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. Practical Example
7. Summary

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

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)

The VAR(p) Model

Definition

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

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

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
- ε_t: $K \times 1$ vector of error terms with $\mathbb{E}[\varepsilon_t] = \mathbf{0}$, $\mathbb{E}[\varepsilon_t \varepsilon_t'] = \mathbf{\Sigma}$

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!

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!

Estimation of VAR

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

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

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

Lag Order Selection

Information Criteria

Choose p that minimizes:

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

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

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

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)

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)

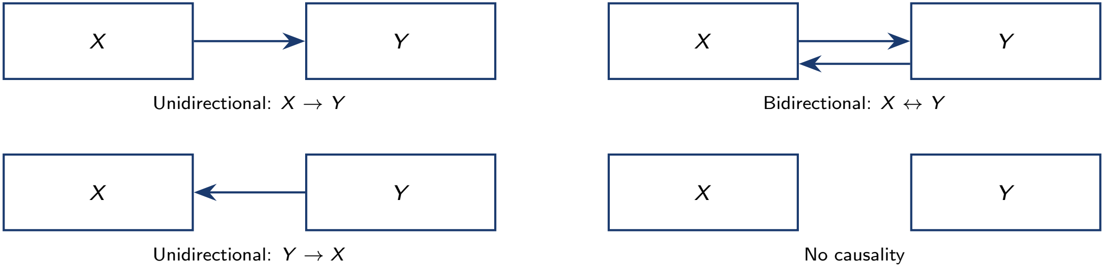
Test Statistic: Wald Test

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

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

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

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): $\mathbf{Y}_t = \mathbf{c} + \mathbf{A}\mathbf{Y}_{t-1} + \varepsilon_t$

The impulse response matrices are:

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

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: $\mathbf{\Phi}_h \to \mathbf{0}$ as $h \to \infty$ (shocks die out)

Orthogonalized IRFs

Problem: Correlated Errors

If $\mathbf{\Sigma}$ 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 $\mathbf{\Sigma} = \mathbf{P}\mathbf{P}'$ where $\mathbf{P}$ is lower triangular.
Define orthogonalized shocks: $\mathbf{u}_t = \mathbf{P}^{-1}\boldsymbol{\varepsilon}_t$ with $\mathbb{E}[\mathbf{u}_t\mathbf{u}_t'] = \mathbf{I}$

Orthogonalized IRFs: $\mathbf{\Theta}_h = \mathbf{\Phi}_h\mathbf{P}$

Ordering Matters!

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

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

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

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)

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?

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**

Fit VAR model model = VAR(data) results = model.fit(maxlags=4, ic='aic')
Granger causality test granger$_t est = grangercausalitytests(data[['Y1',' Y2']], maxlag = 4)$
Impulse response functions irf = results.irf(periods=20) irf.plot()
Variance decomposition fevd = results.fevd(periods=20) fevd.plot()

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 (or differencing)

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)

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

Questions?