Time Series Analysis and Forecasting

Chapter 6: VAR and Granger Causality

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:

1. Understand the **motivation** for multivariate time series analysis
2. Specify and estimate **VAR(p)** models
3. Apply **Granger causality** tests
4. Interpret **Impulse Response Functions (IRFs)**
5. Perform **Forecast Error Variance Decomposition (FEVD)**
6. Select the optimal lag order using information criteria
7. Implement VAR analysis in **Python**

Outline

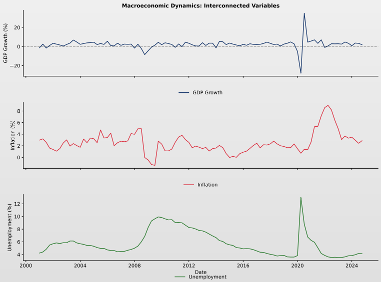
Foundations

Applications

Motivating Example: Macroeconomic Dynamics

Observations

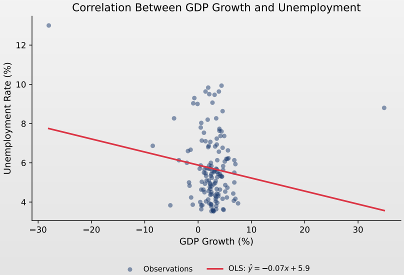
- Economic variables are **interconnected**: GDP affects unemployment, inflation affects interest rates
- Changes in one variable **propagate** through the system
- Understanding these dynamics requires **multivariate** analysis



TSA_ch6_motivation_econ

The Key Insight: Variables Interact

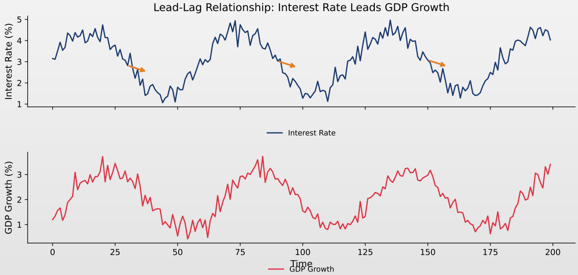
- **Okun's Law**: Higher GDP growth ⇒ lower unemployment
- **Taylor Rule**: Higher inflation ⇒ higher interest rates
- **Phillips Curve**: Unemployment-inflation tradeoff



TSA_ch6_motivation_scatter

Lead-Lag Relationships

- Some variables **lead** others: stock market predicts economic activity
- Cross-correlation reveals the **timing** of relationships
- Peak correlation at lag 4: stock market leads unemployment by ~4 months



TSA_ch6_motivation_leadlag

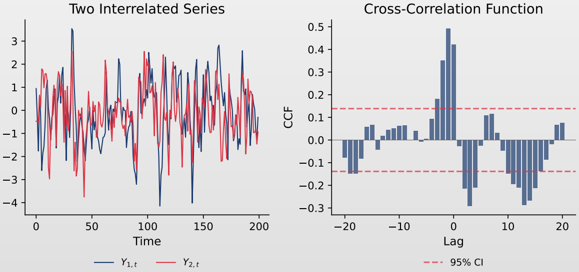
Why Univariate Models Are Not Enough

The Problem

ARIMA models each variable **in isolation**; ignores **interactions** and **feedback effects**

Examples

GDP–Unemployment, Interest Rate–Inflation, Stocks–Volume, Exchange Rate–Trade Balance



Univariate AR(1)

$$Y_{1,t} = \phi_1 Y_{1,t-1} + \varepsilon_t$$

→ Ignores Y_2

VAR(1)

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

→ Captures all cross-dynamics

TSA_ch6_motivation_univariate

What We'll Learn Today

Core Concepts

1. **VAR Models**: How to model multiple time series jointly
2. **Granger Causality**: Does X help predict Y?
3. **Impulse Response Functions**: How do shocks propagate?
4. **Variance Decomposition**: What drives each variable?

Recurring Example: GDP Growth and Unemployment

- Y_{1t}: **GDP Growth** and Y_{2t}: **Unemployment Rate** (*Okun's Law*)
- Central question: Does GDP cause unemployment, or vice versa, or both?

Applications

- Macroeconomic policy • Financial markets • Business cycle • Risk management

Multivariate Time Series Notation

Vector of Variables

Let $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, \ldots, Y_{Kt})'$ be a $K \times 1$ vector of time series.

Example with $K = 2$:

$$\mathbf{Y}_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} \text{GDP growth}_t \\ \text{Inflation}_t \end{pmatrix}$$

Key Questions

1. Does Y_1 help predict Y_2? (Granger causality)
2. How do shocks to Y_1 affect Y_2? (Impulse responses)
3. What proportion of Y_2's variance is due to Y_1? (Variance decomposition)

Multivariate Stationarity

Definition: Weak Stationarity

A K-dimensional time series Y_t is **weakly stationary** if:

1. $\mathbb{E}[\mathrm{Y}_t] = \boldsymbol{\mu}$ (constant mean vector)
2. $\mathrm{Cov}(\mathrm{Y}_t, \mathrm{Y}_{t-h}) = \boldsymbol{\Gamma}(h)$ depends only on h, not t

Autocovariance Matrix

$$\boldsymbol{\Gamma}(h) = \mathbb{E}[(\mathrm{Y}_t - \boldsymbol{\mu})(\mathrm{Y}_{t-h} - \boldsymbol{\mu})'] = \begin{pmatrix} \gamma_{11}(h) & \gamma_{12}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) \end{pmatrix}$$

- Note: $\boldsymbol{\Gamma}(-h) = \boldsymbol{\Gamma}(h)'$ (transpose, not equal!)

Cross-Covariance Properties

Cross-Covariance Function

For variables Y_{it} and Y_{jt}:

$$\gamma_{ij}(h) = \mathrm{Cov}(Y_{it}, Y_{j,t-h}) = \mathbb{E}[(Y_{it} - \mu_i)(Y_{j,t-h} - \mu_j)]$$

Key Difference from Univariate Case

- In general: $\gamma_{ij}(h) \neq \gamma_{ij}(-h)$
- But: $\gamma_{ij}(h) = \gamma_{ji}(-h)$
- The cross-covariance matrix is **not symmetric** for $h \neq 0$

Example

- If Y_1 leads Y_2: $\gamma_{12}(h) > 0$ for $h > 0$ but $\gamma_{12}(h) \approx 0$ for $h < 0$

Correlation Matrix Function

Definition

- The **autocorrelation matrix** at lag h:

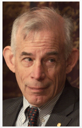
$$\mathrm{R}(h) = \mathrm{D}^{-1}\boldsymbol{\Gamma}(h)\mathrm{D}^{-1}$$

- $\mathrm{D} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_K)$ and $\sigma_i = \sqrt{\gamma_{ii}(0)}$

For Bivariate Case

- **Matrix**: $\mathrm{R}(h) = \begin{pmatrix} \rho_{11}(h) & \rho_{12}(h) \\ \rho_{21}(h) & \rho_{22}(h) \end{pmatrix}$, where $\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sigma_i \sigma_j}$
- **Interpretation**:
  - Diagonal elements: usual ACFs
  - Off-diagonal: cross-correlations

Researcher Spotlight: Sims & Granger



Christopher Sims (*1942)

Nobel Prize 2011

W Wikipedia

Clive Granger (1934–2009)

Nobel Prize 2003

W Wikipedia

Biography

- **Christopher Sims**: American econometrician at Princeton. Nobel Prize (2011) "for empirical research on cause and effect in the macroeconomy"
- **Clive Granger**: British-American economist at UC San Diego. Nobel Prize (2003) "for methods of analyzing economic time series with common trends (cointegration)"

Key Contributions

- **VAR models** (Sims, 1980) — vector autoregression for macroeconomics
- **Granger causality** (Granger, 1969) — predictive causality concept
- **Impulse response functions** and structural VAR identification
- **Cointegration** (Granger, 1981) — long-run equilibrium relationships

The VAR(p) Model

Definition

A **VAR(p)** model for K variables:

$$\mathrm{Y}_t = \mathrm{c} + \mathrm{A}_1 \mathrm{Y}_{t-1} + \mathrm{A}_2 \mathrm{Y}_{t-2} + \cdots + \mathrm{A}_p \mathrm{Y}_{t-p} + \varepsilon_t$$

where:

- Y_t: $K \times 1$ vector of endogenous variables
- c: $K \times 1$ vector of constants
- A_i: $K \times K$ coefficient matrices
- ε_t: $K \times 1$ vector of error terms with $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t \varepsilon_t'] = \boldsymbol{\Sigma}$

VAR(1) with Two Variables

Bivariate VAR(1)

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Equation by Equation

$$Y_{1t} = c_1 + a_{11} Y_{1,t-1} + a_{12} Y_{2,t-1} + \varepsilon_{1t}$$
$$Y_{2t} = c_2 + a_{21} Y_{1,t-1} + a_{22} Y_{2,t-1} + \varepsilon_{2t}$$

Key insight: Each equation includes lags of **all** variables!

Numerical Example: VAR(1)

Specific VAR(1) Model

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.3 \end{pmatrix} + \begin{pmatrix} 0.7 & 0.2 \\ -0.1 & 0.6 \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Interpretation of Coefficients

- $a_{11} = 0.7$: A 1-unit increase in Y_1 at $t-1$ increases Y_1 at t by 0.7
- $a_{12} = 0.2$: A 1-unit increase in Y_2 at $t-1$ increases Y_1 at t by 0.2
- $a_{21} = -0.1$: A 1-unit increase in Y_1 at $t-1$ **decreases** Y_2 at t by 0.1
- $a_{22} = 0.6$:
  - A 1-unit increase in Y_2 at $t-1$ increases Y_2 at t by 0.6

VAR(2): Higher Order Dynamics

VAR(2) Specification

$\mathrm{Y}_t = \mathrm{c} + \mathrm{A}_1\mathrm{Y}_{t-1} + \mathrm{A}_2\mathrm{Y}_{t-2} + \boldsymbol{\varepsilon}_t$. For $K = 2$: $2 + 2 \times 4 + 2 \times 4 = 18$ parameters!

Written Out

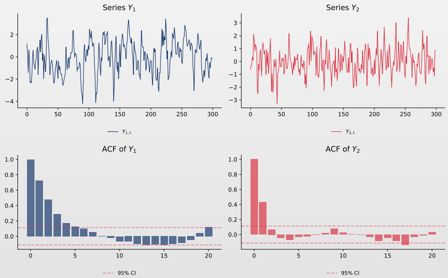
$$Y_{1t} = c_1 + a_{11}^{(1)} Y_{1,t-1} + a_{12}^{(1)} Y_{2,t-1} + a_{11}^{(2)} Y_{1,t-2} + a_{12}^{(2)} Y_{2,t-2} + \varepsilon_{1t}$$
$$Y_{2t} = c_2 + a_{21}^{(1)} Y_{1,t-1} + a_{22}^{(1)} Y_{2,t-1} + a_{21}^{(2)} Y_{1,t-2} + a_{22}^{(2)} Y_{2,t-2} + \varepsilon_{2t}$$

Curse of Dimensionality

VAR(p) with K variables has $K + pK^2$ parameters. With $K = 5$, $p = 4$: $5 + 4 \times 25 = 105$ parameters!

VAR Process: GDP and Unemployment (FRED)

- **Data**: US GDP Growth (GDPC1) and Unemployment Rate (UNRATE) from FRED
- Each variable responds to both its own past and the other variable's past
- Classic example of macroeconomic interdependence (Okun's Law)



TSA_ch6_var_simulation

The Companion Form

Converting VAR(p) to VAR(1)

Any VAR(p) can be written as a VAR(1) in **companion form**: $\xi_t = A\xi_{t-1} + v_t$

For VAR(2)

$$\underbrace{\begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix}}_{\xi_t} = \underbrace{\begin{pmatrix} A_1 & A_2 \\ I_K & 0 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} Y_{t-1} \\ Y_{t-2} \end{pmatrix}}_{\xi_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}}_{v_t}$$

Why Useful?

- Stationarity, forecasting, and IRFs are easier in companion form
- Matrix A is $Kp \times Kp$

Stationarity of VAR

Stability Condition

- VAR(p) is **stable** (stationary) if all roots of:

$$\det(I_K - A_1 z - A_2 z^2 - \cdots - A_p z^p) = 0$$

- Lie **outside** the unit circle (i.e., $|z| > 1$)

For VAR(1)

- The model is stable if all **eigenvalues** of A_1 are less than 1 in absolute value
- Example: For $A_1 = \begin{pmatrix} 0.5 & 0.1 \\ 0.2 & 0.3 \end{pmatrix}$, eigenvalues are $\lambda_1 = 0.6$ and $\lambda_2 = 0.2$
  - Both $< 1 \Rightarrow$ stable!

Stability Condition: Numerical Example

For $A = \begin{pmatrix} 0.7 & 0.2 \\ -0.1 & 0.6 \end{pmatrix}$

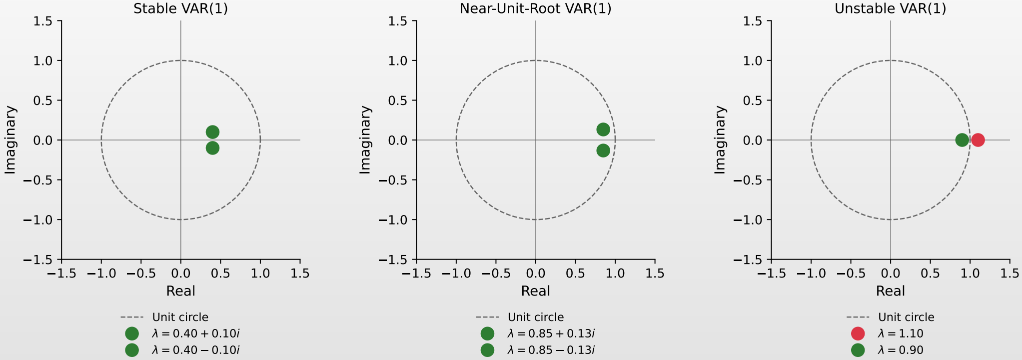
Characteristic polynomial: $\det(A - \lambda I) = 0$

$$\det \begin{pmatrix} 0.7-\lambda & 0.2 \\ -0.1 & 0.6-\lambda \end{pmatrix} = (0.7-\lambda)(0.6-\lambda) + 0.02 = 0 \implies \lambda^2 - 1.3\lambda + 0.44 = 0$$

Solution

$\lambda = \frac{1.3 \pm \sqrt{1.69 - 1.76}}{2} = 0.65 \pm 0.132i, \quad |\lambda| = \sqrt{0.65^2 + 0.132^2} = \sqrt{0.44} = 0.663 < 1 \quad \checkmark$ Stable!

Stability Condition: Visual Interpretation



TSA_ch6_stability_roots

Mean of a Stationary VAR

Unconditional Mean

- For stationary VAR(1): $Y_t = c + AY_{t-1} + \varepsilon_t$
- Taking expectations: $\mathbb{E}[Y_t] = c + A\mathbb{E}[Y_{t-1}]$
- Since $\mathbb{E}[Y_t] = \mathbb{E}[Y_{t-1}] = \boldsymbol{\mu}$ (stationarity):

$$\boldsymbol{\mu} = c + A\boldsymbol{\mu} \quad \Rightarrow \quad \boldsymbol{\mu} = (I_K - A)^{-1}c$$

Example

If $c = \begin{pmatrix} 0.5 \\ 0.3 \end{pmatrix}$ and $A = \begin{pmatrix} 0.7 & 0.2 \\ -0.1 & 0.6 \end{pmatrix}$: $\boldsymbol{\mu} = \begin{pmatrix} 0.3 & -0.2 \\ 0.1 & 0.4 \end{pmatrix}^{-1} \begin{pmatrix} 0.5 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 2.3 \\ 1.0 \end{pmatrix}$

Covariance Structure of VAR(1)

Variance-Covariance Matrix $\boldsymbol{\Gamma}(0)$

For VAR(1), the variance satisfies the **discrete Lyapunov equation**:

$$\boldsymbol{\Gamma}(0) = \mathrm{A}\boldsymbol{\Gamma}(0)\mathrm{A}' + \boldsymbol{\Sigma}$$

Autocovariance at Lag h

$$\boldsymbol{\Gamma}(h) = \mathrm{A}^h\boldsymbol{\Gamma}(0), \quad h \geq 0$$

This shows that autocovariances decay geometrically with the eigenvalues of A.

Solving the Lyapunov Equation

Can solve by vectorization:

$$\mathrm{vec}(\boldsymbol{\Gamma}(0)) = (\mathrm{I}_{K^2} - \mathrm{A} \otimes \mathrm{A})^{-1}\mathrm{vec}(\boldsymbol{\Sigma})$$

where $\otimes$ denotes the Kronecker product.

Estimation of VAR

OLS Estimation

- Each equation can be estimated by **OLS separately**:

$$\hat{A} = \left(\sum_{t=1}^{T} Y_{t-1} Y_t' \right) \left(\sum_{t=1}^{T} Y_{t-1} Y_{t-1}' \right)^{-1}$$

- Efficient because all equations have the **same regressors**

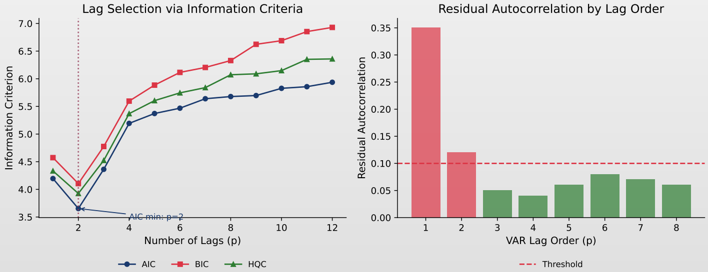
Covariance Matrix

- Estimator: $\hat{\Sigma} = \frac{1}{T-Kp-1} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$
- The errors ε_{1t} and ε_{2t} may be **contemporaneously correlated**

Lag Selection: Example

Observations

- Real US data (FRED): GDP and Unemployment, $T = 140$ quarters
- Information criteria: AIC and BIC for lag $p = 1, \ldots, 10$ (may suggest different orders)
- Interpretation: lower values = better fit; both select $p = 2$



TSA_ch6_lag_selection

Lag Order Selection

Information Criteria

Choose p that minimizes:

$$\text{AIC}(p) = \ln|\hat{\Sigma}_p| + \frac{2pK^2}{T}$$
$$\text{BIC}(p) = \ln|\hat{\Sigma}_p| + \frac{pK^2 \ln T}{T}$$
$$\text{HQ}(p) = \ln|\hat{\Sigma}_p| + \frac{2pK^2 \ln\ln T}{T}$$

where: $\hat{\Sigma}_p$ = residual covariance matrix, K = no. of variables, p = no. of lags, T = sample size

Guidelines

- AIC: larger models (better forecasting); BIC: smaller models (consistent)
- Start with p_{max} based on frequency (4 quarterly, 12 monthly)

Restricted VAR Models

Why Restrict?

Full VAR models can be **overparameterized**:

- Many coefficients may be insignificant
- Poor forecasting performance
- Loss of degrees of freedom

Common Restrictions

- **Zero restrictions**: Set small coefficients to zero
- **Block exogeneity**: Some variables don't affect others
- **Lag exclusion**: Exclude certain lags

Testing Restrictions

- Use likelihood ratio test: $LR = T(\ln|\hat{\Sigma}_R| - \ln|\hat{\Sigma}_U|) \sim \chi^2_r$
- r = number of restrictions

What is Granger Causality?

Clive Granger (1969, Nobel Prize 2003)

"X **Granger-causes** Y" if past values of X help predict Y, **beyond** what past values of Y alone can predict.

Important Distinction

Granger causality $\neq$ True causality

- Granger causality is about **predictive content**
- Does NOT imply economic/structural causation
- "X Granger-causes Y" means:
  - X contains useful information for forecasting Y

Formal Definition

Granger Causality

X **does not Granger-cause** Y if:

$$\mathbb{E}[Y_t | Y_{t-1}, Y_{t-2}, \ldots, X_{t-1}, X_{t-2}, \ldots] = \mathbb{E}[Y_t | Y_{t-1}, Y_{t-2}, \ldots]$$

In other words: adding X's history does not improve the prediction of Y.

In the VAR Context

- For VAR(1): $Y_{1t} = c_1 + a_{11} Y_{1,t-1} + a_{12} Y_{2,t-1} + \varepsilon_{1t}$
- Y_2 does **not** Granger-cause Y_1 if $a_{12} = 0$
- For VAR(p): Y_2 does not Granger-cause Y_1 if $a_{12}^{(1)} = a_{12}^{(2)} = \cdots = a_{12}^{(p)} = 0$

Testing for Granger Causality

Hypothesis Test

- H_0: Y_2 does **not** Granger-cause Y_1: $H_0 : a_{12}^{(1)} = a_{12}^{(2)} = \cdots = a_{12}^{(p)} = 0$
- H_1: At least one $a_{12}^{(i)} \neq 0$ (Granger causality exists)

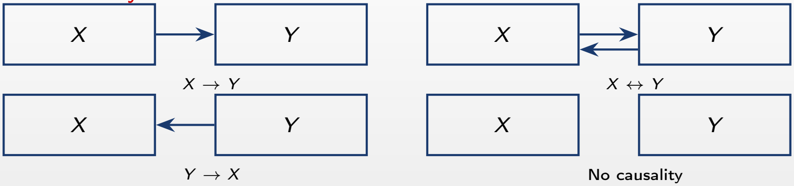
Test Statistic: Wald Test

$$F = \frac{(RSS_R - RSS_U)/p}{RSS_U/(T - 2p - 1)} \sim F_{p, T-2p-1}$$

where:

- RSS_R: Residual sum of squares from restricted model (without Y_2 lags)
- RSS_U: Residual sum of squares from unrestricted model (full VAR)

Types of Granger Causality



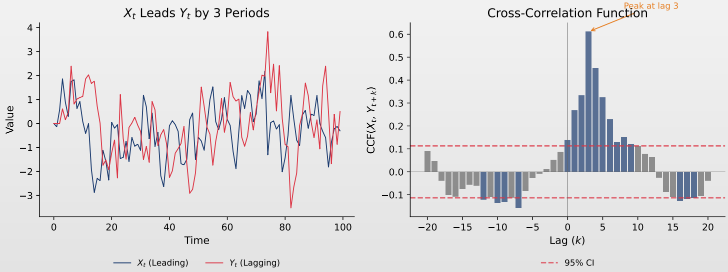
Economic Examples

- Money → Output? (monetarist); Stock prices ↔ Volume (bidirectional)

Cross-Correlation: Visual Illustration

Interpretation

- Left: two related series; Right: CCF reveals that X leads Y (significant correlations at positive lags)



TSA_ch6_def_ccf

Cross-Correlation Function

Definition 1 (Cross-Correlation Function)

The **cross-correlation** between X_t and Y_t at lag k is:

$$\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sigma_X \sigma_Y} = \frac{\text{Cov}(X_t, Y_{t+k})}{\sqrt{\text{Var}(X_t)\text{Var}(Y_t)}}$$

Interpretation

- $\rho_{XY}(k) > 0$ at $k > 0$: X is positively correlated with future Y (X may lead Y)
- $\rho_{XY}(k) > 0$ at $k < 0$: X is positively correlated with past Y (Y may lead X)

Note

Unlike ACF, cross-correlation is **not symmetric**: $\rho_{XY}(k) \neq \rho_{XY}(-k)$ in general.

Granger Causality: Practical Considerations

Common Pitfalls

1. **Omitted variables**: A third variable Z may cause both X and Y
2. **Non-stationarity**: Test requires stationary data (or cointegration)
3. **Lag selection**: Results can be sensitive to p
4. **Sample size**: Need sufficient observations

Best Practices

- Test for unit roots first
- Use multiple lag selection criteria
- Check robustness to different lag lengths
- Report results for both directions

Granger Causality Test: Numerical Example

Testing: Does Money Growth Granger-cause Output?

Unrestricted model (VAR with 2 lags):

$$\Delta Y_t = c + \alpha_1 \Delta Y_{t-1} + \alpha_2 \Delta Y_{t-2} + \beta_1 \Delta M_{t-1} + \beta_2 \Delta M_{t-2} + \varepsilon_t$$

Restricted model (H_0: $\beta_1 = \beta_2 = 0$):

$$\Delta Y_t = c + \alpha_1 \Delta Y_{t-1} + \alpha_2 \Delta Y_{t-2} + \varepsilon_t$$

Test Computation

With $T = 100$, $RSS_U = 45.2$, $RSS_R = 52.8$:

$$F = \frac{(52.8 - 45.2)/2}{45.2/(100-5)} = \frac{3.8}{0.476} = 7.98$$

$F_{0.05}(2, 95) = 3.09 \Rightarrow$ **Reject** H_0: Money Granger-causes output!

The Toda-Yamamoto Procedure

Problem with Non-Stationary Data

Standard Granger test has **non-standard distributions** when:

- Variables have unit roots
- Variables are cointegrated

Toda-Yamamoto Solution (1995)

1. Determine maximum order of integration d_{max}
2. Estimate VAR($p + d_{max}$) in **levels**
3. Test restrictions on first p lags only
4. Extra d_{max} lags are **not** tested (just for correct distribution)

Advantage

Wald test has asymptotic χ^2 distribution regardless of cointegration!

Instantaneous Causality

Definition

- X **instantaneously causes** Y if $\mathbb{E}[Y_t|\Omega_{t-1}, X_t] \neq \mathbb{E}[Y_t|\Omega_{t-1}]$
- Ω_{t-1} contains all past information

Testing in VAR

- Test whether $\sigma_{12} \neq 0$ in the covariance matrix: $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$
- If $\sigma_{12} = 0$: no instantaneous causality

Interpretation

Instantaneous causality often reflects **common shocks** or **data aggregation**, not true contemporaneous effects.

Granger Causality in Multiple Systems

Block Exogeneity Test

- In a VAR with $K > 2$ variables, test whether a **group** of variables Granger-causes another group
- Example: Do financial variables (interest rates, stock prices) Granger-cause real variables (GDP, unemployment)?

Test Statistic

$$\chi^2 = T \cdot K_1 \cdot p \cdot \left(\ln|\hat{\Sigma}_R| - \ln|\hat{\Sigma}_U|\right) \sim \chi^2_{K_1 \cdot K_2 \cdot p}$$

where K_1 = number of "caused" variables, K_2 = number of "causing" variables

What are Impulse Response Functions?

Definition

An **Impulse Response Function (IRF)** traces the effect of a one-time shock to one variable on the current and future values of all variables.

Question IRFs Answer

"If there is an unexpected 1-unit shock to Y_1 today, what happens to Y_1 and Y_2 over the next h periods?"

MA(∞) Representation

A stable VAR(p) can be written as:

$$\mathrm{Y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \boldsymbol{\varepsilon}_{t-i}$$

The matrices $\boldsymbol{\Phi}_i$ are the **impulse responses** at horizon i.

Computing IRFs for VAR(1)

For VAR(1): $\mathrm{Y}_t = \mathrm{c} + \mathrm{A}\mathrm{Y}_{t-1} + \varepsilon_t$

The impulse response matrices are:

$$\mathbf{\Phi}_0 = \mathrm{I}_K, \quad \mathbf{\Phi}_1 = \mathrm{A}, \quad \mathbf{\Phi}_2 = \mathrm{A}^2, \quad \ldots, \quad \mathbf{\Phi}_h = \mathrm{A}^h$$

Interpretation

- $[\mathbf{\Phi}_h]_{ij}$ = Effect on Y_i at time $t + h$ of a unit shock to Y_j at time t
- For stable VAR: $\mathbf{\Phi}_h \to 0$ as $h \to \infty$ (shocks die out)

Computing IRFs for General VAR(p)

Recursive Formula for VAR(p)

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

$$\Phi_h = \sum_{j=1}^{\min(h,p)} A_j \Phi_{h-j}, \quad h = 1, 2, 3, \ldots$$

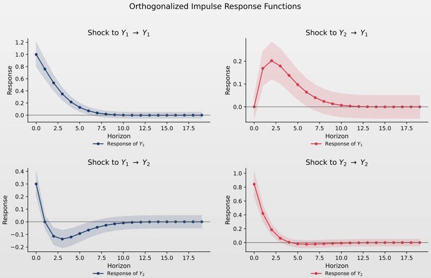
with $\Phi_0 = I_K$ and $\Phi_h = 0$ for $h < 0$.

Example: VAR(2) IRFs

- $\Phi_0 = I_K$
- $\Phi_1 = A_1 \Phi_0 = A_1$
- $\Phi_2 = A_1 \Phi_1 + A_2 \Phi_0 = A_1^2 + A_2$
- $\Phi_3 = A_1 \Phi_2 + A_2 \Phi_1 = A_1(A_1^2 + A_2) + A_2 A_1$

Impulse Response Functions: Example

- IRFs show how each variable responds to a one-unit shock over time
- Shaded regions represent confidence intervals (uncertainty in estimates)
- For stable VAR models, responses converge to zero as the horizon increases



TSA_ch6_irf

Orthogonalized IRFs

Problem: Correlated Errors

- If $\boldsymbol{\Sigma}$ is not diagonal, shocks ε_{1t} and ε_{2t} are correlated
- A shock to "Y_1" also involves a shock to "Y_2"

Solution: Cholesky Decomposition

- Factor $\boldsymbol{\Sigma} = \mathrm{PP}'$ where P is lower triangular
- Orthogonalized shocks: $\mathrm{u}_t = \mathrm{P}^{-1}\boldsymbol{\varepsilon}_t$ with $\mathbb{E}[\mathrm{u}_t\mathrm{u}_t'] = \mathrm{I}$
- Orthogonalized IRFs: $\boldsymbol{\Theta}_h = \boldsymbol{\Phi}_h\mathrm{P}$

Ordering Matters!

- Cholesky assumes variables ordered from "most exogenous" to "most endogenous"
- Results depend on this ordering

Cholesky Decomposition: How It Works

Numerical Example

Suppose the VAR residual covariance is $\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix}$. The Cholesky factor P (lower triangular) satisfying $\boldsymbol{\Sigma} = \mathrm{PP}'$ is:

$$\mathrm{P} = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \quad \Rightarrow \quad \mathrm{PP}' = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix} = \boldsymbol{\Sigma} \checkmark$$

Interpretation: What the Ordering Implies

- $P_{21} = 1 \neq 0$: A unit shock to Y_1 has an **immediate effect on** Y_2 (impact $= 1$)
- $P_{12} = 0$: A shock to Y_2 has **no contemporaneous effect on** Y_1
- Y_1 is “more exogenous” — it affects Y_2 instantly, but not vice versa

Reverse the Ordering (Y_2 First)

Swapping the variables gives a different P and different IRFs. This is why **economic theory** must guide the ordering — e.g., GDP before unemployment (Okun’s law: output shocks affect labor markets, not the reverse on impact).

IRF Numerical Example

For $A = \begin{pmatrix} 0.7 & 0.2 \\ -0.1 & 0.6 \end{pmatrix}$

$$\Phi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Phi_1 = A = \begin{pmatrix} 0.7 & 0.2 \\ -0.1 & 0.6 \end{pmatrix}$$

$$\Phi_2 = A^2 = \begin{pmatrix} 0.47 & 0.26 \\ -0.13 & 0.34 \end{pmatrix}$$

Interpretation

- $[\Phi_2]_{12} = 0.26$: A unit shock to Y_2 increases Y_1 by 0.26 after 2 periods
- $[\Phi_2]_{21} = -0.13$: A unit shock to Y_1 **decreases** Y_2 by 0.13 after 2 periods

Cumulative Impulse Responses

Definition

- The **cumulative IRF** up to horizon H: $\boldsymbol{\Psi}_H = \sum_{h=0}^{H} \boldsymbol{\Phi}_h$
- Measures the **total accumulated effect** of a shock

Long-Run Multiplier

- For stable VAR: $\boldsymbol{\Psi}_\infty = (\mathbf{I}_K - \mathbf{A}_1 - \mathbf{A}_2 - \cdots - \mathbf{A}_p)^{-1}$
- This gives the **permanent effect** of a one-time shock

When to Use

Cumulative IRFs are useful when interested in total impact (e.g., cumulative GDP loss from a shock).

Confidence Intervals for IRFs

Sources of Uncertainty

- IRFs are functions of estimated parameters $\hat{A}_1, \ldots, \hat{A}_p$, so they have **sampling uncertainty**

Methods for Confidence Bands

1. **Asymptotic**: Delta method for standard errors
2. **Monte Carlo**: Simulate from asymptotic distribution of $\hat{A}$
3. **Bootstrap**: Resample residuals and re-estimate VAR

Bootstrap Procedure

1. Estimate VAR, save residuals $\{\hat{\varepsilon}_t\}$
2. Draw with replacement to create $\{\hat{\varepsilon}_t^*\}$
3. Generate bootstrap sample, re-estimate, compute IRFs
4. Repeat B times; use percentiles for CIs

Structural VAR (SVAR)

Motivation

- Standard VAR shocks ε_t are **reduced-form** innovations—linear combinations of structural shocks
- We want to identify economically meaningful **structural shocks**

Structural Form

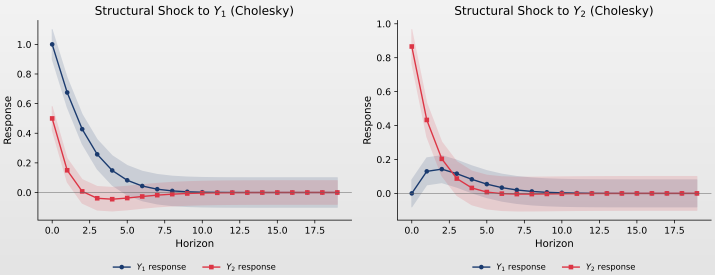
- $B_0 Y_t = \Gamma_0 + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + u_t$
- u_t are **structural shocks** with $\mathbb{E}[u_t u_t'] = I_K$

Relationship to Reduced Form

$\varepsilon_t = B_0^{-1} u_t \quad \Rightarrow \quad \Sigma = B_0^{-1}(B_0^{-1})'$

Structural IRF Example

- Structural IRFs based on Cholesky identification
- Order of variables affects interpretation of shocks
- First variable responds only to own shocks contemporaneously



TSA_ch6_structural_irf

Identification in SVAR

The Identification Problem

- Σ has $K(K+1)/2$ unique elements, but B_0^{-1} has K^2 elements
- Need $K(K-1)/2$ additional restrictions!

Common Identification Schemes

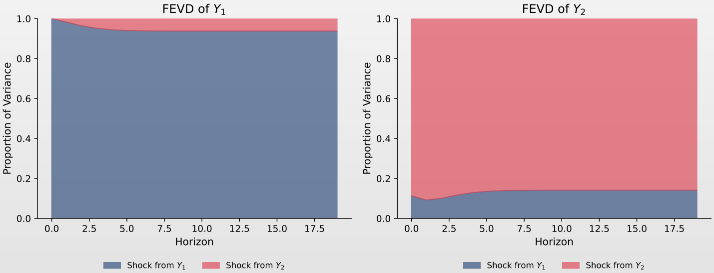
1. **Short-run restrictions**: Zero impact effects (Cholesky)
2. **Long-run restrictions**: Zero long-run effects (Blanchard-Quah)
3. **Sign restrictions**: Inequality constraints on IRFs
4. **External instruments**: Use outside information

Example: Cholesky (Recursive) Ordering

- For $K = 2$: $B_0^{-1} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix}$
- Variable 1 doesn't respond to shock 2 contemporaneously

FEVD: Example

- FEVD shows the proportion of forecast variance attributable to each shock
- At short horizons, own shocks dominate; cross-variable effects grow over time
- Useful for understanding the relative importance of different shocks in the system



TSA_ch6_fevd

Variance Decomposition

Question

What proportion of the forecast error variance of Y_i at horizon h is due to shocks to Y_j?

FEVD Formula

$$\text{FEVD}_{ij}(h) = \frac{\sum_{s=0}^{h-1} [\Theta_s]_{ij}^2}{\sum_{s=0}^{h-1} \sum_{k=1}^{K} [\Theta_s]_{ik}^2}$$

Properties

- $0 \leq \text{FEVD}_{ij}(h) \leq 1$ and $\sum_{j=1}^{K} \text{FEVD}_{ij}(h) = 1$ (sums to 100%)
- At $h = 1$: own shocks dominate (by Cholesky construction)

FEVD: Numerical Example

Computing FEVD for Bivariate VAR

Using orthogonalized IRFs $\boldsymbol{\Theta}_h$, FEVD at horizon H:

$$\text{FEVD}_{11}(H) = \frac{\sum_{h=0}^{H-1} \theta_{11}^2(h)}{\sum_{h=0}^{H-1} [\theta_{11}^2(h) + \theta_{12}^2(h)]}$$

Example Calculation

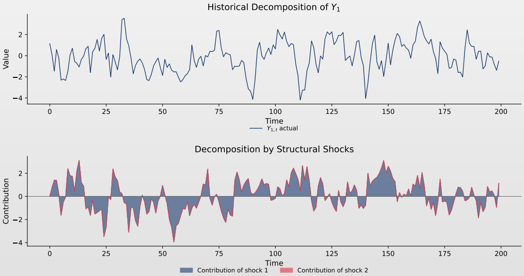
h	$\theta_{11}(h)$	$\theta_{12}(h)$	$\theta_{11}^2(h)$	$\theta_{12}^2(h)$
0	1.00	0.00	1.00	0.00
1	0.70	0.20	0.49	0.04
2	0.47	0.26	0.22	0.07

$\text{FEVD}_{11}(3) = \frac{1.00+0.49+0.22}{1.00+0.49+0.22+0.00+0.04+0.07} = \frac{1.71}{1.82} = 94\%$

Historical Decomposition: Example

Observations

- Structural contributions: each color = a different shock, stacked to sum the deviation from mean
- Useful for identifying shocks behind historical episodes



TSA_ch6_historical_decomp

Historical Decomposition

Definition

- **Historical decomposition** breaks down each observed value into contributions from each structural shock:

$$Y_{it} - \bar{Y}_i = \sum_{j=1}^{K} \sum_{s=0}^{t-1} \theta_{ij}(s) \cdot u_{j,t-s}$$

Application

- "How much of the 2008 GDP decline was due to financial shocks vs. oil shocks?"
  - Attributes historical movements to specific identified shocks
  - Useful for policy analysis and narrative interpretation

Residual Diagnostics

What to Check

After estimating VAR, verify that residuals $\hat{\varepsilon}_t$ behave like white noise:

1. No serial correlation
2. Constant variance (homoskedasticity)
3. Normality (for inference)

Why It Matters

- Autocorrelated residuals $\Rightarrow$ inefficient estimates
- Heteroskedasticity $\Rightarrow$ invalid standard errors
- Non-normality $\Rightarrow$ inference may be unreliable

Testing for Serial Correlation

Portmanteau Test (Ljung-Box)

- $Q_h = T(T+2)\sum_{j=1}^{h} \frac{1}{T-j}\mathrm{tr}(\hat{C}_j'\hat{C}_0^{-1}\hat{C}_j\hat{C}_0^{-1})$
- $\hat{C}_j = \frac{1}{T}\sum_{t=j+1}^{T} \hat{\varepsilon}_t\hat{\varepsilon}_{t-j}'$
- Under H_0 (no autocorrelation): $Q_h \sim \chi^2_{K^2(h-p)}$

Breusch-Godfrey LM Test

1. Regress $\hat{\varepsilon}_t$ on $\hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-h}$ and original regressors
2. $LM = T \cdot R^2 \sim \chi^2_{K^2 h}$ under H_0

If Rejected

- Consider increasing lag order p or adding additional variables

Testing for Heteroskedasticity

ARCH-LM Test

- Test for ARCH in residuals: $\hat{\varepsilon}_{it}^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{i,t-1}^2 + \cdots + \alpha_q \hat{\varepsilon}_{i,t-q}^2 + v_t$
- H_0: $\alpha_1 = \cdots = \alpha_q = 0$ (homoskedasticity)
- $LM = TR^2 \sim \chi_q^2$

Multivariate Version

- Test all equations jointly using:

$$\text{vech}(\hat{\varepsilon}_t \hat{\varepsilon}_t') = \mathrm{c} + \sum_{j=1}^{q} \mathrm{B}_j \text{vech}(\hat{\varepsilon}_{t-j} \hat{\varepsilon}_{t-j}') + \mathrm{v}_t$$

Normality Testing

Jarque-Bera Test (Univariate)

$$JB = \frac{T}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \sim \chi^2_2$$

where S = skewness, K = kurtosis

Multivariate Normality (Doornik-Hansen)

- Transform residuals and test joint skewness and kurtosis:

$$DH = s_1'(\Omega^{-1/2})'(\Omega^{-1/2})s_1 + s_2'(\Omega^{-1/2})'(\Omega^{-1/2})s_2 \sim \chi^2_{2K}$$

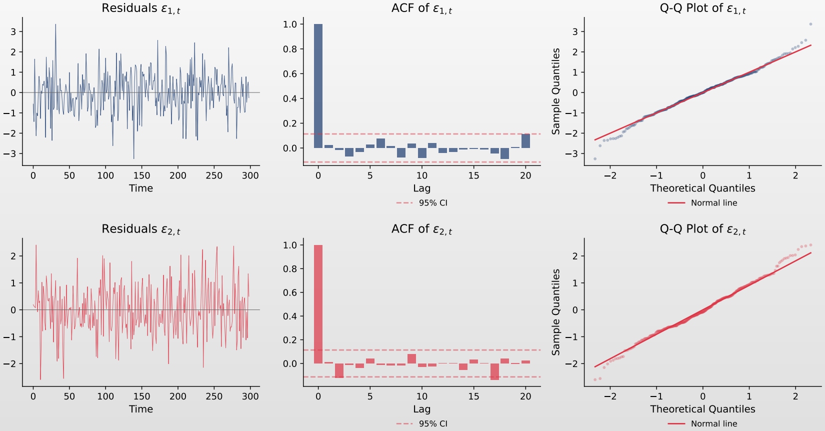
Note

Normality is often rejected in financial data. Consider robust standard errors if non-normality is severe.

Diagnostic Summary Plot

- Residual ACF should show no significant autocorrelation
- Histogram should approximate normal distribution
- Q-Q plot should follow 45-degree line

Diagnostic Summary Plot



TSA_ch6_diagnostics

Point Forecasts from VAR

Iterative Forecasting

For VAR(1): $Y_t = c + AY_{t-1} + \varepsilon_t$

- **1-step forecast**: $\hat{Y}_{T+1|T} = c + AY_T$
- **2-step forecast**: $\hat{Y}_{T+2|T} = c + A\hat{Y}_{T+1|T}$
- ***h*-step forecast**: $\hat{Y}_{T+h|T} = c + A\hat{Y}_{T+h-1|T}$

Direct Formula

- $\hat{Y}_{T+h|T} = (I + A + A^2 + \cdots + A^{h-1})c + A^h Y_T$
- For stable VAR: converges to $\boldsymbol{\mu} = (I - A)^{-1}c$ as $h \to \infty$

Forecast Error and MSE

h-Step Forecast Error

$\mathbf{e}_{T+h|T} = \mathbf{Y}_{T+h} - \hat{\mathbf{Y}}_{T+h|T} = \sum_{j=\mathbf{0}}^{h-\mathbf{1}} \mathbf{A}^j \boldsymbol{\varepsilon}_{T+h-j}$

Mean Squared Error Matrix

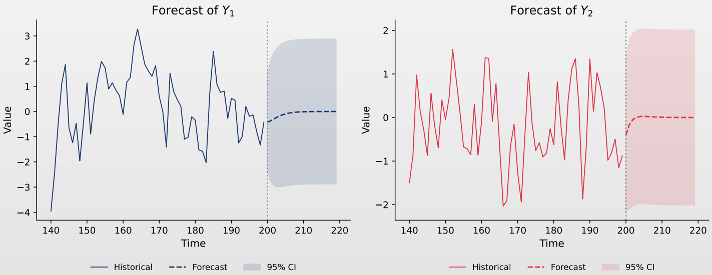
$\text{MSE}(\hat{\mathbf{Y}}_{T+h|T}) = \mathbb{E}[\mathbf{e}_{T+h|T}\mathbf{e}'_{T+h|T}] = \sum_{j=\mathbf{0}}^{h-\mathbf{1}} \mathbf{A}^j \boldsymbol{\Sigma}(\mathbf{A}^j)'$

Key Insight

- MSE increases with h; converges to $\boldsymbol{\Gamma}(0)$ for stable VAR
- Long-horizon forecasts $\to$ unconditional mean

VAR Forecasts: Example

- Point forecasts shown as solid line beyond observed data
- Confidence bands widen as forecast horizon increases
- Forecasts converge to unconditional mean for long horizons



TSA_ch6_var_forecast

Forecast Confidence Intervals

Constructing Intervals

- For normally distributed errors, $(1-\alpha)$ confidence interval: $\hat{Y}_{i,T+h|T} \pm z_{\alpha/2}\sqrt{[\text{MSE}(\hat{\mathbf{Y}}_{T+h|T})]_{ii}}$

Joint Confidence Regions

- For multiple variables, use ellipsoids: $(\mathbf{Y}_{T+h} - \hat{\mathbf{Y}}_{T+h|T})'[\text{MSE}(\hat{\mathbf{Y}}_{T+h|T})]^{-1}(\mathbf{Y}_{T+h} - \hat{\mathbf{Y}}_{T+h|T}) \leq \chi^2_{K,\alpha}$

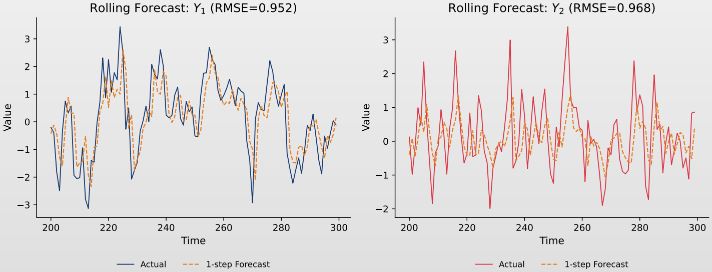
Note

- These assume known parameters. Bootstrap methods account for parameter uncertainty.

Out-of-Sample Evaluation: VAR vs. AR

General Methodology

- Train / Test Split: split data into training + test; estimate on train, evaluate on test
- Why VAR vs. AR? AR ignores other variables; VAR exploits interdependencies
- Metric: RMSE $= \sqrt{\frac{1}{h}\sum e_i^2}$



TSA_ch6_var_rolling_forecast

Forecast Evaluation

Out-of-Sample Evaluation

Split data: estimation sample (1 to T_1) and test sample ($T_1 + 1$ to T). Compute forecast errors: $e_{t+h} = Y_{t+h} - \hat{Y}_{t+h|t}$

Common Metrics

- **RMSE**: $\sqrt{\frac{1}{n}\sum e_{t+h}^2}$ **MAE**: $\frac{1}{n}\sum |e_{t+h}|$ **MAPE**: $\frac{100}{n}\sum \left|\frac{e_{t+h}}{Y_{t+h}}\right|$

Diebold-Mariano Test

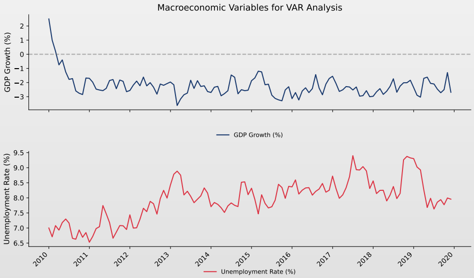
Test whether VAR forecasts are significantly better than alternative: $DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2/n}} \sim N(0,1)$ where

$d_t = L(e_{1t}) - L(e_{2t})$ is the loss differential.

GDP and Unemployment: Quarterly Data

Observations

- GDP growth and unemployment rate: Okun's Law, common cyclical patterns
- Bivariate system ideal for VAR analysis + Granger causality



TSA_ch6_gdp_unemployment

Example: GDP and Unemployment

Okun's Law

- Negative relationship between GDP growth and unemployment:

$$\Delta U_t \approx -\beta(\Delta Y_t - \bar{g})$$

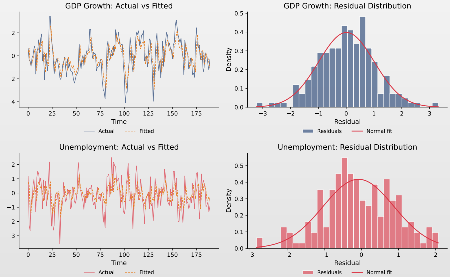
- $\bar{g}$ = trend GDP growth, $\beta \approx 0.4$

VAR Analysis Questions

1. Does GDP growth Granger-cause unemployment changes?
2. Does unemployment Granger-cause GDP growth?
3. How do shocks propagate between variables?

Estimated VAR Results

- Estimated coefficients with standard errors and t-statistics
- Information criteria values for model comparison
- Model diagnostics summary (residual tests)



TSA_ch6_var_results

VAR Workflow

1. **Data preparation**
   - Check for stationarity (unit root tests)
   - Transform if necessary (differences, logs)
2. **Lag selection**
   - Use AIC, BIC, HQ criteria
   - Check residual autocorrelation
3. **Estimation**
   - OLS equation by equation
   - Check stability (eigenvalues)
4. **Analysis**
   - Granger causality tests
   - Impulse response functions
   - Variance decomposition
5. **Forecasting**

Granger Causality Results

Test Results: GDP and Unemployment

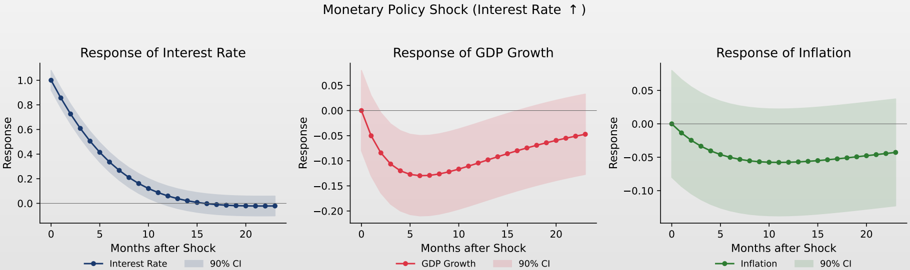
Null Hypothesis	F-statistic	df	p-value	Decision
GDP $\not\to$ Unemployment	8.42	(2, 95)	0.0004	Reject
Unemployment $\not\to$ GDP	2.15	(2, 95)	0.1220	Fail to Reject

Interpretation

- GDP growth Granger-causes unemployment (consistent with Okun's Law)
- Unemployment does not significantly Granger-cause GDP
- Evidence of **unidirectional** causality: GDP $\to$ Unemployment

Monetary Policy VAR: IRFs

- Contractionary monetary policy shock (interest rate increase)
- Output decreases with peak effect after 4-6 quarters (“long and variable lags”)
- Inflation responds more slowly, decreasing after output



TSA_ch6_monetary_irf

Example: Monetary Policy Analysis

Three-Variable VAR

Study the monetary transmission mechanism with:

- Y_1: Output gap (GDP deviation from trend)
- Y_2: Inflation rate
- Y_3: Interest rate (policy instrument)

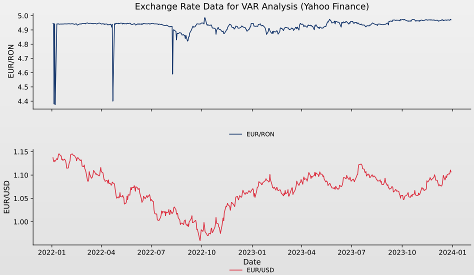
Key Questions

1. How does an interest rate shock affect output and inflation?
2. How long until the maximum effect is felt?
3. What fraction of output variance is due to monetary shocks?

Case Study: The Relationship between GDP and Unemployment

Data

- Real US data (FRED, 1990–2024): GDP Growth and Unemployment Rate ($T = 140$ quarters)
- Visible negative correlation between series (Okun's Law); bidirectional dynamics

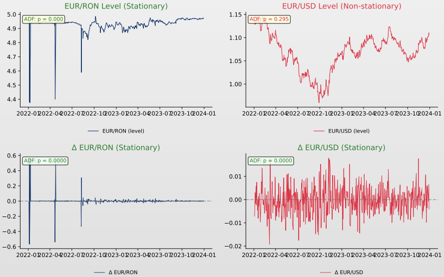


TSA_ch6_case_raw_data

Step 1: Preliminary Analysis

Results

- GDP: ACF decays quickly $\Rightarrow$ stationary; Unemployment: persistent ACF (ADF: $p = 0.02$)
- Negative GDP–Unemployment correlation ($\rho = -0.17$); cross-correlation suggests bidirectional relationships

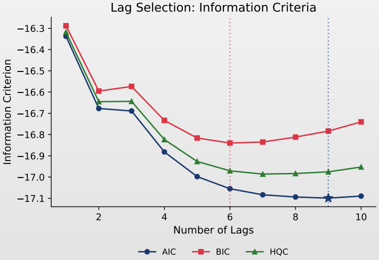


TSA_ch6_case_stationarity

Step 2: VAR Order Selection

Results

- AIC and BIC criteria suggest VAR(2); trade-off between complexity and fit



Optimal Lag Summary

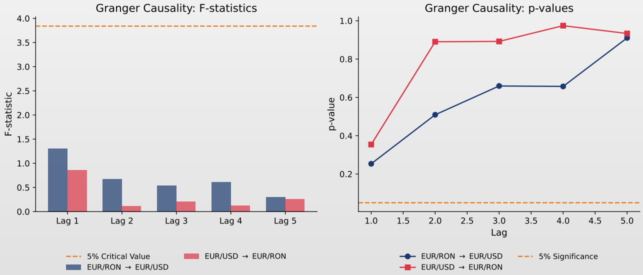
Criterion	Optimal Lag
AIC	p = 9
BIC	p = 6
HQC	p = 7
FPE	p = 9

TSA_ch6_case_lag_selection

Step 3: Granger Causality Test

Results

- GDP $\Rightarrow$ Unemployment: $F = 17.35$, $p < 0.001 \Rightarrow$ GDP "Granger-causes" Unemployment
- Unemployment $\Rightarrow$ GDP: $F = 38.93$, $p < 0.001 \Rightarrow$ bidirectional causality (Okun's Law)

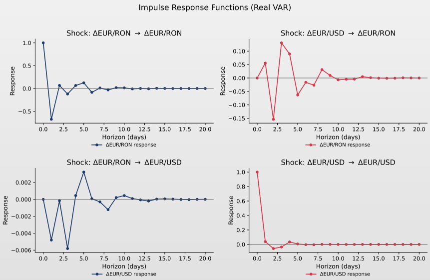


TSA_ch6_case_granger

Step 4: Impulse Response Functions (IRF)

IRF Results

- GDP shock ⇒ persistent negative effect on unemployment (Okun's Law, >20 quarters)
- Unemployment shock ⇒ short-lived positive effect on GDP (recovery, 2–3 quarters)

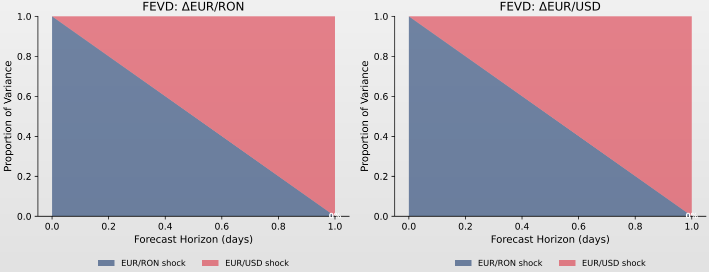


TSA_ch6_case_irf

Step 5: Variance Decomposition (FEVD)

FEVD Results

- GDP: ~65% of variance explained by own shocks, ~35% by Unemployment shocks
- Unemployment: dominated by GDP shocks (~65% at $h = 1$, increasing to ~92% at $h = 20$)

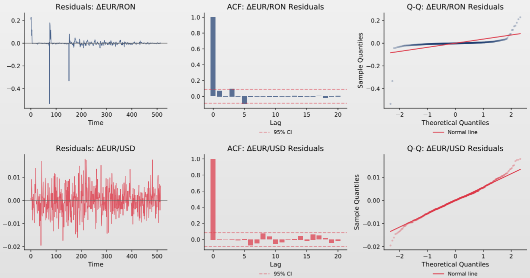


TSA_ch6_case_fevd

Step 6: Residual Diagnostics

Diagnostics

- Residuals show no significant autocorrelation (ACF within bounds)
- Significant non-normality (JB rejected) $\Rightarrow$ due to extreme COVID-19 values

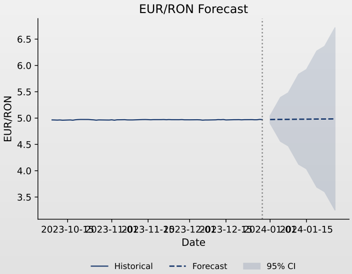


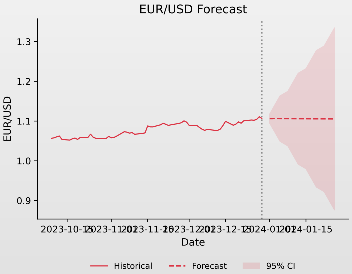
TSA_ch6_case_diagnostics

Step 7: VAR Forecasting

Forecast Results

- 12-quarter forecast; VAR captures interdependencies between series
- Forecasts converge to long-run equilibrium values





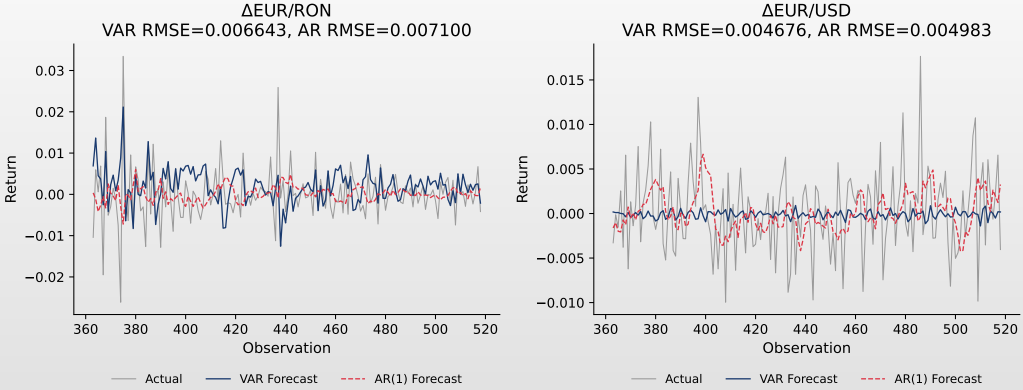
TSA_ch6_case_forecast

Step 8: Rolling Forecast – VAR vs AR

Procedure

- Recursive Train/Test: $W = 80$, re-estimation at each step; AR(2) vs VAR(2) models
- Expanding RMSE: $\text{RMSE}_t = \sqrt{\frac{1}{t}\sum_{s=1}^{t} e_s^2}$; 95% confidence intervals
- Mixed results:
  - Unemployment — VAR -10% RMSE vs AR (information from GDP helps)
  - GDP — VAR $+6\%$ RMSE vs AR (information from unemployment does not help)

Step 8: Rolling Forecast – VAR vs AR



TSA_ch6_case_rolling_forecast

Step 9: Out-of-Sample Comparison – AR vs VAR

Methodology

- Recursive forecast: estimate on $[1, \ldots, t]$, forecast $\hat{y}_{t+1|t}$; AR(2) vs VAR(2) models
- Expanding RMSE: $\text{RMSE}_t = \sqrt{\frac{1}{t}\sum_{s=1}^{t} e_s^2}$
- VAR reduces RMSE for Unemployment ($\sim$10%), not for GDP

TSA_ch6_case_oos_comparison

AI Exercise: Critical Thinking

Prompt to test in ChatGPT / Claude / Copilot

"Download from FRED: quarterly US Real GDP growth rate (A191RL1Q225SBEA) and monthly unemployment rate (UNRATE, aggregated to quarterly) for 2000-Q1 to 2024-Q4 (100 observations). Test Granger causality in both directions, estimate a VAR model, and compute orthogonalized impulse response functions. Give me complete Python code."

Exercise:

1. Run the prompt in an LLM of your choice and critically analyze the response.
2. Does it test stationarity of each variable before estimating the VAR?
3. How does it select the lag order? Does it compare AIC, BIC, HQIC?
4. Are impulse response functions orthogonalized? Does it discuss Cholesky ordering?
5. Does it check the stability condition (eigenvalues inside the unit circle)?

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

Key Takeaways

VAR Models

- Model **multiple** time series jointly
- Each variable depends on its own lags AND lags of other variables
- Estimated by OLS equation by equation; requires stationarity

Granger Causality

- Tests whether X helps predict Y beyond Y's own history
- **Not** the same as true causality; F-test on coefficient restrictions

IRF and FEVD

- IRF: How shocks propagate through the system
- FEVD: What proportion of variance is due to each shock
- Both depend on variable ordering (Cholesky decomposition)

VAR Model Selection Checklist

Before Estimation

- ☐ Test for unit roots
- ☐ Transform if needed
- ☐ Check for breaks

Model Specification

- ☐ Select lag order (AIC/BIC)
- ☐ Estimate VAR by OLS
- ☐ Check stability

Post-Estimation

- ☐ Test autocorrelation
- ☐ Test ARCH effects
- ☐ Test normality
- ☐ Compute IRFs, FEVDs

Common Mistakes to Avoid

Pitfalls in VAR Analysis

1. **Ignoring non-stationarity**: Always test for unit roots first
2. **Overfitting**: Too many lags $\Rightarrow$ poor forecasts
3. **Wrong ordering**: Cholesky results depend on variable order
4. **Confusing correlation with causation**: Granger causality $\neq$ true causality
5. **Ignoring parameter uncertainty**: Use bootstrap CIs for IRFs
6. **Short samples**: VAR requires many observations ($T > 50$)

What's Next?

Topics for Further Study

- **Cointegration**: Long-run relationships between non-stationary variables
- **VECM**: Error correction models for cointegrated systems
- **Structural VAR**: Imposing economic theory restrictions
- **Panel VAR**: VAR for panel data
- **Bayesian VAR**:
  - Shrinkage priors for high-dimensional systems

Questions?

Question 1

Question

For a VAR(1) model with coefficient matrix $A = \begin{pmatrix} 0.8 & 0.3 \\ 0.1 & 0.5 \end{pmatrix}$, is the model stable?

Answer Choices

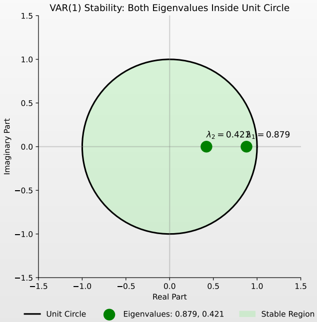
(A) Yes, because all diagonal elements are less than 1

(B) Yes, because all eigenvalues are inside the unit circle

(C) No, because the sum of coefficients exceeds 1

(D) Cannot be determined without knowing $\boldsymbol{\Sigma}$

Question 1: Answer



Answer: (B)

- $\lambda_1 = 0.879$, $\lambda_2 = 0.421$ — both $|\lambda| < 1 \Rightarrow$ Stable!

TSA_ch6_quiz1_var_stability

Question 2

Question

□ If X Granger-causes Y at the 5% significance level, which of the following statements is TRUE?

Answer Choices

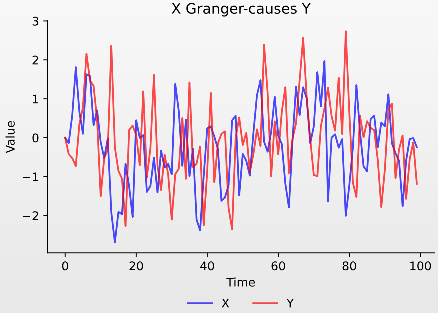
(A) X is the economic cause of Y

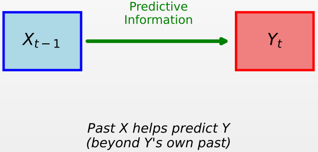
(B) Past values of X contain useful information for predicting Y

(C) Y cannot Granger-cause X

(D) The correlation between X and Y is positive

Question 2: Answer





Answer: (B)

- Granger causality = predictive content, not true economic causation. Past X helps predict Y.

TSA_ch6_quiz2_granger_causality

Question 3

Question

☐ In a VAR with Cholesky-identified IRFs, what does the ordering of variables determine?

Answer Choices

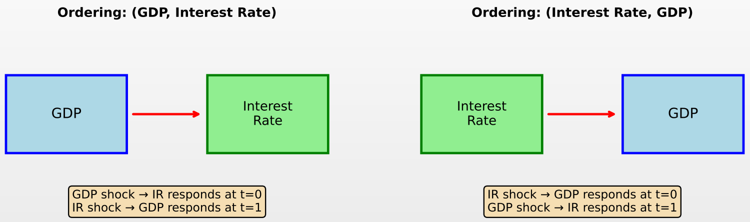
(A) The magnitude of the impulse responses

(B) The speed at which shocks die out

(C) Which variables can respond contemporaneously to which shocks

(D) The number of lags in the VAR

Question 3: Answer



Answer: (C)

- Ordering determines which variables respond immediately to which shocks.

TSA_ch6_quiz3_cholesky_ordering

Question 4

Question

- For a bivariate VAR(1), how many parameters need to be estimated (excluding the error covariance matrix)?

Answer Choices

(A) 4

(B) 6

(C) 8

(D) 10

Question 4: Answer

Answer: (B)

- 6 parameters

Detailed Count

VAR(1) with $K = 2$ variables:

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \underbrace{\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}}_{\text{2 params}} + \underbrace{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}}_{\text{4 params}} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

- Constant vector c: $K = 2$ parameters
- Coefficient matrix A: $K^2 = 4$ parameters
- Total: $K + K^2 = 2 + 4 = 6$ parameters

General Formula

VAR(p) with K variables: $K + pK^2$ parameters (excluding $\boldsymbol{\Sigma}$)

Question 5

Question

- What does $\text{FEVD}_{12}(h) = 0.35$ mean?

Answer Choices

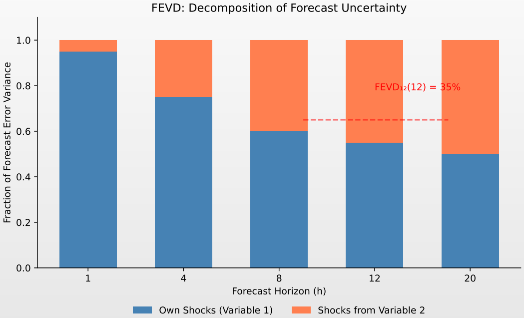
(A) 35% of variable 1's total variance is explained by variable 2

(B) 35% of variable 1's h-step forecast error variance is due to shocks to variable 2

(C) The correlation between variables 1 and 2 at lag h is 0.35

(D) Variable 2 explains 35% of the impulse response of variable 1

Question 5: Answer



Answer: (B)

- 35% of variable 1's h-step forecast error variance is due to shocks from variable 2.

TSA_ch6_quiz5_fevd

Key Formulas – Summary

VAR(p) Model

- $\mathrm{Y}_t = \mathrm{c} + \sum_{i=1}^{p} \mathrm{A}_i \mathrm{Y}_{t-i} + \varepsilon_t$
- $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$, i.i.d.

Granger Causality

- H_0: X does not Granger-cause Y
- F or Wald test on lag coefficients of X

Lag Selection

- $\mathrm{AIC} = \ln|\hat{\Sigma}| + \frac{2pK^2}{T}$
- $\mathrm{BIC} = \ln|\hat{\Sigma}| + \frac{pK^2 \ln T}{T}$

Impulse Response Functions

- $\mathrm{Y}_{t+h} = \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t+h-i}$
- Φ_i = multipliers at horizon i

FEVD

- $\mathrm{FEVD}_{jk}(h) = \frac{\sum_{i=0}^{h-1} (\mathrm{e}_j' \Phi_i \mathrm{P} \mathrm{e}_k)^2}{\sum_{i=0}^{h-1} \mathrm{e}_j' \Phi_i \Sigma \Phi_i' \mathrm{e}_j}$
- Contribution of shock k to variance of j

VAR Stationarity

- All eigenvalues of A inside the unit circle

Bibliography I

Fundamental works on VAR and causality

- Sims, C.A. (1980). Macroeconomics and Reality, *Econometrica*, 48(1), 1–48.
- Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, 37(3), 424–438.
- Toda, H.Y., & Yamamoto, T. (1995). Statistical Inference in Vector Autoregressions with Possibly Integrated Processes, *Journal of Econometrics*, 66(1-2), 225–250.

VAR Textbooks

- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer.
- Kilian, L., & Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Cambridge University Press.

Bibliography II

Impulse response functions and variance decomposition

- Pesaran, H.H., & Shin, Y. (1998). Generalized Impulse Response Analysis in Linear Multivariate Models, *Economics Letters*, 58(1), 17–29.
- Hamilton, J.D. (1994). *Time Series Analysis*, Princeton University Press.
- Tsay, R.S. (2014). *Multivariate Time Series Analysis with R and Financial Applications*, Wiley.

Online resources and code

- **Quantlet**: `https://quantlet.com` – Code platform for quantitative methods
- **Quantinar**: `https://quantinar.com` – Learning platform for quantitative methods
- **GitHub TSA**: `https://github.com/QuantLet/TSA/tree/main/TSA_ch6` – Python code for this chapter

Thank You!

Questions?

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

Quantlet Quantinar