



Time Series Analysis and Forecasting

Seminar 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

Seminar Outline

- ▣ **Review Quiz** – Knowledge check
- ▣ **True/False Questions** – Conceptual checks
- ▣ **Practice Problems** – Applied practice
- ▣ **Worked Examples** – Detailed solutions
- ▣ **Discussion Topics** – Critical thinking
- ▣ **AI-Assisted Exercises** – Applied artificial intelligence

Quiz 1: Cointegration Definition

Question

Two $I(1)$ variables X_t and Y_t are cointegrated if:

Answer choices

- (A) They are both stationary
- (B) Their sum is $I(2)$
- (C) A linear combination of them is $I(0)$
- (D) They have the same mean

Answer on next slide...

Quiz 1: Answer

Answer: C – A linear combination is $I(0)$

Answer choices

- (A) They are both stationary ✗
- (B) Their sum is $I(2)$ ✗
- (C) **A linear combination of them is $I(0)$** ✓
- (D) They have the same mean ✗

- ☐ $Y_t - \beta X_t \sim I(0)$ means they share a common stochastic trend
- ☐ The linear combination (spread) is stationary even though both series are non-stationary

Quiz 2: Spurious Regression

Question

When regressing one independent random walk on another, you typically get:

Answer choices

- (A) Low R^2 and insignificant coefficients
- (B) High R^2 and significant coefficients (spurious!)
- (C) Zero coefficients
- (D) Undefined results

Answer on next slide...

Quiz 2: Answer

Answer: B – High R^2 and significant coefficients (spurious!)

Answer choices

- (A) Low R^2 and insignificant coefficients ✗
- (B) High R^2 and significant coefficients (spurious!) ✓
- (C) Zero coefficients ✗
- (D) Undefined results ✗

- ▣ **Granger-Newbold (1974)**: Regressing unrelated $I(1)$ series gives misleading results
- ▣ Rule of thumb: If $R^2 > DW$, suspect spurious regression!
- ▣ 🌐 Real-world examples: tylervigen.com/spurious-correlations

 TSA_ch7_spurious_regression

Quiz 3: Engle-Granger Test

Question

In the Engle-Granger two-step method, what do you test in step 2?

Answer choices

- (A) Whether the original variables are stationary
- (B) Whether the regression residuals have a unit root
- (C) Whether the coefficients are significant
- (D) Whether the R^2 is high enough

Answer on next slide...

Quiz 3: Answer

Answer: B – Whether residuals have unit root

Answer choices

- (A) Whether the original variables are stationary ✗
- (B) **Whether the regression residuals have a unit root** ✓
- (C) Whether the coefficients are significant ✗
- (D) Whether the R^2 is high enough ✗

- ▣ **Step 1:** Run OLS: $Y_t = \alpha + \beta X_t + e_t$, save residuals $\hat{e}_t$
- ▣ **Step 2:** ADF test on residuals: $\Delta \hat{e}_t = \rho \hat{e}_{t-1} + \dots$
- ▣ 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!

Quiz 4: Johansen Test Advantage

Question

The main advantage of Johansen over Engle-Granger is:

Answer choices

- (A) It's simpler to compute
- (B) It can detect multiple cointegrating relationships
- (C) It doesn't require data
- (D) It always finds cointegration

Answer on next slide...

Quiz 4: Answer

Answer: B – Can detect multiple cointegrating relationships

Answer choices

- (A) It's simpler to compute ✗
- (B) **It can detect multiple cointegrating relationships ✓**
- (C) It doesn't require data ✗
- (D) It always finds cointegration ✗

- ☐ **Johansen advantages:** Tests for $r = 0, 1, \dots, k - 1$ cointegrating vectors
- ☐ MLE (more efficient); no need to choose dependent variable

Quiz 5: Rank of Π

Question

In a VECM with $k = 3$ variables, if $\text{rank}(\Pi) = 2$, this means:

Answer choices

- (A) No cointegration
- (B) One cointegrating relationship
- (C) Two cointegrating relationships
- (D) All variables are stationary

Answer on next slide...

Quiz 5: Answer

Answer: C – Two cointegrating relationships

Answer choices

- (A) No cointegration ✗
- (B) One cointegrating relationship ✗
- (C) **Two cointegrating relationships** ✓
- (D) All variables are stationary ✗

- ☐ $\text{rank}(\mathbf{\Pi}) = 0$: No cointegration (use VAR in differences)
- ☐ $0 < \text{rank}(\mathbf{\Pi}) = r < k$: r cointegrating vectors (use VECM)
- ☐ $\text{rank}(\mathbf{\Pi}) = k$: All variables are $I(0)$ (use VAR in levels)
- ☐ **With $k = 3$ and $r = 2$** : Two equilibrium relationships, only $k - r = 1$ common stochastic trend

Quiz 6: VECM Structure

Question

In the VECM equation $\Delta Y_t = c + \alpha\beta'Y_{t-1} + \dots$, what does α represent?

Answer choices

- (A) The cointegrating vectors
- (B) The adjustment (loading) coefficients
- (C) The short-run dynamics
- (D) The error variance

Answer on next slide...

Quiz 6: Answer

Answer: B – The adjustment (loading) coefficients

Answer choices

- (A) The cointegrating vectors ✗
- (B) **The adjustment (loading) coefficients** ✓
- (C) The short-run dynamics ✗
- (D) The error variance ✗

- ☐ $\Pi = \alpha\beta'$:
- ☐ β = cointegrating vectors (define equilibrium)
- ☐ α = adjustment speeds (how fast each variable corrects)

Quiz 7: Error Correction Term

Question

If $Y_t - \beta X_t$ is the cointegrating relation and this term is positive, what happens?

Answer choices

- (A) Y is above equilibrium; Y should decrease (if $\alpha < 0$)
- (B) Y is below equilibrium; Y should increase
- (C) Nothing, error correction doesn't affect levels
- (D) Both variables increase

Answer on next slide...

Quiz 7: Answer

Answer: $A - Y$ above equilibrium; decreases if $\alpha < 0$

Answer choices

(A) Y is above equilibrium; Y should decrease (if $\alpha < 0$) ✓

(B) Y is below equilibrium; Y should increase ✗

(C) Nothing, error correction doesn't affect levels ✗

(D) Both variables increase ✗

□ **Error correction mechanism:** $\Delta Y_t = \alpha(Y_{t-1} - \beta X_{t-1}) + \dots$

□ If $Y_{t-1} - \beta X_{t-1} > 0$: Y is “too high”

□ With $\alpha < 0$: $\Delta Y_t < 0$ (Y decreases toward equilibrium)

□ **Sign convention:** α should be negative for the dependent variable to move back toward equilibrium

Quiz 8: Weak Exogeneity

Question

If $\alpha_2 = 0$ in a bivariate VECM, this means:

Answer choices

- (A) There is no cointegration
- (B) Variable 2 does not adjust to disequilibrium (weakly exogenous)
- (C) Variable 1 does not adjust
- (D) Both variables are stationary

Answer on next slide...

Quiz 8: Answer

Answer: B – Variable 2 is weakly exogenous

Answer choices

- (A) There is no cointegration ✗
- (B) **Variable 2 does not adjust to disequilibrium (weakly exogenous) ✓**
- (C) Variable 1 does not adjust ✗
- (D) Both variables are stationary ✗

- ▣ **Weak exogeneity:** Variable doesn't respond to disequilibrium
- ▣ **Example:** Long rate (R_t) often weakly exogenous ($\alpha_R \approx 0$)
- ▣ Short rate (r_t) adjusts to spread ($\alpha_r < 0$)
- ▣ **Implication:** Can estimate single equation for the adjusting variable

Quiz 9: Trace Test

Question

The Johansen trace test with $H_0 : r \leq 1$ vs $H_1 : r > 1$ tests whether:

Answer choices

- (A) There is exactly one cointegrating vector
- (B) There are at most one cointegrating vectors
- (C) There are more than one cointegrating vectors
- (D) All eigenvalues are zero

Answer on next slide...

Quiz 9: Answer

Answer: B – There are at most one cointegrating vectors

Answer choices

- (A) There is exactly one cointegrating vector ✗
- (B) **There are at most one cointegrating vectors** ✓
- (C) There are more than one cointegrating vectors ✗
- (D) All eigenvalues are zero ✗

- **Note:** (B) describes the null hypothesis $H_0 : r \leq 1$, which is what the test “tests whether”. (C) describes the alternative $H_1 : r > 1$, not what the test tests.
- **Sequential testing:** Test $H_0 : r = 0$, then $H_0 : r \leq 1$, continue until fail to reject
- **Trace statistic:** $\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i)$
- Reject H_0 if trace statistic > critical value

Quiz 10: VECM vs VAR in Differences

Question

If variables are cointegrated, using VAR in first differences instead of VECM:

Answer choices

- (A) Gives identical results
- (B) Is more efficient
- (C) Loses long-run information (misspecified)
- (D) Is the preferred approach

Answer on next slide...

Quiz 10: Answer

Answer: C – Loses long-run information

Answer choices

- (A) Gives identical results ✗
- (B) Is more efficient ✗
- (C) **Loses long-run information (misspecified)** ✓
- (D) Is the preferred approach ✗

- **Granger Representation Theorem:** If cointegrated, VECM representation exists and should be used

	VAR(Δ)	VECM
Long-run equilibrium	Lost	Preserved
Error correction	No	Yes
Forecasts (long-run)	Poor	Better

- **Bottom line:** Differencing removes the long-run relationship that cointegration represents!

True or False? — Questions

-
- 1. Cointegration requires all variables to be stationary.
 - 2. The cointegrating vector is orthogonal to the space spanned by the cointegration matrix.
 - 3. Spurious regression has low Durbin-Watson statistic.
 - 4. In VECM, both α coefficients must be negative.
 - 5. Johansen test requires choosing a dependent variable.
 - 6. The number of common trends is equal to the rank of the cointegration matrix.
-

True or False? — Answers

Statement	T/F
1. Cointegration requires all variables to be $I(1)$.	T
2. The cointegrating vector is unique.	F
3. Spurious regression has low Durbin-Watson statistic.	T
4. In VECM, both α coefficients must be non-zero.	F
5. Johansen test requires choosing a dependent variable.	F
6. The number of common trends $= k - r$.	T

Exercise 1: Cointegration Identification

Problem

- ▣ **Data:** Quarterly data on consumption (C_t) and income (Y_t). ADF tests show both are $I(1)$. The regression $C_t = 0.85Y_t + e_t$ gives residuals with ADF statistic $= -3.92$. The 5% Engle-Granger critical value for 2 variables is -3.34 .
- ▣ **Question:** Are C_t and Y_t cointegrated?

Solution: Yes, they are cointegrated

- ▣ **Test:** H_0 : No cointegration (residuals have unit root)
- ▣ **ADF statistic:** -3.92 , **Critical value (5%):** -3.34
- ▣ Since $-3.92 < -3.34$, we **reject** H_0 at 5% level
- ▣ **Conclusion:** Residuals are stationary $\Rightarrow$ Cointegration exists!
- ▣ **Interpretation:** Consumption and income share a common trend. The cointegrating vector is approximately $(1, -0.85)$, consistent with permanent income hypothesis

Exercise 2: VECM Interpretation

Problem

- A VECM for short rate (r_t) and long rate (R_t) gives:

$$\Delta r_t = 0.01 - 0.25(r_{t-1} - R_{t-1}) + \dots$$

$$\Delta R_t = 0.005 - 0.02(r_{t-1} - R_{t-1}) + \dots$$

Interpret the adjustment coefficients.

Solution

- **Error correction term:** $(r_{t-1} - R_{t-1}) = \text{spread}$
- **Short rate** ($\alpha_r = -0.25$): When spread is positive, short rate decreases. 25% of disequilibrium corrected per period
- **Long rate** ($\alpha_R = -0.02$): Very small adjustment coefficient. Long rate is nearly weakly exogenous
- **Economic interpretation:** Central bank (short rate) adjusts to maintain yield curve

Exercise 3: Johansen Test Results

Problem

- Johansen trace test for 3 variables gives:

H_0	Trace Stat	5% CV
$r = 0$	45.2	29.8
$r \leq 1$	18.1	15.5
$r \leq 2$	3.2	3.8

What is the cointegrating rank?

Solution: Rank = 2

- $H_0 : r = 0$: $45.2 > 29.8 \Rightarrow$ **Reject** (at least 1)
- $H_0 : r \leq 1$: $18.1 > 15.5 \Rightarrow$ **Reject** (at least 2)
- $H_0 : r \leq 2$: $3.2 < 3.8 \Rightarrow$ **Fail to reject**
- **Conclusion**: $r = 2$, two equilibrium relationships, only $3 - 2 = 1$ common stochastic trend

Example: Term Structure of Interest Rates

Economic Theory

Expectations hypothesis: $R_t^{(n)} = \frac{1}{n} \sum_{i=0}^{n-1} E_t[r_{t+i}] + \text{premium}$
If premium is constant $\Rightarrow$ spread $(R_t - r_t)$ should be stationary.

Typical Findings

- ▣ Both rates are $I(1)$ (confirmed by ADF)
- ▣ Johansen test: $r = 1$ cointegrating vector
- ▣ Cointegrating vector $\approx (1, -1)$: spread is stationary
- ▣ Short rate adjusts ($\alpha_r < 0$), long rate weakly exogenous

Policy Implication

Central bank controls short rate; long rate driven by expectations.

Example: Purchasing Power Parity

PPP Theory

$e_t = p_t - p_t^*$ (log exchange rate = price differential)

Real exchange rate: $q_t = e_t - p_t + p_t^*$ should be stationary (long-run PPP)

Empirical Challenges

- Unit root tests: e_t, p_t, p_t^* all $I(1)$
- Cointegration tests: Mixed results depending on sample
- Half-life of PPP deviations: 3-5 years (slow adjustment)
- Weak exogeneity: Exchange rate often doesn't adjust

PPP Puzzle

Real exchange rate is highly persistent—slow mean reversion is hard to explain with standard models.

Example: Pairs Trading Strategy

Idea

Find cointegrated stocks $\Rightarrow$ trade the stationary spread

Implementation Steps

1. **Identify pairs:** Test cointegration (e.g., Coca-Cola & Pepsi)
2. **Estimate spread:** $z_t = P_A - \beta P_B$
3. **Trading rules:**
 - ▶ $z_t > \mu + 2\sigma$: Sell A, Buy B (spread too wide)
 - ▶ $z_t < \mu - 2\sigma$: Buy A, Sell B (spread too narrow)
 - ▶ Exit when $z_t \approx \mu$

Risks

Cointegration can break down; spread may not revert; transaction costs.

Python Cointegration Analysis: Key Functions

Essential Libraries

```
from statsmodels.tsa.stattools import coint, adfuller
from statsmodels.tsa.vector_ar.vecm import coint_johansen, VECM
```

Workflow

1. Unit root tests: `adfuller(series)`
2. Engle-Granger: `coint(y, x)` returns test stat & p-value
3. Johansen: `coint_johansen(data, det_order, k_ar_diff)`
4. Fit VECM: `model = VECM(data, k_ar_diff=2, coint_rank=1)`
5. Results: `results = model.fit()`

Note

Complete working examples are provided in the Jupyter notebooks.

Discussion: Cointegration vs Correlation

Key Question

Two series are highly correlated. Are they cointegrated?

Answer: Not necessarily!

- ▣ **Correlation:** Measures co-movement (can be spurious for $I(1)$)
- ▣ **Cointegration:** Requires stationary linear combination

Example

- ▣ Two independent random walks can have correlation > 0.9 purely by chance (spurious correlation). But they're NOT cointegrated—their spread is also $I(1)$.
- ▣ **Cointegration** implies a meaningful long-run equilibrium relationship.

Discussion: Choosing Deterministic Components

Key Question

Johansen test has 5 cases for deterministics. Which to choose?

Guidelines

1. **No constant, no trend:** Rarely used (requires mean-zero data)
2. **Constant in CE only:** Level series, no drift
3. **Constant unrestricted:** Most common for economic data
4. **Trend in CE:** Series have deterministic trends
5. **Trend unrestricted:** Trending differences (uncommon)

Practical Advice

- Start with Case 3 (constant unrestricted). Check sensitivity to specification. Use economic reasoning: do levels have trends?

Take-Home Exercises

1. **Theoretical:** Show that if Y_t and X_t are both random walks with the same innovation, they are cointegrated.
2. **Computation:** Given VECM estimates:

$$\Delta Y_t = 0.5 - 0.3(Y_{t-1} - 2X_{t-1}) + 0.2\Delta Y_{t-1}$$

$$\Delta X_t = 0.1 + 0.1(Y_{t-1} - 2X_{t-1}) + 0.4\Delta X_{t-1}$$

- ▶ What is the cointegrating vector?
 - ▶ Which variable adjusts more quickly?
 - ▶ What is the long-run equilibrium relationship?
3. **Applied:** Download 10-year and 3-month Treasury rates:
 - ▶ Test for unit roots; Test for cointegration (Engle-Granger and Johansen)
 - ▶ Estimate VECM; Interpret adjustment coefficients
 4. **Critical Thinking:** Why might PPP hold in the long run but not short run?

Exercise Solutions Hints

Hints

1. If $Y_t = Y_{t-1} + \varepsilon_t$ and $X_t = X_{t-1} + \varepsilon_t$ (same shock), then $Y_t - X_t = Y_0 - X_0$ is constant (stationary).
2. From the VECM:
 - ▶ Cointegrating vector: $(1, -2)$ (normalized on Y)
 - ▶ Y adjusts faster: $|\alpha_Y| = 0.3 > |\alpha_X| = 0.1$
 - ▶ Long-run: $Y = 2X$ (when EC term = 0)
3. For interest rates:
 - ▶ Both typically $I(1)$; spread usually stationary
 - ▶ Expect one cointegrating vector with $(1, -1)$
 - ▶ Short rate typically adjusts; long rate often weakly exogenous
4. PPP deviations: Transportation costs, non-traded goods, sticky prices, tariffs, market segmentation all slow adjustment but don't prevent long-run convergence.

Summary: Chapter 7

Key Concepts

1. **Cointegration:** $I(1)$ variables with stationary linear combination
2. **Spurious regression:** High R^2 without cointegration is meaningless
3. **Engle-Granger:** Simple, but only one cointegrating vector
4. **Johansen:** Multiple vectors, MLE, more powerful

VECM Insights

- ▣ β defines equilibrium; α determines adjustment speed
- ▣ Weak exogeneity ($\alpha = 0$): Variable doesn't respond to disequilibrium
- ▣ Always use VECM (not VAR in differences) when cointegrated

Remember

Cointegration is about **long-run equilibrium**, not just correlation!

Questions?

AI Exercise: Critical Thinking

Prompt to test in ChatGPT / Claude / Copilot

“Download oil spot and futures prices (CL=F, BZ=F). Test for cointegration and estimate a VECM model. Can I use the spread for a pairs trading strategy?”

Exercise:

1. Run the prompt in an LLM of your choice and critically analyze the response.
2. Does the AI test the order of integration of each series before testing for cointegration?
3. Johansen or Engle-Granger test? Does the AI justify the choice?
4. Is the cointegrating rank (r) correctly interpreted?
5. Does the pairs trading strategy account for transaction costs and model risk?

Warning: AI-generated code may run without errors and look professional. *That does not mean it is correct.*

Bibliography I

Fundamental Textbooks

- ▣ Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, 3rd ed., OTexts.
- ▣ Shumway, R.H., & Stoffer, D.S. (2017). *Time Series Analysis and Its Applications*, 4th ed., Springer.
- ▣ Brockwell, P.J., & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*, 3rd ed., Springer.

Financial Time Series

- ▣ Tsay, R.S. (2010). *Analysis of Financial Time Series*, 3rd ed., Wiley.
- ▣ Franke, J., Härdle, W.K., & Hafner, C.M. (2019). *Statistics of Financial Markets*, 4th ed., Springer.

Bibliography II

Modern Approaches and Machine Learning

- ▣ Nielsen, A. (2019). *Practical Time Series Analysis*, O'Reilly Media.
- ▣ Petropoulos, F., et al. (2022). *Forecasting: Theory and Practice*, International Journal of Forecasting.
- ▣ Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition, International Journal of Forecasting.

Online Resources and Code

- ▣ **Quantlet**: <https://quantlet.com> — Code repository for statistics
- ▣ **Quantinar**: <https://quantinar.com> — Platform for learning quantitative methods
- ▣ **GitHub TSA**: https://github.com/QuantLet/TSA/tree/main/TSA_ch7 — Python code for this seminar

Thank You!

Questions?

Seminar materials are available at: <https://danpele.github.io/Time-Series-Analysis/>



Quantlet



Quantinar