Time Series Analysis - Estimation of ARMA models

ECON 722

Elena Pesavento — Emory University

Fall 2024

The identification Problem

Recall the Strictly stationary VAR (ignore the constant for now). This is the *Reduced Form (RF)*

$$\boldsymbol{y}_t = \sum_{\ell=1}^{\infty} \boldsymbol{A}_\ell \boldsymbol{y}_{t-\ell} + \boldsymbol{e}_t.$$

With

$$\boldsymbol{\Sigma} = \mathbb{E}\left(\boldsymbol{e}_t \boldsymbol{e}_t'\right) < \infty.$$

This has a MA representation, which we know this always exist because of the Wold Decomposition that is given by

$$\boldsymbol{y_t} = \boldsymbol{\Theta}(\mathrm{L})\boldsymbol{e_t}$$

The identification Problem

- The RF can estimated and we know how to do inference etc.
- The RF is not very interesting except for forecasting
- As we have seen from the lecture on Dynamic Causal models, what we are really interested is a structural interpretation.
- The RF tells us very little about the structural parameters

Intuition of supply and demand example...

The SVAR

When we think about a structural form, we think about contemporaneous linkages between the variable and structural shocks that are uncorrelated since they are *structural*. So we can write the structural form (SF) VAR as

$$\boldsymbol{B}_0 \boldsymbol{y}_t = \boldsymbol{B}_1 y_{t-1} + \ldots + \boldsymbol{B}_p y_{t-p} + \varepsilon_t$$

$$E\left(\varepsilon_t \varepsilon_s'\right) = \begin{cases} D \text{ or } I & \text{if } t = s \\ 0 & \text{otherwise} \end{cases}$$

or in MA representation

$$\boldsymbol{y_t} = \boldsymbol{\Theta}(\mathrm{L}) \varepsilon_{\boldsymbol{t}}$$

Recall the option for the normalization

The identification Problem

How do we go from the SF to the RF and viceversa?

$$\varepsilon_t = B_0 e_t$$

$$E\left(e_t e_t'\right) = \Sigma = E\left(B_0^{-1}\varepsilon_t\varepsilon_t' B_0^{-1'}\right) = B_0^{-1}E\left(\varepsilon_t\varepsilon_t'\right)B_0^{-1'} = B_0^{-1}B_0^{-'}$$

$$\Sigma = B_0^{-1}B_0^{-1'}$$

- Identification boils down to having n^2 unknown parameters but only $\frac{n(n+1}{2}$ equations. So we need $\frac{n(n-1}{2}$ restrictions for exact identification.

Options for Identification

- Short run/ World ordering /Choleski Decomposition
- Long run
- SR +LR
- Identification via heteroskedasticity
- Alternatively, you identify the shocks externally to the VAR (e.g. High frequency)
- Or you use an IV as we saw. (Proxy VAR or External Instrument)
- Or you set identifying restrictions (signs restrictions)

We will read a few papers on these topics next

Some clarification on language

I make a distinction between

- Partial identification
- Set identification

Partial Identification

Consider $z_t' = \left[\begin{array}{ccc} x_t & y_{1t} & y_{2t} \end{array}\right]'$. We are interested in the response of $y_t' = \left[\begin{array}{cc} y_{1t} & y_{2t} \end{array}\right]'$ to an unexpected shock to x_t The SVAR can be written as

$$B_0 z_t = B(L) y_{t-1} + \varepsilon_t$$

We are going to assume that the VAR is only partially identified or block recursive so

$$B_0 = \left[\begin{array}{ccc} 1 & 0 & 0 \\ -\beta_{21} & 1 & -\beta_{23} \\ -\beta_{31} & -\beta_{32} & 1 \end{array}\right]$$

The inverse B_0^{-1} can be rewritten as follows by partitioning in blocks

$$B_0^{-1} = \left[\begin{array}{cc} B_0^{(1,1)} & 0 \\ B_0^{(2,1)} & B_0^{(2,2)} \end{array}\right].$$

Partial Identification

Notice that $B_0^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{\beta_{21}+\beta_{31}\beta_{23}}{\beta_{23}\beta_{32}-1} & -\frac{1}{\beta_{23}\beta_{32}-1} & -\frac{\beta_{23}}{\beta_{23}\beta_{32}-1} \\ -\frac{\beta_{31}+\beta_{21}\beta_{32}}{\beta_{23}\beta_{32}-1} & -\frac{\beta_{32}}{\beta_{23}\beta_{32}-1} & -\frac{1}{\beta_{23}\beta_{32}-1} \end{bmatrix}$ The SVAR can be estimated via the reduced form

$$z_t = A(L)y_{t-1} + v_t$$

where $A(L) = B_0^{-1}B(L)$ and $v_t = B_0^{-1}\varepsilon_t$. The reduced form can be consistently estimated by OLS.

Partial Identification

Can the IRF be identified from the parameters of the reduced form? MA of the RF is where $\Psi(L) = [I - A(L)L]^{-1}$

$$z_t = \Psi(L)v_t = \Psi(L)B_0^{-1}\varepsilon_t$$

$$z_t = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} B_0^{(1,1)} & 0 \\ B_0^{(2,1)} & B_0^{(2,2)} \end{bmatrix} \varepsilon_t =$$

$$= \begin{bmatrix} \Psi_{11}(L)B_0^{(1,1)} + \Psi_{12}(L)B_0^{(2,1)} & \Psi_{12}(L)B_0^{(2,2)} \\ \Psi_{21}(L)B_0^{(1,1)} + \Psi_{22}(L)B_0^{(2,1)} & \Psi_{22}(L)B_0^{(2,2)} \end{bmatrix} \varepsilon_t.$$

$$IRF_{21} = \Psi_{21}(L)B_0^{(1,1)} + \Psi_{22}(L)B_0^{(2,1)}$$

It can be seen that the block $B_0^{(2,2)}$ does not enter the IRF of interest so assuming that $B_0^{(2,2)}$ is recursive or not recursive will not matter. So for ease of estimation, we can recover B_0^{-1} from the Choleski decomposition of Σ_{vv}.

Identification Example - Blanchard and Perotti (2002)

A broader class of models captured by the equation system

$$\boldsymbol{A}\boldsymbol{e}_t = \boldsymbol{B}\varepsilon_t$$

where (in a 3×3 case)

$$\boldsymbol{A} = \begin{bmatrix} 1 & a_{12} & a_{13} \\ a_{21} & 1 & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

Blanchard and Perotti (2002)

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2.08 \\ a_{31} & a_{32} & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix}$$

IMPORTANT: This is an order condition. This is necessary but not sufficient. You also need a rank condition. check!

Identification Example- Long Run

The long-run structural impulse response is the cumulative sum of all impulse responses

$$\boldsymbol{C} = \sum_{\ell=1}^{\infty} \boldsymbol{\Theta}_\ell \boldsymbol{B} = \boldsymbol{\Theta}(1)\boldsymbol{B} = \boldsymbol{A}(1)^{-1}\boldsymbol{B}.$$

- A long-run restriction is a restriction placed on the matrix $\boldsymbol{C}$.
- Blanchard and Quah (1989) suggest a bivariate VAR for the first-differenced logarithm of real GDP and the unemployment rate. Blanchard-Quah assume that the structural shocks are aggregate supply and aggregate demand.
- They adopt the hypothesis that aggregate demand has no long-run impact on GDP. This means that the long-run impulse response matrix satisfies

$$\boldsymbol{C}\boldsymbol{C}' = \boldsymbol{A}(1)^{-1}\boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}(1)^{-1\prime} = \boldsymbol{A}(1)^{-1}\boldsymbol{\Sigma}\boldsymbol{A}(1)^{-1\prime}.$$

Identification - Long Run

$$\widehat{\boldsymbol{C}}\widehat{\boldsymbol{C}}' = \widehat{\boldsymbol{A}}(1)^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{A}}(1)^{-1\prime}$$

In either case the estimator for $\boldsymbol{B}$ is

$$\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{A}}(1)\widehat{\boldsymbol{C}}$$

and the estimator of the structural impulse response is

$$\widehat{SIRF}(h) = \widehat{\boldsymbol{\Theta}}_h\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{\Theta}}_h\widehat{\boldsymbol{A}}(1)\widehat{\boldsymbol{C}}.$$

If lower triangular this is easy as it is just the choleski decomposition

$$\boldsymbol{C} = \text{chol}\left(\boldsymbol{A}(1)^{-1}\boldsymbol{\Sigma}\boldsymbol{A}(1)^{-1}\right).$$

Foreast Error Decomposition

- An alternative tool to investigate an estimated VAR is the forecast error decomposition (also called variance decomposition).
- which decomposes multi-step forecast error variances by the component shocks.
- The forecast error decomposition indicates which shocks contribute towards the fluctuations of each variable in the system.

Take the moving average representation of the i^{th} variable $y_{i,t+h}$ written as a function of the orthogonalized shocks

$$y_{i,t+h} = \mu_i + \sum_{\ell=0}^{\infty} \boldsymbol{\theta}_i(\ell)' \boldsymbol{B} \varepsilon_{t+h-\ell}.$$

The best linear forecast of y_{t+h} at time t is

$$y_{i,t+h|t} = \mu_i + \sum_{\ell=h}^{\infty} \boldsymbol{\theta}_i(\ell)' \boldsymbol{B} \boldsymbol{\varepsilon}_{t+h-\ell}.$$

Variance Decomposition

Thus the h-step forecast error is the difference

$$y_{i,t+h} - y_{i,t+h|t} = \sum_{\ell=0}^{h-1} \boldsymbol{\theta}_i(\ell)' \boldsymbol{B} \varepsilon_{t+h-\ell}.$$

The variance of this forecast error is

$$\begin{aligned} \text{var}\left(y_{i,t+h} - y_{i,t+h|t}\right) &= \sum_{\ell=0}^{h-1} \text{var}\left(\boldsymbol{\theta}_i(\ell)' \boldsymbol{B} \varepsilon_{t+h-\ell}\right) \\ &= \sum_{\ell=0}^{h-1} \boldsymbol{\theta}_i(\ell)' \boldsymbol{B} \boldsymbol{B}' \boldsymbol{\theta}_i(\ell). \end{aligned}$$

Variance Decomposition

To isolate the contribution of the j^{th} shock, notice that

$$e_t = \boldsymbol{B}\varepsilon_t = b_1\varepsilon_{1t} + \cdots + b_m\varepsilon_{mt}.$$

Thus the contribution of the j^{th} shock is $\boldsymbol{b}_j\varepsilon_{jt}$. Now imagine replacing $\boldsymbol{B}\varepsilon_t$ in the variance calculation by the j^{th} contribution $\boldsymbol{b}_j\varepsilon_{jt}$. This is

$$\operatorname{var}\left(y_{it+h} - y_{i,t+h|t}\right) = \sum_{\ell=0}^{h-1} \operatorname{var}\left(\boldsymbol{\theta}_i(\ell)'\boldsymbol{b}_j\varepsilon_{jt+h-\ell}\right) = \sum_{\ell=0}^{h-1} \left(\boldsymbol{\theta}_i(\ell)'\boldsymbol{b}_j\right)^2.$$

$$\operatorname{var}\left(y_{i,t+h} - y_{i,t+h|t}\right) = \sum_{j=1}^{m}\sum_{\ell=0}^{h-1} \left(\boldsymbol{\theta}_i(\ell)'\boldsymbol{b}_j\right)^2.$$

Variance Decomposition

The forecast error decomposition is defined as the ratio of the j^{th} contribution to the total which is the ratio :

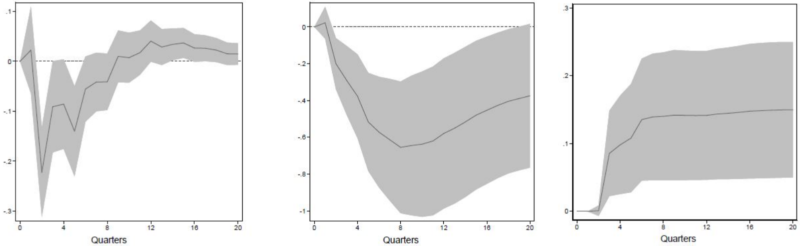
$$FE_{ij}(h) = \frac{\sum_{\ell=0}^{h-1} \left(\boldsymbol{\theta}_i(\ell)' \boldsymbol{b}_j\right)^2}{\sum_{j=1}^{m} \sum_{\ell=0}^{h-1} \left(\boldsymbol{\theta}_i(\ell)' \boldsymbol{b}_j\right)^2}.$$

The $FE_{ij}(h)$ lies in $[0, 1]$ and varies across h. Small values indicate that ε_{jt} contributes only a small amount to the variance of y_{it}. Large values indicate that ε_{jt} contributes a major amount of the variance of ε_{it}.

A forecast error decomposition requires that orthogonalized innovations. There is no nonorthogonalized version.

Example - Hansen 15.21

To illustrate we use the three-variable system. We use the ordering (1) real GDP growth rate, (2) inflation rate, (3) Federal funds interest rate. Graph is the IRF for GDP growth to one SD increase in ffr.



(a) Impulse Response Function

(b) Cumulative IRF

(c) Error Decomposition

Figure 15.1: Response of GDP Growth to Orthogonalized Fed Funds Shock

Historical Decomposition

- How much a given structural shock explains of the historically observed fluctuations in the VAR variables.
- what is the cumulative effect of a given structural shock on each variable at every given point in time?
- Example: Did monetary policy cause the 1982 recession?

$$y_t = \sum_{s=0}^{t-1} \Theta_s \varepsilon_{t-s} + \sum_{s=t}^{\infty} \Theta_s \varepsilon_{t-s}$$

Second term approximately zero.

$$\widehat{y}_t = \sum_{s=0}^{t-1} \Theta_s \varepsilon_{t-s}$$

We start by plotting $\widehat{y}_t$ and the (suitably demeaned and detrended) actual data y_t in the same plot. We discard the initial observations (also known as transients), for which the two series do not effectively coincide. How many

Historical Decomposition

For the remaining sample period, we can decompose the sum in to isolate the cumulative contribution of each shock to each element of $\widehat{y}_t$, as discussed next. For example, let $n = 5$ and suppose that we are interested in the cumulative effect of each of the five structural shocks on the 4th variable of the VAR system. In that case, we compute the following weighted sums for $t = 1, \ldots, T$:

$$\widehat{y}_{4t}^{(1)} = \sum_{i=0}^{t-1} \theta_{41,i} \varepsilon_{1,t-i}, \quad \widehat{y}_{4t}^{(2)} = \sum_{i=0}^{t-1} \theta_{42,i} \varepsilon_{2,t-i}, \widehat{y}_{4t}^{(3)} = \sum_{i=0}^{t-1} \theta_{43,i} \varepsilon_{3,t-i}$$

$$\widehat{y}_{4t}^{(4)} = \sum_{i=0}^{t-1} \theta_{44,i} \varepsilon_{4,t-i} \qquad \widehat{y}_{4t}^{(5)} = \sum_{i=0}^{t-1} \theta_{45,i} \varepsilon_{5,t-i}$$

$\widehat{y}_4^{(j)} = \left(\widehat{y}_{41}^{(j)}, \ldots, \widehat{y}_{4T}^{(j)} \right)'$ shows the cumulative contribution of shock j on the 4th variable in the VAR model over time. By construction,

$$\widehat{y}_{4t} = \sum_{j=1}^{K} \widehat{y}_{4t}^{(j)}.$$

Historical Decomposition

- In practice, replace the unknown quantities $\theta_{jk,i}$ and ε_t by the usual estimates.
- If we plot the historical decomposition along with actual data, we must first remove all deterministic components in the actual data, because the data generated from the structural MA representation are zero mean by construction
- Each time series $\widehat{y}_{kT^*}^{(j)}, \ldots, \widehat{y}_{kT}^{(j)}$, $j = 1, \ldots, K$, shows how our approximate variable $\widehat{y}_{kt}$ would have evolved if all other structural shocks had been turned off.
- Plotting each time series $\widehat{y}_{kT^*}^{(j)}, \ldots, \widehat{y}_{kT}^{(j)}, j = 1, \ldots, K$, against the actual data helps assess which structural shock(s) alone or in combination account for the fluctuations in $\widehat{y}_{kT^*}, \ldots, \widehat{y}_{kT}$ during specific historical episodes of interest.

Historical Decomposition Example

- Figure illustrates the use of historical decompositions in understanding the evolution of the real price of oil from the late 1970s to early 2012.
- The example is based on a global oil market model in Kilian and Lee (2014): attributes variation in the real price of oil to shocks to the flow supply of oil, shocks to the flow demand for oil, a speculative oil demand shock, and a residual shock designed to capture various idiosyncratic shocks.
- Much of this surge (as well as the collapse of the real price of oil in late 2008 and its recovery since then) must be attributed to the effects of flow demand shocks.
- Neither flow supply shocks nor speculative demand shocks are able to explain the surge in the real price of oil during this period.
- This result could not have been inferred from the structural impulse

Historical Decomposition Example

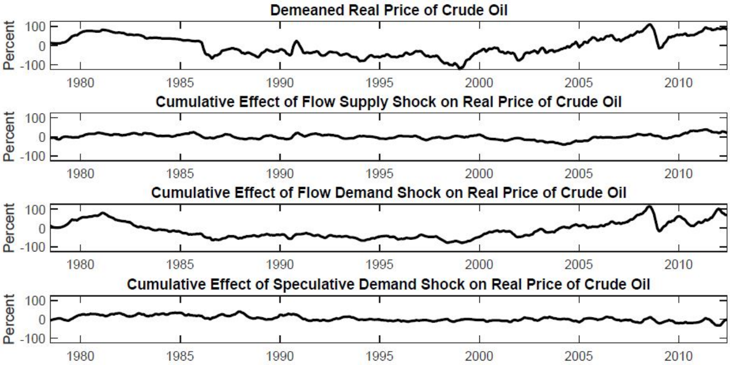


Figure 4.2: Historical decomposition of the real price of crude oil in percent deviations from the mean.
Source: Kilian and Lee (2014).

Historical Decomposition Example

- The importance of conducting historical decompositions is not always appreciated in empirical macroeconomics.
- An occasional mistake in the business cycle literature is to view evidence of large and persistent impulse responses of real output to a structural shock as evidence of this shock's ability to explain the business cycle.
- This conclusion is unwarranted because impulse responses trace out the response to a one-time positive shock only, whereas business cycle variation in real output is driven by a sequence of shocks of different magnitude and signs.
- It is not uncommon for the effects of a positive shock in one period to be eclipsed by negative shocks in subsequent periods.
- Only the historical decomposition allows us to assess the cumulative effect of these shocks on the business cycle and the relative importance of different shocks in explaining particular recessions or expansions.

Let's read some papers

In Canvas you will find a list of papers that we have and we will read on these topics.