

Time Series Analysis

Chapter 6: VAR models 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

Acronyms used in this material

- **AI**: Artificial Intelligence
- **AIC**: Akaike Information Criterion
- **AR**: AutoRegressive (model)
- **ARMA**: AutoRegressive Moving Average (model)
- **BET**: Bucharest Exchange Trading (index)
- **BIC**: Bayesian Information Criterion
- **BNR**: Banca Națională a României (the National Bank of Romania)
- **CCF**: Cross-Correlation Function
- **COVID-19**: COronaVIrus Disease 2019
- **DM**: Diebold–Mariano (test)
- **DSGE**: Dynamic Stochastic General Equilibrium (model)
- **EODHD**: EOD Historical Data (market-data provider)
- **EU**: European Union
- **FEVD**: Forecast Error Variance Decomposition
- **FRED**: Federal Reserve Economic Data
- **GDP**: Gross Domestic Product
- **GLS**: Generalised Least Squares
- **HICP**: Harmonised Index of Consumer Prices
- **HQ**: Hannan–Quinn (information criterion)
- **IRF**: Impulse Response Function
- **MA**: Moving Average
- **MSE**: Mean Squared Error
- **OLS**: Ordinary Least Squares
- **RMSE**: Root Mean Squared Error
- **ROBOR**: Romanian Interbank Offer Rate
- **RSS**: Residual Sum of Squares
- **SE**: Standard Error
- **SVAR**: Structural Vector AutoRegression
- **US**: United States
- **VAR**: Vector AutoRegression
- **VECM**: Vector Error Correction Model

Today's question and route

- **Question**: when several series move together (output, prices, interest rates; three stock markets), how do we model them jointly, and what can we learn about who leads whom?
  - one equation per variable, each with the past of all variables: the vector autoregression (VAR)
- **Route** of the chapter
  - from one series to several: vector processes and cross-correlations; the VAR(p) model and its stability
  - estimation, lag selection and diagnostics; Granger causality and its pitfalls
  - impulse responses, variance decomposition, forecasts; the structural VAR idea of Sims (1980)
- We build on Chapter 2 (AR models) and Chapter 3 (stationarity tests)

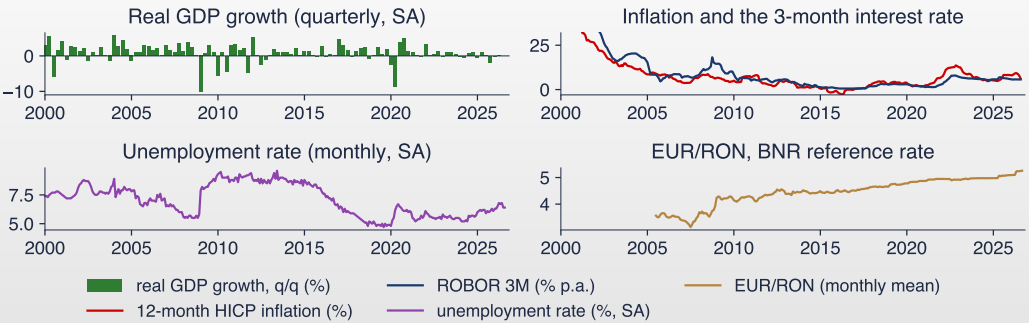
Learning outcomes

- Write a VAR(p) in matrix form and equation by equation, and count its parameters
- Check stability with the eigenvalues of the companion matrix, and compute the mean of a stable VAR
- Estimate a VAR by OLS, choose p with AIC, BIC and HQ, and check the residuals
- Test Granger causality and explain what it does not prove
- Compute and read orthogonalised impulse responses and variance decompositions, and explain why the ordering matters
- Forecast with a VAR and compare it fairly with univariate benchmarks

Reading and tools

- Textbook: Huang and Petukhina (2022), Ch. 7 (multivariate time series)
  - companion, free online: Hyndman and Athanasopoulos (2021), Sec. 12.3 (vector autoregressions)
- Theory: Lütkepohl (2005), Ch. 2–4; Hamilton (1994), Ch. 11; an accessible survey: Stock and Watson (2001)
- Python Quantlets of this chapter: Quantlets/Ch_06
  - `VAR` from `statsmodels`: `select_order`, `fit`, `test_whiteness`, `irf`, `fevd`, `forecast_interval`
- Lecture notebook: open in Google Colab
- Video course: Applied Time Series Analysis with Python

Romania: four macro series that move together



⊡ Real GDP growth (q/q) and unemployment rate, seasonally adjusted; 12-month HICP inflation; ROBOR 3M, the 3-month interbank rate; source: Eurostat; EUR/RON: the BNR reference rate

TSA_ch6_multivariate_data

Interpreting the Romanian series

- Inflation and ROBOR 3M rise and fall together: correlation 0.66 over 2005Q1–2026Q2 (quarterly means)
  - inflation peaked at 13.6% in November 2022; ROBOR 3M peaked at 18.2% in October 2008 (the 2008 liquidity squeeze)
- GDP growth is noisy, with two deep falls: -10.2% in 2009Q1 and -8.6% in 2020Q2
  - latest values: inflation 6.1% (August 2026), ROBOR 3M 5.69%, unemployment 6.4%, EUR/RON 5.26
- Questions for this chapter
  - Does inflation help to forecast the interest rate?
  - How does ROBOR react to an inflation surprise, and for how long?

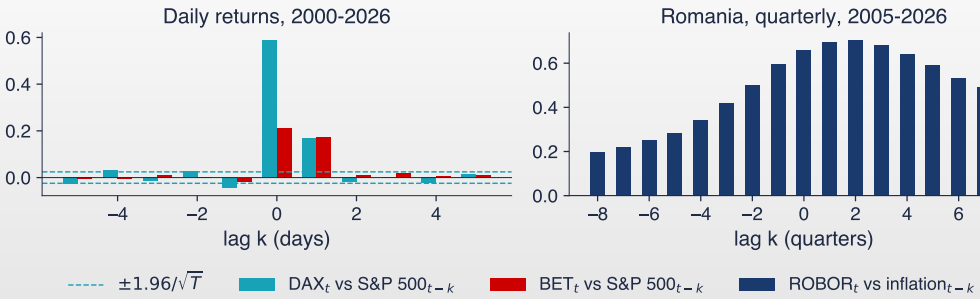
Vector time series

- **Definition**: $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{Kt})'$ collects K series observed at the same dates
  - example: $\mathbf{Y}_t = (g_t, \pi_t, i_t)'$, GDP growth, inflation and ROBOR 3M of quarter t; $K = 3$
  - Y_{jt}: the value of series j at date t; the prime $'$ denotes the transpose (a column vector)
- **Mean vector**: $\boldsymbol{\mu} = E[\mathbf{Y}_t]$, $K \times 1$
  - **autocovariance matrix** at lag h: $\boldsymbol{\Gamma}(h) = \text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-h}) = E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-h} - \boldsymbol{\mu})']$, $K \times K$
  - diagonal: the autocovariances of each series (Chapter 1); off-diagonal: **cross-covariances** $\gamma_{jk}(h) = \text{Cov}(Y_{jt}, Y_{k,t-h})$
- **Weak stationarity**: $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}(h)$ do not depend on t
  - $\boldsymbol{\Gamma}(h)$ is not symmetric, but $\boldsymbol{\Gamma}(-h) = \boldsymbol{\Gamma}(h)'$: "$Y_1$ leads Y_2" and "Y_2 leads Y_1" are different statements

The cross-correlation function

- **Definition**: $\rho_{yx}(k) = \text{Corr}(y_t, x_{t-k}) = \gamma_{yx}(k)/(\sigma_y \sigma_x)$, for $k = 0, \pm 1, \pm 2, \ldots$
  - the cross-correlation function (CCF): $k > 0$: past x with today's y (x leads); $k < 0$: y leads
  - $\gamma_{yx}(k) = \text{Cov}(y_t, x_{t-k})$: the cross-covariance; σ_y, σ_x: the standard deviations of the two series
- Sample version: $\hat{\rho}_{yx}(k)$, the correlation of the pairs (y_t, x_{t-k})
  - if the two series are independent white noises, $\hat{\rho}_{yx}(k) \approx N(0, 1/T)$, with T the number of observations: bands $\pm 1.96/\sqrt{T}$
  - with autocorrelated series the bands are too narrow: spurious lead–lag patterns (compare Chapter 3, spurious regression)
- The CCF looks at one pair and one lag at a time; a VAR looks at all variables and all lags together

Cross-correlations: markets and Romanian macro data



- Left: daily log returns (EODHD), 6 433 common trading days, 2000–2026; bars: Corr(DAX_t, $\text{S\&P}_{t-k}$) and Corr(BET_t, $\text{S\&P}_{t-k}$). Right: quarterly ROBOR 3M against lags of inflation

TSA_ch6_multivariate_data

Interpreting the cross-correlations

- BET with the S&P 500: 0.21 on the same day and 0.17 with the previous day ($k = 1$); with $k = -1$: -0.02
  - the band is ± 0.024: yesterday's New York return is clearly visible in today's Bucharest and Frankfurt returns (DAX at $k = 1$: 0.17); the reverse link is much weaker (DAX at $k = -1$: -0.046, just outside the band)
  - reason: Wall Street closes at 22:00 in Frankfurt, after the European markets; its news reaches them the next morning
- ROBOR with inflation: the largest correlation, 0.70, is at $k = 2$ quarters (inflation leads)
  - but every value is large (0.34 at $k = -4$, 0.64 at $k = 4$): two persistent series; the CCF alone cannot separate lead from persistence

The advantage of one model for all series

- Univariate models (Chapters 2–3) forecast each series from its own past
  - they ignore that last quarter's inflation may help to forecast this quarter's interest rate
- **Feedback**: the central bank reacts to inflation, inflation reacts to the interest rate
  - a single regression of π_t on i_t treats i_t as given (exogenous); here both are **endogenous**, determined inside the system
- A VAR treats all variables symmetrically: each depends on the past of all
  - no economic theory is needed to write it down; theory comes back when we interpret shocks (Section 7)

Recap: From one series to several

- A vector series $\mathbf{Y}_t$ has a mean vector and autocovariance matrices $\mathbf{\Gamma}(h)$, with $\mathbf{\Gamma}(-h) = \mathbf{\Gamma}(h)'$
- The CCF $\rho_{yx}(k)$ shows lead–lag patterns; $k > 0$: x leads
- S&P 500 returns lead DAX and BET returns by one day
- Feedback between variables calls for a joint model

The VAR(p) model

- **Definition**: $\mathbf{Y}_t = \mathbf{c} + \mathbf{A_1}\mathbf{Y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t$
  - $\mathbf{c}$: $K \times 1$ vector of constants; $\mathbf{A_1}, \ldots, \mathbf{A}_p$: $K \times K$ coefficient matrices
  - $\boldsymbol{\varepsilon}_t$: vector white noise: $E[\boldsymbol{\varepsilon}_t] = \mathbf{0}$, $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{\Sigma}$, $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s'] = \mathbf{0}$ for $t \neq s$
- $\boldsymbol{\Sigma}$ is not diagonal in general: shocks of different equations are correlated **at the same date**
  - no lagged errors are correlated: all the dynamics is in the matrices $\mathbf{A}_j$
- With $K = 1$ we recover the AR(p) of Chapter 2
  - this form is called the **reduced form**: it describes the data, not the economic shocks (Section 10)

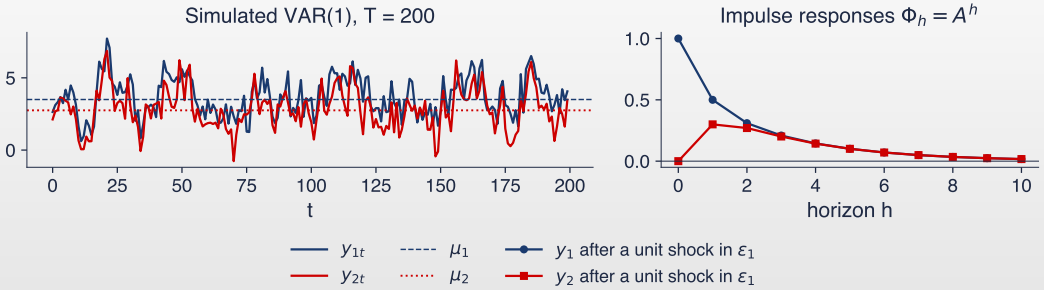
A bivariate VAR(1), equation by equation

- In matrix form:
  - $\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$
- Equation by equation:
  - $y_{1t} = c_1 + a_{11} y_{1,t-1} + a_{12} y_{2,t-1} + \varepsilon_{1t}$
  - $y_{2t} = c_2 + a_{21} y_{1,t-1} + a_{22} y_{2,t-1} + \varepsilon_{2t}$
- Reading the coefficients: a_{jk} = effect of $y_{k,t-1}$ on y_{jt} (row: equation; column: regressor)
  - $a_{12} = a_{21} = 0$: two separate AR(1) models; $a_{12} \neq 0$: the past of y_2 helps to predict y_1

Worked example: a VAR(1)

- Model: $\mathbf{c} = (1.2, 0.6)'$, $\mathbf{A} = \begin{pmatrix} 0.5 & 0.2 \\ 0.3 & 0.4 \end{pmatrix}$, $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$
  - $y_{1t} = 1.2 + 0.5y_{1,t-1} + 0.2y_{2,t-1} + \varepsilon_{1t}$; $\quad y_{2t} = 0.6 + 0.3y_{1,t-1} + 0.4y_{2,t-1} + \varepsilon_{2t}$
- Reading: $a_{21} = 0.3$: one unit more of y_1 last period raises y_2 by 0.3 today, other things fixed
  - the shocks are correlated: $\text{Corr}(\varepsilon_{1t}, \varepsilon_{2t}) = 0.5$
- Forecasts from $\mathbf{Y}_T = (5, 2)'$: $\hat{\mathbf{Y}}_{T+1} = \mathbf{c} + \mathbf{A}\mathbf{Y}_T = \begin{pmatrix} 4.10 \\ 2.90 \end{pmatrix}$, $\quad \hat{\mathbf{Y}}_{T+2} = \mathbf{c} + \mathbf{A}\hat{\mathbf{Y}}_{T+1} = \begin{pmatrix} 3.83 \\ 2.99 \end{pmatrix}$
  - the forecasts move towards the mean $\boldsymbol{\mu} = (3.50, 2.75)'$ (next section)

The worked example, simulated



- Left: 200 periods of the VAR(1) of the previous slide, with its two means; right: the responses of y_1 and y_2 to a unit shock in ε_1 at $h = 0$, the columns of $\mathbf{A}^h$

TSA_ch6_var_stability

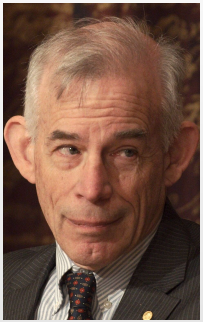
Interpreting the simulated VAR(1)

- The two series wander around their means and move together: sample correlation 0.75
  - two sources of co-movement: the correlated shocks (Corr $= 0.5$) and the cross effects a_{12}, a_{21}
- A unit shock to y_1 reaches y_2 one period later ($a_{21} = 0.3$), then both fade
  - the decay is governed by the largest eigenvalue of $\mathbf{A}$, 0.7: $(\mathbf{A}^h)_{11} = 0.6 \cdot 0.7^h + 0.4 \cdot 0.2^h$, so after 5 periods 0.10 of the shock remains in y_1

The number of parameters

- A VAR(p) with K variables and a constant: $K + pK^2$ coefficients, plus $K(K+1)/2$ in $\boldsymbol{\Sigma}$
  - each equation has $1 + pK$ regressors
  - Romanian VAR: $K = 3$, $p = 2$: $3 + 2 \cdot 9 = 21$ coefficients estimated on 84 quarters
- **The curse of dimensionality**: $K = 5$, $p = 4$: $5 + 4 \cdot 25 = 105$ coefficients
  - with 25 years of quarterly data (100 observations) each equation has 21 regressors and 79 residual degrees of freedom
- Hence small VARs (3–6 variables), few lags, and estimates with large standard errors; individual coefficients are rarely interpreted

Christopher Sims and "Macroeconomics and Reality"



Christopher A. Sims, Nobel Prize in Economics 2011

Photo: Holger Motzkau (2011); CC BY-SA 3.0; Wikimedia Commons

- Sims (1980): the large macro models of the 1970s imposed "incredible" restrictions
  - which variables are exogenous, which lags are zero
  - his proposal: an unrestricted VAR that lets the data speak
  - then only the few assumptions needed to interpret the shocks
- The VAR became the standard tool of central banks for forecasting and for measuring the effects of monetary policy
- Nobel Prize 2011, with Thomas Sargent, "for their empirical research on cause and effect in the macroeconomy"

Recap: The VAR(p) model

- $\mathbf{Y}_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{Y}_{t-j} + \boldsymbol{\varepsilon}_t$, $\boldsymbol{\varepsilon}_t$ white noise with covariance $\boldsymbol{\Sigma}$
- a_{jk}: effect of the lag of variable k in the equation of variable j
- $K + pK^2$ coefficients: keep K and p small
- Sims (1980): VARs as an alternative to "incredible" restrictions

Stability of a VAR(1)

- Substitute repeatedly: $\mathbf{Y}_t = \mathbf{c} + \mathbf{A}\mathbf{c} + \cdots + \mathbf{A}^{h-1}\mathbf{c} + \mathbf{A}^h\mathbf{Y}_{t-h} + \sum_{j=0}^{h-1} \mathbf{A}^j \boldsymbol{\varepsilon}_{t-j}$
  - the past is forgotten only if $\mathbf{A}^h \to \mathbf{0}$
- **Stability condition**: all eigenvalues λ of $\mathbf{A}$ satisfy $|\lambda| < 1$
  - eigenvalues: the roots of $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$; for $K = 2$: $\lambda^2 - \text{tr}(\mathbf{A})\lambda + \det(\mathbf{A}) = 0$
  - $\mathbf{I}$: the identity matrix; $\text{tr}(\mathbf{A}) = a_{11} + a_{22}$: the trace; $\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$: the determinant; $|\lambda|$: the modulus (eigenvalues can be complex)
  - a stable VAR is weakly stationary (if it started long ago); this is the multivariate version of $|\phi| < 1$ (Chapter 2)
- **Example**: $\text{tr}(\mathbf{A}) = 0.9$, $\det(\mathbf{A}) = 0.5 \cdot 0.4 - 0.2 \cdot 0.3 = 0.14$: $\lambda^2 - 0.9\lambda + 0.14 = 0$
  - $\lambda = (0.9 \pm 0.5)/2$: $\lambda_1 = 0.7$, $\lambda_2 = 0.2$; both inside the unit circle: stable

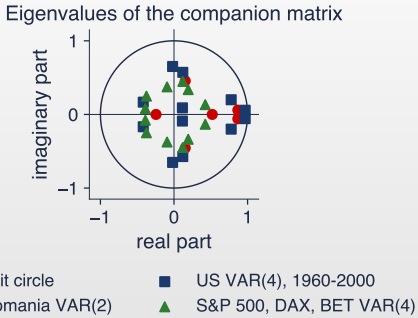
Mean and variance of a stable VAR(1)

- **Mean**: take expectations, $\boldsymbol{\mu} = \mathbf{c} + \mathbf{A}\boldsymbol{\mu}$, so $\boldsymbol{\mu} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{c}$
  - example: $\mathbf{I} - \mathbf{A} = \begin{pmatrix} 0.5 & -0.2 \\ -0.3 & 0.6 \end{pmatrix}$, determinant 0.24; $\boldsymbol{\mu} = \frac{1}{0.24}\begin{pmatrix} 0.6 & 0.2 \\ 0.3 & 0.5 \end{pmatrix}\begin{pmatrix} 1.2 \\ 0.6 \end{pmatrix} = \begin{pmatrix} 3.50 \\ 2.75 \end{pmatrix}$
- **Variance**: $\boldsymbol{\Gamma}(0) = \mathbf{A}\boldsymbol{\Gamma}(0)\mathbf{A}' + \boldsymbol{\Sigma}$ (a discrete Lyapunov equation); $\boldsymbol{\Gamma}(h) = \mathbf{A}^h\boldsymbol{\Gamma}(0)$
  - solved by $\operatorname{vec}\boldsymbol{\Gamma}(0) = (\mathbf{I} - \mathbf{A} \otimes \mathbf{A})^{-1}\operatorname{vec}\boldsymbol{\Sigma}$; example: $\boldsymbol{\Gamma}(0) = \begin{pmatrix} 1.75 & 1.22 \\ 1.22 & 1.73 \end{pmatrix}$
  - $\otimes$: the Kronecker product; vec stacks the columns of a matrix
- The variances 1.75 and 1.73 exceed those of the shocks (1): the dynamics accumulates past shocks

The companion form of a VAR(p)

- Stack p consecutive vectors: $\boldsymbol{\xi}_t = (\mathbf{Y}_t', \mathbf{Y}_{t-1}', \ldots, \mathbf{Y}_{t-p+1}')'$, of size Kp
  - then $\boldsymbol{\xi}_t = \boldsymbol{\nu} + \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$: every VAR($p$) is a VAR(1) in $\boldsymbol{\xi}_t$
  - $\boldsymbol{\nu}$: $\mathbf{c}$ followed by zeros; $\mathbf{v}_t$: ε_t followed by zeros; $\mathbf{F}$: the $Kp \times Kp$ companion matrix
- For $p = 2$: $\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ \mathbf{0} \end{pmatrix} + \underbrace{\begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I}_K & \mathbf{0} \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{Y}_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \mathbf{0} \end{pmatrix}$
  - the second block row is the identity $\mathbf{Y}_{t-1} = \mathbf{Y}_{t-1}$
- **Stability of a VAR**(p): all Kp eigenvalues of the companion matrix $\mathbf{F}$ have modulus below 1
  - equivalently: all roots z (complex numbers) of $\det(\mathbf{I}_K - \mathbf{A}_1 z - \cdots - \mathbf{A}_p z^p) = 0$ lie outside the unit circle
  - the eigenvalues of $\mathbf{A}_1$ alone say nothing about the stability of a VAR(2)

Eigenvalues of three estimated VARs



- Companion matrices: Romanian VAR(2) (GDP growth, inflation, ROBOR 3M; 6 eigenvalues), the US VAR(4) of Section 10 (12), the daily S&P 500, DAX and BET VAR(4) (12)

TSA_ch6_var_stability

Interpreting the eigenvalues

- All three VARs are stable; largest moduli: Romania 0.869, US 0.969, markets 0.464
  - the largest modulus sets the speed at which shocks fade: $0.97^{20} \approx 0.54$ after 20 quarters, $0.46^{2} \approx 0.21$ after 2 days
- Macro VARs have eigenvalues close to 1 (persistent inflation and interest rates); return VARs have small ones
  - complex pairs produce damped oscillations in the responses
- An eigenvalue very close to 1 signals a unit root, perhaps cointegration: Chapter 7

The moving-average representation

- A stable VAR can be written as an infinite moving average of the shocks (Wold, Chapter 1):
  - $\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Phi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Phi}_2\boldsymbol{\varepsilon}_{t-2} + \cdots = \boldsymbol{\mu} + \sum_{h=0}^{\infty} \boldsymbol{\Phi}_h \boldsymbol{\varepsilon}_{t-h}$, $\boldsymbol{\Phi}_0 = \mathbf{I}$
  - $\boldsymbol{\Phi}_h$: the $K \times K$ matrix that weights the shock of h periods ago
- VAR(1): $\boldsymbol{\Phi}_h = \mathbf{A}^h$; VAR($p$): $\boldsymbol{\Phi}_h = \sum_{j=1}^{\min(h,p)} \mathbf{A}_j \boldsymbol{\Phi}_{h-j}$
  - example: $\boldsymbol{\Phi}_2 = \mathbf{A}^2 = \begin{pmatrix} 0.31 & 0.18 \\ 0.27 & 0.22 \end{pmatrix}$
- Element (j, k) of $\boldsymbol{\Phi}_h$: the effect on $Y_{j,t+h}$ of a unit change in ε_{kt}, all other shocks fixed
  - these are the **impulse responses** (Section 7); they also give the forecast errors (Sections 8–9)

Recap: Stationarity and the companion form

- VAR(1): stable if all eigenvalues of $\mathbf{A}$ have $|\lambda| < 1$; VAR(p): the same for the companion matrix
- Mean $\boldsymbol{\mu} = (\mathbf{I} - \mathbf{A_1} - \cdots - \mathbf{A}_p)^{-1}\mathbf{c}$; forecasts converge to it
- Moving-average form $\mathbf{Y}_t = \boldsymbol{\mu} + \sum_h \boldsymbol{\Phi}_h \boldsymbol{\varepsilon}_{t-h}$: the basis of impulse responses
- Estimated macro VARs have eigenvalues close to 1

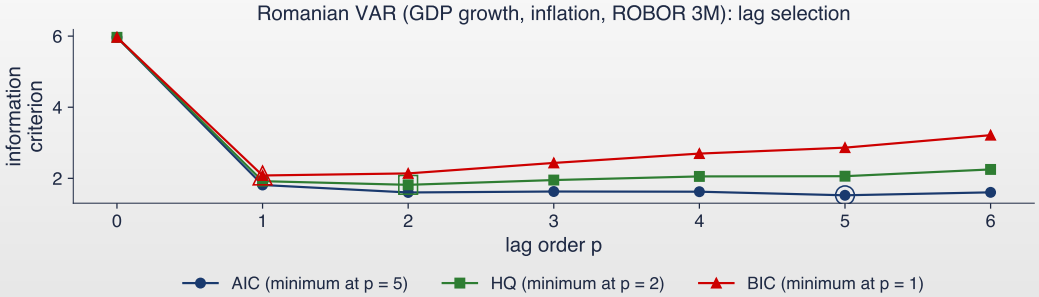
Estimation by OLS, equation by equation

- Each equation is a linear regression of Y_{jt} on a constant and $\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_{t-p}$
  - the same $1 + Kp$ regressors in every equation; OLS (ordinary least squares) on each equation separately
- **OLS is enough**: with identical regressors, system estimation (GLS, generalised least squares, on all equations) gives exactly the OLS estimates
  - the correlation of the shocks does not improve the estimates; under Normal errors OLS is also the maximum likelihood estimator
- **Covariance of the shocks**: $\hat{\mathbf{\Sigma}} = \frac{1}{T-Kp-1} \sum_t \hat{\varepsilon}_t \hat{\varepsilon}_t'$, from the residuals of all equations
  - $\hat{\varepsilon}_t$: the vector of residuals at date t; $T - Kp - 1$: the residual degrees of freedom of each equation
  - stationarity matters: with stationary series the usual t and F tests are valid in large samples

Choosing p: information criteria

- Fit VAR(p) for $p = 0, \ldots, p_{\max}$ on the **same sample**; $\tilde{\mathbf{\Sigma}}(p)$: residual covariance divided by T
  - AIC (Akaike) $= \ln\det\tilde{\mathbf{\Sigma}}(p) + \frac{2}{T}pK^2$ Akaike (1974)
  - BIC (Schwarz) $= \ln\det\tilde{\mathbf{\Sigma}}(p) + \frac{\ln T}{T}pK^2$ Schwarz (1978)
  - HQ (Hannan–Quinn) $= \ln\det\tilde{\mathbf{\Sigma}}(p) + \frac{2\ln\ln T}{T}pK^2$ Hannan and Quinn (1979)
  - $\det\tilde{\mathbf{\Sigma}}(p)$: a generalised residual variance; pK^2: the number of lag coefficients; $p_{\max}$: the largest order tried
- First term: fit (smaller residual "volume"); second: a penalty per coefficient; choose the p with the smallest value
  - penalties: AIC $<$ HQ $<$ BIC (for $T \geq 16$): BIC chooses the smallest p, AIC the largest
  - BIC and HQ are consistent (they find the true p as $T \to \infty$); AIC tends to overfit, but often forecasts well

Romanian VAR: which p?



- GDP growth, 12-month inflation and ROBOR 3M, 2005Q1–2026Q2; VAR(p) for $p = 0, \ldots, 6$ on the common sample of 80 quarters; circles mark the minima

TSA_ch6_estimation_diagnostics

Interpreting the lag selection

- BIC chooses $p = 1$, HQ $p = 2$, AIC $p = 5$: the three criteria disagree, as they often do in short samples
  - AIC: 1.812 at $p = 1$, 1.603 at $p = 2$, 1.523 at $p = 5$: a flat curve after $p = 2$
- $p = 5$ would mean $3 + 5 \cdot 9 = 48$ coefficients for about 80 quarters
  - we take $p = 2$ (HQ), then check whether the residuals are white noise; if not, we add lags
- A rule of thumb: start from the criteria, never stop at them

The estimated Romanian VAR(2)

Regressor	equation g_t	equation π_t	equation i_t
constant	1.09 (2.1)	0.49 (2.0)	−0.11 (−0.6)
g_{t-1}	0.01 (0.1)	0.01 (0.1)	0.14 (3.5)
π_{t-1}	0.53 (2.3)	1.26 (11.7)	0.24 (2.8)
i_{t-1}	−0.73 (−2.7)	0.16 (1.2)	1.04 (10.2)
g_{t-2}	−0.08 (−0.7)	−0.03 (−0.5)	0.07 (1.7)
π_{t-2}	−0.41 (−1.7)	−0.35 (−3.1)	−0.17 (−1.9)
i_{t-2}	0.54 (2.0)	−0.16 (−1.3)	−0.12 (−1.2)
R^2	0.16	0.90	0.93

- OLS on 84 quarters (2005Q3–2026Q2); t-statistics in parentheses; 21 coefficients
- Residual standard deviations: 2.32 (g), 1.09 (π), 0.87 (i); residual correlation of π and i: 0.25

Interpreting the coefficients

- Inflation is very persistent: $1.26\,\pi_{t-1} - 0.35\,\pi_{t-2}$; ROBOR too: $1.04\,i_{t-1}$
  - R^2 of 0.90 and 0.93: most of it is the series' own past
- In the ROBOR equation, past GDP growth ($t = 3.5$) and past inflation ($t = 2.8$) matter: the interest rate follows the economy
  - GDP growth is hard to predict: $R^2 = 0.16$
- Single coefficients of correlated lags are unstable (e.g. i_{t-1} and i_{t-2} with opposite signs in the g equation): read the system through tests, responses and forecasts, not one coefficient at a time

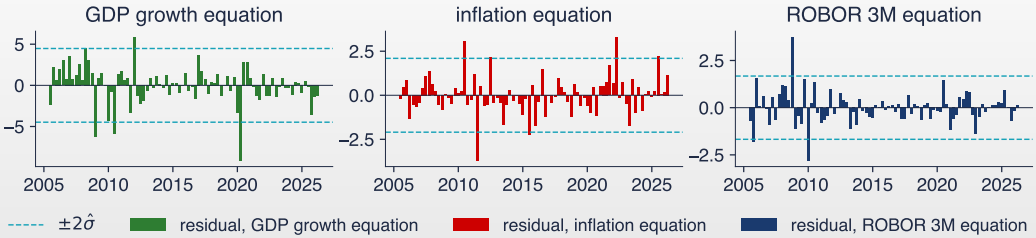
Recap: Estimation and lag selection

- OLS equation by equation is efficient: all equations have the same regressors
- AIC, BIC, HQ: $\ln\det\tilde{\boldsymbol{\Sigma}}(p)$ plus a penalty; compare on a common sample
- BIC $\leq$ HQ $\leq$ AIC in the chosen p; Romanian VAR: $p = 2$
- Interpret the system, not single coefficients

Are the residuals white noise?

- **Multivariate portmanteau test** Hosking (1980)
  - H_0: no autocorrelation and no cross-correlation of the residuals up to lag h
  - $Q_h = T^2 \sum_{j=1}^{h} \frac{1}{T-j} \text{tr}(\hat{\mathbf{C}}_j' \hat{\mathbf{C}}_0^{-1} \hat{\mathbf{C}}_j \hat{\mathbf{C}}_0^{-1})$, $\hat{\mathbf{C}}_j = \frac{1}{T} \sum_t \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_{t-j}'$: the residual autocovariance matrix at lag j; tr: the trace
  - under H_0: χ^2 with $K^2(h-p)$ degrees of freedom; the multivariate Ljung–Box test (Chapter 2); a small p-value: the residuals are still autocorrelated
- **Normality**: a multivariate Jarque–Bera test Jarque and Bera (1980) on the standardised residuals (skewness and kurtosis)
  - Normality is not needed for OLS; it matters for exact intervals and for small samples
- If the portmanteau test rejects: add lags, add a variable, or model breaks and outliers (dummy variables)
  - also check stability (eigenvalues) and, for long samples, parameter stability over time

Residuals of the Romanian VAR(2)



⊡ Residuals of the three equations, 2005Q3–2026Q2, with ±2 residual standard deviations

TSA_ch6_estimation_diagnostics

Interpreting the residuals

- Largest residuals: GDP -9.1 in 2020Q2 (the pandemic), ROBOR $+3.7$ in 2008Q4 (the 2008 liquidity squeeze), inflation -3.7 in 2011Q3
  - kurtosis 6.5 (g) and 7.9 (i), against 3 under the Normal distribution: Jarque–Bera 191 ($p < 0.001$)
- Portmanteau: $Q_8 = 76.7$ (54 d.f., $p = 0.023$); $Q_{12} = 105.3$ (90 d.f., $p = 0.130$)
  - borderline at 8 lags, fine at 12: some short-run correlation is left, partly from overlapping 12-month inflation rates
- Consequence: point estimates are usable, but the intervals of the next sections rely on large-sample approximations; we use bootstrap bands for the responses

Recap: Residual diagnostics

- Portmanteau test: H_0 white-noise residuals, χ^2 with $K^2(h-p)$ d.f.
- Normality fails in macro data: crises produce outliers
- Remedies: more lags, more variables, dummies for crisis quarters

Granger causality: the definition



Clive Granger (1934–2009), Nobel Prize in Economics 2003

Photo: Olaf Storbeck (2008); CC BY-SA 2.0; Wikimedia Commons

- **Definition** Granger (1969): x **Granger-causes** y
  - if the past of x improves the forecast of y made from the past of y (and of the other variables)
  - formally: the MSE (mean squared error) of the optimal forecast of y_{t+1} is smaller when $x_t, x_{t-1}, \ldots$ are used
- A statement about **predictability**, not about cause and effect
  - the cause comes before the effect, but "before" does not imply "because"
- Four cases: no causality, $x \to y$, $y \to x$, feedback $x \leftrightarrow y$

Testing Granger causality in a VAR

- In a VAR(p), x does not Granger-cause y if all coefficients of $x_{t-1}, \ldots, x_{t-p}$ in the equation of y are zero
  - bivariate VAR(2): H_0: $a_{12}^{(1)} = a_{12}^{(2)} = 0$ in $y_t = c + a_{11}^{(1)} y_{t-1} + a_{12}^{(1)} x_{t-1} + a_{11}^{(2)} y_{t-2} + a_{12}^{(2)} x_{t-2} + \varepsilon_t$
  - $a_{12}^{(j)}$: the coefficient of x_{t-j} in the equation of y (the superscript is the lag)
- F **test**: $F = \dfrac{(RSS_R - RSS_U)/p}{RSS_U/(T - Kp - 1)} \sim F(p, T - Kp - 1)$ under H_0
  - RSS_U: residual sum of squares of the equation of y; RSS_R: the same without the lags of x
  - a large F (a small p-value): dropping the lags of x raises the RSS too much, so H_0 is rejected
- **Wald test**: $W = (\mathbf{R}\hat{\boldsymbol{\beta}})'[\mathbf{R}\hat{\mathbf{V}}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}) \sim \chi^2(p)$; here $W = pF$
  - $\hat{\boldsymbol{\beta}}$: the estimated coefficients of the equation of y; $\mathbf{R}$ selects the tested ones; $\hat{\mathbf{V}}$: the estimated covariance of $\hat{\boldsymbol{\beta}}$; both versions need stationary variables

Worked example: does GDP growth Granger-cause ROBOR?

- Equation of i_t in the Romanian VAR(2), $T = 84$, $K = 3$, $p = 2$
  - unrestricted (with g_{t-1}, g_{t-2}): $RSS_U = 58.27$; restricted (without them): $RSS_R = 69.27$
- $F = \dfrac{(69.27 - 58.27)/2}{58.27/77} = \dfrac{5.498}{0.7568} = 7.26$
  - degrees of freedom $(2, 84 - 6 - 1) = (2, 77)$; 5% critical value 3.12; p-value 0.001
- **Decision**: reject H_0: past GDP growth helps to forecast ROBOR 3M, given past inflation and past ROBOR
  - consistent with a central bank that tightens when the economy grows fast

Granger causality in the Romanian VAR(2)

Cause	Effect	F	p
GDP growth	inflation	0.16	0.853
GDP growth	ROBOR 3M	7.26	0.001
inflation	GDP growth	2.94	0.059
inflation	ROBOR 3M	4.74	0.011
ROBOR 3M	GDP growth	4.02	0.022
ROBOR 3M	inflation	0.89	0.417

- $F(2, 77)$ tests in the VAR(2), 2005Q3–2026Q2; 5% critical value 3.12
- Significant at 5%: GDP growth $\to$ ROBOR, inflation $\to$ ROBOR, ROBOR $\to$ GDP growth

TSA_ch6_granger

Interpreting the Granger tests

- The interest rate follows the economy: past growth and past inflation help to forecast ROBOR
  - the pattern of a reaction function (the central bank responds to inflation and activity); the test alone cannot prove it is the BNR's rule
- ROBOR does not Granger-cause inflation ($p = 0.417$): no evidence that past rates improve inflation forecasts
  - this does not say that monetary policy is ineffective: the effect may be slow, small relative to the noise, or already anticipated
- ROBOR $\rightarrow$ GDP growth ($p = 0.022$): higher rates come before weaker growth; in the bivariate VAR (GDP, ROBOR) alone the p-value is 0.063

Markets: does New York lead Frankfurt and Bucharest?



Trading floor of the Frankfurt Stock Exchange (DAX)

Photo: Ank Kumar (2015); CC BY-SA 4.0; Wikimedia Commons

Cause	Effect	F	p
S&P 500	DAX	94.6	< 0.001
S&P 500	BET	37.8	< 0.001
DAX	S&P 500	6.9	< 0.001
DAX	BET	3.2	0.012
BET	S&P 500	1.0	0.416
BET	DAX	2.6	0.032

- Daily log returns, VAR(4) (the order of Diebold and Yilmaz (2012)), 6 429 days, 2000–2026; $F(4, \cdot)$ tests; HQ would choose $p = 2$, AIC $p = 8$
- Coefficient of the S&P 500 of yesterday in the BET equation: 0.218 (SE 0.018); in the DAX equation: 0.346

TSA_ch6_granger

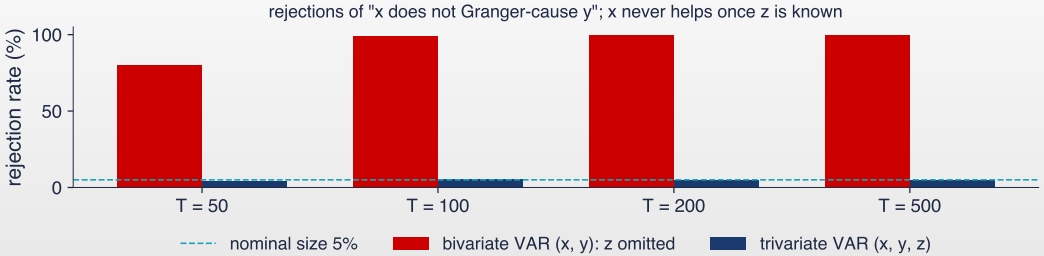
Interpreting the market tests

- Strong one-way causality from the S&P 500 to the DAX and the BET: a 1% S&P return yesterday adds about 0.218% to the BET today
  - the BET does not Granger-cause the S&P 500 ($p = 0.416$)
- Cause: the closing times (non-synchronous trading), not a slow market
  - New York keeps trading after Europe closes; "yesterday's" S&P return partly happened while Europe was closed
  - on weekly data the effect is much weaker (Seminar 6, B2)
- Can you trade on it? You would buy at the Bucharest open, after prices have already adjusted: Granger causality in closing prices is not a free profit

Limitations of Granger causality: omitted variables

- Granger causality is relative to the information set: adding or removing a variable can create or destroy it Lütkepohl (1982)
  - a common driver z that affects x earlier than y makes x look like a cause of y
- Simulation: $z_t = 0.5z_{t-1} + e_t$, $x_t = 0.8z_{t-1} + u_t$, $y_t = 0.8z_{t-2} + v_t$
  - e_t, u_t, v_t: independent standard Normal white noises
  - x never enters the equation of y; but x_{t-1} is a noisy measure of z_{t-2}, which drives y_t
- Example: ice-cream sales Granger-cause drownings in a bivariate model; temperature, the omitted driver, explains both

Spurious Granger causality from an omitted variable



- 1 000 simulations per sample size of the system of the previous slide; F test of "x does not Granger-cause y" at 5% in a VAR(2), with and without z

TSA_ch6_granger

Interpreting the simulation

- Without z, the test rejects in 80% of samples with $T = 50$ and 100% with $T = 200$: almost always a "cause"
  - more data makes it worse: the test becomes more sure of a wrong conclusion
- With z in the VAR, the rejection rate is 4.3%–5.2%, close to the nominal 5%
  - Granger causality is a property of the chosen system, not of the world
- Romanian example: inflation $\rightarrow$ GDP growth has p = 0.059 in the three-variable VAR and 0.173 in the bivariate one

Limitations of Granger causality: instantaneous causality

- Granger tests only use lags; effects within the same period are hidden in the correlation of the shocks $\boldsymbol{\Sigma}$
  - **instantaneous causality** between x and y: $\mathrm{Cov}(\varepsilon_{xt}, \varepsilon_{yt}) \neq 0$; tested by a Wald test on the off-diagonal elements of $\boldsymbol{\Sigma}$
- It has **no direction**: the data cannot tell whether x moved y or y moved x within the quarter
  - Romania: ROBOR shocks are instantaneously correlated with the others ($p = 0.041$); the residual correlation of inflation and ROBOR is 0.25
- With quarterly data, many true effects happen within the quarter: a VAR in which nothing is Granger-significant can still have strong links (Section 7)

Limitations of Granger causality: trends, expectations and the meaning of "cause"

- **Non-stationary variables**: with $I(1)$ series the F and Wald tests do not have their usual distributions Sims, Stock and Watson (1990)
  - remedies: test on differences (if there is no cointegration), or the Toda–Yamamoto approach Toda and Yamamoto (1995): estimate a VAR($p + d_{\max}$) in levels and test only the first p lags
  - $d_{\max}$: the highest order of integration of the series (usually 1); the extra lags are estimated but not tested
  - cointegrated series need a VECM: Chapter 7
- **Expectations**: forward-looking prices move before the events they anticipate
  - stock prices Granger-cause GDP, and weather forecasts "cause" rain; neither is a cause in the everyday sense
- Report: "x helps to predict y, given these variables and this sample"; never "x causes y"

Recap: Granger causality

- $x \to y$ (Granger): the lags of x improve the forecast of y; F or Wald test of zero restrictions
- Romania: GDP growth and inflation $\to$ ROBOR; markets: S&P 500 $\to$ DAX, BET (time zones)
- Pitfalls: omitted variables, instantaneous effects, non-stationarity, expectations
- Granger causality is predictability, not causation

Impulse responses and why we orthogonalise

- **Impulse response function** (IRF): the path of $Y_{j,t+h}$, $h = 0, 1, 2, \ldots$, after a one-time shock to variable k at time t
  - from the moving-average form: the response to a unit change in ε_{kt} is the column k of $\mathbf{\Phi}_h$
- **Problem**: the shocks are correlated ($\mathbf{\Sigma}$ not diagonal)
  - "a shock to inflation with no shock to ROBOR" is an experiment that the data rarely show: the two shocks usually come together (correlation 0.25)
- **Solution**: rewrite the shocks as $\boldsymbol{\varepsilon}_t = \mathbf{P}\mathbf{u}_t$, with $\mathbf{u}_t$ uncorrelated, unit variance: $\mathbf{\Sigma} = \mathbf{P}\mathbf{P}'$
  - the **orthogonalised** IRF: $\mathbf{\Theta}_h = \mathbf{\Phi}_h\mathbf{P}$; element (j, k): the response of Y_j to a one-standard-deviation shock u_k
  - $\mathbf{P}$: a $K \times K$ matrix that turns the uncorrelated shocks $\mathbf{u}_t$ into the observed shocks $\boldsymbol{\varepsilon}_t$

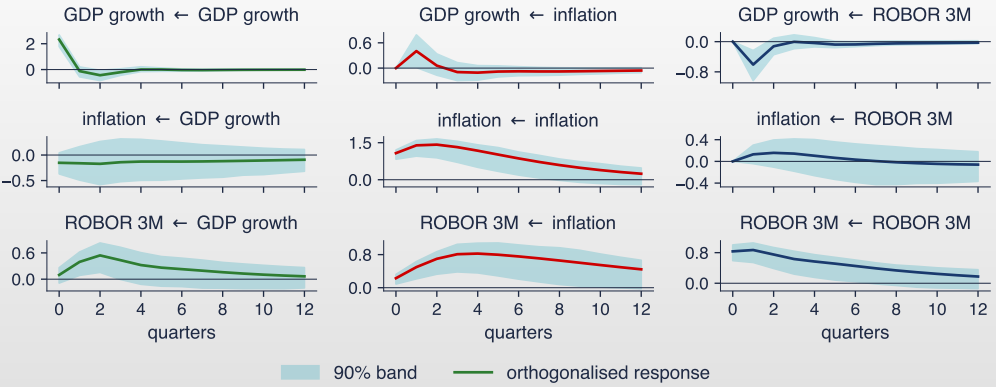
The Cholesky decomposition and the ordering

- **Cholesky**: the unique lower triangular $\mathbf{P}$ with positive diagonal such that $\mathbf{\Sigma} = \mathbf{P}\mathbf{P}'$
  - example: $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$: $\mathbf{P} = \begin{pmatrix} 1.000 & 0.000 \\ 0.500 & 0.866 \end{pmatrix}$, since $p_{11} = 1$, $p_{21} = 0.5/1$, $p_{22} = \sqrt{1 - 0.25}$
- **Recursive structure**: $\varepsilon_{1t} = p_{11} u_{1t}$, $\varepsilon_{2t} = p_{21} u_{1t} + p_{22} u_{2t}$
  - the first variable is not moved by u_2 in the same period; the second reacts to both at once
  - the whole correlation is attributed to the variable ordered first
- **The ordering matters**: with y_2 first, $\mathbf{P} = \begin{pmatrix} 0.866 & 0.500 \\ 0.000 & 1.000 \end{pmatrix}$ (upper triangular in the original order)
  - rule: order from the slowest variable (does not react within the period) to the fastest; Romania: GDP growth, inflation, ROBOR

Worked example: orthogonalised responses

- VAR(1) of Section 2, order (y_1, y_2): $\boldsymbol{\Theta}_0 = \mathbf{P} = \begin{pmatrix} 1.000 & 0.000 \\ 0.500 & 0.866 \end{pmatrix}$
  - a shock $u_1 = 1$ moves y_1 by 1 and y_2 by 0.500 at once; $u_2 = 1$ moves only y_2, by 0.866
- One period later: $\boldsymbol{\Theta}_1 = \mathbf{AP} = \begin{pmatrix} 0.600 & 0.173 \\ 0.500 & 0.346 \end{pmatrix}$
  - y_2 responds to u_1 by 0.500: $0.3 \cdot 1 + 0.4 \cdot 0.5$, the direct effect a_{21} plus its own persistence
- The **cumulative** response $\sum_{s \le h} \boldsymbol{\Theta}_s$ gives the effect on the level when the VAR is in growth rates

Impulse responses of the Romanian VAR(2)



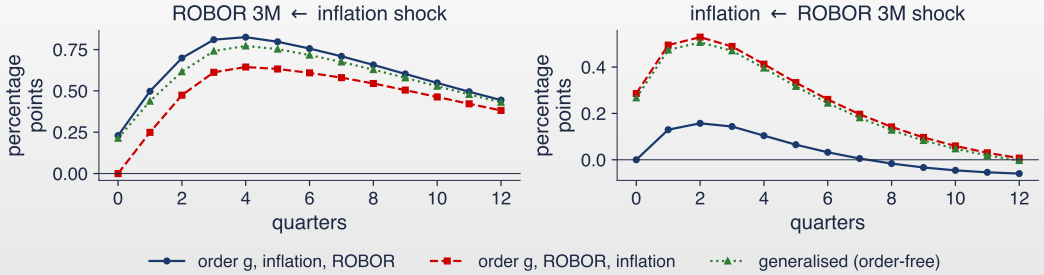
- Row: responding variable; column: shock of one standard deviation; Cholesky order g, π, i; 90% bands from a residual bootstrap (500 replications) Kilian (1998)

TSA_ch6_irf_fevd

Interpreting the Romanian responses

- An inflation shock (1.08 pp on impact) raises ROBOR by 0.23 pp at once and by up to 0.82 pp after 4 quarters; still 0.44 pp after 3 years
  - the band excludes zero for 11 of the 13 horizons: the clearest result of the VAR
- A ROBOR shock lowers GDP growth by 0.61 pp the next quarter, then the effect vanishes
  - but inflation **rises** after a ROBOR shock (up to 0.16 pp, not significant): the "price puzzle" Sims (1992)
  - a likely reason: the rate rises when the central bank expects inflation; information omitted from the VAR (exchange rate, energy prices) shows up as a "shock"
- Inflation shocks themselves are persistent: 1.18 pp after a year, 0.24 pp after three

Does the ordering change the answer?



- Same VAR(2), two Cholesky orderings (inflation before ROBOR, ROBOR before inflation) and the generalised responses of Pesaran and Shin (1998), which do not depend on the order

TSA_ch6_irf_fevd

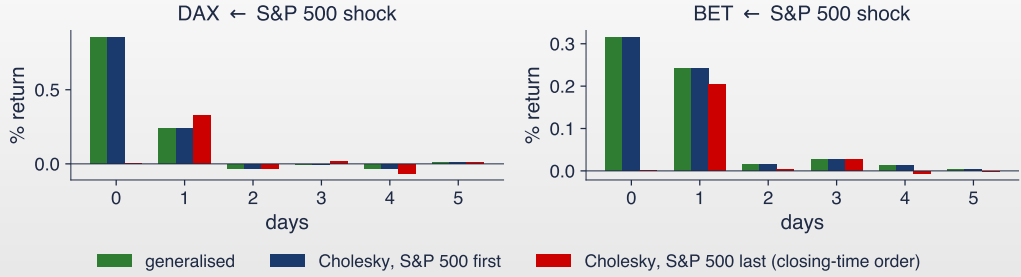
Interpreting the two orderings

- ROBOR after an inflation shock: peak 0.82 pp (inflation first) against 0.64 pp (ROBOR first)
  - with ROBOR first, the impact response is zero by construction
- Inflation after a ROBOR shock: 0.16 pp against 0.53 pp, with 0.29 pp already on impact
  - the second ordering assigns the common part of the two shocks (correlation 0.25) to ROBOR, so the "price puzzle" gets larger
- The ordering is an assumption about which variable reacts within the quarter; it must be stated and defended, and results that change with it are fragile

Generalised impulse responses

- **Generalised IRF** Pesaran and Shin (1998) (after Koop, Pesaran and Potter (1996)): shock variable k by one standard deviation and let the other shocks move as they usually do with it
  - $\psi_k(h) = \mathbf{\Phi}_h \mathbf{\Sigma} \mathbf{e}_k / \sqrt{\sigma_{kk}}$; $\mathbf{e}_k$: column k of the identity matrix; σ_{kk}: the variance of shock k
- Properties
  - no ordering is needed; for the variable ordered first, it equals the Cholesky response
  - the shocks are not orthogonal: the responses to different shocks overlap and cannot be added up
- Useful for markets, where no ordering is natural; it describes typical co-movements, not structural shocks

Markets: responses to an S&P 500 shock



- VAR(4) of daily S&P 500, DAX and BET returns; S&P shock of one standard deviation (1.21%); generalised responses, Cholesky with the S&P 500 first, and Cholesky in the order of the closing times (BET, DAX, S&P 500)

TSA_ch6_market_spillovers

Interpreting the market responses

- Generalised: DAX 0.85% on the same day and 0.24% the next; BET 0.32% and 0.24%
  - identical to the Cholesky responses with the S&P 500 first, as the theory says
- With the S&P 500 last (it closes last), the same-day response is zero by assumption and the next-day responses are 0.32% (DAX) and 0.20% (BET)
  - the same-day correlation (0.63 with the DAX) is attributed to Europe in this order
- Either way, the effect is over after two days: markets absorb the news quickly

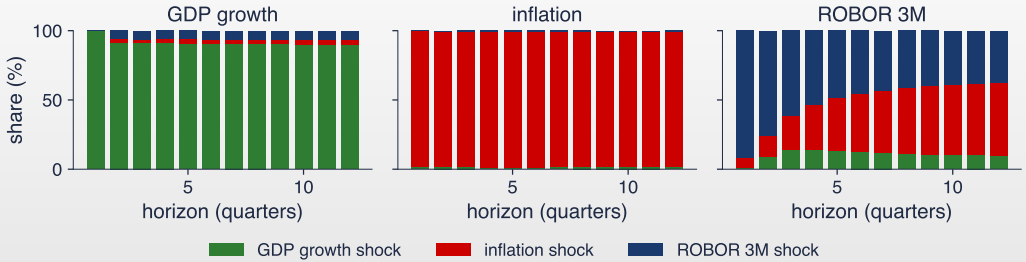
Recap: Impulse responses

- IRF: columns of $\mathbf{\Phi}_h$; orthogonalised: $\mathbf{\Theta}_h = \mathbf{\Phi}_h \mathbf{P}$, $\mathbf{\Sigma} = \mathbf{P}\mathbf{P}'$
- Cholesky = recursive ordering; the ordering is an assumption and can change the results
- Romania: ROBOR rises for years after an inflation shock; a "price puzzle" appears
- Generalised IRF: order-free, but not additive; bootstrap bands for inference

Forecast error variance decomposition

- h-step forecast error: $\mathbf{Y}_{t+h} - \hat{\mathbf{Y}}_{t+h|t} = \sum_{s=0}^{h-1} \mathbf{\Phi}_s \boldsymbol{\varepsilon}_{t+h-s} = \sum_{s=0}^{h-1} \mathbf{\Theta}_s \mathbf{u}_{t+h-s}$
  - the orthogonal shocks u_k are uncorrelated, so the variance splits into K parts
- **FEVD**: $\omega_{jk}(h) = \dfrac{\sum_{s=0}^{h-1} \theta_{jk,s}^2}{\sum_{s=0}^{h-1} \sum_{m=1}^{K} \theta_{jm,s}^2}$, the share of shock k in the h-step error variance of Y_j
  - $\theta_{jk,s}$: element (j,k) of $\mathbf{\Theta}_s$; $\hat{\mathbf{Y}}_{t+h|t}$: the forecast of $\mathbf{Y}_{t+h}$ made at t
  - each row sums to 100%; it depends on the Cholesky ordering; at $h = 1$ the first variable is explained 100% by its own shock
- Example (VAR(1) of Section 2), $h = 2$: the variance of the error of y_2 is 1.37; shock u_1 explains 36.5%
  - $(0.5^2 + 0.5^2)/1.37$: the impact $p_{21} = 0.5$ and the next-period response 0.500

FEVD of the Romanian VAR(2)



- Share of each orthogonal shock in the forecast error variance, horizons 1–12 quarters, Cholesky order g, π, i

TSA_ch6_irf_fevd

Interpreting the variance decomposition

- ROBOR: own shocks explain 92% at one quarter, but only 37% after three years; inflation shocks explain 53%, GDP shocks 10%
  - in the long run, the interest rate is mostly driven by inflation
- Inflation: 97% own shocks even after 12 quarters; GDP growth: 90% own shocks
  - in this small VAR, inflation is driven by forces outside the model (energy, food, exchange rate, taxes)
- FEVD and Granger tests agree: the information flows from inflation and output to the interest rate

Case study: Diebold and Yilmaz (2009, 2012)

- **Question**: how much of the forecast uncertainty of one market comes from shocks in other markets, and how does it change over time?
  - Diebold and Yilmaz (2009): Cholesky FEVD of weekly returns of 19 stock markets; Diebold and Yilmaz (2012): generalised FEVD, which removes the dependence on the ordering
- **Spillover index**: $S(H) = \frac{100}{K} \sum_{j \neq k} \tilde{\omega}_{jk}(H)$, the average share of other markets' shocks
  - $\tilde{\omega}_{jk}$: generalised FEVD rescaled so that each row sums to 1; the specification of Diebold and Yilmaz (2012): VAR(4), $H = 10$ days, rolling 200-day windows
- **Finding**: spillovers rise sharply in crises; markets become more connected exactly when diversification is most needed
  - next: the index for the S&P 500, the DAX and the BET

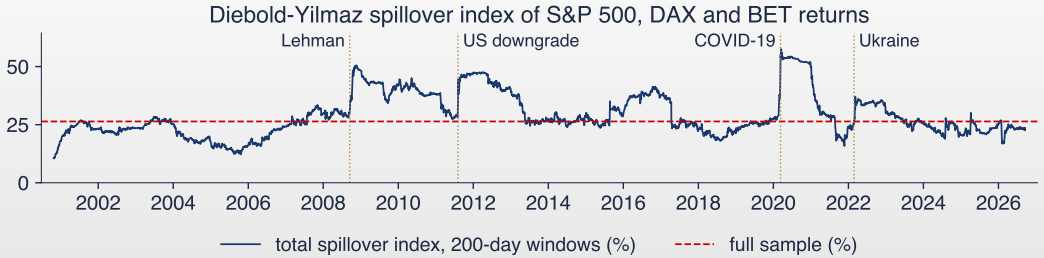
Spillover table: S&P 500, DAX, BET

To ← from (%)	S&P 500	DAX	BET	**from others**
S&P 500	69.2	27.1	3.7	30.8
DAX	28.2	66.4	5.4	33.6
BET	7.2	7.5	85.3	14.7
to others	35.4	34.6	9.1	index 26.4

- Row j: shares of the 10-day forecast error variance of market j due to shocks in each market (generalised FEVD, full sample 2000–2026)
- The S&P 500 receives 27.1% from the DAX and the DAX 28.2% from the S&P 500; the BET receives 14.7% and gives only 9.1%: a small, net receiving market

TSA_ch6_market_spillovers

The spillover index over time



- Total spillover index of S&P 500, DAX and BET daily returns: generalised FEVD at 10 days from a VAR(4), rolling windows of 200 trading days (29.0% on average)

TSA_ch6_market_spillovers

Interpreting the spillover index

- From 28% in 2007 to 47% in 2008Q4 (after Lehman); maximum 57.5% on 17 March 2020, in the COVID-19 crash
  - the same pattern as in Diebold and Yilmaz (2009): connectedness jumps in crises and decays slowly
- Most recent value: 22.7%, below the full-sample index (26.4%)
  - the jumps are partly mechanical: a 200-day window that contains a crash is dominated by it for the next 200 days
- For a portfolio holding the BET: diversification across these markets helps least when it is needed most

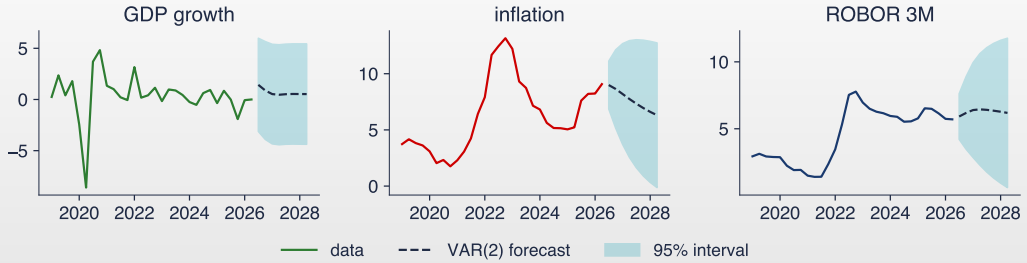
Recap: Variance decomposition

- FEVD: share of each orthogonal shock in the h-step forecast error variance; rows sum to 100%
- Romania: ROBOR is driven by inflation at long horizons; inflation mainly by its own shocks
- Diebold–Yilmaz spillover index: the off-diagonal mass of the generalised FEVD
- Spillovers among S&P 500, DAX and BET peak in crises (2008, 2020)

VAR forecasts and their uncertainty

- **Point forecasts** by recursion: $\hat{\mathbf{Y}}_{T+h|T} = \mathbf{c} + \mathbf{A}_1\hat{\mathbf{Y}}_{T+h-1|T} + \cdots + \mathbf{A}_p\hat{\mathbf{Y}}_{T+h-p|T}$, with $\hat{\mathbf{Y}}_{T+j|T} = \mathbf{Y}_{T+j}$ for $j \leq 0$
  - for a stable VAR they converge to the mean $\boldsymbol{\mu}$ as h grows
- **Error covariance**: $\boldsymbol{\Sigma}(h) = \sum_{s=0}^{h-1} \boldsymbol{\Phi}_s \boldsymbol{\Sigma} \boldsymbol{\Phi}_s'$; it grows with h towards $\boldsymbol{\Gamma}(0)$
  - 95% interval for Y_j: $\hat{Y}_{j,T+h|T} \pm 1.96\sqrt{\sigma_{jj}(h)}$, $\sigma_{jj}(h)$: element (j,j) of $\boldsymbol{\Sigma}(h)$ (Normal shocks; parameter uncertainty ignored)
  - example (Section 2): $\boldsymbol{\Sigma}(2) = \boldsymbol{\Sigma} + \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$, with variances 1.39 and 1.37
- Same logic as the ARMA forecasts of Chapter 2, with matrices in place of numbers

Romanian VAR(2): forecasts for two years



- Data until 2026Q2; forecasts 2026Q3–2028Q2 with 95% intervals

TSA_ch6_forecasting

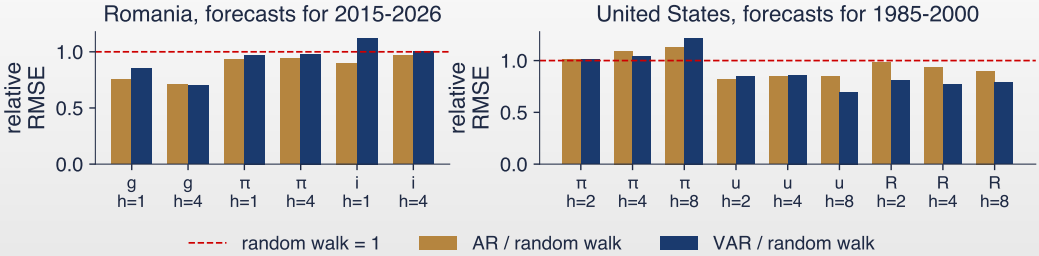
Interpreting the forecasts

- Inflation: from 9.1% in 2026Q2 to 7.7% after a year and 6.3% after two (95%: $[-0.1, 12.7]$)
  - the forecasts move towards the VAR mean (inflation 4.8%, ROBOR 4.8%), slowly because the largest eigenvalue is 0.869
- ROBOR: 6.4% after a year, 6.2% after two; GDP growth: 0.5% per quarter, with an interval $[-4.4, 5.3]$
  - the GDP interval is wide because the 2009 and 2020 outliers inflate $\hat{\sigma}_g$
- A mechanical forecast: it knows nothing of the budget, energy prices or BNR decisions; its value is as a transparent benchmark

Evaluating VAR forecasts fairly

- **Pseudo out-of-sample**: re-estimate on data up to t only, forecast $t + h$, move t forward (expanding window)
  - never evaluate on the data used for estimation (Chapter 0)
- **Benchmarks**: the random walk (no change) and a univariate AR(p) for each variable, p by AIC
  - the question is not "is the VAR good?", but "do the other variables add anything to the series' own past?"
- Measures: RMSE relative to the random walk (below 1: better); the Diebold–Mariano test Diebold and Mariano (1995) for equal accuracy
  - DM statistic: the mean difference of squared errors divided by its standard error; approximately $N(0, 1)$ under H_0

VAR against univariate benchmarks



⊡ Relative RMSE (model / random walk). Left: Romania, 46 forecasts per horizon, origins 2014Q4–2026Q1; right: the US VAR of Section 10, forecasts for 1985Q1–2000Q4, the evaluation period of Stock and Watson (2001)

TSA_ch6_forecasting

Interpreting the forecast comparison

- Romania, one quarter ahead: the VAR loses to the AR for GDP growth (0.85 against 0.75; DM = 2.59, p = 0.010) and is no better for inflation or ROBOR
  - 21 coefficients on 38 to 84 quarters: the estimation noise eats the information of the other variables
- United States, 8 quarters ahead: the VAR beats the AR for unemployment (0.70 against 0.85) and the fed funds rate (0.79 against 0.90)
  - for inflation, nothing beats the random walk (1.22) at this horizon in this period
- VARs help when the sample is long and the links are strong; in short, noisy samples a simple AR is hard to beat

Recap: Forecasting with a VAR

- Recursive point forecasts; error covariance $\sum_s \mathbf{\Phi}_s \mathbf{\Sigma} \mathbf{\Phi}_s'$
- Evaluate out of sample against the random walk and univariate AR models
- Romania: no gain from the VAR; US 1985–2000: gains for unemployment and the interest rate

From reduced form to structural VAR

- **Structural VAR** (SVAR): $\mathbf{B}_0\mathbf{Y}_t = \mathbf{d} + \mathbf{B}_1\mathbf{Y}_{t-1} + \cdots + \mathbf{B}_p\mathbf{Y}_{t-p} + \mathbf{u}_t$, $\mathbf{u}_t$ uncorrelated **structural shocks**
  - $\mathbf{d}$: constants; $\mathbf{B}_1, \ldots, \mathbf{B}_p$: structural lag coefficients; $\mathbf{B}_0$: the contemporaneous effects (e.g. the policy rate reacts to this quarter's inflation)
  - multiplying by $\mathbf{B}_0^{-1}$ gives the reduced form, with $\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\mathbf{u}_t$
- **Identification**: $\boldsymbol{\Sigma}$ has $K(K+1)/2$ distinct elements, $\mathbf{B}_0^{-1}$ has K^2; we need $K(K-1)/2$ restrictions
  - Cholesky: $\mathbf{B}_0^{-1} = \mathbf{P}$ lower triangular: exactly $K(K-1)/2$ zeros; the recursive SVAR of Sims (1980)
  - other schemes: long-run restrictions, sign restrictions, external instruments (beyond this course)
- The data identify the reduced form; the structural interpretation always rests on assumptions that must be argued with economics

Case study: Stock and Watson (2001)

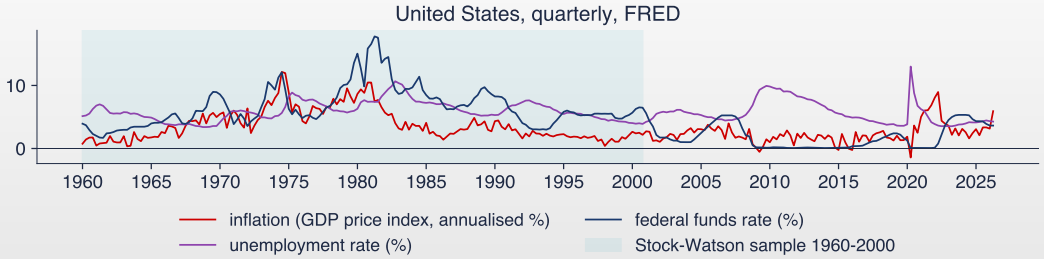


Eccles Building, Washington: the Federal Reserve Board

Photo: Federal Reserve (2011); public domain; Wikimedia Commons

- Stock and Watson (2001): "Vector autoregressions", a widely used survey of the method
  - three variables: inflation π_t, unemployment u_t, the federal funds rate R_t
  - $\pi_t = 400 \ln(P_t / P_{t-1})$, P_t: the GDP price index; 400 annualises, in %
  - quarterly, 1960Q1–2000Q4, VAR(4), Cholesky order π, u, R
- Four tasks: data description, forecasting, structural inference, policy analysis
  - their verdict: VARs are good at the first two; the last two depend on the identifying assumptions
- We re-estimate it with today's FRED data (series GDPCTPI, UNRATE, FEDFUNDS): the numbers differ slightly from the published ones because the data have been revised

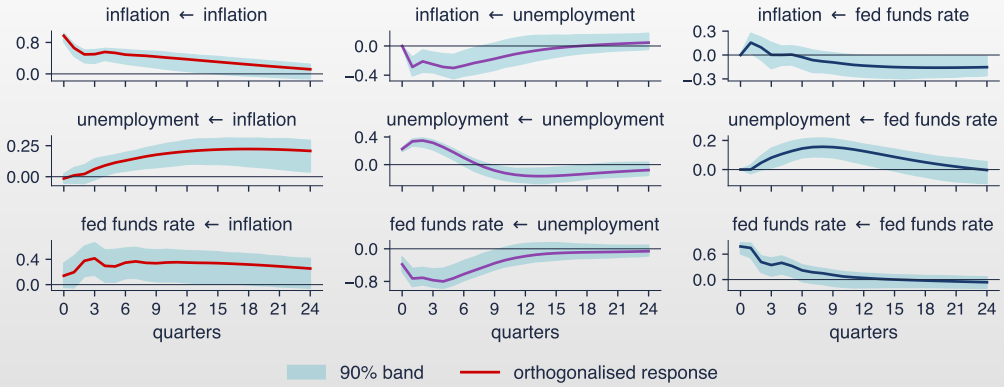
The three US series



⊡ Quarterly, 1960Q1–2026Q2 (FRED); shaded: the sample of Stock and Watson (2001), 164 quarters

TSA_ch6_stock_watson

Stock–Watson VAR: impulse responses



- VAR(4), 1960Q1–2000Q4, 39 coefficients; Cholesky order inflation, unemployment, fed funds rate; 24 quarters; 90% bootstrap bands

TSA_ch6_stock_watson

Interpreting the Stock–Watson VAR

- A monetary shock ($+0.78$ pp in the fed funds rate) raises unemployment by up to 0.16 pp after 8 quarters
  - inflation first rises (0.16 pp, a price puzzle), then falls slowly (-0.13 pp after 3 years)
- Granger tests: unemployment $\to$ inflation ($p = 0.036$), inflation and unemployment $\to$ fed funds ($p < 0.001$, < 0.001), fed funds $\to$ inflation not significant ($p = 0.402$)
  - FEVD after 12 quarters: monetary shocks explain 19% of unemployment and 2% of inflation
- On the extended sample 1960–2026Q2 ($T = 262$) the unemployment response has the wrong sign in the first quarters (-0.13 pp): zero rates after 2008 and the pandemic change the system

The legacy of the VAR



Christopher Sims giving his Nobel lecture, Stockholm, 8 December 2011

Photo: Holger Motzkau (2011); CC BY-SA 3.0; Wikimedia Commons

- Central banks use VARs for forecasts and policy scenarios
  - and their descendants: Bayesian VARs, DSGE models compared with VARs
- The price puzzle Sims (1992)
  - led to adding commodity prices and expectations to monetary VARs
- Non-stationary variables that move together need the error-correction form
  - Chapter 7: cointegration and VECM Engle and Granger (1987)

Recap: Structural VARs

- SVAR: $\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\mathbf{u}_t$; $K(K-1)/2$ restrictions identify the structural shocks
- Cholesky is the simplest (recursive) identification
- Stock and Watson (2001): a monetary shock raises unemployment; inflation reacts slowly
- Results depend on the sample and on the identification

Possible contribution of AI

- **Code**: a first draft of a script that downloads several series, aligns their dates, selects p and estimates a VAR
- **Explanation**: a second explanation of the Cholesky decomposition or of a FEVD table
- **Exploration**: Granger tests and spillover indices for many pairs of markets or EU countries at once
- Example prompt
  - `Write Python code that builds a quarterly data set for Romania (real GDP growth from Eurostat namq_10_gdp, 12-month HICP inflation from prc_hicp_minr, ROBOR 3M from irt_st_m), chooses the VAR lag order by AIC, BIC and HQ, checks stability and residuals, runs Granger tests and plots orthogonalised impulse responses with bootstrap bands.`

Checks you must run

- The dates: series aligned on the same quarters, no forward-filled or overlapping values that create spurious leads
- Stationarity of each variable (Chapter 3) before Granger tests; with $I(1)$ variables, differences, Toda–Yamamoto or a VECM
- Stability from the companion matrix, not from $\mathbf{A}_1$ alone
- The Cholesky ordering used and its justification; the same results with another ordering or with generalised responses
- That "Granger-causes" is never reported as "causes"
- Every cited reference: it must exist; check the DOI

Key takeaways

- A VAR(p) regresses each variable on the past of all; OLS equation by equation; few variables and lags
- Stability: all eigenvalues of the companion matrix inside the unit circle; then a mean and a moving-average form exist
- Choose p with AIC, BIC, HQ, then check the residuals with the portmanteau test
- Granger causality = incremental predictability, relative to the chosen variables
- IRF and FEVD need orthogonal shocks: the Cholesky ordering is an identifying assumption
- VAR forecasts must beat AR and random-walk benchmarks out of sample; often they do not

Key formulas

Quantity	Formula
VAR(p)	$\mathbf{Y}_t = \mathbf{c} + \mathbf{A_1}\mathbf{Y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'] = \boldsymbol{\Sigma}$
Stability, mean	$\lvert\lambda_i(\mathbf{F})\rvert < 1; \quad \boldsymbol{\mu} = (\mathbf{I} - \mathbf{A_1} - \cdots - \mathbf{A}_p)^{-1}\mathbf{c}$
Criteria	$\ln\det\tilde{\boldsymbol{\Sigma}}(p) + c_T\, pK^2/T, \quad c_T = 2,\ \ln T,\ 2\ln\ln T$
Portmanteau	$Q_h \sim \chi^2(K^2(h-p))$
Granger F	$\dfrac{(RSS_R - RSS_U)/p}{RSS_U/(T-Kp-1)} \sim F(p, T-Kp-1)$
MA form, IRF	$\boldsymbol{\Phi}_h = \sum_{j=1}^{\min(h,p)} \mathbf{A}_j\boldsymbol{\Phi}_{h-j}, \quad \boldsymbol{\Theta}_h = \boldsymbol{\Phi}_h\mathbf{P}, \quad \boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}'$
FEVD	$\omega_{jk}(h) = \sum_{s<h}\theta_{jk,s}^2 / \sum_{s<h}\sum_m \theta_{jm,s}^2$
Forecast error	$\boldsymbol{\Sigma}(h) = \sum_{s=0}^{h-1} \boldsymbol{\Phi}_s\boldsymbol{\Sigma}\boldsymbol{\Phi}_s'$

Self-assessment

- **Question**: is the VAR(1) with $\mathbf{A} = \begin{pmatrix} 0.9 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}$ stable?
  - **Answer**: no: $\lambda^2 - 1.7\lambda + 0.66 = 0$ gives $\lambda = 1.1$ and 0.6
- **Question**: a Granger test gives $\mathrm{p} = 0.30$ for $x \to y$. Does x have no effect on y?
  - **Answer**: no such conclusion: the lags of x do not improve the forecast; effects within the period, slow effects or low power remain possible
- **Question**: in a Cholesky FEVD, what share of the 1-step error variance of the first variable comes from its own shock?
  - **Answer**: 100%, by construction of the ordering
- Next: Chapter 7, cointegration and the vector error correction model (VECM)

References (1/3)

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. doi:10.1109/TAC.1974.1100705
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. doi:10.1080/07350015.1995.10524599
- Diebold, F. X., & Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534), 158–171. doi:10.1111/j.1468-0297.2008.02208.x
- Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1), 57–66. doi:10.1016/j.ijforecast.2011.02.006
- Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2), 251–276. doi:10.2307/1913236
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438. doi:10.2307/1912791
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press. doi:10.2307/j.ctv14jx6sm
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B*, 41(2), 190–195. doi:10.1111/j.2517-6161.1979.tb01072.x

References (2/3)

- Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, 75(371), 602–608. doi:10.1080/01621459.1980.10477520
- Huang, C., & Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python*. Springer. doi:10.1007/978-3-031-13584-2
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.), Section 12.3, Vector autoregressions. OTexts. otexts.com/fpp3/VAR.html
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255–259. doi:10.1016/0165-1765(80)90024-5
- Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics*, 80(2), 218–230. doi:10.1162/003465398557465
- Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1), 119–147. doi:10.1016/0304-4076(95)01753-4
- Lütkepohl, H. (1982). Non-causality due to omitted variables. *Journal of Econometrics*, 19(2–3), 367–378. doi:10.1016/0304-4076(82)90011-2

References (3/3)

- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer. doi:10.1007/978-3-540-27752-1
- Pesaran, M. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1), 17–29. doi:10.1016/S0165-1765(97)00214-0
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. doi:10.1214/aos/1176344136
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1), 1–48. doi:10.2307/1912017
- Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36(5), 975–1000. doi:10.1016/0014-2921(92)90041-T
- Sims, C. A., Stock, J. H., & Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, 58(1), 113–144. doi:10.2307/2938337
- Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15(4), 101–115. doi:10.1257/jep.15.4.101
- Toda, H. Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, 66(1–2), 225–250. doi:10.1016/0304-4076(94)01616-8