VARs, LPs and Identification in Macro

Giovanni Ricco

13th October 2023

Structural and Reduced Form Representations

What are macro shocks?

Macroeconomic shocks are best understood as structural disturbances in a simultaneous equation system

Structural shocks (Frisch-Slutsky paradigm)

Shocks are primitive exogenous disturbances to multivariate stochastic dynamic equations that are uncorrelated with each other and through time and economically meaningful (e.g. Ramey 2016)

What are macro shocks?

Macroeconomic shocks are best understood as structural disturbances in a simultaneous equation system

Structural shocks (Frisch-Slutsky paradigm)

Shocks are primitive exogenous disturbances to multivariate stochastic dynamic equations that are uncorrelated with each other and through time and economically meaningful (e.g. Ramey 2016)

instruments $\neq$ VAR innovations $\neq$ structural shocks

The Economy as a Stochastic Process

Frisch-Slutsky Paradigm

- Data Generating Process (DGP)

$$f(Y_t, u_t) = 0$$

- For example, a Dynamic Stochastic General Equilibrium model (DSGE), that is usually written as a VARMA model

$$\Psi(L)Y_t = \kappa + \Theta(L)u_t \qquad u_t \sim \mathcal{WN}(0, I_q)$$

VARs and SVARs

- A DSGE usually admits a Structural Vector Moving Average (SVMA) representation

$$Y_t = \mu + F_0 u_t + F_1 u_{t-1} + F_2 u_{t-2} + \dots$$

- and if invertible a Structural Vector Autoregression (SVAR) form

$$B_0 Y_t = K + B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + u_t$$

VARs and SVARs

- The (approximated) structural model is a SVAR(p)

$$B_0 Y_t = K + B_1 Y_{t-1} + B_2 Y_{t-2} + \ldots + B_p Y_{t-p} + u_t \qquad u_t \sim \mathcal{WN}(0, I)$$

- The econometrician can estimate a pth-order Vector Autoregression VAR(p)

$$Y_t = C + A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + e_t \qquad e_t \sim \mathcal{WN}(0, \Sigma)$$

hence

$$C = {B_0}^{-1} K \qquad A_i = {B_0}^{-1} B_i \qquad \Sigma = {B_0}^{-1} {B_0}^{-1\prime}$$

- To find B_0 is the problem of 'structural identification'

The Identification Problem

Remark: Given B_0 one orthogonal decomposition of the variance covariance matrix of the residuals

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

and H any orthogonal matrix (rotations)

$$HH' = H'H = I$$

we obtain another orthogonal decomposition of the residuals as $\widetilde{B}_0^{-1} = B_0{}^{-1} H$ since

$$\widetilde{B}_0^{-1} \widetilde{B}_0^{-1\prime} = B_0{}^{-1} HH' B_0{}^{-1\prime} = B_0{}^{-1} B_0{}^{-1\prime} = \Sigma$$

The orthogonal matrices span the class of the orthonormal representation

Structural Models and the Wold Decomposition

The Rational Expectations Interpretation of VARs

Structural Vector Moving Average SVMA(∞)

$$Y_t = \mu + F_0 u_t + F_1 u_{t-1} + F_2 u_{t-2} + \dots \quad u_t \sim \mathcal{WN}(0, I_q)$$

where

$$F_0 u_t = Y_t - E(Y_t | \mathcal{I}_{t-1})$$

and

$$\mathcal{I}_t = span\{u_t, u_{t-1}, \dots\}$$

is (usually assumed to be) the Economic Agents' Information set

Wold Decomposition

Wold representation: Any zero-mean covariance stationary vector process Y_t admits is unique Vector Moving Average (VMA) representation

$$Y_t = C(L)e_t + \mu_t \tag{1}$$

where $C(L)e_t$ is the stochastic component with $C(L) = \sum_{i=0}^{\infty} C_i L^i$ and μ_t a purely deterministic component and the following properties

(a) e_t is the innovation for Y_t, i.e., $e_t = Y_t - \text{Proj}(Y_t | Y_{t-1}, Y_{t-2}, ...)$, i.e., the innovation is Y_t-fundamental.

(b) e_t is white noise, $Ee_t = 0$, $Ee_t e'_\tau = 0$, for $t \neq \tau$, $Ee_t e'_t = \Sigma$

(c) The coefficients are square summable $\sum_{j=0}^{\infty} \|C_j\|^2 < \infty$.

(d) $C_0 = I$

If $\mu_t = 0$ the process is said regular.

Wold Decomposition

- Very powerful result: it holds for any covariance stationary process
- It provides a justification to the use of VARs
- However, the theorem does not implies that the **Wold representation** is the true representation of the process
- For instance the process could be stationary but non-linear or non-invertible

Approximating the Wold Decomposition

The Rational Expectations Interpretation of VARs

Approximate the Wold Decomposition by inverting a finite order VAR(p)

$$Y_t = C + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + \tilde{e}_t \qquad \tilde{e}_t \sim \mathcal{WN}(0, \Sigma)$$

hence

$$\tilde{e}_t = Y_t - \left(C + \sum_{s=1}^{p} A_s Y_{t-s} \right)$$

and $\tilde{e}_t \perp Y_{t-s}$ for $s = 1, 2, \ldots, p$

Structural Models and the Wold Decomposition

The Rational Expectations Interpretation of VARs

If the information set of the economic agents (in the model) $\mathcal{I}_{t-1}$ is aligned to the information set of the econometrician $\mathcal{M}_{t-1}$, i.e. $\mathcal{M}_{t-1} \equiv \mathcal{I}_{t-1}$

$$Y_t - \widehat{P}(Y_t|\mathcal{M}_{t-1}) \approx Y_t - E(Y_t|\mathcal{I}_{t-1})$$

$$\Downarrow$$

$$\boxed{e_t \approx {B_0}^{-1} u_t}$$

$$\Downarrow$$

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

SVAR: Estimation and Identification

SVAR: Estimation and Identification

1. Estimate VAR(p): $\widehat{C}, \widehat{A}_1, \ldots, \widehat{A}_p, \widehat{\Sigma}$
2. Invert the VAR(p) to derive the coefficients of the Wold representation
   $\widehat{C}_1, \widehat{C}_2, \ldots$ (MA representation)
3. Use Economic arguments to identify $\widehat{B}_0$. Derive structural shocks $\widehat{u}_t = B_0 \widehat{e}_t$
4. Derive Impulse Response Functions $\widehat{F}_j = \widehat{C}_j {B_0}^{-1} \cancel{\widehat{u}_t}$
   (and Variance Decomposition and Historical Decomposition)
5. Inference on the distribution of the IRFs can be derived using:
   - Delta method (see Hamilton, 1995)
   - Montecarlo simulations
   - Bootstrap

   The IRFs are complicated functions of the estimated parameters

VAR(p): Estimation

In the Seemingly Unrelated Regressions (SUR) representation, the VAR(p) can be stacked as

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

where:

$$\mathbf{X} = [X_1, ..., X_T]'$$

$$X_t = [1, Y'_{t-1}, Y'_{t-2}..., Y'_{t-p}]'$$

$$\mathbf{Y} = [Y_1, ..., Y_T]'$$

$$\epsilon = [e_1, ..., e_T]'$$

and

$$\beta = [C, A_1, ..., A_p]'$$

VAR(p): Estimation

Remark: In general SUR models should be estimated by (F)GLS to take into account the correlation of the errors across equations.

There are two exceptions:

- Σ is diagonal
- The regressors X are the same in all equations (see previous slides)

In these cases Maximum Likelihood $\Longrightarrow$ OLS equation by equation:

- $\widehat{\beta} = (X'X)^{-1}X'Y$
- $E(vec(\widehat{\beta} - \beta)vec(\widehat{\beta} - \beta)') = \Sigma \otimes (X'X)^{-1}$
- $\widehat{e}_t = Y_t - x_t'\widehat{\beta} = Y_t - \left(\widehat{c} + \widehat{A}_1 Y_{t-1} + \cdots + \widehat{A}_p Y_{t-p}\right)$
- $\widehat{\Sigma} = \frac{1}{T}\sum_{t=p+1}^{T} \widehat{e}_t \widehat{e}_t'$

From the VAR(p) to the Wold Decomposition

Let us first rewrite the VAR(p) into a VAR(1) companion form (skipping $\widehat{\ }$'s):

$$Y_t = c + \sum_{s=1}^{p} A_s Y_{t-s} + e_t$$

Define

$$\underbrace{\begin{pmatrix} Y_t \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p+1} \end{pmatrix}}_{\mathcal{Y}_t} = \underbrace{\begin{pmatrix} C \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}}_{\mathcal{C}} + \underbrace{\begin{pmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_n & 0_{n\times n} & \dots & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & I_n & \dots & 0_{n\times n} & 0_{n\times n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{n\times n} & 0_{n\times n} & \dots & I_n & 0_{n\times n} \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ Y_{t-3} \\ \vdots \\ Y_{t-p} \end{pmatrix}}_{\mathcal{Y}_{t-1}} + \underbrace{\begin{pmatrix} e_t \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}}_{E_t}$$

$$\mathcal{Y}_t = \mathcal{C} + A\mathcal{Y}_{t-1} + E_t$$

From the VAR(p) to the Wold Decomposition

- Define selector matrix

$$J = (I_n \; \underbrace{0_{n\times n} \; \dots 0_{n\times n}}_{p-1 \text{ times}}) \Longrightarrow Y_t = J\mathcal{Y}_t \text{ and } e_t = JE_t$$

- Solve $\mathcal{Y}_t = \mathcal{C} + A\mathcal{Y}_{t-1} + E_t$ backward:

$$\mathcal{Y}_t = (\mathcal{C} + A\mathcal{C} + A^2\mathcal{C} \cdots + A^k\mathcal{C}) + \\ (E_t + AE_{t-1} + A^2E_{t-2} + \cdots + A^kE_{t-k}) + A^{k+1}\mathcal{Y}_{t-k-1}$$

Stability $\Longrightarrow \max|\lambda(A)| < 1$ then $A^{k+1}\mathcal{Y}_{t-k-1} \longrightarrow 0$

From the VAR(p) to the Wold decomposition

$$\mathcal{Y}_t = (I_{np} - A)^{-1}\mathcal{C} + \sum_{j=1}^{\infty} A^j E_{t-j} \Longrightarrow Y_t = J\left(\sum_{j=0}^{\infty} A^j\right)\mathcal{C} + \sum_{j=1}^{\infty} JA^j J' e_{t-j}$$

Equating term by term with the Wold decomposition

$$C_j = JA^j J'$$

$$\mu = J\left(\sum_{j=0}^{\infty} A^j\right)\mathcal{C} = J(I_{np} - A)^{-1}\mathcal{C}$$

The Identification Problem

Remark:

- ${B_0}^{-1}$ has n^2 elements
- while Σ has $\frac{n(n+1)}{2}$ elements.
- Therefore

$$n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}$$

additional restrictions are needed to map e_t into u_t.

Structural Vector Autoregressions

We would like to have:

1. Orthogonal shock
2. Shocks with economic meaning

Let's start from the orthogonalisation. Two 'standard' statistical ways:

1. Cholesky decomposition
2. Spectral Decomposition

Cholesky Decomposition

The Cholesky factor, S, of Σ is defined as the unique lower triangular matrix such that

$$SS' = \Sigma$$

This implies that we can rewrite the VAR in terms of orthogonal shocks

$$\eta_t = S^{-1} e_t$$

with identity covariance matrix

$$A(L)Y_t = S\eta_t$$

Spectral Decomposition

Let V and be a matrix containing the eigenvectors of Σ and Λ a diagonal matrix with the eigenvalues of Σ on the main diagonal. Then we have that

$$V \Lambda V' = \Sigma$$

This implies that we can rewrite the VAR in terms of orthogonal shocks

$$\xi_t = (V \Lambda^{1/2})^{-1} e_t$$

with identity covariance matrix

$$A(L) Y_t = V \Lambda^{1/2} \xi$$

The Class of Orthonormal Representations

From the class of invertible MA representation of Y_t we can derive the class of orthonormal representation, i.e. the class of representations of Y_t in term of orthonormal shocks.

Let H any orthogonal matrix (rotations)

$$HH' = H'H = I$$

Defining

$$u_t = (SH)^{-1}e_t$$

we recover the general class of the orthonormal representation of Y_t

$$Y_t = C(L)SHu_t = F(L)u_t$$

The Class of Orthonormal Representations

where

$$F(L) = C(L)SH$$

and

$$u_t \sim \mathcal{WN}(0, I_n)$$

we get

$$\begin{aligned} E(u_t u_t') &= E\left[(SH)^{-1} e_t e_t' \left((SH)^{-1}\right)'\right] \\ &= H'S^{-1}E(e_t e_t')(H'S^{-1})' = H'S^{-1}\Sigma(S')^{-1}H \\ &= H'S^{-1}SS'(S')^{-1}H = I \end{aligned}$$

Problem: H can be any, so how should we choose one?

The Identification Problem

- Identifying the VAR means fixing a particular matrix H
- Structural shocks are obtained as linear combinations of the VAR innovations
- In order to choose a matrix H we have to fix $n(n-1)/2$ parameters since there is a total of n^2 parameters and a total of $n(n+1)/2$ restrictions implied by orthonormality

Identification

Find a B_0

(Not easy...)

Statistical Identification

- Recursive Identification
- Non-recursive Identification
- Sign Restrictions
- Long Run Zero Restrictions
- Medium Run
- Maximum Variance
- Identification by Heteroskedasticity
- ...

IV Identification

- **Narrative Methods**:
  - Romer & Romer (1989, 2004) monetary shock series
  - Ramey and Shapiro (1998) and Ramey (2011) series of expected changes in future government defence spending
  - Romer and Romer (2010) narrative tax shocks
- **Information from markets**:
  - High frequency identification (HFI) of MP shocks
    Gürkaynak et al. (2005), Piazzesi and Swanson (2008), Gertler and Karadi (2015)
  - Leeper, Richter, Walker (2011): tax shocks from spreads between federal and municipal bonds
- **Information from surveys**

Reporting SVAR Results

SVARs Results

Results from identified SVARs are often reported in the form of:

1. Impulse response functions
2. Variance decomposition
3. Historical decomposition

Structural Impulse Response Functions

Structural Identified Model

$$Y_t = \mu + C(L){B_0}^{-1}u_t =$$
$$\mu + {B_0}^{-1}u_t + C_1{B_0}^{-1}u_{t-1} + C_2{B_0}^{-1}u_{t-2} + C_3{B_0}^{-1}u_{t-3} + \dots$$

The **Impulse Response Functions** to identified shocks are

$$IRFs_{t+j} = \frac{\partial Y_{t+j}}{\partial u_t} = C_j{B_0}^{-1}$$

Interpretation: $IRFs_{t+j}$ row i, column k is the 'causal effects' of a unit increase in u_{kt} for the value of the ith variable at time $t+j$ holding all other shocks constant

Variance Decomposition

- Decompose the total variance of a time series into the percentages attributable to each structural shock
- Variance decomposition analysis is useful in order to address questions like:
  - What are the sources of the business cycle?
  - Is the shock important for economic fluctuations?

Variance Decomposition

Let us consider the MA representation of an identified SVAR

$$Y_t = F(L)u_t$$

The variance of Y_{it} is given by

$$\begin{aligned} var(Y_{it}) &= \sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik,j}^2 var(u_{k,t-j}) \\ &= \sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik,j}^2 \end{aligned}$$

where $\sum_{j=0}^{\infty} F_{ik,j}^2$ is the variance of Y_{it} generated by the k-th shock.

Variance Decomposition

This implies that

$$\frac{\sum_{j=0}^{\infty} F_{ik,j}^2}{\sum_{k=1}^{n} \sum_{j=0}^{\infty} F_{ik,j}^2}$$

is the percentage of variance of Y_{it} explained by the kth shock.

Variance Decomposition

It is also possible to study the percentage of variance of the series explained by the shock at different horizons, i.e. short vs. long run.

Consider the forecast error in terms of structural shocks.

The horizon h forecast error is given by

$$Y_{t+h} - Y_{t+h|t} = F_{h-1}u_{t+1} + F_{h-2}u_{t+2} + \cdots + F_0 u_{t+h}$$

Variance Decomposition

The variance of the forecast error of the i-th variable is thus

$$\begin{aligned} var(Y_{it+h} - Y_{it+h|t}) &= \sum_{k=1}^{n} \sum_{j=0}^{h-1} F_{ik,j}^2 var(u_{k,t-j}) \\ &= \sum_{k=1}^{n} \sum_{j=0}^{h-1} F_{ik,j}^2 \end{aligned}$$

Thus the percentage of variance of Y_{it} explained by the kth shock is

$$\frac{\sum_{j=0}^{h-1} F_{ik,j}^2}{\sum_{k=1}^{n} \sum_{j=0}^{h-1} F_{ik,j}^2}$$

Historical Decomposition

Decomposition of time series of interest in the contributions of structural shocks

- Neglect deterministic terms and consider MA representations
- Iterate the (estimated) VAR backward to the initial conditions
- Decompose each series $Y_{i,t}$ into

$$Y_{it}^{(k)} = \sum_{i=0}^{t-1} F_{ik,j} u_{k,t-j} + (A_1^{(t)} Y_0 + \cdots + A_p^{(t)} Y_{-p+1})_i$$

- The series $Y_{j,t}^{(k)}$ represents the contribution of the k-th structural shock to the j-the component of the vector Y_t, given initial conditions

2. VARs and Local Projections

The Estimation of the IRFs – VAR

- Let's consider a VAR(1) or a VAR(p) in companion form
- VAR(1):

$$y_t = Ay_{t-1} + e_t$$

The Estimation of the IRFs – VAR

- Let's consider a VAR(1) or a VAR(p) in companion form
- VAR(1):

$$y_t = Ay_{t-1} + e_t$$

- Iterating (MA Representation):

$$
\begin{aligned}
y_{t+1} &= Ay_t + e_{t+1} \\
y_{t+2} &= A^2 y_t + e_{t+2} + A^1 e_{t+1} \\
&\vdots \\
y_{t+h} &= A^h y_t + e_{t+h} + \cdots + A^{h-2} e_{t+2} + A^{h-1} e_{t+1}
\end{aligned}
$$

The Estimation of the IRFs – VAR

- Identifying Shocks:

$$e_t = {B_0}^{-1} u_t$$

- Responses to the identified shocks are obtained by retrieving the VAR moving average representation

- Impulse Response Functions:

$$\boxed{IRF_h^{VAR} = A^h B_0^{-1}}$$

Is the VAR correctly specified?

Potential misspecifications (Braun and Mittnik, 1993):

- Too few lags?
- MA(q) component missing?
- Non-linearities?
- Non-invertible shocks?
- ...

- A non-parametric alternative to VARs are **Local Projections**
  - Conditioning set as in a VAR (Jorda, 2005)
  - Direct regression on some 'instrument' for the shock of interest (Ramey, 2016)

Local Projections IRFs

Jordà (2005)

- Local Projection/Direct Forecast method

$$
\begin{aligned}
y_{t+1} &= \widetilde{A}^{(1)} y_t + v_{t+1} \\
y_{t+2} &= \widetilde{A}^{(2)} y_t + v_{t+2} \\
&\vdots \\
y_{t+h} &= \widetilde{A}^{(h)} y_t + v_{t+h}
\end{aligned}
$$

- Identifying Shocks:

$$v_t \equiv e_t = {B_0}^{-1} u_t$$

- Impulse Response Functions:

$$\boxed{IRF_h^{LP} = \widetilde{A}^{(h)} B_0^{-1}}$$

Local Projections IRFs

Jordà (2005)

Remark:

- Error term is **serially correlated** (except for horizon h = 0)
- v_{t+h} is a moving average of the forecast errors from t to $t+h$
- In our VAR(1), estimated by local projections, we would have had:

$$v_{t+h} = e_{t+h} + \cdots + A^{h-2}e_{t+2} + A^{h-1}e_{t+1}$$

- Standard errors needs correction for serial correlation (Newey-West, 1987)

Local Projections – Direct Regression

Ramey (2016)

- A **direct measure** z_t of the shock u_t^1 is observed (an **instrument**)
- Direct regression at horizon h

$$Y_{i,t+h} = F_{h,i1}\widehat{Y}_{1,t} + \gamma_h' W_t + \nu_{i,t+h}^h ,$$

  - $\widehat{Y}_{1,t}$ is the fitted value of $Y_{1,t}$ (an 'indicator' variable) from a first-stage regression on the external instrument z_t
  - $F_{h,i1}$ are the causal responses of $Y_{i,t+h}$ to u_t^1 at horizon h
  - W_t is a generic set of control variables
  - $\nu_{i,t+h}^h$ are serially correlated projection residuals
  - Also possible to directly regress on the shock with **no controls**

VARs vs Local Projections

- VAR-IRFs are optimal and consistent only if the VAR adequately captures the DGP
- LP-IRFs are more **robust to model misspecification** but suffer from **high estimation uncertainty**
- While direct methods should in principle be preferable, their advantages are hardly realised in practice
(see Marcellino, Stock, and Watson, 2006, and Kilian and Kim, 2011)
- Selecting between the two methods: empirical problem choosing between **bias** and **estimation variance**...

VARs vs Local Projections

Plagborg-Møller and Wolf 2021

VARs and LPs estimate the same IRFs when

- weakly stationary process
- unrestricted lag order
- and same information set