Vector Autoregressions (VARs)

Wouter J. Den Haan
London School of Economics

Wouter J. Den Haan

March 23, 2018

Overview

- Impulse Response Functions
- Reduced form & Structural VARs
    - Short-term restrictions
    - Long-term restrictions
    - Sign restrictions
- Estimation
- Problems/topics

How to estimate/evaluate models?

- Full information methods like ML and its Bayesian version take *every* aspect of the model as truth

- A less ambitious approach is to focus on just some "key properties"
  - *both in the model and in the data*

- What properties?
  - means, standard deviations, cross-correlations
  - but propagation of shocks is key aspect of economic models $\Longrightarrow$ autocovariance say something about this but not in the most intuitive way
  - IRFs are better for this

General definition IRFs

- Suppose

  $y_t = f(y_{t-1}, y_{t-2}, \cdots, y_{t-p}, \varepsilon_t)$ and ε_t has a variance equal to σ^2

- The IRF gives the j^{th}-period response when the system is shocked by a one-standard-deviation shock.

General definition IRFs

- Consider a sequence of shocks $\{\bar{\varepsilon}_t\}_{t=1}^{\infty}$. $\{\bar{y}_t\}_{t=1}^{\infty}$ are the generated series

- Consider an *alternative* series of shocks such that

$$\tilde{\varepsilon}_t = \begin{cases} \bar{\varepsilon}_t + \sigma & \text{if } t = \tau \\ \bar{\varepsilon}_t & \text{o.w.} \end{cases}$$

- The IRF is then defined as

$$IRF(j) = \tilde{y}_{\tau-1+j} - \bar{y}_{\tau-1+j}$$

IRFs for linear processes

- Linear processes: The IRF is independent of the particular draws for $\bar{\varepsilon}_t$

- Thus we can simply start at the steady state (that is when $\bar{\varepsilon}_t$ has been zero for a very long time)

- The effect of a shock of size $\Lambda\sigma$ is Λ times the effect of a shock of size σ

IRFs for linear processes

- For example, if

$$y_t = \rho y_{t-1} + \varepsilon_t$$

then

$$IRF(j) = \sigma \rho^{j-1}$$

- Often you can *not* get an analytical formula for the impulse response function, but simple iteration on the law of motion (driving process) gives you the exact same answer

- Note that this IRF is not stochastic

IRFs for nonlinear processes

- IRF depends on
  1. state in the period when shock occur $(y_{t-1}, y_{t-2}, \cdots y_{t-p})$
  2. subsequent shocks
- Moreover, the effect of a shock of size $\Lambda\sigma$ is not Λ times the effect of a shock of size σ

IRFs in theoretical models

- When you have solved for the policy functions, then it is trivial to get the IRFs by simply giving the system a one standard deviation shock and iterating on the policy functions.

- Shocks in the model are *structural* shocks, such as
    - productivity shock
    - preference shock
    - monetary policy shock

IRFs in the data

The big question

- Can we estimate IRFs from the data **without** specifying an explicit theoretical model
- That is what *structural* VARs attempt to do

VARs & IRFs

What we are going to do?

- Describe an empirical model that has turned out to be very useful (for example for forecasting)
  - *Reduced*-form VAR

- Describe a way to back out structural shocks (this is the hard part)
  - *Structural*-VAR

Reduced Form VARs

- Let y_t be an $n \times 1$ vector of n variables (typically in logs)

$$y_t = \sum_{j=1}^{J} A_j y_{t-j} + u_t$$

where A_j is an $n \times n$ matrix.

- Wold representation is a justification for the linearity.

Reduced Form Vector AutoRegressive models (VARs)

- constants and trend terms are left out to simplify the notation
- This system can be estimated by OLS (equation by equation) even if y_t contains $I(1)$ variables

Estimation of VARs

$$y_t = \sum_{j=1}^{J} A_j y_{t-j} + u_t$$

Claim:

- You can simply estimate a VAR in (log) levels even if variables are I(1) (and even when you have higher-order integration as long as you have enough lags)
- Why?

Spurious regression

- Let z_t and x_t be $I(1)$ variables that have nothing to do with each other
- Consider the regression equation

$$z_t = ax_t + u_t$$

- The least-squares estimator is given by

$$\hat{a}_T = \frac{\sum_{t=1}^{T} x_t z_t}{\sum_{t=1}^{T} x_t^2}$$

- Problem:

$$\lim_{T \longrightarrow \infty} \hat{a}_T \neq 0$$

Source of spurious regressions

- The problem is not that z_t and x_t are I(1)
- The problem is that there is not a single value for a such that u_t is stationary
- If z_t and x_t are cointegrated then there is a value of a such that

$$z_t - ax_t \text{ is stationary}$$

- Then least-squares estimates of a are consistent
- but you have to change formula for standard errors

How to avoid spurious regressions?

Answer: Add enough lags.

- Consider the following regression equation

$$z_t = ax_t + bz_{t-1} + u_t$$

- Now there are values of the regression coefficients so that u_t is stationary, namely

$$a = 0 \text{ and } b = 1$$

- So as long as you have enough lags in the VAR you are fine (but be careful with inferences)

How to get standard errors?

- If all data series are stationary you can get standard errors using the usual formulas (see Hamilton 1994).
- If they are not you can use bootstrapping

Bootstrapping

- Suppose

$$y_t = ay_{t-1} + \varepsilon_t$$

$$\widehat{a}_T = \frac{\sum y_t y_{t-1}}{\sum y_{t-1} y_{t-1}}$$

- How to get standard errors for IRF?
  technique easily generates for more complex VAR and other statistics

Bootstrapping

1. Estimate model and IRF

2. Calculate residuals, $\{\widehat{\varepsilon}_t\}_{t=2}^{T} = \Theta$

3. Generate J new sample of length T from

$$\begin{aligned} z_t &= \widehat{a}_T z_{t-1} + e_t \\ z_1 &= y_1 \end{aligned}$$

e_t is drawn from Θ

Bootstrapping

4. In each sample j calculate statistics of interest,
   e.g., 4^{th} and 6^{th}-period IRF, $IRF(4, j)$ and $IRF(6, j)$

5. Order statistics across all J samples from small to large

6. Use this distribution to calculate confidence intervals
   e.g., 90% confidence goes from 5^{th} to 95^{th} percentile

Structural VARs

Consider the reduced-form VAR

$$y_t = \sum_{j=1}^{J} A_j y_{t-j} + u_t$$

- For example suppose that y_t contains
  - the interest rate set by the central bank
  - real GDP
  - residential investment
- What affects
  - the error term in the interest rate equation?
  - the error term in the output equation?
  - the error term in the housing equation?

Structural shocks

- Suppose that the economy is being hit by "structural shocks", that is shocks that are not responses to economic events
- Suppose that there are 10 structural shocks. Thus

$$u_t = Be_t$$

where B is a 3×10 matrix.

- Without loss of generality we can assume that

$$\mathrm{E}[e_t e_t'] = I$$

Structural shocks

- Can we identify B from the data?

$$\mathsf{E}[u_t u_t'] = B\mathsf{E}[e_t e_t']B' = BB'$$

- We can get an estimate for $\mathsf{E}[u_t u_t']$ using

$$\hat{\Sigma} = \sum_{t=J+1}^{T} \hat{u}_t \hat{u}_t' / (T - J)$$

- But B contains 30 unknowns and

$$\mathsf{E}\left[u_t u_t'\right] = BB'$$

has only 9 equations

Identification of B

- Can we identify B if there are only three structural shocks?
- B has 9 distinct elements
- But $\hat{\Sigma}$ is symmetric, so we only have 6 (not 9) equations

- Answer is still NO

Identification of B

- Reason for lack of identification:
  Not all equations are independent. $\Sigma_{1,2} = \Sigma_{2,1}$. For example

$$\Sigma_{1,2} = b_{11}b_{21} + b_{12}b_{22} + b_{13}b_{23}$$

  but also

$$\Sigma_{2,1} = b_{21}b_{11} + b_{22}b_{12} + b_{23}b_{13}$$

- In other words, different B matrices lead to the same Σ matrix

Identification of B

- To identify B we need additional restrictions
  - short-term restrictions: direct restrictions on B
  - long-term restrictions: restrictions on B such that long-term responses have a certain value (typically zero)
  - sign restrictions: restrictions on B such that IRFs have certain signs at certain horizons

Identification of B

$$\begin{bmatrix} u_t^i \\ u_t^y \\ u_t^r \end{bmatrix} = B \begin{bmatrix} e_t^1 \\ e_t^2 \\ e_t^{\text{mp}} \end{bmatrix}$$

- Suppose we impose

$$B = \begin{bmatrix} & 0 & 0 \\ & & 0 \\ & & \end{bmatrix}$$

- Then I can solve for the remaining elements of B from

$$\hat{B}\hat{B}' = \hat{\Sigma}$$

Matlab commands

- If

$$B = \begin{bmatrix} & 0 & 0 \\ & & 0 \\ & & \end{bmatrix}$$

  use $B =$chol(Σ)'

- If

$$B = \begin{bmatrix} & & \\ 0 & & \\ 0 & 0 & \end{bmatrix}$$

  use $B = \left[\text{chol}(\Sigma^{-1})\right]^{-1}$

Identification of B

- Suppose instead we use

$$\begin{bmatrix} u_t^y \\ u_t^i \\ u_t^r \end{bmatrix} = D \begin{bmatrix} e_t^1 \\ e_t^2 \\ e_t^{\text{mp}} \end{bmatrix}$$

- And that we impose

$$D = \begin{bmatrix} & 0 & 0 \\ & & 0 \\ & & \end{bmatrix}$$

- This corresponds with imposing

$$B = \begin{bmatrix} & & 0 \\ & 0 & 0 \\ & & \end{bmatrix}$$

- This does not affect the IRF of e_t^{mp}. All that matters for the IRF is whether a variable is ordered before or after r_t

Calculating IRFs from (structural) VAR

1. Calculation IRFs from first-order VAR is trivial
2. Calculation IRFs from higher-order VAR is also trivial, since higher-order VARs can be written as first-order system (or you simply iterate on the system)

First-order VAR

$$y_t = A_1 y_{t-1} + B e_t$$

- IRFs, variances, etc. can be calculated analytically, because you can easily calculate the MA representation:

$$y_t = B e_t + A_1 B e_{t-1} + A_1^2 B e_{t-2} + \cdots$$

State-space notation

Every VAR can be presented as a first-order VAR. For example let

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = A_1 \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + A_2 \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + B \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ I_{2\times 2} & 0_{2\times 2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} B & 0_{2\times 2} \\ 0_{2\times 2} & 0_{2\times 2} \end{bmatrix} \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ 0 \\ 0 \end{bmatrix}$$

State-space notation

- $$Y_t = AY_{t-1} + E_t,$$

  where Y_t is an $n \times 1$ vector and E_t is serially uncorrelated. This AR(1) structure allows for analytical results. For example, let

  $$\mathrm{E}\left[Y_t Y_t'\right] = \Sigma_Y \text{ and } \mathrm{E}\left[E_t E_t'\right] = \Sigma_Y.$$

- Then

  $$vec\left(\Sigma_Y\right) = \left(I - A \otimes A\right)^{-1} vec\left(\Sigma_E\right),$$

  which uses that

  $$vec\left(TVR\right) = R' \otimes Tvec\left(V\right)$$

  for conformable matrices T, V, R

Alternative identification assumptions

- restrictions do not have to be zero restrictions
- you can impose restrictions on B such that IRFs have certain properties
  then restrictions imposed depend on rest of the VAR

Identifying assumption (Blanchard-Quah)

VAR used by Gali (1999)

$$
\begin{aligned}
z_t &= \sum_{j=1}^{J} A_j z_{t-j} + B\varepsilon_t \\
&\quad \text{with} \\
z_t &= \begin{bmatrix} \Delta \ln(y_t/h_t) \\ \Delta \ln(h_t) \end{bmatrix} \\
\varepsilon_t &= \begin{bmatrix} \varepsilon_{t,\text{technology}} \\ \varepsilon_{t,\text{non-technology}} \end{bmatrix}
\end{aligned}
$$

Identifying assumption (Blanchard-Quah)

- Non-technology shock does not have a long-run impact on productivity
- Long-run impact is zero if
  - Response of the *level* goes to zero
  - Responses of the *differences* sum to zero

Get MA representation

$$
\begin{aligned}
z_t &= A(L)z_t + B\varepsilon_t \\
&= (I - A(L))^{-1}B\varepsilon_t \\
&= D(L)\varepsilon_t \\
&= D_0\varepsilon_t + D_1\varepsilon_{t-1} + \cdots
\end{aligned}
$$

Note that $D_0 = B$

Sum of responses

$$\sum_{j=0}^{\infty} D_j = D(1) = (I - A(1))^{-1} B$$

Blanchard-Quah assumption:

$$\sum_{j=0}^{\infty} D_j = \begin{bmatrix} & 0 \\ & \end{bmatrix}$$

Sign restrictions

$$BB' = \Sigma$$

General idea of sign restrictions:

- Try **"all"** matrices B such that the IRFs satisfy certain properties

Sign restrictions - example

- Try **"all"** matrices B such that the IRFs satisfy certain properties such as
    - In response to an expansionary monetary policy shock, the interest rate falls while money and prices rise.
    - In response to a positive shock to money demand, both the interest rate and money increase.
    - In response to a positive demand shock, both output and prices rise.
    - In response to a positive supply shock, output rises but prices fall.
    - In response to a positive external shock, the exchange rate devaluates and output increases.
- You would have to specify the horizon for which this should hold

These examples are from Rubio-Ramirez, Waggoner, Zha (2005).

Sign restrictions - General Idea

How to search for "all" B that satisfy $BB' = \Sigma$ and the sign restrictions?

- Let $\overline{B}$ be the Cholesky decomposition of Σ
- Bs satisfying $BB' = \Sigma$ can be expressed as

$$B = \overline{B}Q$$

with Q being an orthogonal matrix, that is

$$QQ' = I.$$

Sign restrictions - In practice

"Systematically" look for Q such that

1.

$$QQ' = I.$$

2.

$$B = Q\overline{B} \text{ satisfies the sign restricions}$$

Givens matrices - Example

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

- Note that

$$\begin{aligned} \sum_{j=1}^{n} Q_{ij}^2 &= 1 \; \forall i \\ &\implies \\ |Q_{ij}| &\leq 1 \end{aligned}$$

Sign restrictions - Givens matrices

- Suppose that B is a 2×2 Matrix
- Then **all** Qs satisfying $QQ' = I$ can be represented with the following *Givens* matrices

$$\text{rotation} \quad : \quad Q^{\text{rot}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, -\pi \le \theta \le \pi$$

$$\text{reflection} \quad : \quad Q^{\text{ref}} = \begin{bmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, -\pi \le \theta \le \pi$$

- In practice you can use a grid for θ or draw θ from a uniform distribution

Number of Givens matrices

- Let's index Q by the Q_{21} element, that is,

$$Q_{21} = \omega \text{ with } -1 \leq \omega \leq 1$$

- For each ω there are (at most) four different solutions for Q_{11}, Q_{12}, and Q_{22}

$$\begin{aligned} Q_{11}^2 + Q_{12}^2 &= 1 \\ Q_{11}\omega + Q_{12}Q_{22} &= 0 \\ \omega + Q_{22}^2 &= 1 \end{aligned}$$

- Thus, focusing on $QQ' = I$ equation indicates there are 4 Qs for every ω.
- $\omega = \sin\theta$ has two solutions for $\theta \Longrightarrow$ again 4 Qs (two Q^{rot}s and two Q^{ref}s).

Givens matrices - Third Order

$$Q_1^{\text{rot}} = \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 & 0 \\ \sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad Q_2^{\text{rot}} = \begin{bmatrix} \cos\theta_2 & 0 & -\sin\theta_2 \\ 0 & 1 & 0 \\ \sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix}$$

$$Q_3^{\text{rot}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_3 & -\sin\theta_3 \\ 0 & \sin\theta_3 & \cos\theta_3 \end{bmatrix}$$

Givens matrices - Third Order

$$Q_1^{\text{ref}} = \begin{bmatrix} -\cos\theta_1 & \sin\theta_1 & 0 \\ \sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$Q_2^{\text{ref}} = \begin{bmatrix} -\cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ \sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix} \quad Q_3^{\text{ref}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\cos\theta_3 & \sin\theta_3 \\ 0 & \sin\theta_3 & \cos\theta_3 \end{bmatrix}$$

Givens matrices - Third Order

For each combination of θ_1, θ_2, and θ_3 consider

$$Q = \prod_{i=1}^{3} Q_i^r(\theta_i) \;\text{ for } r \in \{\text{rot,ref}\}$$

QR Decomposition

Rubio-Ramirez, Waggoner, and Zha (2005) propose the following alternative to find orthogonal $n \times n$ matrices, which is computationally more efficient for large VARs:

1. Let W be an $n \times n$ matrix, each element is an i.i.d. draw from a $N(0,1)$
2. Decompose W using the QR decomposition (Householder transformation)
$$W = QR,$$

where Q is the orthogonal matrix we are looking for

QR Decomposition - Matlab

1. `W = randn(3,3);`
2. [Q,R]=qr(W);

QR Decomposition - example

1.
$$\mathtt{W} = \begin{bmatrix} -0.0551 & 0.1992 & 0.8829 \\ -1.0717 & -0.4964 & 0.7643 \\ -0.3729 & -1.6501 & 0.2373 \end{bmatrix}$$

2.
$$\mathtt{Q} = \begin{bmatrix} -0.0485 & 0.174 & 0.174 \\ -0.9433 & 0.3156 & -0.1027 \\ -0.3283 & -0.9327 & 0.1496 \end{bmatrix}$$

Sign restrictions - comments

- Sign restrictions give you a set of IRFs.
  If you would plot the median at each horizon then this typically would be a *combination* of different IRFs, that is, there may not be one IRF that is close to what you are plotting

- When using sign restrictions in a Bayesian framework, then you should be careful that drawing from the posterior does not impose additional restrictions (See Arias, Rubio-Ramirez and Waggoner 2014 discuss this and provide a mechanism to do this right)

If you ever feel bad about getting too much criticism ....

-

If you ever feel bad about getting too much criticism ....

-
- be glad you are not a structural VAR

Structural VARs & critiques

- From MA to AR
  - Lippi & Reichlin (1994)
- From prediction errors to structural shocks
  - Fernández-Villaverde, Rubio-Ramirez, Sargent, Watson (2007)
- Problems in finite samples
  - Chari, Kehoe, McGratten (2008)

From MA to AR

Consider the two following *different* MA(1) processes

$$
\begin{aligned}
y_t &= \varepsilon_t + \frac{1}{2}\varepsilon_{t-1}, \;\; \mathsf{E}_t\left[\varepsilon_t\right] = 0,\ \mathsf{E}_t\left[\varepsilon_t^2\right] = \sigma^2 \\
x_t &= e_t + 2e_{t-1}, \;\; \mathsf{E}_t\left[e_t\right] = 0,\ \mathsf{E}_t\left[e_t^2\right] = \sigma^2/4
\end{aligned}
$$

- Different IRFs
- Same variance and covariance

$$
\mathsf{E}\left[y_t y_{t-j}\right] = \mathsf{E}\left[x_t x_{t-j}\right]
$$

From MA to AR

- AR representation:

$$\begin{aligned} y_t &= (1+\theta L)\,\varepsilon_t \\ \frac{1}{(1+\theta L)} y_t &= \varepsilon_t \\ \frac{1}{(1+\theta L)} &= \sum_{j=0}^{\infty} a_j L^j \end{aligned}$$

- Solve for a_js from

$$1 = a_0 + (a_1 + a_0\theta)\,L + (a_2 + a_1\theta)\,L^2 + \cdots$$

From MA to AR

Solution:

$$\begin{aligned} a_0 &= 1 \\ a_1 &= -a_0\theta \\ a_2 &= -a_1\theta = a_0\theta^2 \\ &\dots \end{aligned}$$

You need

$$|\theta| < 1$$

Prediction errors and structural shocks

Solution to economic model

$$
\begin{aligned}
x_{t+1} &= Ax_t + B\varepsilon_{t+1} \\
y_{t+1} &= Cx_t + D\varepsilon_{t+1}
\end{aligned}
$$

- x_t: state variables
- y_t: observables (used in VAR)
- ε_t: structural shocks

Prediction errors and structural shocks

- From the VAR you get prediction error e_{t+1}

$$\begin{aligned} e_{t+1} &= y_{t+1} - \mathsf{E}_t\left[y_{t+1}\right] \\ &= Cx_t + D\varepsilon_{t+1} - \mathsf{E}_t\left[Cx_t\right] \\ &= C\left(x_t - \mathsf{E}_t\left[x_t\right]\right) + D\varepsilon_{t+1} \end{aligned}$$

- Problem: Not guaranteed that

$$x_t = \mathsf{E}_t\left[x_t\right]$$

Prediction errors and structural shocks

- Suppose: $y_t = x_t$
  - that is, all state variables are observed
- Then

$$x_t = \mathsf{E}_t [x_t]$$

Prediction errors and structural shocks

- Suppose: $y_t \neq x_t$
- Has y_t has enough info to uncover x_t and, thus, ε_t?

Prediction errors and structural shocks

- Suppose D is invertible

$$\varepsilon_t = D^{-1}\left(y_{t+1} - Cx_t\right)$$
$$\Longrightarrow$$
$$x_{t+1} = Ax_t + BD^{-1}\left(y_{t+1} - Cx_t\right)$$
$$\Longrightarrow$$
$$x_{t+1}\left(I - \left(A + BD^{-1}C\right)L\right) = y_{t+1}$$

- $\Longrightarrow$

$$x_t = E_t\left[x_t\right] \text{ if}$$

the eigenvalues of $A - BD^{-1}C$

must be strictly less than 1 in modulus

- See F-V,R-R,S, W (2007)

Finite sample problems

- Summary of discussion above
  - Life is excellent if you observe all state variables
  - But,
    - we don't observe capital (well)
    - even harder to observe news about future changes
  - If ABCD condition is satisfied, you are still ok *in theory*
- Problem: you may need ∞-order VAR for observables
  - recall that k_t has complex dynamics

Finite sample problems

1. Bias of estimated VAR
   - apparently bigger for VAR estimated in first differences
2. Good VAR may need many lags

Alleviating finite sample problems

Do with model exactly what you do with data:

- NOT: compare data results with model IRF
- YES:
  - generate N samples of length T
  - calculate IRFs as in data
  - compare average across N samples with data analogue

This is how Kydland & Prescott calculated business cycle stats

References

- Arias, J.E., J.F. Rubio-Ramirez, D.F. Waggoner, Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications, Federal Reserve Board International Finance Discussion Paper 2014-1100. Available at
  - http://www.federalreserve.gov/pubs/ifdp/2014/1100/default.htm.
- Chari, V.V., P.J. Kehoe, E.R. McGrattan, 2008, Are structural VARs with long-run restrictions useful in developing business cycle theory?, Journal of Monetary Economics, 55, 1337-52.
- Fernandez-Villaverde, J., J.F. Rubio-Ramirez, T.J. Sargent, and M.Watson, 2007, ABCs (and Ds) of Understanding VARs, Econometrica, 97, 1021-26.
  Gives conditions whether a particular VAR can infer structural shocks.
- Fry, R. and A. Pagan, 2011, Sign Restrictions in Structural Vector Autoregressions: A Critical Review, Journal of Economic Literature, 49, 938-960.
  - overview of sign restrictions in VARs and detailed discussion of its weaknesses

- Kilian, Lutz, 2011, Structural Vector Autogressions.
    - Overview paper that gives several examples of identification choices for different theoretical models. Available at
    - http://www-personal.umich.edu/~lkilian/elgarhdbk_kilian.pdf.
- Lippi, M., and L. Reichlin, 1994, VAR analysis, nonfundamental representations, Blaschke matrices, Journal of Econometrics, 63, 307-325.
- Luetkepohl, H., 2011, Vector Autoregressive Models, EUI Working Papers ECO2011/30.
    - detailed paper on estimating and working with VARs. Available at
    - cadmus.eui.eu/bitstream/handle/1814/19354/ECO_2011_30.pdf

- Rubio-Ramirez, Juan F., D.F. Waggoner, and T. Zha, 2005, Markov-Switching Structural Vector Autogressions: Theory and Applications, Federal Reserve Bank of Atlanta Working Paper 2005-27.
  - contains a detailed discussion of different identification schemes and sign restrictions in particular. Available at
  - http://www.frbatlanta.org/filelegacydocs/wp0527.pdf.
- Whelan, K.,2014 MA Advanced Macroeconomics.
  - Set of slides with more detailed info and a discussion of several empirical examples. Available at
  - http://www.karlwhelan.com/MAMacro/part2.pdf
  - http://www.karlwhelan.com/MAMacro/part3.pdf
  - http://www.karlwhelan.com/MAMacro/part4.pdf