LECTURE 1.2

Vector Autoregressions

ECON42240 Advanced Macroeconomics

Sam Deegan
University College Dublin

Spring 2027

Where We Are

Last Week's Lecture

Last week built the evidence a model has to match:

1. **Syllabus.**
   What the module covers, and how you are assessed.
2. **Trends and Cycles.**
   Why a trend is a modelling choice.
3. **Shocks and Propagation.**
   Where a series gets its volatility.
4. **From One Series to Many.**
   What a vector autoregression adds.

Lecture Plan

1. **Specifying a VAR.**
   Which variables, how many lags, on what data.

2. **Estimating a VAR.**
   Equation by equation, by least squares.

3. **Reading the Output.**
   Impulse responses, error bands and variance shares.

4. **Residuals and Structural Shocks.**
   Why every mix of shocks fits the data.

SECTION 1

Specifying a VAR

What Sims Objected To

Christopher Sims attacked large macro models for the restrictions they imposed (Sims, 1980):

- **The Charge:** Each equation usually excluded most other variables, a strong restriction.
- **The Alternative:** Let every variable depend on the lags of all variables.

A VAR Is Not Theory Free

A VAR still rests on identifying assumptions, stated openly rather than buried.

The VAR in Matrix Form

Stacked into vectors, Lecture 1.1's two-variable system with one lag is one equation:

$$Y_t = AY_{t-1} + e_t, \qquad Y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad e_t = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \tag{1}$$

One Matrix Holds Every Coefficient

Row i of A is equation i. Column j holds the coefficients on variable j's lag.

A Constant Is Just Another Variable

An intercept needs no new notation, only one more variable:

- **Add a Variable:** Set it to one each period, equal to its own lag.
- **The Intercept Returns:** Its coefficients in the other equations are the intercepts.
- **Nothing Else Changes:** The system keeps the form $Y_t = AY_{t-1} + e_t$.

Longer Lags Fit the Same Form

Each extra lag adds one coefficient per variable to every equation:

$$\begin{aligned} y_{1t} &= a_{11}y_{1,t-1} + a_{12}y_{1,t-2} + a_{13}y_{2,t-1} + a_{14}y_{2,t-2} + e_{1t} \\ y_{2t} &= a_{21}y_{1,t-1} + a_{22}y_{1,t-2} + a_{23}y_{2,t-1} + a_{24}y_{2,t-2} + e_{2t} \end{aligned} \tag{2}$$

Two Lags Renumber the Coefficients

With one lag, a_{12} multiplied $y_{2,t-1}$. Here a_{12} multiplies $y_{1,t-2}$.

The Companion Form Absorbs Extra Lags

Stacking current and lagged values makes the two-lag system first order, $Z_t = AZ_{t-1} + e_t$:

$$Z_t = \begin{pmatrix} y_{1t} \\ y_{1,t-1} \\ y_{2t} \\ y_{2,t-1} \end{pmatrix}, \qquad A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 1 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{3}$$

Stacking Lags Makes Any VAR First Order

An identity row carries each lag forward. Powers of one matrix then cover any lag length.

Two of the Four Rows Are Bookkeeping

Multiplying $Z_t = AZ_{t-1} + e_t$ out returns both equations and two identities:

$$\begin{aligned}
y_{1t} &= a_{11}y_{1,t-1} + a_{12}y_{1,t-2} + a_{13}y_{2,t-1} + a_{14}y_{2,t-2} + e_{1t} \\
y_{1,t-1} &= y_{1,t-1} \\
y_{2t} &= a_{21}y_{1,t-1} + a_{22}y_{1,t-2} + a_{23}y_{2,t-1} + a_{24}y_{2,t-2} + e_{2t} \\
y_{2,t-1} &= y_{2,t-1}
\end{aligned} \tag{4}$$

Stability Is an Eigenvalue Condition

A VAR is stable when a shock's effect eventually dies away:

- **The Condition:** Every eigenvalue of A lies inside the unit circle, so $A^h \to 0$.
- **One Variable:** A is then ρ, and the condition is last week's $|\rho| < 1$.

> **A Unit Root Leaves a Shock in the Level**
> An eigenvalue of one leaves a shock's effect in the level permanently. Differencing is one response, and differencing is a choice.

Which Variables Go In

A VAR's variable list is an economic choice rather than a statistical one:

> **Leaving a Variable Out Is an Assumption**
> A variable left out has zero coefficients in every equation. Omitting it treats it as irrelevant to the variables inside.

The VAR here holds output, a price level and a short-term policy rate.

The Series a Baseline VAR Uses

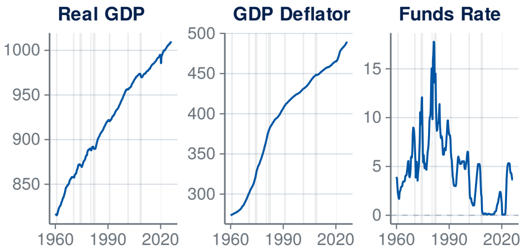

Figure 1: Real GDP and deflator (100 × log), funds rate (%), 1960–2026; recessions shaded.

> **Each Series Needs a Transformation Decision**
> Output and the deflator trend upward. The VAR's sample ends in 2007, before the funds rate neared zero.

Detrending Determines What the VAR Estimates

Each way of handling a trend carries a cost inside a VAR:

- **Levels:** Keeps the long-run relationships, at the cost of non-stationary data.
- **Differences:** Removes a unit root, and discards the level relationships.
- **Filtered:** An HP cycle puts the filter inside every impulse response.

The baseline VAR here uses levels. Its largest eigenvalue has modulus 0.997.

How Many Lags a VAR Needs

Lag length sets which response shapes a VAR can produce:

- **Too Few Lags:** Serial correlation stays in the residuals and biases the responses.
- **Too Many Lags:** Spends degrees of freedom and widens the error bands.
- **The Criteria:** AIC, HQ, BIC and FPE trade fit against parameter count.

> **BIC's Lag Length Is Never Longer Than AIC's**
> BIC's penalty per coefficient is $\log T$ for T observations, against AIC's 2. From $T = 8$, BIC's is heavier.

Information Criteria at One to Eight Lags

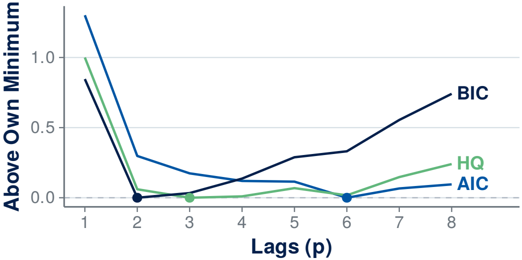

Figure 2: AIC, HQ and BIC by lag length, each relative to its minimum, 1960–2007.

> **The Criteria Span Two to Six Lags**
> BIC reaches its minimum at two lags, HQ at three, AIC at six. FPE agrees with AIC, and the VAR uses six.

What the Specification Costs

A VAR with n variables and k lags carries n^2k lag coefficients:

9
three variables, one lag

216
six variables, six lags

36
regressors per equation at 6×6

Shrinkage Is the Alternative to Cutting
A large VAR can overfit, forecast badly and give wide bands. Cutting variables or lags risks misspecification.

Which Three Variables Would You Choose?

You want the dynamic effect of a rate change on output.

> **Question**
> Name three variables, a sample and a transformation. What does your choice rule out?

Take answers from the room.

SECTION 2

Estimating a VAR

Least Squares, One Equation at a Time

A VAR is a set of linear equations. Least squares is an obvious estimator:

- **Same Regressors Everywhere:** Every equation carries the identical lagged variables.
- **Equation by Equation:** Each equation alone gives the system estimate.
- **The Catch:** Least squares estimates of a VAR's coefficients are biased.

Least Squares Estimates Persistence as a Ratio

Least squares estimates the AR(1) persistence (ρ) as a ratio of sums:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} y_{t-1} y_t}{\sum_{t=2}^{T} y_{t-1}^2} \tag{5}$$

Put Lecture 1.1's AR(1), $\rho y_{t-1} + \epsilon_t$, in place of y_t in the numerator:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} y_{t-1} (\rho y_{t-1} + \epsilon_t)}{\sum_{t=2}^{T} y_{t-1}^2} \tag{6}$$

The Estimate Is the Truth Plus an Error

Multiply out the numerator; ρ is constant, so it factors out of its sum:

$$\hat{\rho} = \frac{\rho \sum_{t=2}^{T} y_{t-1}^2 + \sum_{t=2}^{T} y_{t-1}\epsilon_t}{\sum_{t=2}^{T} y_{t-1}^2} \tag{7}$$

Split the fraction in two; the first part cancels to ρ:

$$\hat{\rho} = \frac{\rho \cancel{\sum y_{t-1}^2}}{\cancel{\sum y_{t-1}^2}} + \frac{\sum_{t=2}^{T} y_{t-1}\epsilon_t}{\sum_{t=2}^{T} y_{t-1}^2} = \rho + \frac{\sum_{t=2}^{T} y_{t-1}\epsilon_t}{\sum_{t=2}^{T} y_{t-1}^2} \tag{8}$$

Why the Error Does Not Average Out

The numerator has expectation zero. The ratio's expectation is not zero:

- **The Numerator:** ϵ_t is independent of y_{t-1}, so $E(y_{t-1}\epsilon_t) = 0$.
- **The Denominator:** With $\rho > 0$, a positive ϵ_t raises y_t and every later value.
- **The Result:** A large ϵ_t comes with a large denominator, so $E(\hat{\rho}) < \rho$.

> **Persistence Is Estimated Too Low**
> With $\rho > 0$, estimated responses therefore decay faster than the true ones.

What Makes the Bias Large

In a VAR as in an AR(1), two factors set the bias:

- **Persistence:** Larger own-lag coefficients strengthen the correlation and the bias.
- **Sample Length:** A longer sample weakens the correlation and shrinks the bias.

> **Macro Series Are Persistent and Macro Samples Are Short**
> Quarterly macro series in levels have own-lag coefficients near one. Both conditions for a large bias hold.

The Sampling Distribution of the Estimate

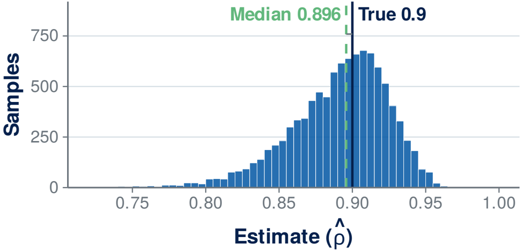

Figure 3: Least squares estimates of ρ from 10,000 samples, $\rho = 0.9$, $T = 200$.

> **The Average Estimate Falls Short of 0.9**
> The mean estimate is 0.891 and the median 0.896. Of the estimates, 55 per cent fall below 0.9.

The Bootstrap Bias Adjustment

The bootstrap measures the bias by treating the estimate $\hat{A}$ as the truth:

1. **Resample:** Draw from the saved residuals $(\hat{e}_t)$, building many series from $\hat{A}$.
2. **Re-Estimate:** Fit a VAR to each artificial series and keep each estimate $\hat{A}^*$.
3. **The Median:** Call the median of those estimates $\bar{A}$.

The Adjustment Subtracts the Measured Bias

The gap between $\bar{A}$ and $\hat{A}$ estimates the bias, so subtract it:

$$\hat{A}_{\text{BOOT}} = \hat{A} - \underbrace{\left(\bar{A} - \hat{A}\right)}_{\text{the estimated bias}} \qquad (9)$$

Positive Own Lags Mean an Upward Adjustment
With positive own lags $\bar{A}$ sits below $\hat{A}$. The adjustment then raises the persistence estimate.

Maximum Likelihood Justifies Least Squares

Least squares also has a large-sample defence in the likelihood:

> **The Likelihood and Its Maximiser**
> The likelihood is the sample's probability, given the parameters. Its maximiser is consistent and asymptotically efficient.

With normal errors that maximiser is least squares, still biased in short samples.

Shrinking the Coefficients Instead

Bayesian estimation pulls coefficients toward a prior, a belief set before the data:

- **Minnesota Prior:** Prior mean large and positive on own first lag, zero elsewhere.
- **The Estimate:** A weighted average of least squares and the prior mean.
- **The Tightness:** Your confidence in the prior sets its weight.

> **Shrinkage Avoids Setting Coefficients to Zero**
> Dropping a variable sets its coefficients to zero with certainty. A prior only pulls them toward zero.

Would You Correct the Bias?

Your VAR's own-lag coefficients sit near one, on a short quarterly sample.

> **Question**
> The bootstrap raises your persistence estimate. What happens to the impulse responses, and to the bands?

Take answers from the room.

SECTION 3

Reading the Output

Today's Value Sums Every Residual So Far

Substituting the VAR into itself repeatedly leaves current and past residuals alone:

$$Y_t = e_t + Ae_{t-1} + A^2 e_{t-2} + \cdots + A^t e_0 \tag{10}$$

A Vector Moving Average

The moving average writes today's values as a weighted sum of residuals. A residual j quarters old carries weight A^j.

The Impulse Response Is a Power of the Matrix

In the moving average, set every residual to zero, except one unit in the first equation:

$$e_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \Longrightarrow \quad Y_h = \underbrace{A^h}_{\text{the response}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{11}$$

A Matrix Power Replaces the Scalar Decay
Column j of A^h holds every variable's response to equation j's residual. The AR(1) had ρ^h.

A Forecast Needs Only the Estimated Coefficients

Setting future residuals to their mean of zero gives the forecast:

$$E_t Y_{t+1} = AY_t, \qquad E_t Y_{t+2} = A^2 Y_t, \qquad E_t Y_{t+h} = A^h Y_t \tag{12}$$

Responses to a Funds Rate Residual

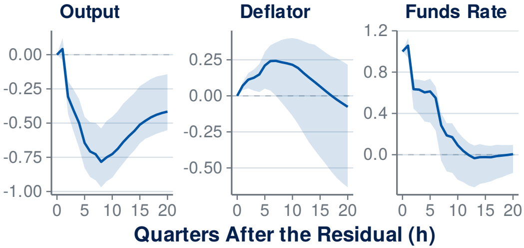

Figure 4: Responses to a one-point funds rate residual, 1960–2007; 10th–90th percentile bands.

> **Output's Band Stays Below Zero**
> Output's band lies below zero from two quarters to twenty. The deflator's lies above zero for seven.

An Error Band Answers a Sign Question

An estimated response is a point estimate, so a band belongs beside it:

> **Percentile Bands Sign a Response**
> Report the tenth and ninetieth percentiles, or the fifth and ninety-fifth. Plus or minus one standard deviation covers only about 68 per cent.

Asymptotic formulas describe short samples poorly, so bootstrap bands are now common.

Bootstrapping the Error Bands

A band needs a distribution, and the bootstrap supplies one:

1. **Resample:** Draw from the fitted residuals, building many artificial series.
2. **Re-Estimate:** Fit a VAR to each series and compute its impulse responses.
3. **Take Quantiles:** At each horizon, take the tenth and ninetieth percentiles.

> **One Simulation Gives the Bias and the Bands**
> Take the median for the bias, and the quantiles for the band.

The Forecast Error Is the Residuals Still to Come

Iterating the VAR h periods ahead splits the future value in two:

$$Y_{t+h} = \underbrace{A^h Y_t}_{\text{the forecast}} + \underbrace{\sum_{j=0}^{h-1} A^j e_{t+h-j}}_{\text{the forecast error}} \tag{13}$$

A Decomposition Splits the Error's Variance

A variance decomposition gives each shock's share of the forecast error's variance:

> **Shares Need Uncorrelated Shocks**
> The error sums the residuals arriving between t and $t + h$. Its variance splits into shares only if those residuals are uncorrelated.

Shares of Output's Forecast Error Variance

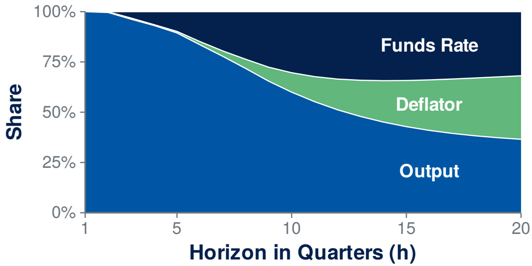

Figure 5: Output's forecast error variance by shock; ordering output, deflator, funds rate.

> **Output's Own Shock Explains Less at Long Horizons**
> Output's own shock explains 93 per cent at four quarters, 37 at twenty. The funds rate's share reaches a third.

How Would You Make the Shocks Uncorrelated?

The funds rate residual correlates 0.23 with output's and 0.20 with the deflator's.

> **Question**
> A decomposition needs uncorrelated shocks. How would you build them from these residuals, and what would you assume?

Take answers from the room.

SECTION 4

Residuals and Structural Shocks

A Residual Is Not a Shock

The VAR so far is a reduced form, with no theory in it. Its residuals have two readings:

Each Residual Is Its Own Shock
e_{1t} enters only y_{1t} on impact, and e_{2t} only y_{2t}.

Each Residual Mixes Structural Shocks
Supply and demand shocks each enter both variables on impact.

A structural shock has one economic source. Identification usually takes such shocks as uncorrelated.

Any Mixing Matrix Fits the Data

Substitute $C\varepsilon_t$ for e_t, with shocks (ε_t) mixed by an unknown matrix (C):

$$e_t = \underbrace{C\varepsilon_t}_{\text{the residual, as a mix of shocks}} \implies Y_t = C\varepsilon_t + AC\varepsilon_{t-1} + A^2 C\varepsilon_{t-2} + \cdots \tag{14}$$

The Data Cannot Distinguish One C from Another

Every invertible C fits the sample equally well, however long the sample. The structural responses $A^h C$ change with C.

Each Variable Can Enter the Other's Equation

A central bank sets this quarter's rate knowing this quarter's inflation:

$$\begin{aligned} y_{1t} &= a_{12}y_{2t} + b_{11}y_{1,t-1} + b_{12}y_{2,t-1} + \varepsilon_{1t} \\ y_{2t} &= a_{21}y_{1t} + b_{21}y_{1,t-1} + b_{22}y_{2,t-1} + \varepsilon_{2t} \end{aligned} \tag{15}$$

Moving the same-quarter terms left gives the matrices A_0 and B:

$$A_0 Y_t = BY_{t-1} + \varepsilon_t, \qquad A_0 = \begin{pmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{pmatrix} \tag{16}$$

The Reduced Form Mixes the Structural Shocks

Multiplying through by the inverse of A_0 returns the system already estimated:

$$Y_t = \underbrace{A_0^{-1}B}_{A}\, Y_{t-1} + \underbrace{A_0^{-1}\varepsilon_t}_{e_t} \tag{17}$$

Every Residual Is a Combination of Shocks

A_0^{-1} plays the role of C, so every A_0 fits equally well. The structural responses are $A^h A_0^{-1}$.

Correlated Residuals Split One Question into Two

Reduced-form residuals are usually correlated, so a what-if question has two versions:

- **First Question:** What follows a shock to the first variable alone?
- **Second Question:** What usually follows, with the second variable's shock alongside?

> **A What-If Needs a Shock That Arrives Alone**
> Structural questions usually concern uncorrelated shocks. A forecast needs only the reduced form.

Identification Needs $2n^2$ Restrictions

The general structural VAR, $A_0 Y_t = BY_{t-1} + C\varepsilon_t$, has more parameters than the data reveal:

$$\underbrace{3n^2 + \tfrac{n(n+1)}{2}}_{\text{structural}} - \underbrace{n^2 + \tfrac{n(n+1)}{2}}_{\text{reduced form}} = 2n^2 \tag{18}$$

> **Doing Nothing Is Itself an Identification**
> Reading the reduced form as structural imposes all of them: $A_0 = C = I$.

Theory Enters Through the Identifying Restrictions

Identification supplies the restrictions that turn correlated residuals into uncorrelated shocks:

> **A Cholesky Ordering Is a Causal Assumption**
> The common choice is a recursive ordering, built by the Cholesky decomposition. The ordering fixes which shocks reach which variables within the quarter.

Three variables allow $3! = 6$ orderings, and no sample can test one against another.

Two Orderings Give Different Output Paths

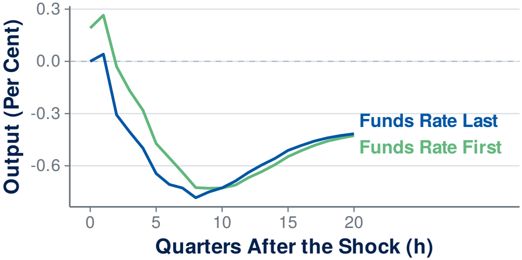

Figure 6: Output's response to a one-point funds rate shock, two orderings, 1960–2007.

> **Both Orderings Fit the Data Identically**
> Ordered first, the rate shock raises output on impact. Ordered last, the shock reproduces the residual response.

Which Variable Comes First?

You must order output, prices and the policy rate for a recursive VAR.

> **Question**
> Which variable would you put first, and what would you be assuming?

Take answers from the room.

Before Lecture 1.3

1. **Play:**
   Fit a three-variable VAR in R with `vars`, and plot the responses.

2. **Apply:**
   Simulate an AR(1) at $\rho = 0.9$, estimate it, and measure the bias.

3. **Read:**
   Stock and Watson on vector autoregressions (Stock & Watson, 2001).

Next Lecture

The next lecture supplies the restrictions this one showed were needed:

1. **The Identification Problem.**
   Why residuals are not shocks.
2. **Recursive Identification.**
   What a Cholesky ordering assumes.
3. **Monetary Policy Shocks.**
   Why the ordering can decide the answer.
4. **Long-Run Restrictions.**
   Identifying technology shocks instead.

Thank you

sam.deegan@ucdconnect.ie
sam-deegan.com

BACKUP

Appendix

Notation

Symbol	What it is	First met
y_{it}	variable i of the VAR at date t	Lecture 1.1
a_{ij}	coefficient j in equation i	Lecture 1.1
e_{it}	the residual in equation i	Lecture 1.1
t	the date, a quarter	Lecture 1.1
h	periods since the shock, the horizon	Lecture 1.1
ρ	the AR(1) persistence	Lecture 1.1
y_t	the AR(1) series	Lecture 1.1
ϵ_t	the AR(1) shock	Lecture 1.1
E	the expectation	Lecture 1.1
Y_t	the vector of the VAR's variables	The VAR in Matrix Form
A	the matrix of reduced-form VAR coefficients	The VAR in Matrix Form
e_t	the vector of reduced-form residuals	The VAR in Matrix Form
j	an equation, a column or a lag in a sum	The VAR in Matrix Form
i	an equation, or a row of A	The VAR in Matrix Form
Z_t	the stacked vector in the companion form	The Companion Form Absorbs Extra Lags
T	the number of observations in the sample	How Many Lags a VAR Needs

Notation, Concluded

Symbol	What it is	First met
n	the number of variables in the VAR	What the Specification Costs
k	the number of lags in the VAR	What the Specification Costs
$\hat{\rho}$	the least squares estimate of ρ	Least Squares Estimates Persistence as a Ratio
$\hat{A}$	the least squares estimate of A	The Bootstrap Bias Adjustment
$\hat{e}_t$	the saved least squares residuals	The Bootstrap Bias Adjustment
$\hat{A}^*$	an estimate from one bootstrap sample	The Bootstrap Bias Adjustment
$\bar{A}$	the median bootstrap estimate of A	The Bootstrap Bias Adjustment
$\hat{A}_{\mathrm{BOOT}}$	the bias-adjusted estimate of A	The Adjustment Subtracts the Measured Bias
E_t	the expectation given information at t	A Forecast Needs Only the Estimated Coefficients
ε_t	the vector of structural shocks	Any Mixing Matrix Fits the Data
C	the matrix mixing shocks into residuals	Any Mixing Matrix Fits the Data
A_0	the matrix of contemporaneous coefficients	Each Variable Can Enter the Other's Equation
B	the matrix of structural lag coefficients	Each Variable Can Enter the Other's Equation
b_{ij}	coefficient j in structural equation i	Each Variable Can Enter the Other's Equation
I	the identity matrix	Identification Needs $2n^2$ Restrictions

References

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48. https://doi.org/10.2307/1912017

Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, *15*(4), 101–115.