like

$$\Delta y_{it} = \cdots + \alpha_i(\beta' \mathrm{y}_{t-1}) + \varepsilon_{it}.$$

- If $\alpha_i = 0$, variable i does not respond directly to the disequilibrium error.
- If $\alpha_i < 0$ and the equilibrium error is positive, variable i tends to fall, helping to restore equilibrium.
- Large $|\alpha_i|$ means faster speed of adjustment.

Applied language

The loading coefficients measure the **speed of error correction**.

Single-equation intuition: the ECM form

A simple scalar error-correction equation is

$$\Delta y_t = \gamma_0 + \gamma_1 \Delta x_t + \gamma_2 \Delta y_{t-1} + \lambda(y_{t-1} - \beta x_{t-1}) + u_t.$$

- Δx_t and Δy_{t-1} capture short-run effects.
- $y_{t-1} - \beta x_{t-1}$ is the lagged equilibrium error.
- λ governs how fast the variable corrects the disequilibrium.

Interpretation

If the system was above equilibrium last period, the sign of λ tells us whether the current change moves it back.

Weak exogeneity inside the VECM

Suppose one variable has an insignificant adjustment coefficient but another has a significant one.

Interpretation

- The non-adjusting variable can often be treated as **weakly exogenous** for the long-run parameters.
- The adjusting variable bears the burden of restoring the equilibrium relation.

Typical macro example

Consumption and income may be cointegrated, but the data may suggest that income adjusts to restore the relation while consumption behaves more like the driving trend variable.

Deterministic terms in cointegrated systems

In practice we must decide where deterministic components enter:

- unrestricted intercept outside the cointegration space;
- restricted intercept inside the cointegration relation;
- restricted or unrestricted trends.

Why this matters

The deterministic specification changes:

1. the asymptotic critical values of rank tests,
2. the interpretation of the equilibrium relation,
3. and whether the long-run relation includes a constant or a trend.

Rule of thumb

Choose deterministic terms from economic logic and the appearance of the data, not by default.

Forecasting and impulse responses in cointegrated systems

A cointegrated VECM still implies a VAR representation in levels, but with reduced-rank long-run structure.

- Forecasts must respect the equilibrium relation.
- Impulse responses should be interpreted relative to both short-run changes and long-run restrictions.
- In software, one common practice is to estimate the Johansen object and then convert it to a level VAR representation for IRF computation.

R practice

After `ca.jo(...)` and choosing rank r, a convenient workflow is

```
vec2var(joh_object, r = r)
```

followed by `irf(...)`.

Why Johansen's method is the natural system approach

Johansen's framework tests and estimates cointegration *inside the full VAR system*.

- It does not require us to guess the normalization in advance.
- It can detect more than one cointegrating relation.
- It estimates the long-run and short-run structure jointly.

Set-up

Begin with the VAR(p) in levels and write it in error-correction form. The null concerns the rank of Π.

Residual matrices behind the Johansen test

The textbook defines:

- R_{0t} as the residuals from regressing Δy_t on lagged differences,
- R_{kt} as the residuals from regressing the lagged levels term on the same lagged differences.

Then form the sample covariance blocks

$$S_{00} = \frac{1}{T}\sum_{t=1}^{T} R_{0t}R_{0t}', \quad S_{0k} = \frac{1}{T}\sum_{t=1}^{T} R_{0t}R_{kt}',$$

$$S_{k0} = \frac{1}{T}\sum_{t=1}^{T} R_{kt}R_{0t}', \quad S_{kk} = \frac{1}{T}\sum_{t=1}^{T} R_{kt}R_{kt}'.$$

Interpretation

This partials out the short-run dynamics so the rank test focuses on the long-run relation in the levels.

Johansen eigenvalue problem

The canonical correlations $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_n$ are obtained from

$$\left|\lambda S_{kk} - S_{k0} S_{00}^{-1} S_{0k}\right| = 0.$$

- Large eigenvalues indicate stronger long-run relations.
- Small eigenvalues indicate weak or absent equilibrium restrictions.

Rank logic

The number of statistically important eigenvalues is the estimated cointegration rank.

Johansen trace statistic

To test

$$H_0: \ \text{rank}(\Pi) \leq r \qquad \text{against} \qquad H_1: \ \text{rank}(\Pi) > r,$$

use the trace statistic

$$\mathcal{LR}_{\text{trace}} = -T \sum_{i=r+1}^{n} \log(1 - \widehat{\lambda}_i).$$

- Start with $r = 0$.
- If the null is rejected, move to $r = 1$, then $r = 2$, and so on.

Decision rule

The first non-rejected rank often becomes the working estimate of the cointegration rank.

Johansen maximum-eigenvalue statistic

To test

$$H_0: \ \mathrm{rank}(\Pi) = r \qquad \text{against} \qquad H_1: \ \mathrm{rank}(\Pi) = r + 1,$$

use

$$\mathcal{LR}_{\max} = -T \log(1 - \widehat{\lambda}_{r+1}).$$

Difference from the trace test

- The trace test accumulates the evidence in all remaining eigenvalues.
- The maximum-eigenvalue test focuses only on the next candidate relation.

What the rank outcomes mean

Rank outcome	Interpretation
$r = 0$	no cointegration, work with a VAR in differences
$0 < r < n$	cointegrated system, estimate a VECM
$r = n$	all variables effectively stationary in levels

Important applied point

Rank selection is not purely mechanical. It should agree with unit-root evidence, theory, deterministic terms, and sample size.

Normalization after rank selection

Once r is chosen, the cointegration space is estimated, but the matrix β is still only identified up to nonsingular linear transformations.

- In practice, we normalize one coefficient to 1.
- For example, with $(\log C_t, \log Y_t^d)'$, a common normalization is

$$\beta' \mathrm{y}_t = \log C_t - b \log Y_t^d.$$

Interpretation

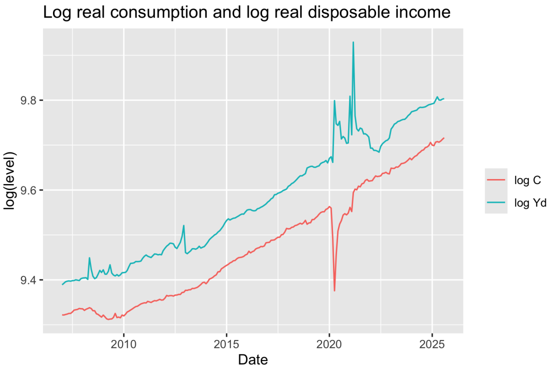
The normalization is not an econometric truth; it is a convenient way of expressing the same equilibrium space.

A textbook case study: consumption and income

The chapter illustrates cointegration using monthly FRED data on:

- real personal consumption expenditures, and
- real disposable personal income.

After taking logarithms, the two series move closely together over time.



Why this example is pedagogically useful

The pair $(\log C_t, \log Y_t^d)$ is useful because:

- each series is strongly persistent in levels;
- theory suggests a stable long-run relation;
- the equilibrium-error interpretation is economically natural;
- the example demonstrates how univariate unit-root testing and multivariate system testing fit together.

Economic intuition

Consumption and disposable income can drift over time with growth, inflation adjustment, and macro shocks, but the gap between them should not explode forever.

R workflow, part 1: get the data and build the time series

```
library(fredr); library(urca)
library(vars); library(dplyr)

fredr_set_key("YOUR_KEY")

cons_raw <- fredr(series_id = "PCEC96",
                observation_start = as.Date("1960-01-01"))
inc_raw <- fredr(series_id = "DSPIC96",
                observation_start = as.Date("1960-01-01"))

df <- inner_join(
  cons_raw %>% select(date, cons = value),
  inc_raw %>% select(date, inc = value),
  by = "date"
) %>%
  mutate(lcons = log(cons), linc = log(inc)) %>%
  filter(!is.na(lcons), !is.na(linc))

y_ts <- ts(df[, c("lcons", "linc")], start = c(1960, 1),
          frequency = 12)
```

Purpose

Create the level vector $(\log C_t, \log Y_t^d)'$ in a monthly ts object for the cointegration tests and VECM estimation.

R workflow, part 2: verify that the variables look $I(1)$

```
adf_lcons <- ur.df(y_ts[, "lcons"], type = "trend",
                  lags = 12, selectlags = "AIC")
adf_linc <- ur.df(y_ts[, "linc"], type = "trend",
                  lags = 12, selectlags = "AIC")

dlcons <- diff(y_ts[, "lcons"])
dlinc <- diff(y_ts[, "linc"])

adf_dlcons <- ur.df(dlcons, type = "drift",
                   lags = 12, selectlags = "AIC")
adf_dlinc <- ur.df(dlinc, type = "drift",
                   lags = 12, selectlags = "AIC")

summary(adf_lcons)
summary(adf_linc)
summary(adf_dlcons)
summary(adf_dlinc)
```

Logic

First diagnose the integration order. Cointegration testing is meaningful only after the level variables look plausibly $I(1)$.

R workflow, part 3: Engle-Granger and Phillips-Ouliaris

```
eg_reg <- lm(lcons ~ linc, data = df)
eg_res <- resid(eg_reg)

eg_adf <- ur.df(eg_res, type = "none",
               lags = 12, selectlags = "AIC")
summary(eg_adf)

po_test <- ca.po(z = df[, c("lcons", "linc")],
               demean = "constant",
               lag = "long", type = "Pz")
summary(po_test)
```

What these commands do

- Estimate the long-run regression in levels.
- Extract the equilibrium error.
- Test whether that residual is stationary.

R workflow, part 4: Johansen test and VECM estimation

```
lag_sel <- VARselect(y_ts, lag.max = 12, type = "trend")
p_opt <- as.numeric(lag_sel$selection["SC(n)"])

joh_trace <- ca.jo(y_ts, type = "trace",
                  ecdet = "trend", K = p_opt,
                  spec = "transitory")
summary(joh_trace)

joh_eigen <- ca.jo(y_ts, type = "eigen",
                  ecdet = "trend", K = p_opt,
                  spec = "transitory")
summary(joh_eigen)

vecm_fit <- cajorls(joh_trace, r = 1)
summary(vecm_fit$rlm)
```

Interpretation

Use `VARselect` for the lag order, `ca.jo` for the rank tests, and `cajorls` for the error-correction equations.

R workflow, part 5: from VECM to IRFs

```r
# Convert the Johansen object to a level VAR representation
vec_as_var <- vec2var(joh_trace, r = 1)

# Compute impulse responses
irf_obj <- irf(vec_as_var,
               impulse = "linc",
               response = "lcons",
               n.ahead = 24,
               ortho = TRUE,
               boot = TRUE)

plot(irf_obj)
```

Why this step matters

Lecture 7 taught us how to interpret IRFs in stationary systems. In a cointegrated system, the same dynamic questions remain relevant, but the estimated VAR must respect the equilibrium rank.

Reported textbook results for the consumption-income example

The chapter reports a coherent empirical picture:

- ADF tests on the log levels are close to the nonstationary boundary, while first differences reject strongly, so both series behave empirically as $I(1)$.
- The Engle-Granger regression gives a slope around 0.895 and an R^2 around 0.92.
- The ADF statistic on the residual is about -3.19, supporting residual stationarity.
- The Phillips-Ouliaris P_z statistic is also strongly supportive of cointegration.
- Johansen trace testing indicates a cointegration rank of $r = 1$.

Economic reading

The data are consistent with one stable long-run relation between log consumption and log disposable income.

Adjustment coefficients in the textbook example

The chapter also reports that:

- the loading coefficient in the income equation is significantly negative, around -0.06;
- the loading coefficient in the consumption equation is insignificant.

Interpretation

- income appears to adjust to restore the equilibrium relation;
- consumption behaves more like a weakly exogenous driver of the long-run relation.

Pedagogical lesson

Cointegration is not just about whether an equilibrium exists. It is also about *who adjusts* when that equilibrium is disturbed.

How to interpret VECM output in practice

When students read a VECM printout, they should look in this order:

1. Are the variables plausibly $I(1)$?
2. What lag order was selected, and by what criterion?
3. What deterministic specification was chosen?
4. What rank is suggested by the trace and maximum-eigenvalue tests?
5. How is the cointegrating vector normalized?
6. Which adjustment coefficients are significant, and with what sign?
7. Do the residual diagnostics suggest a sensible dynamic specification?

IRFs in a cointegrated system: what changes conceptually?

The idea of an impulse response remains:

response of variable i at horizon h to a shock in variable j.

But now the dynamics contain both:

- **short-run propagation through differences and lagged changes**, and
- **long-run equilibrium restoration through the error-correction term**.

Interpretive point

A shock may move the system away from equilibrium at first, but the VECM embeds forces that gradually pull it back.

Common pitfalls in empirical cointegration work

- Testing for cointegration before verifying that the variables are plausibly $I(1)$.
- Mixing incompatible deterministic specifications across unit-root and rank tests.
- Over-interpreting one normalization of β as if it were uniquely meaningful.
- Ignoring the possibility of more than one cointegrating relation in larger systems.
- Using a VAR in differences and then talking about long-run equilibrium.
- Using a VAR in levels and ignoring the nonstationary nature of the data.

Good practice

Let the economics suggest the equilibrium relation, let the data suggest the rank, and let the VECM connect the long run to the short run.

Summary

- Nonstationary variables can still be tied together by stationary linear combinations.
- The rank of Π in the differenced VAR determines whether the system is random-walk-like, stationary, or cointegrated.
- Cointegration leads naturally to the VECM:

$$\Delta \mathrm{y}_t = \Gamma(L)\Delta \mathrm{y}_{t-1} + \alpha\beta'\mathrm{y}_{t-1} + \varepsilon_t.$$

- β contains the long-run equilibria; α contains the speeds of adjustment.
- Engle-Granger is simple but limited; Johansen is the full system method.
- In applied work, lag selection, deterministic terms, rank choice, normalization, and interpretation of loading coefficients all matter.

Preview of Lecture 9

Lecture 9 continues the Chapter 3 block and then transitions into Chapter 4:

- structural VAR identification through short-run, long-run, and A/B/AB restrictions;
- then a move from dynamic mean equations to dynamic variance equations;
- ARCH and GARCH models as the natural next step once we realize that shocks can be serially uncorrelated but their *volatility* can still be highly persistent.