Time Series Analysis and Forecasting

Chapter 7: Cointegration and VECM

Daniel Traian PELE

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

Learning Objectives

By the end of this chapter, you will be able to:

- Understand the problem of spurious regression with non-stationary data
- Test for cointegration using Engle-Granger and Johansen methods
- Estimate Vector Error Correction Models (VECM)
- Interpret error correction mechanisms and adjustment speeds

Outline

Why Cointegration Matters

The Challenge

- Many economic/financial time series are **non-stationary** (I(1))
- GDP, stock prices, exchange rates, interest rates all have unit roots
- Standard regression with I(1) variables $\Rightarrow$ **spurious results**
- Differencing removes non-stationarity but loses **long-run information**

The Solution: Cointegration

Some non-stationary series share a **common stochastic trend**—they move together in the long run.

Nobel Prize 2003

Granger & Engle received the Nobel Prize for "methods for analyzing economic time series with common trends."

Real-World Applications

Finance

- **Pairs Trading**: Cointegrated stocks
- **Term Structure**: Interest rates
- **Spot-Futures**: Arbitrage

Macroeconomics

- **Consumption & Income**
- **Money & Prices**
- **PPP**: Exchange rates

Policy Analysis

- **Fiscal**: Spending & taxes
- **Monetary**: Rate pass-through
- **Labor**: Wages & productivity

The Spurious Regression Problem

Granger & Newbold (1974)

Regressing one random walk on another **independent** random walk: $Y_t = \alpha + \beta X_t + u_t$ where Y_t and X_t are independent I(1) processes.

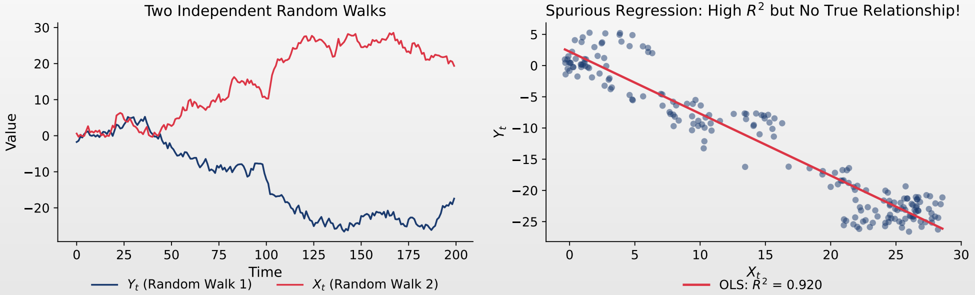
Symptoms of Spurious Regression

- High R^2 (often > 0.9) even though variables are **unrelated**
- Highly significant t-statistics (reject $H_0 : \beta = 0$)
- Very low Durbin-Watson statistic ($DW \approx 0$)
- Residuals are non-stationary (have unit root)

Rule of Thumb

If $R^2 > DW$, suspect spurious regression!

Spurious Regression: Visual Example



Warning: Two independent random walks show high correlation ($R^2 > 0.8$) by chance! This is why we need cointegration analysis.

TSA_ch7_spurious_regression

Spurious Correlations in the Real World

Data Mining Can Produce Meaningless Correlations

With enough variables and long time series, purely coincidental patterns emerge:

- Distance between Neptune and Uranus ↔ SAP SE stock price (2002–2023)
- GMO corn use in South Dakota ↔ Google searches for "i cant even" (2004–2023)
- *Two and a Half Men* season ratings ↔ Jet fuel used in Serbia (2006–2015)
- "Its Wednesday my dudes" meme popularity ↔ Boeing stock price (2006–2023)

Lesson

High correlation ≠ causation. Non-stationary series with common trends produce high R^2 by construction. Always test for stationarity and cointegration before interpreting regression results!

Explore more examples: `tylervigen.com/spurious-correlations`

Definition of Cointegration

Definition 1 (Cointegration (Engle & Granger, 1987))

Variables $Y_{1t}, Y_{2t}, \ldots, Y_{kt}$ are **cointegrated of order** (d, b), written $CI(d, b)$, if:

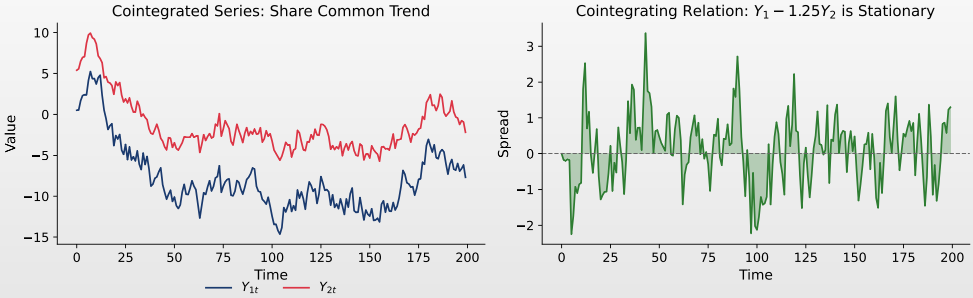
1. All variables are integrated of order d: $Y_{it} \sim I(d)$
2. There exists a linear combination $\boldsymbol{\beta}'\mathrm{Y}_t = \beta_1 Y_{1t} + \cdots + \beta_k Y_{kt}$ that is integrated of order $(d - b)$, where $b > 0$

Most Common Case: $CI(1, 1)$

- Variables are $I(1)$ (have unit roots)
- Linear combination is $I(0)$ (stationary)
- Vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is the **cointegrating vector**

The cointegrating vector is unique only up to scalar multiplication. Usually normalized: $\beta_{\mathbf{1}} = 1$.

Cointegration: Visual Example



Key insight: Both series are I(1) and trend together, but their linear combination (spread) is stationary—this is cointegration!

TSA_ch7_cointegrated_series

Intuition: Common Stochastic Trends

Why Does Cointegration Occur?

Cointegrated variables share **common stochastic trends**: $Y_{1t} = \gamma_1\tau_t + S_{1t}$, $Y_{2t} = \gamma_2\tau_t + S_{2t}$ where τ_t is a common random walk and S_{it} are stationary.

Linear Combination Eliminates the Trend

$\gamma_2 Y_{1t} - \gamma_1 Y_{2t} = \gamma_2 S_{1t} - \gamma_1 S_{2t} \sim I(0)$

Economic Interpretation

- Cointegration = **long-run equilibrium relationship**
- Variables may deviate in the short run, but are "pulled back"
- The cointegrating vector defines the equilibrium

Cointegrating Rank

How Many Cointegrating Relationships?

For k variables that are $I(1)$:

- Maximum possible cointegrating relationships: $r = k - 1$
- If $r = 0$: No cointegration (variables drift apart)
- If $r = k$: All variables are $I(0)$ (contradiction)

Example: 3 Variables

- $r = 0$: No cointegration
- $r = 1$: One cointegrating relationship
- $r = 2$: Two cointegrating relationships (only 1 common trend)

The number of common stochastic trends $= k - r$

Engle-Granger Two-Step Method

Step 1: Estimate Cointegrating Regression

Run OLS: $Y_t = \alpha + \beta X_t + e_t$. Save residuals: $\hat{e}_t = Y_t - \hat{\alpha} - \hat{\beta} X_t$

Step 2: Test Residuals for Stationarity

Test if $\hat{e}_t$ is $I(0)$ using ADF: $\Delta \hat{e}_t = \rho \hat{e}_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta \hat{e}_{t-j} + v_t$

- H_0: $\rho = 0$ (unit root $\Rightarrow$ no cointegration)
- H_1: $\rho < 0$ (stationary $\Rightarrow$ cointegration)

Important

Use **Engle-Granger critical values**, not standard ADF! (More negative because residuals are estimated)

Engle-Granger Critical Values

Critical Values for Cointegration Test

Variables	1%	5%	10%
2	−3.90	−3.34	−3.04
3	−4.29	−3.74	−3.45
4	−4.64	−4.10	−3.81
5	−4.96	−4.42	−4.13

MacKinnon (1991), $T = 100$

Limitations of Engle-Granger

- Only tests for **one** cointegrating relationship
- Results depend on choice of dependent variable
- Small sample bias; cannot test hypotheses on cointegrating vector

Johansen Cointegration Test

Advantages over Engle-Granger

- Tests for **multiple** cointegrating relationships
- Maximum likelihood estimation (more efficient)
- Can test restrictions on cointegrating vectors
- Does not require choosing a dependent variable

Starting Point: VAR in Levels

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t$$

Rewrite in **Vector Error Correction** form...

VECM Representation

Vector Error Correction Model

$\Delta Y_t = c + \mathbf{\Pi} Y_{t-1} + \sum_{j=1}^{p-1} \mathbf{\Gamma}_j \Delta Y_{t-j} + \varepsilon_t$

- $\mathbf{\Pi} = \sum_i A_i - I$ (long-run impact); $\mathbf{\Gamma}_j$ (short-run dynamics)

Key Insight: Rank of $\mathbf{\Pi}$

The **rank of $\mathbf{\Pi}$** determines cointegration:

- $\text{rank}(\mathbf{\Pi}) = 0$: No cointegration (VAR in differences)
- $\text{rank}(\mathbf{\Pi}) = k$: All variables are $I(0)$ (VAR in levels)
- $0 < \text{rank}(\mathbf{\Pi}) = r < k$: r cointegrating vectors

Derivation: From VAR to VECM

Starting Point: VAR(p) in Levels

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t$$

Step 1: Subtract Y_{t-1} from Both Sides

$$Y_t - Y_{t-1} = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} - Y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = (A_1 - I) Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t$$

Goal

Rewrite so that all terms are either in **levels** (Y_{t-1}) or **differences** (ΔY_{t-j}).

Derivation: From VAR to VECM (cont.)

Step 2: Add and Subtract Terms Strategically

Add $A_2 Y_{t-1}$ and subtract $A_2 Y_{t-1}$:

$$\Delta Y_t = (A_1 + A_2 - I)Y_{t-1} - A_2(Y_{t-1} - Y_{t-2}) + A_3 Y_{t-3} + \cdots + \varepsilon_t$$

Continue adding $A_3 Y_{t-1}$, etc., until all lagged **levels** are collected in one term.

Step 3: General Pattern

After algebraic manipulation, we obtain:

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \varepsilon_t$$

The Key Matrices

$$\boxed{\Pi = \sum_{i=1}^{p} A_i - I = -(I - A_1 - A_2 - \cdots - A_p)}$$

Derivation: Verifying the $\boldsymbol{\Gamma}_j$ Formula

Example: VAR(2)

Starting from: $Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \varepsilon_t$

Subtract Y_{t-1}:

$$\Delta Y_t = (A_1 - I)Y_{t-1} + A_2 Y_{t-2} + \varepsilon_t$$

Add and subtract $A_2 Y_{t-1}$:

$$\Delta Y_t = (A_1 + A_2 - I)Y_{t-1} + A_2(Y_{t-2} - Y_{t-1}) + \varepsilon_t$$

$$\Delta Y_t = \underbrace{(A_1 + A_2 - I)}_{\boldsymbol{\Pi}} Y_{t-1} \underbrace{-A_2}_{\boldsymbol{\Gamma}_1} \Delta Y_{t-1} + \varepsilon_t$$

Verification

For VAR(2): $\boldsymbol{\Pi} = A_1 + A_2 - I$ and $\boldsymbol{\Gamma}_1 = -A_2$

Using our formula: $\boldsymbol{\Gamma}_1 = -\sum_{i=2}^{2} A_i = -A_2 \quad \checkmark$

Economic Interpretation of Error Correction

The VECM with Cointegration

When $\text{rank}(\mathbf{\Pi}) = r$, we write $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$:

$$\Delta Y_t = \boldsymbol{\alpha} \underbrace{(\boldsymbol{\beta}' Y_{t-1})}_{\text{equilibrium error}} + \sum_{j=1}^{p-1} \mathbf{\Gamma}_j \Delta Y_{t-j} + \varepsilon_t$$

Economic Interpretation

- $\boldsymbol{\beta}' Y_{t-1}$ = **equilibrium error**: deviation from long-run relationship
- $\boldsymbol{\alpha}$ = **adjustment speeds**: how fast variables correct deviations
- $\mathbf{\Gamma}_j$ = **short-run dynamics**: transitory effects

Error Correction Mechanism

If $\boldsymbol{\beta}' Y_{t-1} > 0$ (above equilibrium) and $\alpha_i < 0$, then ΔY_{it} decreases.

The system self-corrects toward equilibrium!

Decomposition of $\mathbf{\Pi}$

When rank($\mathbf{\Pi}$) $= r < k$

$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ where $\boldsymbol{\beta}$ $(k \times r)$ = **cointegrating vectors**, $\boldsymbol{\alpha}$ $(k \times r)$ = **adjustment coefficients**

Interpretation

- $\boldsymbol{\beta}'\mathbf{Y}_{t-1}$ = deviations from equilibrium (error correction terms)
- $\boldsymbol{\alpha}$ = speed of adjustment; rows show each variable's response

VECM: $\Delta\mathbf{Y}_t = \mathbf{c} + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{Y}_{t-1}) + \sum_{j=1}^{p-1} \mathbf{\Gamma}_j \Delta\mathbf{Y}_{t-j} + \boldsymbol{\varepsilon}_t$

Johansen Test Statistics

Two Test Statistics

Based on eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_k$ of a certain matrix:

Trace Test:

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^{k} \ln(1 - \hat{\lambda}_i)$$

Tests H_0: rank $\leq r$ vs H_1: rank $> r$

Maximum Eigenvalue Test:

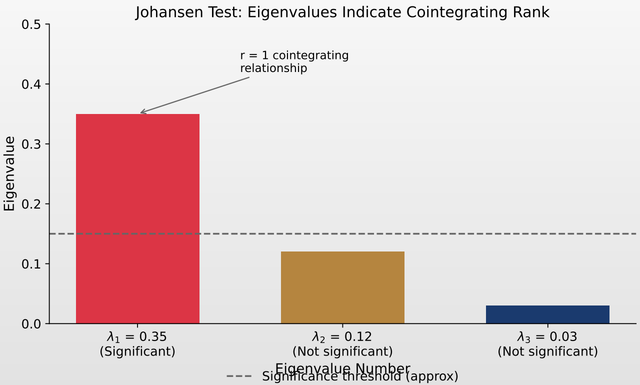
$$\lambda_{\max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1})$$

Tests H_0: rank $= r$ vs H_1: rank $= r + 1$

Critical values from Johansen & Juselius (1990), depend on:

- Number of variables k
- Deterministic components (constant, trend)

Johansen Test: Visual Interpretation



Significant eigenvalues (above threshold) indicate cointegrating relationships. First eigenvalue significant $\Rightarrow r = 1$.

TSA_ch7_johansen_eigenvalues

Testing Procedure

Sequential Testing (Trace Test)

1. Test H_0: $r = 0$. If rejected $\Rightarrow$ continue
2. Test H_0: $r \leq 1$. If not rejected $\Rightarrow r = 1$
3. Continue until H_0 is not rejected

Deterministic Components

- No constant, no trend (rarely used)
- Constant in cointegrating relation only
- **Constant in both** (most common)
- Constant + trend in cointegrating relation

VECM Structure

Full VECM Specification

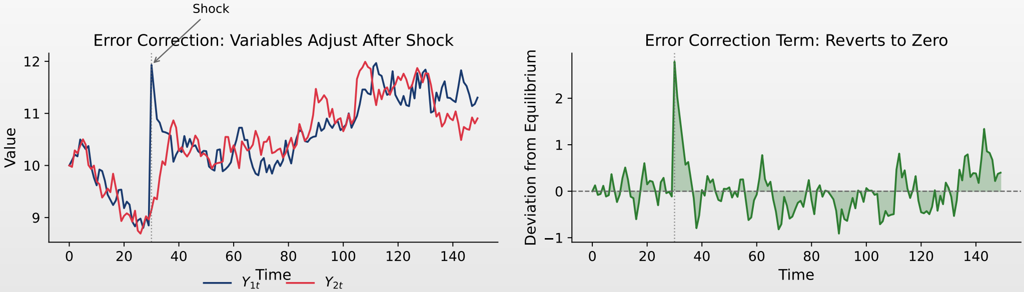
For $k = 2$ variables with $r = 1$ cointegrating relation:

$$\Delta Y_{1t} = c_1 + \alpha_1(Y_{1,t-1} - \beta Y_{2,t-1}) + \gamma_{11}\Delta Y_{1,t-1} + \gamma_{12}\Delta Y_{2,t-1} + \varepsilon_{1t}$$
$$\Delta Y_{2t} = c_2 + \alpha_2(Y_{1,t-1} - \beta Y_{2,t-1}) + \gamma_{21}\Delta Y_{1,t-1} + \gamma_{22}\Delta Y_{2,t-1} + \varepsilon_{2t}$$

Components

- $(Y_{1,t-1} - \beta Y_{2,t-1})$ = error correction term (deviation from equilibrium)
- α_1, α_2 = adjustment speeds (should have opposite signs)
- γ_{ij} = short-run dynamics
- ε_{it} = innovations

Error Correction Mechanism: Visual



Error correction in action: When series deviate from equilibrium (shaded regions), the adjustment mechanism pulls them back. Positive deviations lead to downward adjustment, negative deviations lead to upward adjustment. TSA_ch7_error_correction

Interpreting Adjustment Coefficients

The α Coefficients

If the cointegrating relation is $Y_1 - \beta Y_2 = 0$ (equilibrium):

- $\alpha_1 < 0$: Y_1 adjusts downward when above equilibrium
- $\alpha_2 > 0$: Y_2 adjusts upward when Y_1 is above equilibrium

Weak Exogeneity

If $\alpha_i = 0$, variable Y_i does **not** respond to disequilibrium.

- Y_i is **weakly exogenous** for the long-run parameters
- The other variable does all the adjusting
- Can simplify estimation (single-equation approach)

Test weak exogeneity: $H_0 : \alpha_i = 0$ using likelihood ratio test.

VECM vs VAR in Differences

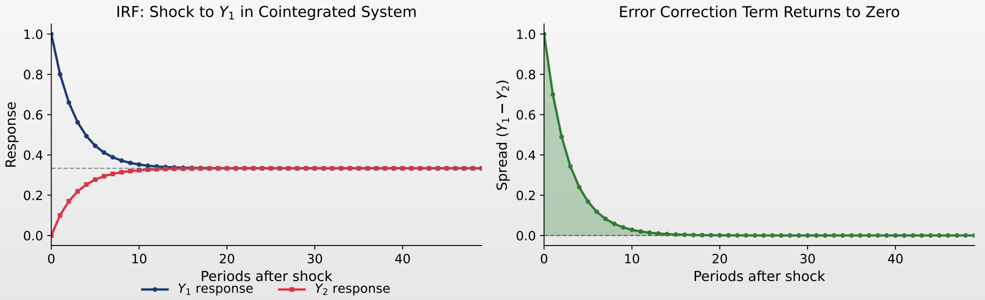
When Variables are Cointegrated

	VAR in Differences	VECM
Long-run info	Lost	Preserved
Short-run dynamics	Yes	Yes
Error correction	No	Yes
Forecasting	Poor (long-run)	Better
IRF interpretation	Short-run only	Both

Granger Representation Theorem

If variables are cointegrated, there **must** exist an error correction representation. Ignoring cointegration = model misspecification!

VECM Impulse Response Functions



IRF interpretation: In a cointegrated system, shocks have **permanent effects** on levels but the system returns to equilibrium. Unlike stationary VAR, effects don't decay to zero—they converge to a new long-run value.

TSA_ch7_vecm_irf

Practical Workflow

Step-by-Step Procedure

1. **Unit Root Tests**: Verify all variables are $I(1)$
   - ADF, KPSS on levels and first differences
2. **Lag Length Selection**: Choose p for VAR in levels
   - Use AIC, BIC, or sequential LR tests
3. **Cointegration Test**: Johansen trace/max-eigenvalue tests
   - Determine cointegrating rank r
4. **Estimate VECM**: If $0 < r < k$
   - Estimate $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}_j$
5. **Diagnostics**: Check residuals for autocorrelation, normality
6. **Analysis**: IRF, FEVD, hypothesis tests

Common Pitfalls

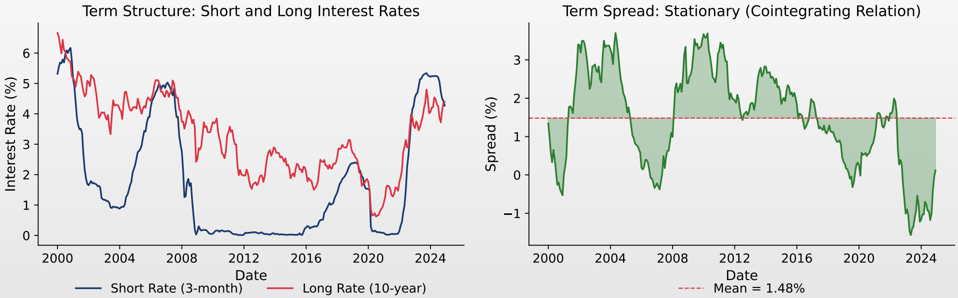
Things to Watch Out For

- **Structural breaks**: Cause spurious unit roots or cointegration
- **Near-unit-root**: Tests have low power
- **Lag selection**: Too many/few lags bias results
- **Small samples**: Johansen test oversized

Recommendation

Always check: residual diagnostics, stability of cointegrating relationship, sensitivity to specification

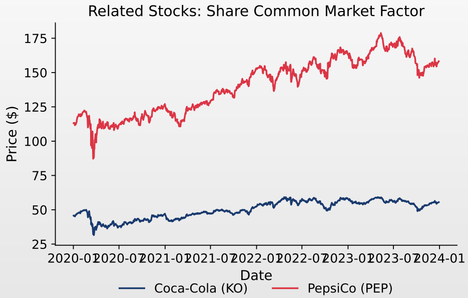
Example 1: Term Structure of Interest Rates

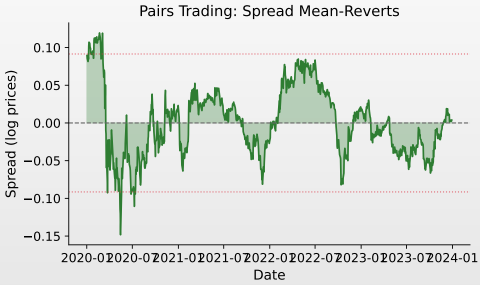


Expectations Hypothesis: Short and long rates share common trend. The spread (term premium) is stationary—evidence of cointegration!

TSA_ch7_interest_rates_coint

Example 2: Pairs Trading in Finance

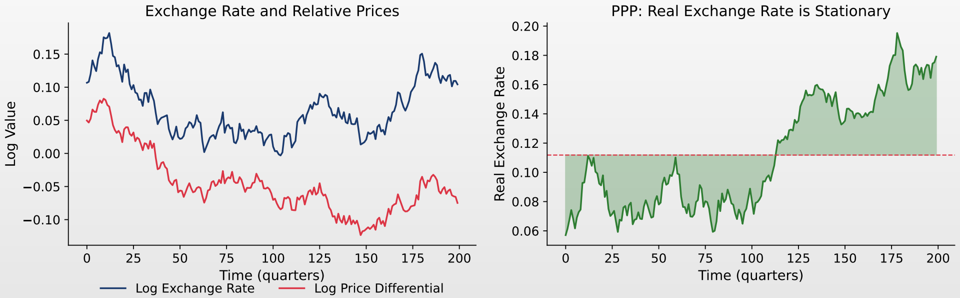




Strategy: Find cointegrated stock pairs (e.g., Coca-Cola & Pepsi). When spread deviates from mean, trade expecting mean reversion. Sell spread when high, buy when low.

TSA_ch7_pairs_trading

Example 3: Purchasing Power Parity (PPP)



PPP Theory: $e_t = p_t - p_t^*$ (log exchange rate equals price differential). Real exchange rate should be stationary in the long run.

TSA_ch7_ppp_cointegration

Case Study: Cointegration of Interest Rates

Research Question

Are short-term and long-term interest rates cointegrated? Does the expectations hypothesis of the term structure hold?

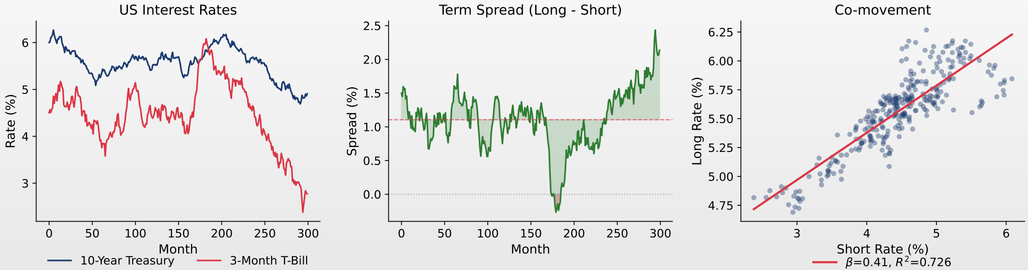
Data

- US Monthly Data (1962-2023)
- 3-Month Treasury Bill Rate
- 10-Year Treasury Bond Yield
- Source: FRED Database

Methodology

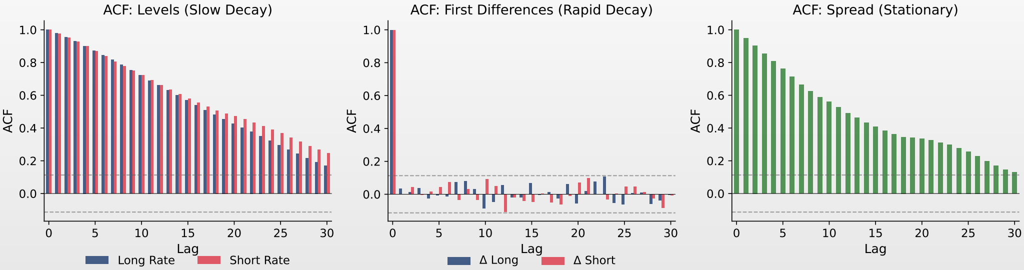
- Unit root tests (ADF, PP)
- Engle-Granger cointegration test
- Johansen procedure
- VECM estimation
- Impulse response analysis

Step 1: Data Visualization



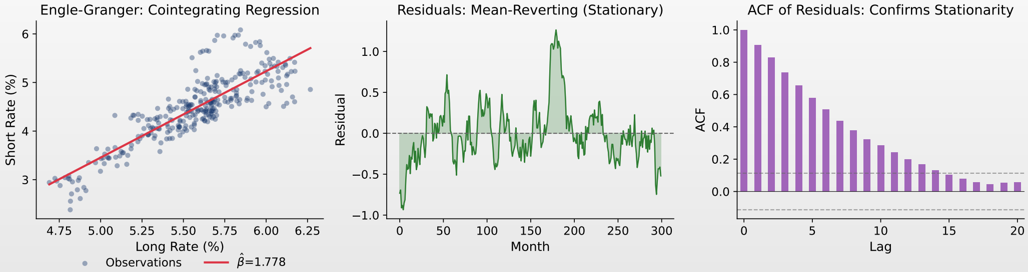
TSA_ch7_case_raw_data

Step 2: Unit Root Tests



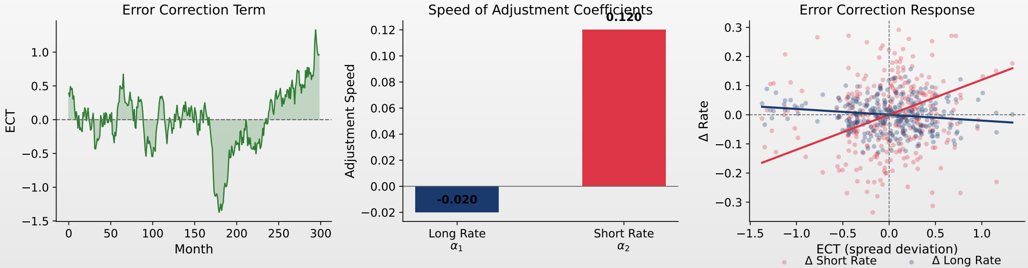
TSA_ch7_case_unit_root

Step 3: Engle-Granger Cointegration Test



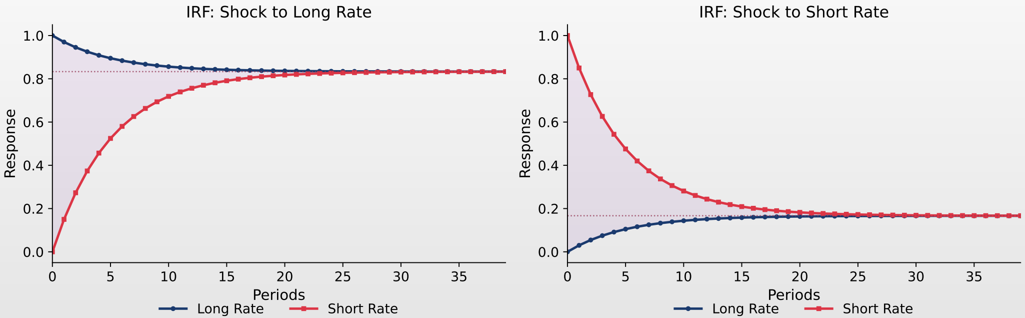
TSA_ch7_case_cointegration

Step 4: VECM Estimation



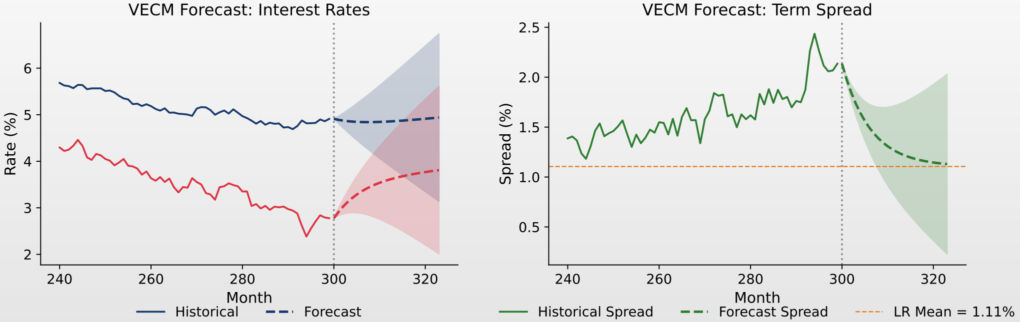
TSA_ch7_case_vecm

Step 5: Impulse Response Functions



TSA_ch7_case_irf

Step 6: Forecasting



TSA_ch7_case_forecast

Key Takeaways

Main Concepts

- **Cointegration**: $I(1)$ variables with stationary linear combination
- **Spurious regression**: High R^2 with unrelated $I(1)$ variables
- **VECM**: VAR with error correction for cointegrated systems

Testing Methods

- **Engle-Granger**: Simple, one vector only
- **Johansen**: Multiple vectors, MLE-based

Remember

Tests have low power in small samples. Theory should guide specification.

What's Next?

Extensions and Related Topics

- **Structural VECM**: Identifying structural shocks
- **Threshold cointegration**: Nonlinear adjustment
- **Panel cointegration**: Multiple cross-sections
- **Fractional cointegration**: Long memory
- **Time-varying cointegration**: Regime changes

Questions?

Quick Quiz

1. What does it mean for two $I(1)$ variables to be cointegrated?
2. What is the "spurious regression" problem?
3. In a VECM, what do the α coefficients represent?
4. What is the main advantage of Johansen over Engle-Granger?
5. If $\alpha_i = 0$ for variable Y_i, what does this imply?

Quiz Answers

1. **Cointegration**: A linear combination of the variables is $I(0)$ (stationary). They share a common stochastic trend.
2. **Spurious regression**: Regressing one $I(1)$ variable on another unrelated $I(1)$ variable gives high R^2 and significant coefficients even though there's no true relationship.
3. α **coefficients**: Speed of adjustment—how quickly each variable responds to deviations from long-run equilibrium.
4. **Johansen advantage**: Can test for multiple cointegrating relationships, uses MLE (more efficient), doesn't require choosing dependent variable.
5. $\alpha_i = 0$: Variable Y_i is weakly exogenous—it doesn't respond to disequilibrium. Other variables do all the adjusting.

Online Resources and Code

- **Quantlet**: https://quantlet.com → Code repository for statistics
- **Quantinar**: https://quantinar.com → Learning platform for quantitative methods
- **GitHub TSA_ch7**: https://github.com/QuantLet/TSA/tree/main/TSA_ch7

Thank You!

Questions?

Course materials available at: https://danpele.github.io/Time-Series-Analysis/

Quantlet Quantinar