



Time Series Analysis and Forecasting

Chapter 5: VAR Models & Granger Causality

Multivariate Time Series



Lecture Outline

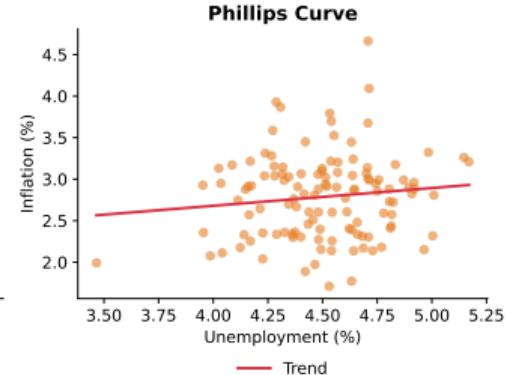
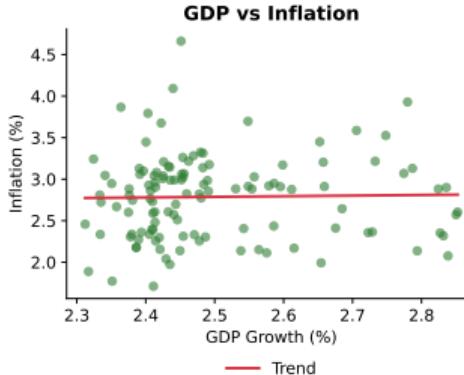
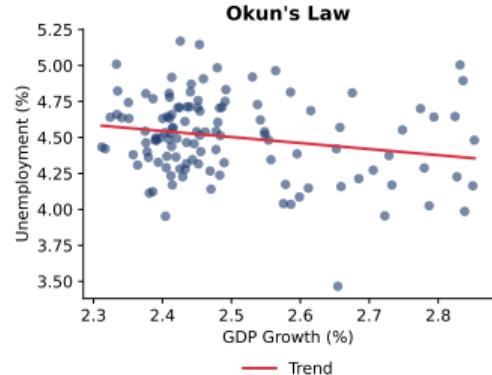
- 1 Introduction to Multivariate Time Series
- 2 Vector Autoregression (VAR)
- 3 Granger Causality
- 4 Impulse Response Functions
- 5 Forecast Error Variance Decomposition
- 6 VAR Diagnostics
- 7 VAR Forecasting
- 8 Practical Example
- 9 Summary
- 10 Quiz

Motivating Example: Macroeconomic Dynamics



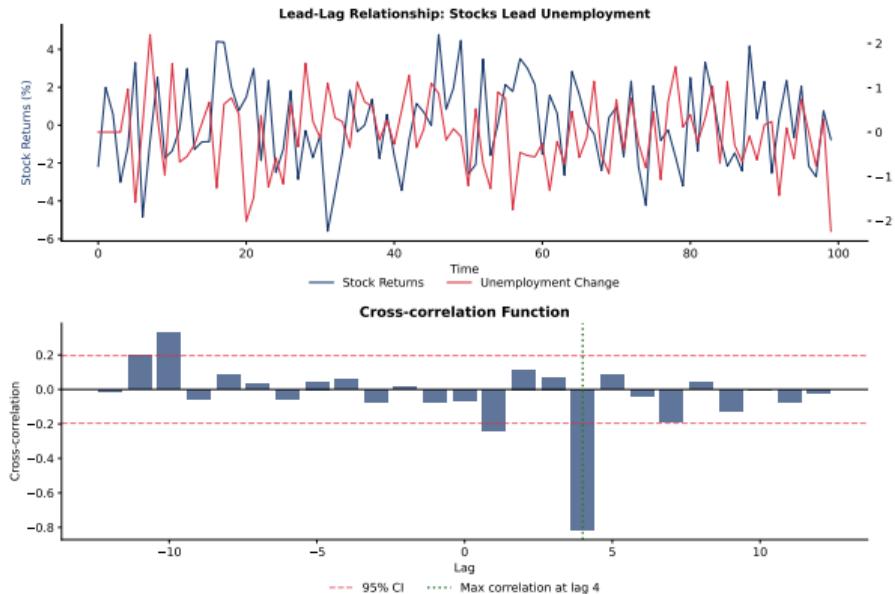
- Economic variables are **interconnected**: GDP affects unemployment, inflation affects interest rates
- Changes in one variable **propagate** through the system
- Understanding these dynamics requires **multivariate** analysis

The Key Insight: Variables Interact



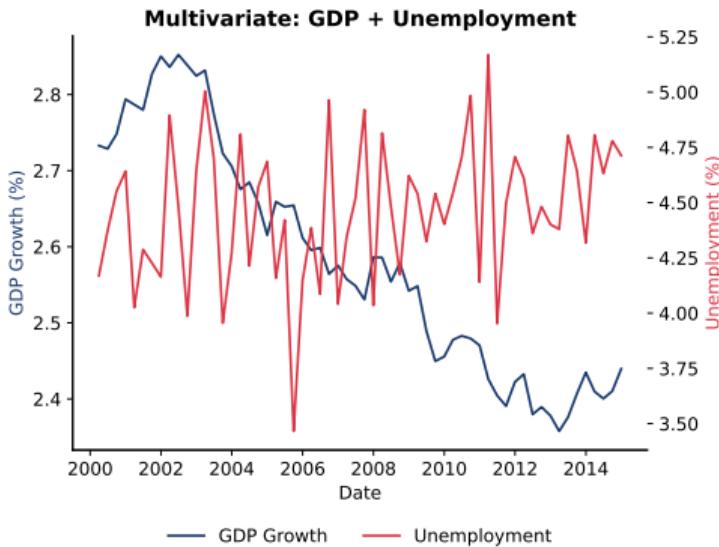
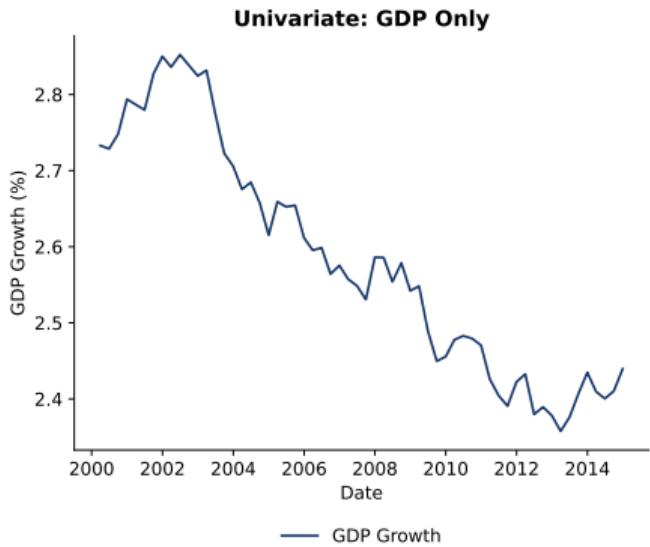
- **Okun's Law:** Higher GDP growth \Rightarrow lower unemployment
- **Taylor Rule:** Higher inflation \Rightarrow higher interest rates
- **Phillips Curve:** Unemployment-inflation tradeoff

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

Why Univariate Models Are Not Enough



The Problem

ARIMA models each variable **in isolation**—ignoring valuable information from other variables!

The Solution

VAR models capture the joint dynamics and feedback effects between multiple time series.

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?

Applications

- Macroeconomic policy analysis (monetary policy effects)
- Financial market dynamics (stock-bond relationships)
- Business cycle analysis (leading indicators)
- Risk management (volatility transmission)

Why Multivariate Analysis?

Limitations of Univariate Models

- ARIMA models each variable **in isolation**
- Ignores potential **interactions** between variables
- Cannot capture **feedback effects**

Economic Examples of Interdependence

- GDP and unemployment (Okun's law)
- Interest rates and inflation (Taylor rule)
- Stock prices and trading volume
- Exchange rates and trade balance

Vector of Variables

Let $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, \dots, Y_{Kt})'$ 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)

Definition: Weak Stationarity

A K -dimensional time series \mathbf{Y}_t is **weakly stationary** if:

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

Autocovariance Matrix

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

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

Cross-Covariance Function

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

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

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 :

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

where $\mathbf{D} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_K)$ and $\sigma_i = \sqrt{\gamma_{ii}(0)}$

For Bivariate Case

$$\mathbf{R}(h) = \begin{pmatrix} \rho_{11}(h) & \rho_{12}(h) \\ \rho_{21}(h) & \rho_{22}(h) \end{pmatrix}$$

where $\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sigma_i \sigma_j}$

Diagonal elements: usual ACFs; Off-diagonal: cross-correlations

The VAR(p) Model

Definition

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

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

where:

- \mathbf{Y}_t : $K \times 1$ vector of endogenous variables
- \mathbf{c} : $K \times 1$ vector of constants
- \mathbf{A}_i : $K \times K$ coefficient matrices
- $\boldsymbol{\varepsilon}_t$: $K \times 1$ vector of error terms with $\mathbb{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t] = \Sigma$

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

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{Y}_{t-1} + \mathbf{A}_2 \mathbf{Y}_{t-2} + \boldsymbol{\varepsilon}_t$$

For $K = 2$, the full model has $2 + 2 \times 4 + 2 \times 4 = 18$ parameters!

Written Out

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

Curse of Dimensionality

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

The Companion Form

Converting VAR(p) to VAR(1)

Any VAR(p) can be written as a VAR(1) in **companion form**:

$$\xi_t = \mathbf{A}\xi_{t-1} + \mathbf{v}_t$$

For VAR(2)

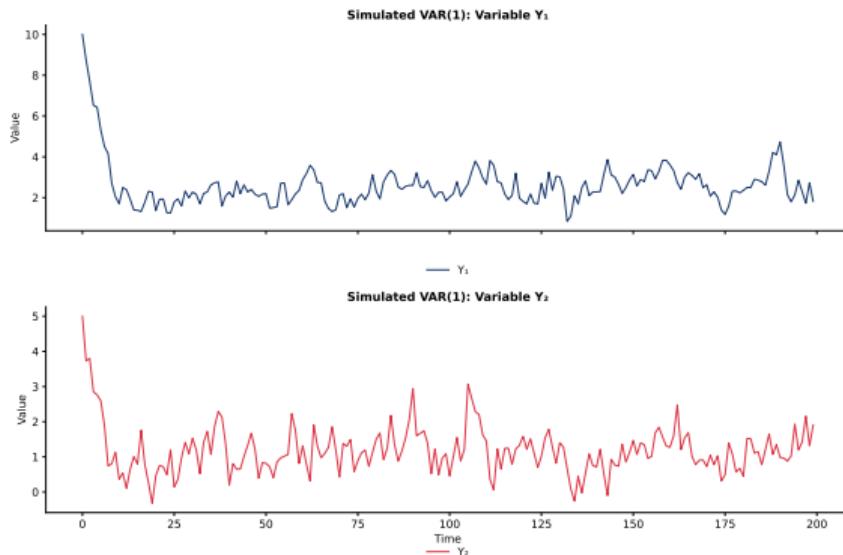
$$\underbrace{\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \end{pmatrix}}_{\xi_t} = \underbrace{\begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I}_K & \mathbf{0} \end{pmatrix}}_A \underbrace{\begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{Y}_{t-2} \end{pmatrix}}_{\xi_{t-1}} + \underbrace{\begin{pmatrix} \mathbf{e}_t \\ \mathbf{0} \end{pmatrix}}_{\mathbf{v}_t}$$

The companion matrix \mathbf{A} is $Kp \times Kp$.

Why Useful?

Stationarity, forecasting, and IRFs are easier to analyze in companion form.

Simulated VAR Process



- Two simulated series from a bivariate VAR(1) process showing interdependence
- Each variable responds to both its own past and the other variable's past
- Notice how the series co-move due to cross-equation dynamics

Stationarity of VAR

Stability Condition

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

$$\det(\mathbf{I}_K - \mathbf{A}_1 z - \mathbf{A}_2 z^2 - \cdots - \mathbf{A}_p z^p) = 0$$

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

For VAR(1)

The model is stable if all **eigenvalues** of \mathbf{A}_1 are less than 1 in absolute value.

Example: For $\mathbf{A}_1 = \begin{pmatrix} 0.5 & 0.1 \\ 0.2 & 0.3 \end{pmatrix}$, eigenvalues are $\lambda_1 = 0.6$ and $\lambda_2 = 0.2$.

Both $< 1 \Rightarrow$ stable!

Computing Eigenvalues: Example

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

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

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

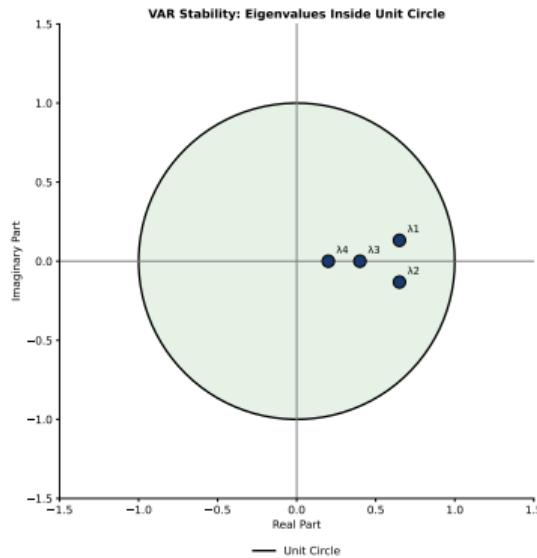
Solution

Using quadratic formula:

$$\lambda = \frac{1.3 \pm \sqrt{1.69 - 1.76}}{2} = \frac{1.3 \pm \sqrt{-0.07}}{2} = 0.65 \pm 0.132i$$

$$|\lambda| = \sqrt{0.65^2 + 0.132^2} = \sqrt{0.44} = 0.663 < 1 \quad \checkmark \text{ Stable!}$$

Stability Condition: Visual Interpretation



- Eigenvalues of the companion matrix must lie inside the unit circle
- Complex eigenvalues come in conjugate pairs
- If any eigenvalue is outside the circle, the VAR is explosive (non-stationary)

Mean of a Stationary VAR

Unconditional Mean

For a stationary VAR(1): $\mathbf{Y}_t = \mathbf{c} + \mathbf{A}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$

Taking expectations:

$$\mathbb{E}[\mathbf{Y}_t] = \mathbf{c} + \mathbf{A}\mathbb{E}[\mathbf{Y}_{t-1}]$$

Since $\mathbb{E}[\mathbf{Y}_t] = \mathbb{E}[\mathbf{Y}_{t-1}] = \boldsymbol{\mu}$ (stationarity):

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

Example

If $\mathbf{c} = \begin{pmatrix} 0.5 \\ 0.3 \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} 0.7 & 0.2 \\ -0.1 & 0.6 \end{pmatrix}$:

$$\boldsymbol{\mu} = \begin{pmatrix} 0.3 & -0.2 \\ 0.1 & 0.4 \end{pmatrix}^{-1} \begin{pmatrix} 0.5 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 2.3 \\ 1.0 \end{pmatrix}$$

Covariance Structure of VAR(1)

Variance-Covariance Matrix $\Gamma(0)$

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

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

Autocovariance at Lag h

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

This shows that autocovariances decay geometrically with the eigenvalues of \mathbf{A} .

Solving the Lyapunov Equation

Can solve by vectorization:

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

where \otimes denotes the Kronecker product.

OLS Estimation

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

$$\hat{\mathbf{A}} = \left(\sum_{t=1}^T \mathbf{Y}_{t-1} \mathbf{Y}'_{t-1} \right)^{-1} \left(\sum_{t=1}^T \mathbf{Y}_{t-1} \mathbf{Y}'_t \right)$$

This is efficient because all equations have the **same regressors**.

Covariance Matrix

$$\hat{\Sigma} = \frac{1}{T - Kp - 1} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}'_t$$

The errors ε_{1t} and ε_{2t} may be **contemporaneously correlated**.

Information Criteria

Choose p that minimizes:

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

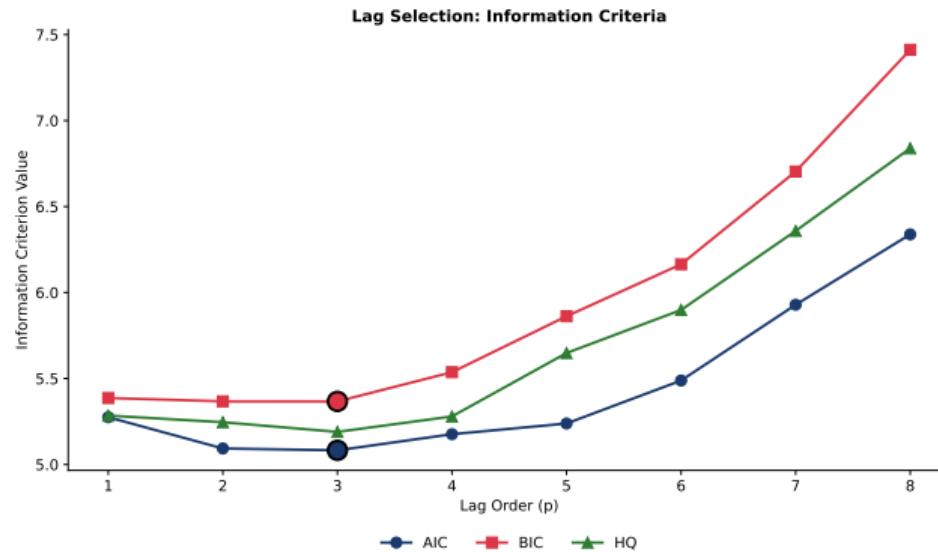
$$BIC(p) = \ln |\hat{\Sigma}_p| + \frac{pK^2 \ln T}{T}$$

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

Guidelines

- AIC tends to select **larger** models (better for forecasting)
- BIC tends to select **smaller** models (consistent selection)
- Start with maximum p_{max} based on data frequency (e.g., 4 for quarterly, 12 for monthly)

Lag Selection: Example



- Information criteria values for different lag orders
- AIC and BIC may suggest different optimal lags
- Lower values indicate better model fit (penalized by complexity)

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: $LR = T(\ln |\hat{\Sigma}_R| - \ln |\hat{\Sigma}_U|) \sim \chi_r^2$
where 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:

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

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 $a_{12}^{(1)} = a_{12}^{(2)} = \dots = a_{12}^{(p)} = 0$.

Testing for Granger Causality

Hypothesis Test

H_0 : Y_2 does **not** Granger-cause Y_1

$$H_0 : a_{12}^{(1)} = a_{12}^{(2)} = \dots = a_{12}^{(p)} = 0$$

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



Unidirectional: $X \rightarrow Y$



Bidirectional: $X \leftrightarrow Y$



Unidirectional: $Y \rightarrow X$



No causality

Economic Examples

- Money \rightarrow Output? (monetarist view)
- Stock prices \leftrightarrow Trading volume (bidirectional)
- Weather \rightarrow Crop yields (unidirectional, obvious)

Cross-Correlation Function

Definition 1 (Cross-Correlation Function)

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

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

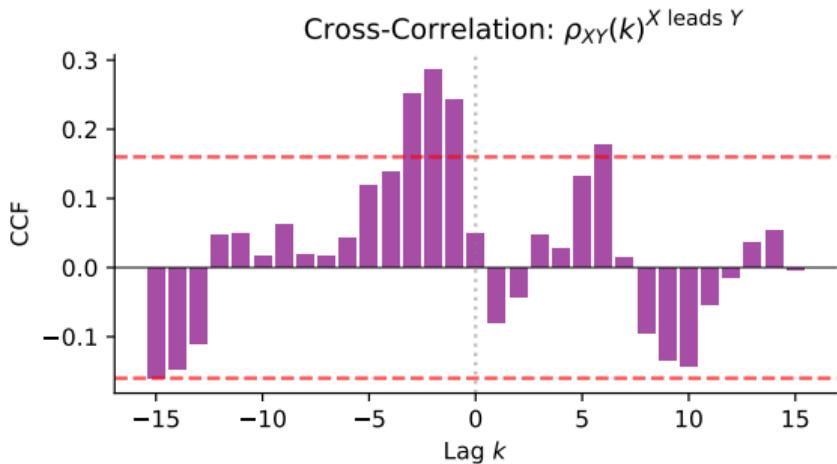
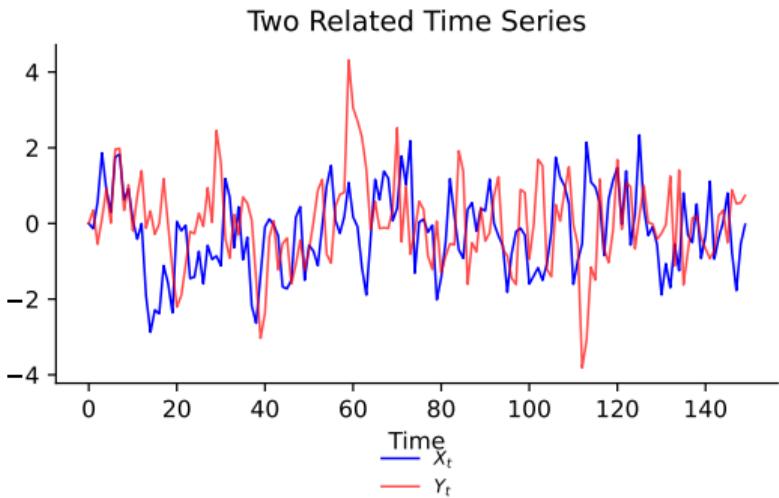
Interpretation

- $\rho_{XY}(k) > 0$ 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.

Cross-Correlation: Visual Illustration



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

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

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

$F_{0.05}(2, 95) = 3.09 \Rightarrow \text{Reject } H_0$: 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 $\text{VAR}(p + d_{max})$ 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:

$$\mathbb{E}[Y_t | \Omega_{t-1}, X_t] \neq \mathbb{E}[Y_t | \Omega_{t-1}]$$

where Ω_{t-1} contains all past information.

Testing in VAR

Test whether $\sigma_{12} \neq 0$ in the covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

If $\sigma_{12} = 0$: no instantaneous causality

Interpretation

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

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:

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

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

Computing IRFs for VAR(1)

For VAR(1): $\mathbf{Y}_t = \mathbf{c} + \mathbf{A}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$

The impulse response matrices are:

$$\Phi_0 = \mathbf{I}_K, \quad \Phi_1 = \mathbf{A}, \quad \Phi_2 = \mathbf{A}^2, \quad \dots, \quad \Phi_h = \mathbf{A}^h$$

Interpretation

$[\Phi_h]_{ij}$ = Effect on Y_i at time $t+h$ of a unit shock to Y_j at time t

For stable VAR: $\Phi_h \rightarrow \mathbf{0}$ as $h \rightarrow \infty$ (shocks die out)

Computing IRFs for General VAR(p)

Recursive Formula for VAR(p)

For $\mathbf{Y}_t = \mathbf{c} + \mathbf{A}_1\mathbf{Y}_{t-1} + \mathbf{A}_2\mathbf{Y}_{t-2} + \cdots + \mathbf{A}_p\mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t$:

$$\Phi_h = \sum_{j=1}^{\min(h,p)} \mathbf{A}_j \Phi_{h-j}, \quad h = 1, 2, 3, \dots$$

with $\Phi_0 = \mathbf{I}_K$ and $\Phi_h = \mathbf{0}$ for $h < 0$.

Example: VAR(2) IRFs

- $\Phi_0 = \mathbf{I}_K$
- $\Phi_1 = \mathbf{A}_1 \Phi_0 = \mathbf{A}_1$
- $\Phi_2 = \mathbf{A}_1 \Phi_1 + \mathbf{A}_2 \Phi_0 = \mathbf{A}_1^2 + \mathbf{A}_2$
- $\Phi_3 = \mathbf{A}_1 \Phi_2 + \mathbf{A}_2 \Phi_1 = \mathbf{A}_1(\mathbf{A}_1^2 + \mathbf{A}_2) + \mathbf{A}_2 \mathbf{A}_1$

Orthogonalized IRFs

Problem: Correlated Errors

If Σ is not diagonal, shocks ε_{1t} and ε_{2t} are correlated.

A shock to "Y₁" also involves a shock to "Y₂".

Solution: Cholesky Decomposition

Factor $\Sigma = \mathbf{P}\mathbf{P}'$ where \mathbf{P} is lower triangular.

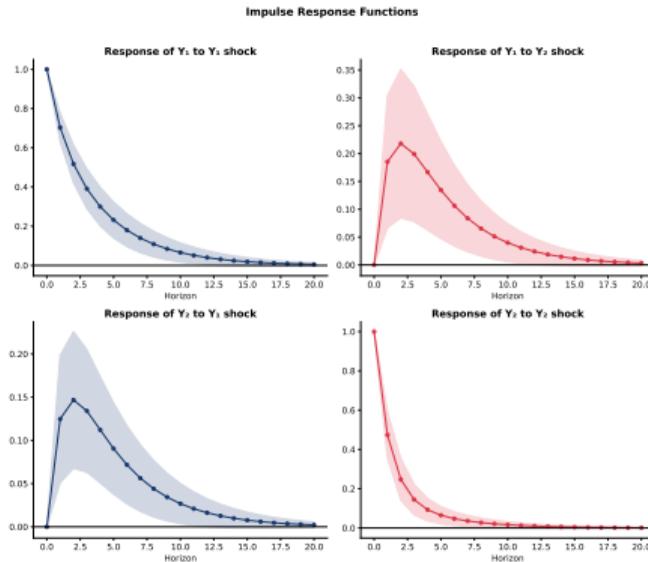
Define orthogonalized shocks: $\mathbf{u}_t = \mathbf{P}^{-1}\varepsilon_t$ with $\mathbb{E}[\mathbf{u}_t\mathbf{u}_t'] = \mathbf{I}$

Orthogonalized IRFs: $\Theta_h = \Phi_h \mathbf{P}$

Ordering Matters!

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

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

IRF Numerical Example

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

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

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

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

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 :

$$\Psi_H = \sum_{h=0}^H \Phi_h$$

Measures the **total accumulated effect** of a shock.

Long-Run Multiplier

For stable VAR: $\Psi_\infty = (\mathbf{I}_K - \mathbf{A}_1 - \mathbf{A}_2 - \cdots - \mathbf{A}_p)^{-1}$

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 $\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_p$, so they have **sampling uncertainty**.

Methods for Confidence Bands

- ① **Asymptotic:** Use delta method to derive standard errors
- ② **Monte Carlo:** Simulate from asymptotic distribution of $\hat{\mathbf{A}}$
- ③ **Bootstrap:** Resample residuals and re-estimate VAR

Bootstrap Procedure

- ① Estimate VAR, save residuals $\{\hat{\epsilon}_t\}$
- ② Draw with replacement to create $\{\hat{\epsilon}_t^*\}$
- ③ Generate bootstrap sample using estimated VAR
- ④ Re-estimate and compute IRFs
- ⑤ 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

$$\mathbf{B}_0 \mathbf{Y}_t = \boldsymbol{\Gamma}_0 + \mathbf{B}_1 \mathbf{Y}_{t-1} + \cdots + \mathbf{B}_p \mathbf{Y}_{t-p} + \mathbf{u}_t$$

where \mathbf{u}_t are **structural shocks** with $\mathbb{E}[\mathbf{u}_t \mathbf{u}_t'] = \mathbf{I}_K$

Relationship to Reduced Form

$$\varepsilon_t = \mathbf{B}_0^{-1} \mathbf{u}_t \quad \Rightarrow \quad \boldsymbol{\Sigma} = \mathbf{B}_0^{-1} (\mathbf{B}_0^{-1})'$$

Identification in SVAR

The Identification Problem

Σ has $K(K + 1)/2$ unique elements, but \mathbf{B}_0^{-1} has K^2 elements.

Need $K(K - 1)/2$ additional restrictions!

Common Identification Schemes

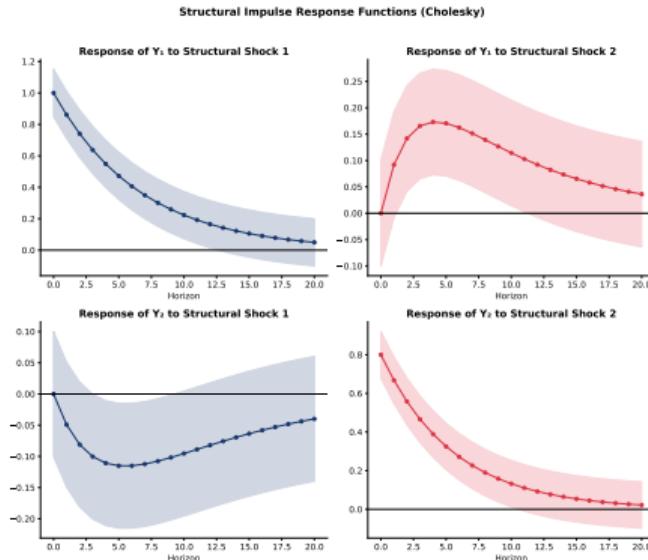
- ① **Short-run restrictions:** Zero impact effects (Cholesky)
- ② **Long-run restrictions:** Zero long-run effects (Blanchard-Quah)
- ③ **Sign restrictions:** Inequality constraints on IRFs
- ④ **External instruments:** Use outside information

Example: Cholesky (Recursive) Ordering

$$\text{For } K = 2: \mathbf{B}_0^{-1} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix}$$

Variable 1 doesn't respond to shock 2 contemporaneously.

Structural IRF Example



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

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

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

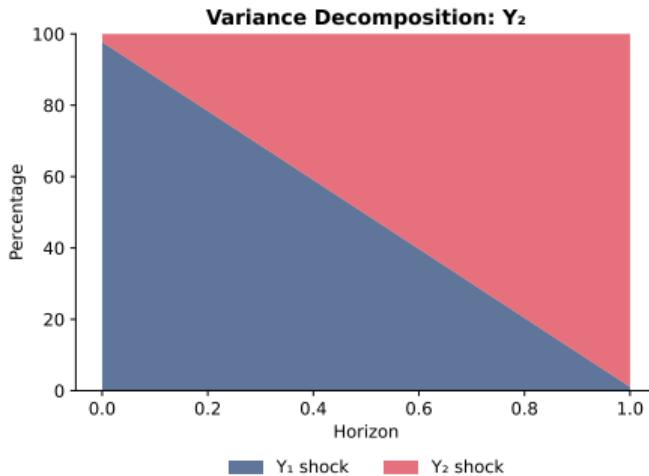
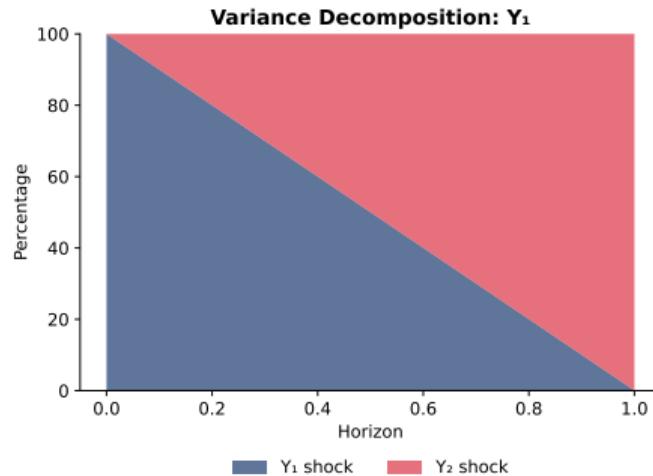
This gives the **percentage** of Y_i 's h -step forecast variance explained by shocks to Y_j .

Properties

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

FEVD: Example

Forecast Error Variance Decomposition



- 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

FEVD: Numerical Example

Computing FEVD for Bivariate VAR

Using orthogonalized IRFs Θ_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

Definition

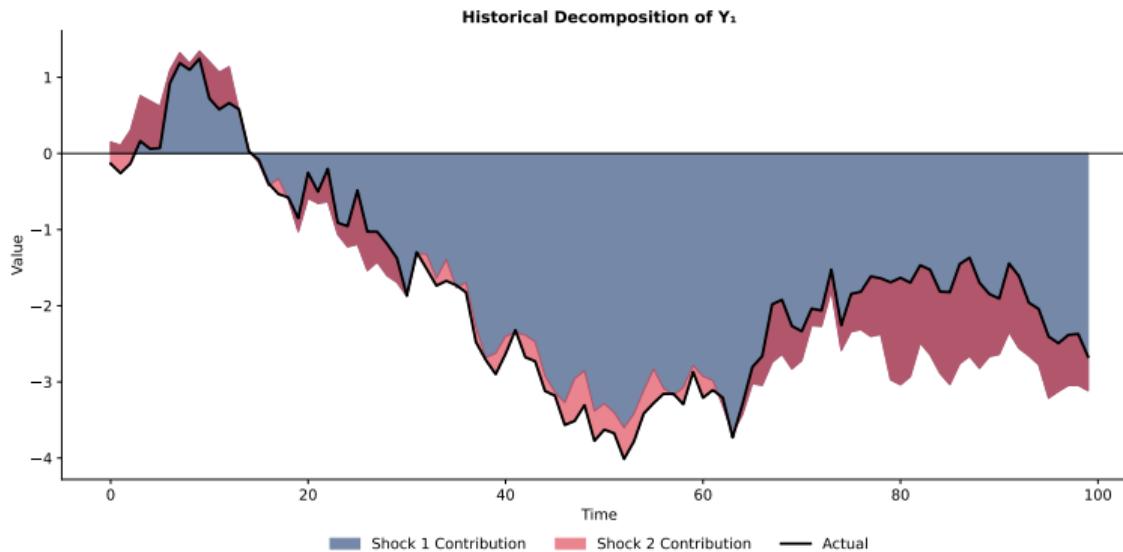
Historical decomposition breaks down each observed value into contributions from each structural shock:

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

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

Historical Decomposition: Example



- Each color represents the contribution of a different structural shock
- Stacked contributions sum to the actual observed deviation from mean
- Helps identify which shocks drove historical episodes

What to Check

After estimating VAR, verify that residuals $\hat{\epsilon}_t$ 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)

$$Q_h = T(T+2) \sum_{j=1}^h \frac{1}{T-j} \text{tr}(\hat{\mathbf{C}}_j' \hat{\mathbf{C}}_0^{-1} \hat{\mathbf{C}}_j \hat{\mathbf{C}}_0^{-1})$$

where $\hat{\mathbf{C}}_j = \frac{1}{T} \sum_{t=j+1}^T \hat{\varepsilon}_t \hat{\varepsilon}'_{t-j}$

Under H_0 (no autocorrelation): $Q_h \sim \chi^2_{K^2(h-p)}$

Breusch-Godfrey LM Test

- ① Regress $\hat{\varepsilon}_t$ on $\hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-h}$ 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 autoregressive conditional heteroskedasticity 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:

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

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)

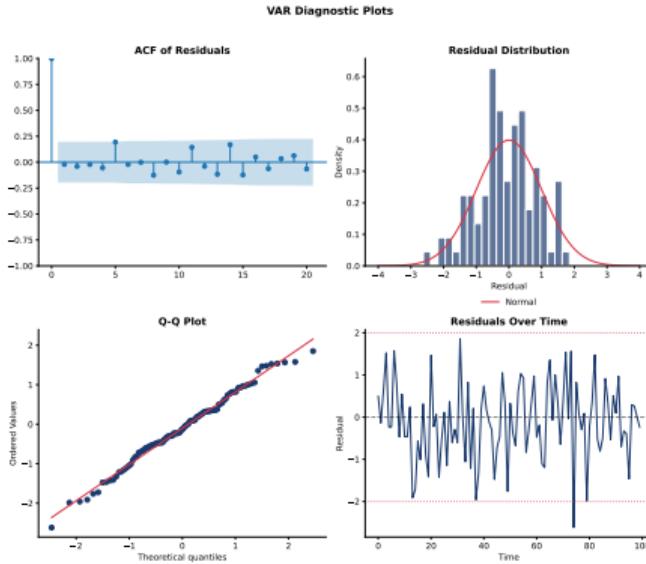
Transform residuals and test joint skewness and kurtosis:

$$DH = s'_1(\Omega^{-1/2})'(\Omega^{-1/2})s_1 + s'_2(\Omega^{-1/2})'(\Omega^{-1/2})s_2 \sim \chi^2_{2K}$$

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

Point Forecasts from VAR

Iterative Forecasting

For VAR(1): $\mathbf{Y}_t = \mathbf{c} + \mathbf{A}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$

1-step forecast: $\hat{\mathbf{Y}}_{T+1|T} = \mathbf{c} + \mathbf{A}\mathbf{Y}_T$

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

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

Direct Formula

$$\hat{\mathbf{Y}}_{T+h|T} = (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^{h-1})\mathbf{c} + \mathbf{A}^h \mathbf{Y}_T$$

For stable VAR: converges to $\mu = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{c}$ as $h \rightarrow \infty$

h -Step Forecast Error

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

Mean Squared Error Matrix

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

Key Insight

- MSE increases with horizon h
- For stable VAR: MSE converges to unconditional variance $\boldsymbol{\Gamma}(0)$
- Long-horizon forecasts \rightarrow unconditional mean with uncertainty $= \boldsymbol{\Gamma}(0)$

Forecast Confidence Intervals

Constructing Intervals

For normally distributed errors, $(1 - \alpha)$ confidence interval:

$$\hat{Y}_{i,T+h|T} \pm z_{\alpha/2} \sqrt{[\text{MSE}(\hat{Y}_{T+h|T})]_{ii}}$$

Joint Confidence Regions

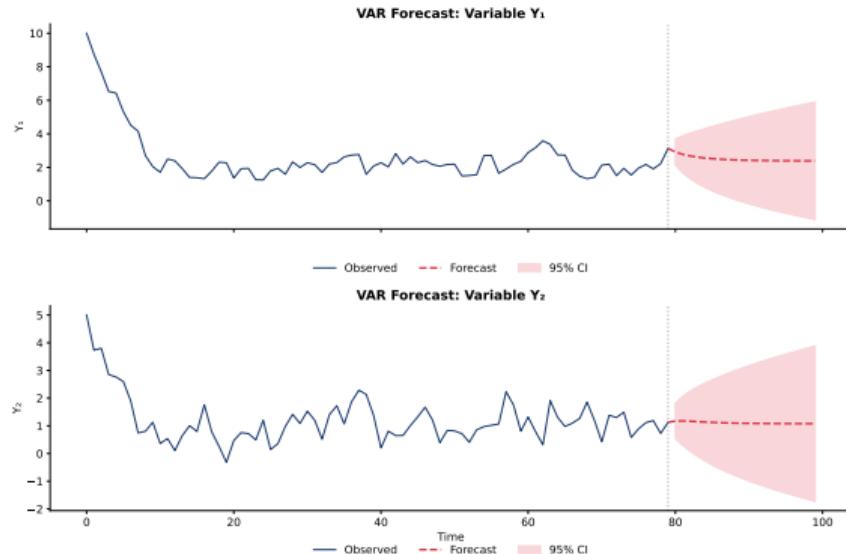
For multiple variables, use ellipsoids:

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

Note

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

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

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.

Example: GDP and Unemployment

Okun's Law

There is a negative relationship between GDP growth and unemployment:

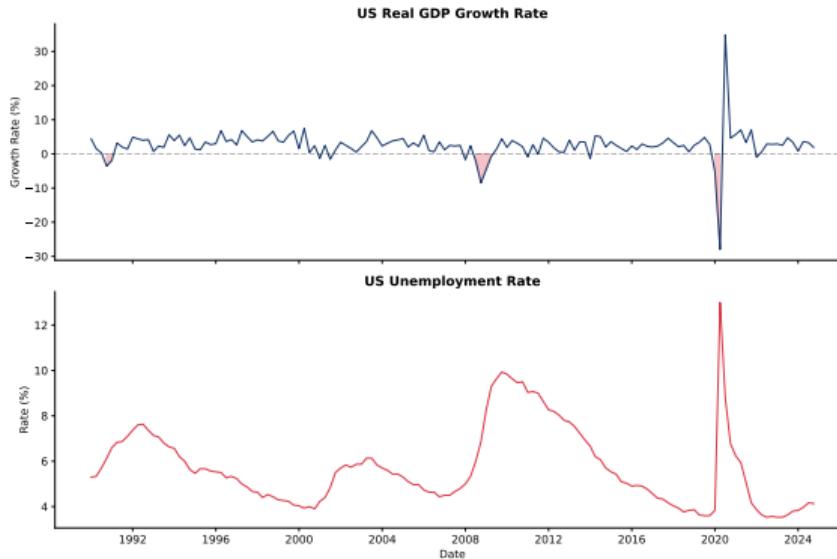
$$\Delta U_t \approx -\beta(\Delta Y_t - \bar{g})$$

where \bar{g} is trend GDP growth and $\beta \approx 0.4$.

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?

GDP and Unemployment: Data



- GDP growth and unemployment rate show clear negative correlation (Okun's Law)
- Both series exhibit cyclical patterns related to business cycle fluctuations
- This bivariate system is ideal for VAR analysis and Granger causality testing

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

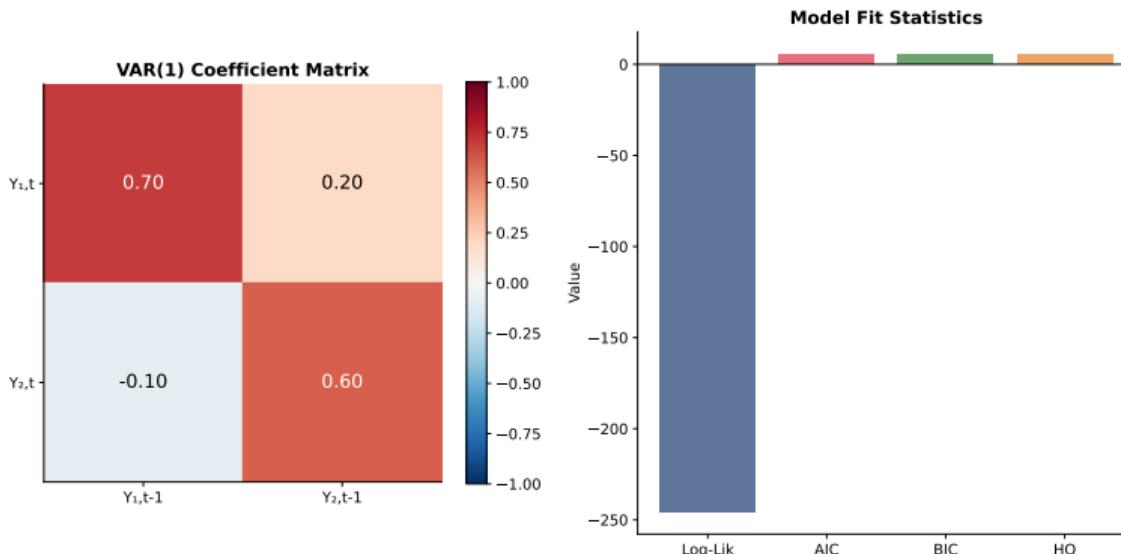
- OLS equation by equation
- Check stability (eigenvalues)

④ Analysis

- Granger causality tests
- Impulse response functions
- Variance decomposition

⑤ Forecasting

Estimated VAR Results



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

Granger Causality Results

Test Results: GDP and Unemployment

Null Hypothesis	F-statistic	df	p-value	Decision
GDP $\not\rightarrow$ Unemployment	8.42	(2, 95)	0.0004	Reject
Unemployment $\not\rightarrow$ 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: $\text{GDP} \rightarrow \text{Unemployment}$

VAR in Python (statsmodels)

```
from statsmodels.tsa.api import VAR
from statsmodels.tsa.stattools import grangercausalitytests

# Fit VAR model
model = VAR(data)
results = model.fit(maxlags=4, ic='aic')

# Granger causality test
granger_test = grangercausalitytests(data[['Y1', 'Y2']],
                                      maxlag=4)

# Impulse response functions
irf = results.irf(periods=20)
irf.plot()

# Variance decomposition
fevd = results.fevd(periods=20)
fevd.plot()
```

VAR in R (vars package)

```
library(vars)

# Select optimal lag order
lag_select <- VARselect(data, lag.max = 8)
print(lag_select$selection)

# Fit VAR model
var_model <- VAR(data, p = 2, type = "const")
summary(var_model)

# Granger causality test
causality(var_model, cause = "GDP")

# Impulse response functions
irf_results <- irf(var_model, n.ahead = 20, boot = TRUE)
plot(irf_results)

# Forecast error variance decomposition
fevd_results <- fevd(var_model, n.ahead = 20)
plot(fevd_results)
```

Example: Monetary Policy Analysis

Three-Variable VAR

Study the monetary transmission mechanism with:

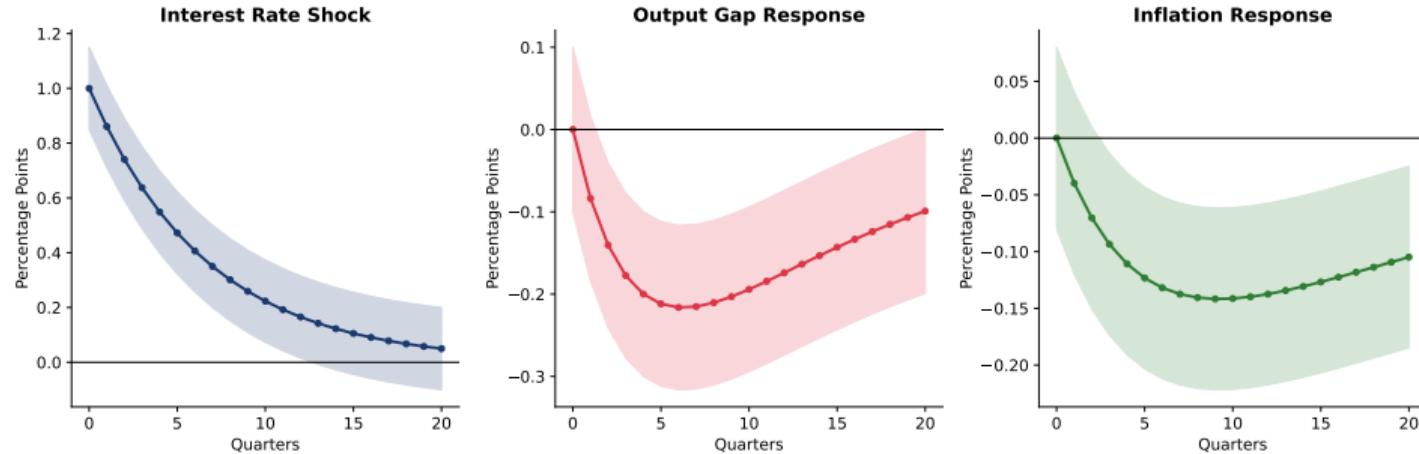
- Y_1 : Output gap (GDP deviation from trend)
- Y_2 : Inflation rate
- Y_3 : Interest rate (policy instrument)

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?

Monetary Policy VAR: IRFs

Monetary Policy Transmission: Response to 1pp Interest Rate Shock



- 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

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 in each variable
- Transform to stationary if needed (differences, logs)
- Check for outliers and structural breaks

Model Specification

- Select lag order using AIC/BIC
- Estimate VAR by OLS
- Check stability (eigenvalues inside unit circle)

Post-Estimation

- Test residuals for autocorrelation
- Test for ARCH effects
- Test for normality
- Compute IRFs, FEVDs, Granger tests

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 ($T > 50$)

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?

Quiz Question 1

Question

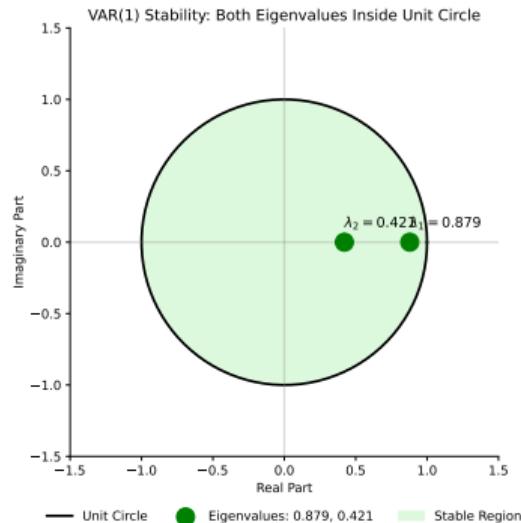
For a VAR(1) model with coefficient matrix $\mathbf{A} = \begin{pmatrix} 0.8 & 0.3 \\ 0.1 & 0.5 \end{pmatrix}$, is the model stable?

- 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 Σ

Quiz Question 1: Answer

Correct Answer: (B) Eigenvalues inside unit circle

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



Quiz Question 2

Question

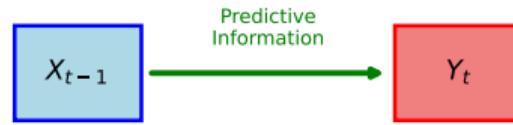
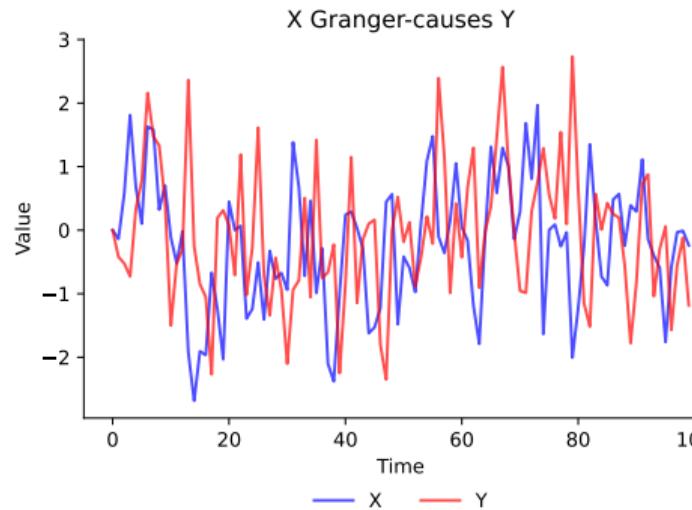
If X Granger-causes Y at the 5% significance level, which of the following statements is TRUE?

- 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

Quiz Question 2: Answer

Correct Answer: (B) Predictive information

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



Past X helps predict Y
(beyond Y's own past)

Quiz Question 3

Question

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

- 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

Quiz Question 3: Answer

Correct Answer: (C) Contemporaneous responses

Ordering determines which variables respond immediately to which shocks.

Ordering: (GDP, Interest Rate)



GDP shock → IR responds at t=0
IR shock → GDP responds at t=1

Ordering: (Interest Rate, GDP)



IR shock → GDP responds at t=0
GDP shock → IR responds at t=1

Quiz Question 4

Question

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

- A 4
- B 6
- C 8
- D 10

Quiz Question 4: Answer

Correct Answer: (B) 6 parameters

Detailed Count

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

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

- Constant vector \mathbf{c} : $K = 2$ parameters
- Coefficient matrix \mathbf{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 Σ)

Quiz Question 5

Question

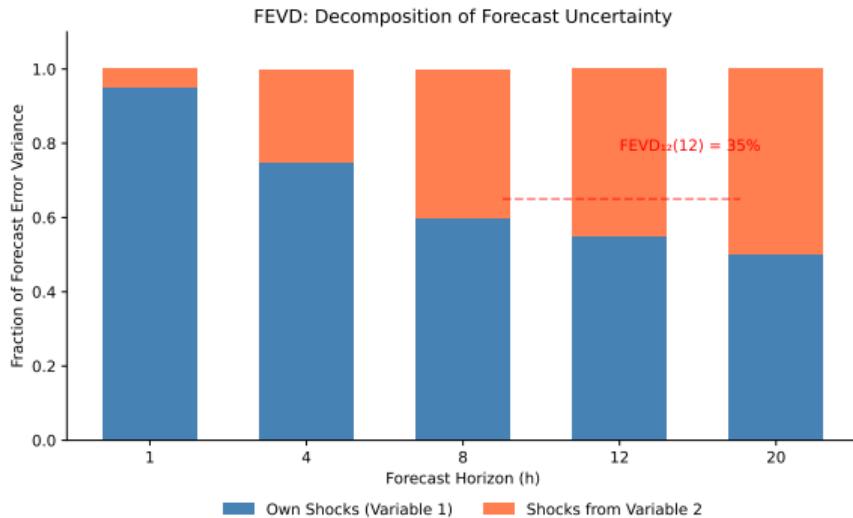
What does $\text{FEVD}_{12}(h) = 0.35$ mean?

- 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

Quiz Question 5: Answer

Correct Answer: (B) Forecast error variance decomposition

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



References

-  Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.
-  Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
-  Sims, C.A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1-48.
-  Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models. *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.