(SEMI-)STRUCTURAL MACROECONOMETRICS

(On SVARs, their identification and estimation)

Academic Year 2023-24

Master Program in Economics and Econometrics

LM(EC)^2

Department of Economics

University of Bologna

Luca Fanelli

Department of Economic Sciences

University of Bologna

luca.fanelli@unibo.it

https://www.unibo.it/sitoweb/luca.fanelli/en

SYLLABUS

1- Structural shocks

2- From small-scale monetary DSGE models to SVARs: Why Choleski-SVARs are not enough?

3- A few words on reduced form VAR representations and estimation issues (this can be skipped) $\longleftarrow$ Supplementary Material

4- Using VARs to estimate single equations from structural econometric models: the New Keynesian Phillips Curve example (intuition)

5- Overview of Rothenberg's (1971, ECMA) general approach to the identification in structural parametric models;

6 - SVARs, Structural IRFs and the identification of SVARs

SYLLABUS (cont'd)

7 - AB-SVARs: specification, identification and estimation

7.1 - Case study: Blanchard and Perotti's (2002, QJE) fiscal model

8 - FEVDs: short account

9 - Analytic confidence bands for IRFs

10 - Bootstrap confidence bands for IRFs

10.1 - The notion of boostrap consistency

11- Some current - novel - identification schemes:

11.1- The sign-restrictions approach: a frequentist view (short account);

11.2 -The "statistical" ICA approach (only mention, or additional material provided by teacher);

11.3 - Heteroskedasticity approach: constant IRFs;

11.4 - Heteroskedasticity approach: regime-dependent IRFs;

11.5 - The "external variables" approach:

11.5.1-Proxy-SVARs, identification and estimation

11.5.2-Local Projections - only partially developed, work in progress,

Supplementary Materials (separate files posted on Virtuale)

A. Representations of interest of reduced form VARs

B. VAR Information Matrix

C. Estimation issues (short account)

D. Large sample properties of estimators and main inferential issues

E. A few words on unit roots and cointegration in (S)VARs

F. A few words on identification through long run restrictions

Important topics not covered

- Non-stationary SVARs (see, partially, Supplementary Materials)

- Bayesian approach (the slides provide a frequentist view)

- Time-varying SVARs (TV-SVARs).

Some prerequisites (not in order of importace):

- Policy oriented empirical macroeconomics

- Basic probability and statistical inference

- Linear algebra and calculus

- The econometrics of OLS and IV regressions

Some supporting readings

- Bunch of papers provided by the teacher

- Amisano, G. and Giannini, C. (1997). Topics in Structural VAR Econometrics, 2nd edn, Springer, Berlin.

- Kilian, L. and Lütkepohl, H. (2017), Structural Vector Autoregressive Analysis, Cambridge University Press

- Lütkepohl, H. (2015), New Intoduction to multivariate time series analysis, Springer. (free).

- Ramey, V. (2016), Handbook of Macroeconomics.

1. Structural shock?

Modern dynamic macroeconomics studies **the impact and propagation of structural shocks.**

Some crucial questions we attempt to answer are:

(a) What are the sources of business cycle fluctuations?

(b) How does monetary policy affect the economy?

(c) How does fiscal policy affect the economy?

(d) How important are technology shocks?

(e) What is the effect of an exogenous oil supply shock on macro activity/inflation?

(f) Is uncertainty an exogenous driver of business cycle fluctuations?

The word "structural" means something like: theory-oriented, theory-driven, theory-informed.

We take causal stands by relying on theory (at most, good economic sense).

Definition 0. Shocks. *Shocks are primitive exogenous forces that are uncorrelated with each other and should be economically meaningful.*

The empirical counterpart of the shocks we attempt to identify must satisfy three conditions:

1.They must be **exogenous** with respect to the other current and lagged endogenous variables in the model.

2. They must be **uncorrelated** with other exogenous shocks; otherwise, we cannot identify the unique causal effects of one exogenous shock relative to another. Here I admit some form of nonlinear dependence. Question: can structural shock can be actually **independent**?

3. They must be **unanticipated**. This means that agents' information set at time t must not contain shocks that will materialize at time $t+q$, $q>0$. We shall assume 3 throughout our course.

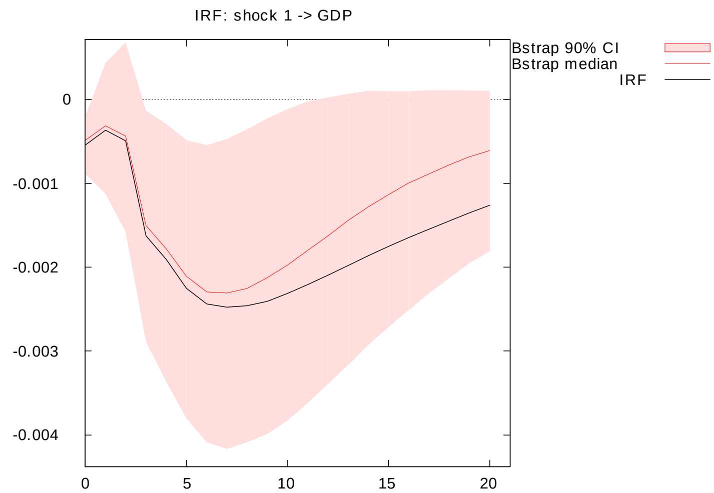
How do we **quantify** the **impact and propagation** of structural shocks?

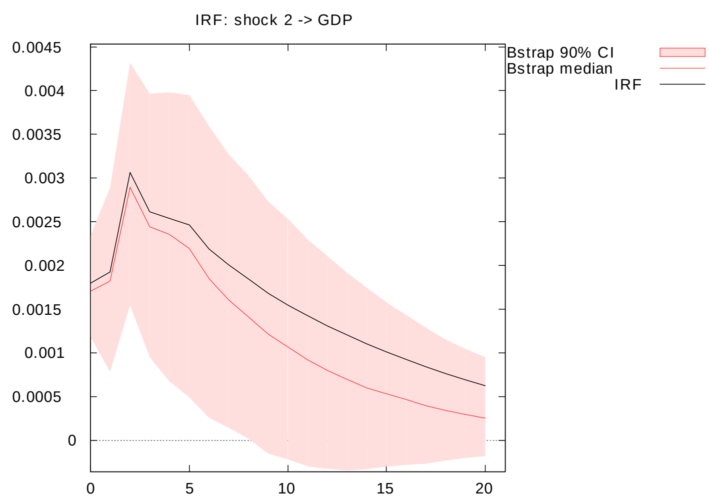
Our primary analytical instruments are **Impulse Response Functions (IRFs)**.

IRFs trace the **dynamic causal effects** of the structural shocks on the variables of interest.

This course can, to some extent, be regarded as a course on the estimation of IRFs and the quantification of the uncertainty surrounding their estimates.

Towards the end of the course, we will also delve into **Local Projections (LPs)**, which share similarities with IRFs.


IRF: shock 1 -> GDP
Bstrap 90% CI
Bstrap median
IRF
0
-0.001
-0.002
-0.003
-0.004
0
5
10
15
20


IRF: shock 2 -> GDP
Bstrap 90% CI
Bstrap median
IRF
0.0045
0.004
0.0035
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
0
5
10
15
20

Even from some media....

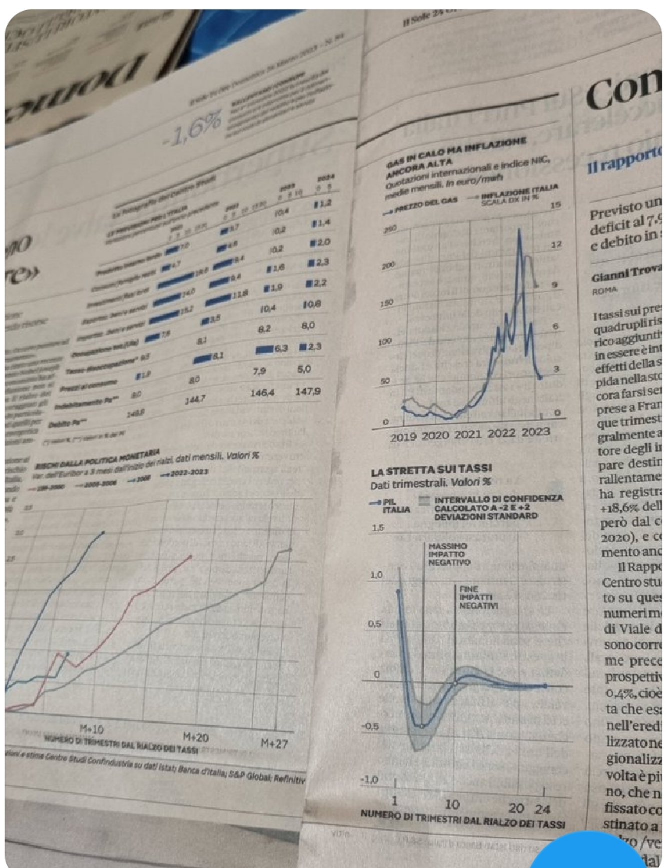
09:41 • 59%

Luca Fanelli
753 post

Post Risposte Highlight Contenuti

Luca Fanelli @lucafanelli5 · 26 Mar

Wow! Today @sole24ore published the quarterly response of Italian GDP to a one standard deviation monetary policy shock, as implied by Confindustria's macro model. With 90% confidence intervals! I was not used to see IRFs published on popular newspapers. Respect!



2 18 165 22,

Objective

The objective of this course is to acquaint students with the principal econometric methods **for identifying and estimating structural shocks**, and the resulting IRFs through **Structural Vector Autoregressions (SVARs)**.

The heuristic path involves commencing with a DSGE-based rationale for non-recursive SVARs.

Subsequently, we delve into the technicalities associated with “traditional” SVARs.

Finally, we transition to exploring more recent advancements in this field.

2. From small-scale monetary DSGE models to SVARs: Why Choleski-SVARs are not enough?

An example of a small-scale monetary DSGE model (a standard New Keynesian DSGE model):

$$\tilde{y}_t = \gamma E_t \tilde{y}_{t+1} + (1-\gamma)\tilde{y}_{t-1} - \delta(R_t - E_t \pi_{t+1}) + \zeta_{\tilde{y},t}$$
$$\pi_t = \underbrace{\frac{\varrho}{1+\varrho\varkappa}}_{\beta_f} E_t \pi_{t+1} + \underbrace{\frac{\varkappa}{1+\varrho\varkappa}}_{\beta_b} \pi_{t-1} + \kappa \tilde{y}_t + \zeta_{\pi,t}$$
$$R_t = \rho R_{t-1} + (1-\rho)\varphi_\pi (E_t \pi_{t+1} - \pi_t^*) + (1-\rho)\varphi_y \tilde{y}_t + \zeta_{R,t}$$

$$\begin{aligned}
\zeta_{\tilde{y},t} &= \phi_{\tilde{y}} \zeta_{\tilde{y},t-1} + \sigma_{\tilde{y}} \varepsilon_{\tilde{y},t} \quad , \quad \varepsilon_{\tilde{y},t} \sim WN(0,1) \\
\zeta_{\pi,t} &= \phi_{\pi} \zeta_{\pi,t-1} + \sigma_{\pi} \varepsilon_{\pi,t} \quad , \quad \varepsilon_{\pi,t} \sim WN(0,1) \\
\zeta_{R,t} &= \phi_{R} \zeta_{R,t-1} + \sigma_{R} \varepsilon_{R,t} \quad , \quad \varepsilon_{R,t} \sim WN(0,1)
\end{aligned}$$

The model is microfounded; we skip that part.

$M = 3$ dimension of the system;

$\tilde{y}_t =$ output gap (measure of real economic activity relative to potential);

$\pi_t =$ inflation rate; $\pi_t^* =$ target inflation rate (might be constant);

$R_t =$ short term interest rate (or policy rate);

$E_t \cdot \equiv E(\cdot \mid \mathcal{I}_t)$ conditional expectations operator;

$\zeta_{x,t}$, $x = \tilde{y}, \pi, R$ disturbance terms (with diagonal VAR structure)

$\varepsilon_t = (\varepsilon_{\tilde{y},t}, \varepsilon_{\pi,t}, \varepsilon_{R,t})'$ vector of **structural shocks:** $\varepsilon_t \sim WN(0, I_M)$;

$\theta := (\gamma, \delta, \varrho, \beta_f, \beta_b, \kappa, \rho, \varphi_\pi, \varphi_y, \phi_{\tilde{y}}, \phi_\pi, \phi_R, \sigma_{\tilde{y}}, \sigma_\pi, \sigma_R)'$ structural parameters that capture agents preferences and technology, policy parameters, etc.

Let $W_t = \begin{pmatrix} \tilde{y}_t \\ \pi_t \\ R_t \end{pmatrix} = (\tilde{y}_t, \pi_t, R_t)'$;

$M = \dim(W_t) = 3$.

In compact matrix form:

$$\Gamma_0 W_t = \Gamma_f E_t W_{t+1} + \Gamma_b W_{t-1} + \Theta \zeta_t$$

$$\zeta_t = \Phi \zeta_{t-1} + \underbrace{\Sigma_\varepsilon^{1/2} \varepsilon_t}_{v_t}$$

$$E(\varepsilon_t) = 0,\ E(\varepsilon_t \varepsilon_t') = I_M$$

where Γ_0, Γ_f, Γ_b, $\Theta = I_M$, $\Phi \equiv dg(\Phi)$ and $\Sigma_\varepsilon \equiv dg(\Sigma_\varepsilon)$ depend on θ.

A remarkable feature of this **structural model** is that it accounts for both **forward-looking behavior** (agents expectations on future economic development) through the term $\Gamma_f E_t W_{t+1}$, and **backward-looking** behavior (inertia) through $\Gamma_b W_{t-1}$.

ζ_t has the role of inducing persistence to the system (law of motion of ζ_t is typically assumed).

Diagonal VAR structure for ζ_t :

$$\zeta_t = \Phi \zeta_{t-1} + \underbrace{\Sigma_\varepsilon^{1/2} \varepsilon_t}_{v_t}$$

is typically postulated had-hoc.

It is possible to prove that if a **unique stable** solution to

$$\Gamma_0 W_t = \Gamma_f E_t W_{t+1} + \Gamma_b W_{t-1} + \Theta\zeta_t$$

$$\zeta_t = \Phi\zeta_{t-1} + \Sigma_\varepsilon^{1/2}\varepsilon_t \quad , \ \varepsilon_t \sim WN(0, I_M)$$

exists, it can be represented in the form

$$W_t = \Upsilon_1 W_{t-1} + \Upsilon_2 W_{t-2} + \Psi\varepsilon_t$$

$$(I_M - \Upsilon_1 L - \Upsilon_2 L^2)W_t = \Psi\varepsilon_t$$

(recall L is the lag operator) where Υ_1 Υ_1 and Ψ depend nonlinearly and **uniquely** on the parameters θ through the matrices Γ_0, Γ_f, Γ_b, $\Phi \equiv dg(\Phi)$ and $\Sigma_\varepsilon \equiv dg(\Sigma_\varepsilon^{1/2})$ and the so-called **cross-equation restrictions (CER).**

The solution to the New Keynesian DSGE model reads as what throughout we shall refer to as "SVAR, B-model".

Importantly, Ψ is a **full matrix** !!!!!

If the unique and stable (asymptotically stationary) solution exists, it can be proved that for $s \in \mathbb{C}$:

$$\det(I_M - \Upsilon_1 s - \Upsilon_2 s^2) = 0 \Leftrightarrow \mid s \mid > 1$$

which implies that the inverse of the matrix polynomial $\Upsilon(L) := (I_M - \Upsilon_1 L - \Upsilon_2 L^2)$ has the representation:

$$\Upsilon(L)^{-1} := (I_M - \Upsilon_1 L - \Upsilon_2 L^2)^{-1}$$

$$= I_M + C_1 L + C_2 L^2 + ... + C_\ell L^\ell + ...$$

where

$$\sum_{j=0}^{\infty} C_j < \infty \quad , C_0 = I_M \ , \ \ C_\ell \underset{\ell \to \infty}{\longrightarrow} 0$$

(absolute summability condition).

Thus, from the unique stable solution

$$\Upsilon(L)W_t = \Psi\varepsilon_t$$

we obtain the **Structural VMA representation** associated with the small-scale DSGE:

$$\begin{aligned} W_t &= \sum_{j=0}^{\infty} C_j \Psi \varepsilon_t \\ &= \Psi\varepsilon_t + C_1\Psi\varepsilon_{t-1} + ... + C_h\Psi\varepsilon_{t-h} + ... \end{aligned}$$

which implies that the IRFs are:

$$IRF_{\bullet,\bullet}(h) := \frac{\partial W_{t+h}}{\partial \varepsilon_t'} = C_h\Psi \quad , \quad h = 0, 1, ...$$

Therefore, the response of variable i at horizon h to a standard deviation shock to the j-th shock in ε_t is:

$$IRF_{i,j}(h) := e_i' C_h \Psi e_j$$

where e_i is $M \times 1$ vector that takes value 1 in position i and zero elsewhere.

Note that

$$IRF_{\bullet,\bullet}(0) := \frac{\partial W_t}{\partial \varepsilon_t'} = \mathbf{\Psi} = \mathbf{\Psi}(\theta)$$

hence we here notice that output and inflation **re-spond instantaneously to a MP shock!**

$$\begin{pmatrix} \tilde{y}_t \\ \pi_t \\ R_t \end{pmatrix}, \quad \mathbf{\Psi} = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \psi_{1,3} \\ \psi_{2,1} & \psi_{2,2} & \psi_{2,3} \\ \psi_{3,1} & \psi_{3,2} & \psi_{3,3} \end{pmatrix}, \begin{pmatrix} \varepsilon_{\tilde{y},t} \\ \varepsilon_{\pi,t} \\ \varepsilon_{R,t} \end{pmatrix} \begin{pmatrix} \text{shock 1} \\ \text{shock 2} \\ \text{shock 3} \end{pmatrix}$$

where, recall

$$\psi_{i,j} = \psi(\theta).$$

Observe that, given the SVMA representation:

$$
\begin{aligned}
W_{t+h} &= \Psi\varepsilon_{t+h} + C_1\Psi\varepsilon_{t+h-1} + C_2\Psi\varepsilon_{t+h-2} \\
&\quad ... + C_{h-1}\Psi\varepsilon_{t++1} + C_h\Psi\varepsilon_t + C_{h+1}\Psi\varepsilon_{t-1} + ...
\end{aligned}
$$

for a vector of shocks of size $\varepsilon_t := \delta$ we have:

$$
E(W_{t+h} \mid \mathcal{I}_{t-1}, \varepsilon_t = \delta) - E(W_{t+h} \mid \mathcal{I}_{t-1}) = C_h\Psi\delta.
$$

For $\delta = e_j$ (meaning that we only perturb the j-th element in ε_t of one standard deviation):

$$
IRF_{\bullet,j}(h) := C_h\Psi e_j.
$$

Take-away from previous considerations:

structural policy-oriented macro models suggest that for a (S)VAR model for W_t to be consistent with theory, representations based on, e.g.

$$\Upsilon(L)W_t = \underbrace{B}_{\text{on-impact}} \varepsilon_t,$$

the matrix B should not be subject, say, to many zero restrictions.

Paradox: we shall see that it is much easier to deal with SVARs featuring many zeros (or other point- restrictions) on B.

But what does it mean “computing the solution of a DSGE (or RE) model”?

Simple example.

Consider famous **Cagan's RE model** (with W_t scalar, $M = 1$):

$$W_t = \alpha E_t W_{t+1} + u_t \quad , \quad \alpha \in \mathbb{R}^+ \ , \ \ u_t \sim \text{MDS} \quad (E_t u_{t+1} = 0)$$

where u_t is a **fundamental shock** (i.e. the "economic" shock driving W_t).

A solution to the RE model above **is any stochastic process $\{W_t\}$ such that once the realization W_t is substituted into the equation, the equation is verified.**

For instance, $W_t := u_t$ is a solution to Cagan's model because for any $\alpha \in \mathbb{R}^+$:

$$u_t = \alpha \underbrace{E_t u_{t+1}}_{0} + u_t \Longleftrightarrow u_t = u_t.$$

The problem is that, in general, there are **multiple solutions to a RE model** (indeterminacy), in the sense that there can be many stochastic processes that verify the equation. Not necessarily these solutions need to be stable (asymptotically stationary).

To see this, use the decomposition:

$$W_{t+1} = E_t W_{t+1} + \eta_{t+1}$$

where $\eta_{t+1} = W_{t+1} - E_t W_{t+1}$ is a RE **error** (or forecast error) such that $E_t \eta_{t+1} = 0$ (MDS).

This allows us to rewrite Cagan's RE model as

$$W_t = \alpha W_{t+1} + u_t - \alpha \eta_{t+1} \quad , \quad \alpha \in \mathbb{R}^+$$

and, by rearranging:

$$W_{t+1} = \frac{1}{\alpha} W_t + \eta_{t+1} - \frac{1}{\alpha} u_t$$

The RE error η_{t+1} can be **linearly projected** onto the fundamental shock, obtaining:

$$\eta_{t+1} = \underbrace{\Pr oj(\eta_{t+1} \mid u_{t+1})}_{\text{part explained by the fundamental shock}} + \underbrace{s_{t+1}}_{\text{unexplained}}$$

where

$$\Pr oj(\eta_{t+1} \mid u_{t+1}) = \tau u_{t+1}, \text{ where } \tau \text{ arbitrary param.}$$

$$s_{t+1}{=}? \ \ldots \ \text{ sunspot shock!}$$

Thus, by using the expression

$$\eta_{t+1} = \tau u_{t+1} + s_{t+1}$$

in Cagan's model, the solution reads:

$$W_{t+1} = \frac{1}{\alpha} W_t + \tau u_{t+1} - \frac{1}{\alpha} u_t + s_{t+1}$$

Aside from the term s_{t+1} (note that $E_t\eta_{t+1} = 0 \Rightarrow E_t s_{t+1} = 0$), the model

$$W_{t+1} = \frac{1}{\alpha}W_t + \tau u_{t+1} - \frac{1}{\alpha}u_t + s_{t+1}$$

reads as an ARMA(1,1)-type model:

$$(1 - \frac{1}{\alpha}L)W_{t+1} = (\tau - \frac{1}{\alpha}L)u_{t+1} + s_{t+1}$$

You can notice that for $\alpha > 1$, the model

$$(1 - \frac{1}{\alpha}L)W_{t+1} = (\tau - \frac{1}{\alpha}L)u_{t+1} + s_{t+1}$$

is stationary because the autoregressive polynomial has solution $|z| = |\alpha| > 1$.

However, since τ is arbitrary and also the sunspot shock is an arbitrary MDS, we have **multiple stable (asymptotically stationary) solutions.**

Always for $\alpha > 1$, in the special case in which $s_{t+1} = 0$ a.s., the stationary ARMA(1,1) solution:

$$(1 - \frac{1}{\alpha}L)W_{t+1} = (\tau - \frac{1}{\alpha}L)u_{t+1}$$

is indeterminate as τ can be any parameter (unrelated to α).

So, there are **multiple stable solutions.**

For $\alpha > 1$ and $s_{t+1} = 0$ a.s., the solution

$$(1 - \frac{1}{\alpha}L)W_{t+1} = (\tau - \frac{1}{\alpha}L)u_{t+1}$$

collapses, **in the very special case in which $\tau = 1$,** to (the two polynomials have common roots, hence cancel)

$$W_{t+1} = u_{t+1}$$

known as **Minimum State Variable** (MSV) solution in the RE literature.

Always keeping $s_{t+1} = 0$ a.s., for $0 < \alpha < 1$, the model generates **explosive solutions** because the representation is:

$$\underbrace{(1 - \frac{1}{\alpha}L)}_{\text{solution } |z|<1} W_{t+1} = (\tau - \frac{1}{\alpha}L)u_{t+1}.$$

So, we have **multiple (**recall τ is indeterminate**) explosive solutions.**

However, **in the even more special case in which** $\tau = 1$, we still have canceling roots which implt that the MSV solution

$$W_{t+1} = u_{t+1}$$

is the only (unique) stable solution consistent with $0 < \alpha < 1$.

3. A few words on the reduced form VAR representations

Given the $M \times 1$ vector of variables W_t, a **Vector Autoregressive (VAR), or reduced form VAR** is described by the system of equations:

$$W_t = \mu + \Pi_1 W_{t-1} + \Pi_2 W_{t-2} + ... + \Pi_p W_{t-p} + u_t$$

where:

$\mu = (\mu_1, ..., \mu_M)'$ is an $M \times 1$ constant (intercept);

Π_j, $j = 1, ..., p$ are $M \times M$ matrices of parameters;

p is the VAR lag order (the VAR is also called VAR(p));

$u_t = (u_{1,t}, ..., u_{M,t})'$ is a disturbance term, **called reduced form disturbances** (or VAR innovations or forecast errors) which satisfies White Noise conditions:

$$\begin{aligned} E(u_t) &= 0_{M\times 1}; \\ E(u_t u_t') &= \Sigma_u \quad , \quad M \times M \quad \text{pos. def.} \\ E(u_t u_{t-k}') &= 0_{M\times M} \quad \text{for} \quad k = 1, 2, ... \end{aligned}$$

The **minimal requirement** on u_t is to be **uncorrelated over time (innovation)**, meaning that u_t must be interpreted as a **forecast error**, i.e. the part of W_t which can not be explained (forecasted) by past information:

$$u_t = W_t - E(W_t \mid W_{t-1}, W_{t-2}, ..., W_1).$$

This implies that $u_t \sim$MDS. Why?

Our course will be based on the assumption that $\Sigma_u := E(u_t u_t')$ is an $M \times M$ covariance matrix, is symmetric and positive definite.

It is seen that in a VAR model each variable in W_t depends on p lags of itself but also of all other variables.

The total number of free parameters in $(\mu, \Pi_1, ..., \Pi_p, \Sigma_u)$ is $f = M + M^2 p + \frac{1}{2}M(M+1)$.

In the following

A reduced form VAR model can be also represented as follows (we use all of them):

(i) Multi-equational reg. model$\leftarrow$ Supplementary

(ii) Vector Moving Average (VMA)$\leftarrow$ Supplementary

(iii) Companion form rep.$\leftarrow$ Supplementary

The important things to know for the rest of our course are the following.

A stationary VAR can be represented in VMA form:

$$W_t = v + C(L)u_t = v + \sum_{h=0}^{\infty} C_h u_{t-h}$$

where $v = \underbrace{\left(\sum_{h=0}^{\infty} C_h\right)}_{<\infty} \mu = C(1)\mu$ and $C_h \rightarrow 0_{M\times M}$ as $h \rightarrow \infty$,

$$C(L) = \sum_{h=0}^{\infty} C_h = \Pi(L)^{-1}$$

$$\Pi(L) := (I_M - \Pi_1 L - \Pi_2 L^2 - ... - \Pi_p L^p).$$

The VMA parameters in C_h can be expressed as:

$$C_h = R\,(\mathcal{C})^h\,R' \quad , \quad h = 0, 1, 2, ...$$

where $\mathcal{C}$ is the **companion matrix**:

$$\underset{Mp\times Mp}{\mathcal{C}} := \begin{pmatrix} \Pi_1 & \Pi_2 & \cdots & \Pi_{p-1} & \Pi_p \\ I_M & 0_{M\times M} & \cdots & 0_{M\times M} & 0_{M\times M} \\ 0_{M\times M} & I_M & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0_{M\times M} & 0_{M\times M} \\ 0_{M\times M} & 0_{M\times M} & \cdots & I_M & 0_{M\times M} \end{pmatrix}$$

and R is a **selection matrix**:

$$\underset{M\times Mp}{R} := \begin{pmatrix} I_M & , & 0_{M\times(Mp-M)} \end{pmatrix}.$$

4- Using VARs to estimate single equations from structural econometric models: the New Keynesian Phillips Curve example (short account)

Reduced form VARs can be conveniently used to estimate the parameters of interest of single structural equations taken from, e.g., DSGE-type models.

At the same time, they allow us to assess their empirical validity in a convenient way through a set of cross-equation restrictions.

In this respect, they serve as the "statistical platform" upon which we can compare our theory.

As an example we now briefly discuss how the NKPC featured by our small-scale monetary DSGE model can be analyzed by a VAR model.

Let's turn on the small-scale DSGE model:

$$
\begin{aligned}
\tilde{y}_t &= \gamma E_t \tilde{y}_{t+1} + (1-\gamma)\tilde{y}_{t-1} - \delta(R_t - E_t \pi_{t+1}) + \zeta_{\tilde{y},t} \\
\pi_t &= \underbrace{\frac{\varrho}{1+\varrho\varkappa}}_{\beta_f} E_t \pi_{t+1} + \underbrace{\frac{\varkappa}{1+\varrho\varkappa}}_{\beta_b} \pi_{t-1} + \kappa \tilde{y}_t + \zeta_{\pi,t} \\
R_t &= \rho R_{t-1} + (1-\rho)\varphi_\pi (E_t \pi_{t+1} - \pi_t^*) + (1-\rho)\varphi_y \tilde{y}_t + \zeta_{R,t}
\end{aligned}
$$

and "isolate", e.g. the NKPC (to simplify we change name to κ: from κ to λ):

$$
\pi_t = \beta_f E_t \pi_{t+1} + \beta_b \pi_{t-1} + \lambda \tilde{y}_t + \zeta_{\pi,t}
$$

Object: we want to use a VAR for estimating the NKPC.

NKPC: it's a micro-funded macroeconomic model of inflation dynamics (now we use x_t in place of $\tilde{y}_t$):

$$\underbrace{\pi_t}_{\text{infl. time } t} = \underbrace{\beta_b \pi_{t-1}}_{\text{previous period infl.}} + \underbrace{\beta_f E_t \pi_{t+1}}_{\text{expceted future (next period) infl.}} + \underbrace{\lambda x_t}_{\text{economic activity}} + \underbrace{\zeta_{\pi,t}}_{\text{cost-push transitory shocks}} \tag{1}$$

where

β_b: backward looking parameter (> 0)

β_f: forward looking parameter (> 0)

stability restriction: $\beta_b + \beta_f < 1$

λ (κ): slope parameter (> 0) (determines the 'flatness' of the NKPC).

Throughout we assume that x_t =output gap, i.e. a measure of the deviation of real output from potential output;

x_t could also be the unemployment rate: in this case the sign of the slope parameter λ must be reverted in (1);

$\theta := (\beta_b, \beta_f, \lambda)'$ are the structural parameters of interest; in particular, the slope parameter λ is of crucial interest for the conduct of monetary policy as it links inflation to real economic activity.

Also β_f is of particular interest as its magnitude indicates how important are inflation expectations in determining current inflation.

Note that in the NKPC model above, $E_t\pi_{t+1} \equiv E(\pi_{t+1} \mid \mathcal{F}_t^a)$, where $\mathcal{F}_t^a$ denotes the agents' information set available at time t.

Assumption (questionable): $\zeta_{\pi,t}$ ∼MDS ≠AR(1) ⇒ $E_t\zeta_{\pi,t+1} = 0$

Under this assumption, conditioning both sides of the NKPC (1) with respect to the information set at time $t-1$, yields:

$$E_{t-1}\pi_t = \beta_b\pi_{t-1} + \beta_f E_{t-1}\pi_{t+1} + \lambda E_{t-1}x_t \quad (2)$$

where we have applied the MDS property for the cost-push shock, the property $E_{t-1}\pi_{t-1} = \pi_{t-1}$ ($\pi_{t-1} \in \mathcal{F}_t^a$) and the property $E_{t-1}E_t\pi_{t+1} = E_{t-1}\pi_{t+1}$.

Temptation: use GMM approach because model (1) implies:

$$E(\pi_t - \beta_b\pi_{t-1} - \beta_f\pi_{t+1} - \lambda x_t \mid \mathcal{F}_{t-1}^a) = 0$$

so that, for any $r \times 1$ vector of instruments $Z_t \in \mathcal{F}_{t-1}^a$ such that $r \geq \dim(\theta) = 3$, it holds:

$$E\left\{(\pi_t - \beta_b\pi_{t-1} - \beta_f\pi_{t+1} - \lambda x_t)Z_t\right\} = 0_{r\times 1}$$

Are you familiar with GMM estimation theory?

Alternative: use a VAR.

VAR dynamics.

Let $W_t := \begin{pmatrix} \pi_t \\ x_t \end{pmatrix}$ $(M = 2)$.

Note that we can reformulate our exercise by also including R_t in W_t.

We posit that W_t is well approximated by the VAR(1) $(p = 1)$:

$$W_t = \Pi_1 \; W_{t-1} + u_t \qquad u_t \sim WN(0, \Sigma_u). \tag{3}$$

Assume that the VAR(1) in (3) is stationary (asymptotically stable), which means that the matrix Π_1 has all eigenvalues less than one in modulus.

This implies the forecasting formula:

$$E_t W_{t+h} \equiv E(W_{t+h} \mid \mathcal{F}_t^e) = (\Pi_1)^h \, W_t, \tag{4}$$

where $\mathcal{F}_t^e := \sigma(W_t, W_{t-1}, ...,)$ is the filtration or the econometrician's information set (potentially $\mathcal{F}_t^e \neq \mathcal{F}_t^a$).

Hence, from the stationary VAR(1) we can compute conditional forecasts:

$$
\begin{aligned}
E_t W_{t+1} &\equiv E_{t-1} W_t = \Pi_1 W_t; \\
E_{t-1} W_{t+1} &= (\Pi_1)^2 \, W_{t-1};
\end{aligned}
$$

where note that:

$$
\Pi_1 := \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} ;
$$

$$
(\Pi_1)^2 := \begin{pmatrix} a_{11}^2 + a_{12}a_{21} & a_{11}a_{12} + a_{12}a_{22} \\ a_{11}a_{21} + a_{21}a_{22} & a_{22}^2 + a_{12}a_{21} \end{pmatrix}. \tag{5}
$$

Also note that, since

$$\pi_t = (1, 0) W_t \quad ; \quad x_t = (0, 1) W_t$$

and, if we want to forecast future values of inflation and the output gap from the VAR, the formula in (4) implies:

$$\begin{aligned} E_t \pi_{t+h} &\equiv (1, 0) E_t W_{t+h} = (1, 0) \left(\Pi_1\right)^h W_t; \\ E_t x_{t+h} &\equiv (0, 1) E_t W_{t+h} = (0, 1) \left(\Pi_1\right)^h W_t. \end{aligned}$$

Our **crucial assumption** is: agents compute their forecasts taking expectations from the VAR in (3), i.e. they have 'VAR-based' expectations:

$$\mathcal{F}_t^a \equiv \mathcal{F}_t^e \ , \ \Rightarrow \ E(\cdot \mid \mathcal{F}_t^a) = E(\cdot \mid \mathcal{F}_t^e)$$

for any random variable.

Under this assumption, the NKPC in (2) can be reformulated as

$$\begin{aligned}(1,0)\Pi_1 W_{t-1} \ &= \ \beta_b(1,0)W_{t-1} \\ &\quad + \beta_f(1,0)\,(\Pi_1)^2\, W_{t-1} + \lambda(0,1)\Pi_1 W_{t-1}\end{aligned}$$

that is

$$\underbrace{\left\{(1,0)\Pi_1 - \beta_b(1,0) - \beta_f(1,0)\,(\Pi_1)^2 - \lambda(0,1)\Pi_1\right\}}_{1\times 2}\underbrace{W_{t-1}}_{2\times 1} = 0. \tag{6}$$

Now, since $W_{t-1} \neq 0_{2\times 1}$ a.s., the way for equation (6) to hold is that it holds the relationship:

$$(1,0)\Pi_1 - \beta_b(1,0) - \beta_f(1,0)\,(\Pi_1)^2 - \lambda(0,1)\Pi_1 = (0,0)$$

(i.e. only the row vector within the parentheses in (6) can be zero).

Recalling the expressions in (5), the relationship above can be re-written as:

$$(a_{11}, a_{12}) - \beta_b(1,0) - \beta_f(a_{11}^2 + a_{12}a_{21},\ a_{11}a_{12} + a_{12}a_{22}) - \lambda(a_{21}, a_{22}) = (0,0)$$

namely:

$$\begin{aligned} a_{11} - \beta_b - \beta_f(a_{11}^2 + a_{12}a_{21}) - \lambda a_{21} &= 0; \\ a_{12} - \beta_f(a_{11}a_{12} + a_{12}a_{22}) - \lambda a_{22} &= 0. \end{aligned}$$

After some simple (but tedious) algebraic manipulations, these can be expressed as (please, check):

$$a_{21} = \frac{a_{11}(1-\beta_f a_{11}) - \beta_b}{\beta_f a_{12} + \lambda}; \quad (7)$$

$$a_{22} = \frac{a_{12}(1-\beta_f a_{11})}{\beta_f a_{12} + \lambda} \quad (8)$$

so that, provided $\gamma_f a_{12} + \lambda \neq 0$ (condition guaranteed by $a_{12} \neq -\frac{\lambda}{\gamma_f}$), we have derived the restrictions that the NKPC implies on the VAR parameters.

Under the restrictions the VAR depends on the following parameters:

$$a_{11}, a_{12}, \beta_b, \beta_f, \lambda$$

that therefore include also the structural parameters of the NKPC $\theta := (\beta_b, \beta_f, \lambda)'$ through a set of CER.

So, imagine that we make a guess:

$$H_0 : \begin{cases} \beta_b = \beta_b^0 \\ \beta_f = \beta_f^0 \\ \lambda = \lambda^0 \end{cases}$$

where β_b^0, β_f^0 and λ^0 are hypothesized values of the structural parameters that we take e.g. from economic theory. Then, under H_0 the VAR(1) model incorporates under the restrictions (7)-(8) the free VAR parameters a_{11} and a_{12} and the structural parameters under H_0, i.e. the matrix Π_1 is such that $\Pi_1^{nkpc} = \Pi_1(a_{11}, a_{12}, \beta_b^0, \beta_f^0, \lambda^0)$.

Hence, imagine a procedure by which one tests the empirical validity of the hypothesis H_0 by estimating VAR model without any restriction (simply estimate the VAR (3)) and then **under** H_0 ... It turns out that the log-likelihood functions of the restricted and unrestricted VAR can be compared... (MORE TO SAY ABOUT THIS).

Exercise 1: try to find the restrictions for the case of a VAR(2); generalize to a VAR(p) ! Hint: write the VAR in companion form ....

Exercise 2: repeat Exercise 2 by including R_t in W_t.

5- Overview of Rothenberg's (1971, ECMA) general approach to the identification problem in structural parametric models

In general, the word "identification" is abused by economists/econometricians.

It is not uncommon for the term "identification" to be used liberally in economics/econometrics, sometimes leading to confusion or imprecision in discussions.

We discuss a very technical - but also general - characterization of "identification" valid for parametric models.

Parametric models are those described by a probability distribution (up to the knowledge of parameters) and a sampling scheme. We often use the term "parametric model" to actually refer to semi-parametric ones, in the sense that we do not need necessarily know the probability distribution (even if we know it is there).

Let Y be a vector-valued random variable in $\mathbb{R}^n$ representing the outcome of an experiment, or a realization in the non-experimental context (recall, macroeconomics is by its very nature a non-experimental discipline).

Henceforth, we denote a realization of Y as y

You can generally consider y to represent the **observed data**.

We assume that the distribution of Y has a parametric representation.

Specifically, we assume that given an open set $\mathbb{A} \subset \mathbb{R}^p$ representing the parameter space that describe our theory-driven structure, we associate with each α in $\mathbb{A}$ a continuos probability density function $f(y, \alpha)$ which, except for the parameter α, is known to the econometrician.

Thus, in this framework the problem of distinguishing between structures (models) is reduced to the problem of distinguishing between parameter points.

We have the following definitions:

Definition 1: Two parameter points (structures) α^1 and α^2 are said to be *observationally equivalent* if $f(y, \alpha^1) = f(y, \alpha^2)$ for all y in $\mathbb{R}^n$.

Definition 2: A parameter point α^0 in $\mathbb{A}$ is said to be *identifiable* (or *globally identified*) if there is no other α in $\mathbb{A}$ which is observationally equivalent.

Definition 3: A parameter point α^0 is said to be *locally identifiable* if there exists an open neighborhood of α^0 containing no other α in $\mathbb{A}$ which is observationally equivalent.

Now we consider the definition of **information matrix** associated with parametric model:

$$I(\alpha) := E\left[\frac{\partial \log f(y,\alpha)}{\partial \alpha} \times \left(\frac{\partial \log f(y,\alpha)}{\partial \alpha}\right)'\right]$$

$$\equiv E\left[\underbrace{\frac{\partial \log f(y,\alpha)}{\partial \alpha}}_{p\times 1} \times \underbrace{\left(\frac{\partial \log f(y,\alpha)}{\partial \alpha}\right)'}_{1\times p}\right]$$

which is $p \times p$ and **symmetric**
(indeed, for $s(y,\alpha) := \frac{\partial \log f(y,\alpha)}{\partial \alpha}$,
$I(\alpha)' = \left(E\left[s(y,\alpha)s(y,\alpha)'\right]\right)'$
$= E\left[s(y,\alpha)s(y,\alpha)'\right] = I(\alpha)$.

Our last definition:

Definition 4: The point α^0 in $\mathbb{A}$ is said to be a *regular point* of the information matrix $I(\alpha)$ if there exists an open neighborhood of α^0 in which $I(\alpha)$ ha constant rank.

Criteria for local identification

Theorem 1 in Rothenberg (1971, ECMA): Let α^0 be a regular point of $I(\alpha)$. Then α^0 is *locally identifiable* if and only if (iff) $I(\alpha)$ is *nonsingular.*

(It is instructive to look at the proof in the paper: basic elements of calculus/topology needed).

Theorem 1 in Rothemberg (1971, ECMA) offers an implicit method for verifying local identification: **one constructs the information matrix associated with the model and checks its rank locally** (which we will elaborate on). We will rely extensively on this result in later sections of the presentation.

The full-rank condition of the information matrix is **a necessary other than sufficient condition** for identification, which means that if we can establish that $rank[I(\alpha)] < p$ in neighborhood of α^0, we are aware that given the data y and the distribution $f(y, \alpha)$, we are not in the position to establish whether our inference is on the structure α^0 or the structure $\alpha^1 \neq \alpha^0$.

Local identifiability is the **minimal requirement** for safe inference on α. It does not rule out, however, the possibility that there exists an observationally equivalent point outside the neighborhood of α^0.

Global identification would be ideal since the econometrician in that case knows that their inference is safe almost everywhere in the parameter space. However, in the majority of cases, we have to be satisfied with local identification.

Criteria for global identification

In very exceptional cases, the structure of the matrix $I(\alpha)$ permits the assertion that $I(\alpha)$ is nonsingular "almost everywhere" in $\mathbb{A}$.

In general, it is difficult to obtain global identification results in the absence of additional information.

Typically, when lucky, we may solely rely on **sufficient conditions** for global identification.

For instance, when one knows that $f(y,\alpha)$ is, e.g. Gaussian, on can rely on the next theorem.

Theorem 3 in Rothenberg (1971, ECMA): let $f(y,\alpha)$ be a member of the **exponential family**. If $I(\alpha)$ is nonsingular in a convex set containing $\mathbb{A}$, then every α in $\mathbb{A}$ is globally identifiable.

Note that $f(y, \alpha)$ is a member of the exponential family when:

$$f(y, \alpha) = h(y) \exp \left\{ \eta(\alpha)' C(y) - D(\alpha) \right\}$$

that is

$$\log f(y, \alpha) = \log h(y) + \eta(\alpha)' C(y) - D(\alpha)$$

Outside the exponential family it does not seen possible to prove conditions for global identifiability using only the information matrix.

Note, for example, that the Student-t **does not** belong to the exponential family.

Spoiler: a reduced form VAR with Gaussian innovations u_t:

$$W_t = \Pi X_t + u_t \quad , \quad u_t \sim WNN(0, \Sigma_u)$$

has likelihood function that belongs to the exponential family, which implies that the parameters (Π, Σ_u) are **identified globally**.

Let $\pi = vec(\Pi')$, $\Pi := (\Pi_1, \Pi_2, \ \ldots, \ \Pi_p, \mu)$ be the parameters associated with the VAR dynamics and intercept (constant);

$\sigma_u^+ = vech(\Sigma_u)$ are the "free" covariance and variance parameters.

The vector

$$\delta = \begin{pmatrix} \pi \\ \sigma_u^+ \end{pmatrix} \begin{matrix} g \times 1 \\ m \times 1 \end{matrix}$$

contains $f = g + m$ elements; $g = Mk = M + M^2 p$ ($k = Mp + 1$) is the number of free autoregressive coefficients in $\Pi := (\Pi_1, \Pi_2, \ \ldots, \ \Pi_p, \mu)$ (including the constant) and $m = \frac{1}{2}M(M+1)$ is the number of free variance-covariance elements in Σ_u.

Theorem 3 in Rothenberg (1971, ECMA) can be applied for $f(w, \delta)$.

VARs are globally identified !

To further see this, it can be proved (see Supplementary Material) that in the VAR with Gaussian innovations above, the Information Matrix reads:

$$I_T(\delta) = T \begin{pmatrix} (\mathbf{\Sigma}_u^{-1} \otimes V_{XX}) & 0_{g \times m} \\ 0_{m \times g} & \frac{1}{2} D_M' (\mathbf{\Sigma}_u^{-1} \otimes \mathbf{\Sigma}_u^{-1}) D_M \end{pmatrix}$$

where $V_{XX} := E(X_t X_t') < \infty$ is $k \times k$ positive definite (hence nonsingular) and D_M is **the duplication matrix**, i.e. the $M^2 \times m$ matrix (of full column rank m) that solves:

$$D_M vech(\mathbf{\Sigma}_u) = vec(\mathbf{\Sigma}_u).$$

It turns out that:

$$rank[I_T(\delta)] = rank[\mathbf{\Sigma}_u^{-1} \otimes V_{XX}]$$

$$+ rank[\frac{1}{2} D'_M(\mathbf{\Sigma}_u^{-1} \otimes \mathbf{\Sigma}_u^{-1}) D_M] = g + m = f$$

and this result is valid almost everywhere in the parameter space of σ_u^+ (i.e. the space of models with positive definite covariance matrix).

Structure and reduced form

In many situations of interest, including our own, it will often be the case that the econometrician does not have direct knowledge of $f(y, \alpha)$.

Instead, they know that the distribution of the data, Y, depends on the p structural parameters in α only through a set of r reduced form parameters in ψ, where $r \geq p$.

That is, we know there exists a mapping of the form:

$$\underbrace{\psi}_{r \times 1} = g(\alpha)$$

where $g(\cdot)$ is a continuos differentiable function, and the econometrician knows the probability distribution function $f^*(y, \psi)$ (up to parameters values).

Hence we have:

$$\underbrace{f^*(y, \psi)}_{\text{known to the econom. (up to } \psi)} = f^*(y, g(\alpha)) = \underbrace{f(y, \alpha)}_{\text{of interest}} .$$

Intuitive argument: imagine that the econometrician knows that ψ is identified globally (e.g. $\psi = \delta$ for VARs).

It means they can safely make inference on ψ (δ) using the data.

Then, it would be natural asking the question: given that the data inform us on ψ and given the mapping

$$\psi = g(\alpha)$$

can we recover α from the knowledge of ψ by exploiting the function $g(\cdot)$?

The question can be rephrased as: can we sort of "invert" the mapping above and recover α from ψ, indirectly benefiting from the knowledge on the reduced form parameters?

To answer questions of this type one can rely on the **Implicit Function Theorem.**

I will try to skip it (see Supplementary Material).

Let

$$J(\alpha) := \frac{\partial g(\alpha)}{\partial \alpha'} \qquad\qquad r \times p$$

be **the Jacobian matrix** associated with the mapping above.

The next theorem provides us with an implicit solution to "the inversion" problem of using the knowledge of the reduced form to recover the structure.

Theorem 6 in Rothenberg (1971, ECMA) (adapted): Assume that ψ is globally identified. If α^0 is a regular point of $J(\alpha)$, then α^0 is *locally identifiable* if and only if (iff) $J(\alpha^0)$ has column rank p.

Thus, the conditions

$$rank[J(\alpha)] = p \quad \text{in a neighborhood of } \alpha^0$$

is a necessary and sufficient condition for (local) identifiability.

It follows that necessary order condition for identification is

$$r \geq p$$

(indeed, if $r < p$, it would be impossible to recover the structural parameters α from the reduced form parameters ψ).

Theorem 6 in Rothenberg (1971, ECMA) is extremely useful to discuss the identifiability of SVARs and we shall use it in this course intensively.

Why?

From the next section onwards we'll face situations represented by mappings of the form:

$$\sigma_u^+ = g(\gamma)$$

where:

$\sigma_u^+ = vech(\mathbf{\Sigma}_u)$ are the m "free" VAR covariance and variance parameters;

γ are p free non-zero structural parameters contained in the matrix H, where H enters expressions like:

$$\mathbf{\Sigma}_u = HH' \quad , \quad H = H(\gamma)$$

so that $g(\gamma) = vech(HH')$.

By implicitly relying on the **Theorem 6** in Rothenberg (1971, ECMA), in this course we derive **the necessary and sufficient conditions** for the identification of SVARs by studying the rank properties of the $m \times p$ Jacobian matrix:

$$J(\gamma) = \frac{\partial g(\gamma)}{\partial \gamma'} \qquad m \times p$$

The **necessary order condition** will be $p \leq m$ (the number of structural parameters cannot exceed the number of reduced form parameters that can be easily recovered from the data)

When the rank conditions holds and $p = m$ we say that the SVAR is **exactly identified** ($g(\cdot)$ can be inverted - locally- hence $\gamma = g^{-1}(\sigma_u^+)$).

When the rank conditions holds and $p < m$ we say that the model is **overidentified**. This means that we can still recover γ from σ_u^+, but to do to the m parameters in σ_u^+ must be subject to $m - p$ restrictions.

We know how to test restrictions on reduced form parameters !

6-SVARs, Structural IRFs and the Identification of SVARs

As seen, we call VAR models **reduced forms.**

A reduced form model is a model (a statistical-like model) that is **identified globally** and can always be estimated from the data.

Economic theory provides either **simultaneous relationship** between variables and/or **on which variables are contemporaneously affected by the shocks of interest**.

When we introduce simultaneous relationships, we switch from VARs to SVARs.

The identification of SVARs is no longer automatic but, when possible, we frame it within Rothenberg's (1971) approach

Consider the VAR model:

$$\Pi(L)W_t = \mu + u_t \quad , \quad \Sigma_u = E(u_t u_t').$$

We know how to estimate the parameters (Π, Σ_u) and we know the properties of the estimators under regularity conditions.

There are **two main ways** by which we can switch from a VAR to a SVAR.

Conventionally, and for teaching purposes, we distinguish between "**B-model**" and "**A-model**" or, B-SVAR and A-SVAR.

B-model

In the B-model, we endow the VAR with an additional set of equations in which the VAR innovations u_t are expressed as linear combination of a vector of **latent (not directly observed) structural shocks** ε_t:

$$u_t = B\varepsilon_t.$$

Here ε_t is $M \times 1$ and such that $E(\varepsilon_t) = 0_{M\times 1}$, $E(\varepsilon_t\varepsilon_t') = I_M$. This relationship **maps the structural shocks to the VAR disturbances** through the nonsingular matrix B.

We call B the **matrix of structural parameters**. Actually, under a proper set of identification restrictions (that we discuss below), the matrix B captures the instantaneous impact of the structural shocks on the the variables.

In the B-model:

- ε_t is the vector of **structural shocks** and satisfies by construction the properties 1 and 2 of our definition, $E(\varepsilon_t) = 0_{M\times 1}$, $Var(\varepsilon_t) = E(\varepsilon_t\varepsilon_t') = I_M$.

- B is an $M \times M$ **nonsingular (full rank)** and contains parameters that maps the structural shocks ε_t onto the VAR innovations u_t (but also on the variables W_t, see VMA representation below).

-B is the matrix of **structural parameters** (or **on-impact coefficients**).

This matrix can not be "any matrix" featuring M^2 free elements but must be properly restricted to ensure identification.

Given the VMA representation of the VAR:

$$W_t = v + \sum_{h=0}^{\infty} C_h u_{t-h}$$

and the **B-model:**

$$u_t = B\varepsilon_t,$$

we now posit that B incorporates the restrictions that ensure identification.

The implied **SVMA** representation is:

$$W_t = v + \sum_{h=0}^{\infty} C_h B \varepsilon_{t-h}$$

Thus, the **structural IRFs** (to one-standard deviation) shocks are:

$$\underbrace{IRF_{\bullet,\bullet}(h)}_{M\times M} := \frac{\partial W_t}{\partial \varepsilon'_{t-h}} \equiv \frac{\partial W_{t+h}}{\partial \varepsilon'_t} = C_h B = \left(R\,(\mathcal{C})^h\,R'\right)B \quad ,$$

Let e_i be the $M \times 1$ vector containing "1" in the position i and zero elsewhere.

The **structural IRF** of one-standard deviation shock to variable j on variable i is defined by:

$$IRF_{i,j}(h) = e_i' C_h B e_j = \underbrace{e_i' C_h}_{\text{row } i} \underbrace{B e_j}_{\text{column } j}$$

and quantifies the impact of shock.

Accordingly,

$$\underbrace{IRF_{\bullet,j}(h)}_{M \times 1} = C_h \underbrace{B e_j}_{\text{column } j} = C_h b_j \equiv \left(R \left(\mathcal{C} \right)^h R' \right) b_j$$

captures the impact of one-standard deviation shock to variable j to all M variables.

The **estimation** of $IRF_{i,j}(h)$ ($IRF_{\bullet,j}(h)$), for $h = 0, 1, .., h_{\text{max}}$ and the **quantification of the uncertainty** surrounding the point estimates is one important task of our course.

A-model

In the A-model, we **pre**-multiply system

$$W_t = \mu + \Pi_1 W_{t-1} + ... + \Pi_p W_{t-p} + u_t$$

by the **nonsingular** $M \times M$ matrix A, obtaining

$$AW_t = \tau + \Upsilon_1 W_{t-1} + ... + \Upsilon_p W_{t-p} + Au_t$$

where $\tau = A\mu$, $\Upsilon_i = A\Pi_i$, $i = 1, ..., p$.

Here Au_t represents a linear combination of VAR innovations. Then, we endow the model by the additional set of equations

$$Au_t = \varepsilon_t$$

where, again, ε_t is $M \times 1$ and such that $E(\varepsilon_t) = 0_{M\times 1}$, $E(\varepsilon_t \varepsilon_t') = I_M$ and is the vector of **structural shocks.**

The relationship $Au_t = \varepsilon_t$ defines a **system of simultaneous equations** in the variables u_t.

Simultaneous systems of equations are, in a sense, DSGE models without expectations terms.

They have been the counterpart of today's DSGE models until the eighties.

The A-model allows to model the interdependence between the variables u_t (that is, W_t "polished from the dynamics of the system; recall that $u_t := W_t - E(W_t \, |\text{past})$ ") through the matrix A.

In the A-model $Au_t = \varepsilon_t$:

- ε_t is the vector of **orthogonal structural shocks** and satisfies by construction the properties 1 and 2 of our definition, $E(\varepsilon_t) = 0_{M\times 1}$, $Var(\varepsilon_t) = E(\varepsilon_t \varepsilon_t') = I_M$.

- A is an $M \times M$ **nonsingular (full rank)** matrix of parameters that captures the simultaneous relationship that characterize the variables in W_t.A is called matrix of **structural parameters** and must be properly selected according to some rules we have to study.

- By inverting the A matrix we obtain the **B-model implied by the A-model**: $B = A^{-1}$.

- Conversely, by inverting the B matrix we obtain the **A-model implied by the B-model**: $A = B^{-1}$.

Given the VMA representation of the VAR:

$$W_t = v + \sum_{h=0}^{\infty} C_h u_{t-h}$$

and the A-model

$$Au_t = \varepsilon_t$$

we now posit that A incorporates the restrictions that ensure identification.

The implied **SVMA** representation is:

$$W_t = v + \sum_{h=0}^{\infty} C_h A^{-1} \varepsilon_{t-h}$$

with implied **structural** IRFs:

$$IRF_{\bullet,\bullet}(h) = \frac{\partial W_t}{\partial \varepsilon'_{t-h}} = \frac{\partial W_{t+h}}{\partial \varepsilon'_t} = C_h A^{-1} \equiv \left(R\,(\mathcal{C})^h\, R' \right) A^{-1}$$

Identification issues

We take the B-model as reference (set $A = B^{-1}$ for the A-model).

We show that in the B-model (A-model) the matrix B (A) can not be "any matrix" but must be properly selected by respecting some rules.

The rules to be respected will be called:

- **necessary order condition (**the # of structural parameters cannot exceed the # of reduced form parameters**)**

- **necessary and sufficient rank condition** (which involves the rank of a Jacobian matrix as we have met in Rothenberg's theory).

Take VAR model:

$$\Pi(L)W_t = \mu + u_t$$

add structural equations

$$\begin{aligned} u_t &= B\varepsilon_t \\ E(\varepsilon_t) &= 0_{M\times 1} \quad , \quad E(\varepsilon_t\varepsilon_t') = I_M \end{aligned}$$

From system $u_t = B\varepsilon_t$ we obtain:

$$u_t u_t' = B\varepsilon_t(B\varepsilon_t)'$$

$$= B\varepsilon_t\varepsilon_t' B'.$$

By taking expectations and using $\Sigma_u = E(u_t u_t')$ and $I_M = E(\varepsilon_t\varepsilon_t')$, we obtain the **second-order moment conditions:**

$$\Sigma_u = BB'.$$

We'll soon see that the $M \times M$ matrix B cannot contain M^2 free parameters.

Let β the $b \times 1$ vector of free parameters contained in B.

We may use the notation: $B = B(\beta)$ to indicate that the nonzero and free elements of B are in the vector β.

Then, from the second-order moment conditions:

$$\mathbf{\Sigma}_u = BB'$$

apply the $vech(\cdot)$ operator to both sides, obtaining:

$$\underbrace{vech(\mathbf{\Sigma}_u)}_{\sigma_u^+} = \underbrace{vech(BB')}_{g(\beta)}.$$

Thus, in our SVAR we have a function mapping that links reduced form **(those in Σ_u)** and structural parameters β **(contained in** B):

$$\sigma_u^+ = g(\beta)$$

where $g(\cdot)$ is nonlinear but continuos and differentiable.

From Rothenberg's theory, we have all the ingredients to understand that the identification problem of our SVAR now amounts to problem **of uniquely recovering the** b elements in β **from the** m elements in σ_u^+.

A straightforward application of **Theorem 6** in Rothenberg (1971, ECMA) suggests that **necessary order condition for identification** will be:

$$p = \mathsf{dim}(\beta) \leq m = \underbrace{\frac{1}{2}M(M+1)}_{\#\text{ free elements in } \Sigma_u \ (\sigma_u^+)},$$

while the **necessary and sufficient rank condition** will be:

$$rank\left[J_B(\beta)\right] = p$$

in a neighborhood of β_0 ($J_B(\beta)$ regular), where

$$J_B(\beta) := \underbrace{\frac{\partial g(\beta)}{\partial \beta'}}_{m \times p}.$$

To compute the Jacobian matrix $J_B(\beta)$ algebraically, we need to find one way to formalize the link between the matrix B and its free elements in the vector β.

We consider the **explicit form representation** of linear restrictions on B **:**

$$vec(B) = S_B\beta + s_B$$

where S_B is an $M^2 \times b$ selection matrix and s_B is $M^2 \times 1$ and contains known elements (e.g. "0" or "1" etc.).

Example of **identifying restrictions on** B.

Consider ($M = 3$)

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

and the three restrictions $b_{21} = 0$, $b_{32} = 0$ and $b_{13} = 0.5$.

The matrix B under the restrictions is

$$B = \begin{pmatrix} b_{11} & b_{12} & 0.5 \\ 0 & b_{22} & b_{23} \\ b_{31} & 0 & b_{33} \end{pmatrix}$$

and we realize that the vector of unrestricted (free) elements of B is $\beta = (b_{11}, b_{31}, b_{12}, b_{22}, b_{23}, b_{33})'$, hence $B = B(\beta)$.

The restrictions $vec(B) = S_B\beta + s_B$ correspond to

$$\begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \\ b_{12} \\ b_{22} \\ b_{32} \\ b_{13} \\ b_{23} \\ b_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_{11} \\ b_{31} \\ b_{12} \\ b_{22} \\ b_{23} \\ b_{33} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.5 \\ 0 \\ 0 \end{pmatrix}$$

By applying some matrix differential calculus rules (the non-interested student can skip the next three lines and go directly to the fourth) the Jacobian $J_B(\beta)$ is given by:

$$J_B(\beta) = \frac{\partial \sigma_u^+}{\partial \beta'} = \frac{\partial g(\beta)}{\partial \beta'} = \frac{\partial vech(B(\beta)B(\beta)')}{\partial \beta'}$$

$$= \frac{\partial vech(BB')}{\partial vec(B)'} \times \frac{\partial vec(B)}{\partial \beta'} = \frac{\partial vech(BB')}{\partial vec(B)'} \times S_B$$

$$= D_M^+ \frac{\partial vec(BB')}{\partial vec(B)'} \times S_B$$

$$= D_M^+ \left\{ 2N_M(B \otimes I_M) \right\} \times S_B$$

$$= 2D_M^+(B \otimes I_M)S_B$$

where D_M^+ is the $m \times M^2$ Moore-Penrose pseudo-inverse of the Duplication matrix (D_M), defined by $D_M^+ := (D_M^{'} D_M)^{-1} D_M'$

We have derived the necessary order condition by counting the # of free parameters in B.

Alternatively, we can count the # **of restrictions to impose on** B.

Let l be the number of restrictions we need to place on B.

Then, $l = M^2 - b$.

Therefore, from the inequality $b \leq \frac{1}{2}M(M+1) = m$ we obtain;

$$l = M^2 - b \geq M^2 - \frac{1}{2}M(M+1) = \frac{1}{2}M(M-1)$$

so that the necessary order condition for identification can be equivalently rephrased as the condition

$$l \geq \frac{1}{2}M(M-1)$$

which can be interpreted by observing that **one must necessarily place at least $\frac{1}{2}M(M-1)$ restrictions on the elements of the matrix B.**

Observe that the moment conditions $\Sigma_u = BB'$ arising from the symmetry of the covariance matrix impose some restrictions on B !

Specifically, the symmetry of $\Sigma_u = BB'$ places $\frac{1}{2}M(M+1)$ "natural" restrictions on the elements of B!

Hence, there are

$$\underbrace{M^2}_{\#\text{ total elements in } B} - \underbrace{\frac{1}{2}M(M+1)}_{\#\text{ restrictions implied by symmetry}}$$

$$= \underbrace{\frac{1}{2}M(M-1)}_{\#\text{ param. in } B \text{ to be deter}}$$

elements in B which are not directly recoverable from Σ_u!

Summing up, we have the following **result on the identification of the B-model**.

Consider the SVAR model based on

$$u_t = B\varepsilon_t \quad , \quad \mathbf{\Sigma}_u = BB'$$

with l ($l = M^2 - b$) linear restrictions on B written in the form

$$vec(B) = S_B\beta + s_B.$$

Then, **necessary order condition for identification** is

$$b \leq \frac{1}{2}M(M+1) \text{ or, alternatively, } l \geq \frac{1}{2}M(M-1).$$

Necessary and sufficient rank condition for **local** identification is

$$rank\left\{2D_n^+(B_0 \otimes I_M)S_B\right\} = b$$

where $B_0 = B(\beta_0)$.

Exactly identified SVAR. If for a specified matrix $B = B(\beta)$ the necessary and sufficient rank condition is satisfied and the necessary order condition is satisfied with $b = \frac{1}{2}M(M+1)$ (or alternatively $l = \frac{1}{2}M(M-1)$), then the SVAR is said to be **exactly identified**.

Overidentified SVAR. If for a specified matrix $B = B(\beta)$ the necessary and sufficient rank condition is satisfied and the necessary order condition is satisfied with $b < \frac{1}{2}M(M+1)$ (or alternatively $l > \frac{1}{2}M(M-1)$), then the SVAR is said to be **overidentified** with $\frac{1}{2}M(M+1) - b$ $(l - \frac{1}{2}M(M-1))$ overidentification restrictions.

The special feature of overidentified SVARs is that they **can be tested against the data**, meaning that we have a test of their empirical validity.

As the next Example will shown, overidentification restrictions in a SVAR reflect in restrictions on the reduced form parameters.

After we study the estimation of SVAR, we'll see an easy way to test overidentification restrictions.

Example, bi-variate SVAR. Assume $M = 2$ and

$$\begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \quad \begin{array}{l} \text{shock1} \\ \text{shock2} \end{array} .$$

$\mathbf{\Sigma}_u = BB'$ corresponds to

$$\begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix}$$

$$= \begin{pmatrix} b_{11}^2 + b_{12}^2 & b_{11}b_{21} + b_{12}b_{22} \\ b_{11}b_{21} + b_{12}b_{22} & b_{21}^2 + b_{22}^2 \end{pmatrix}$$

so that

$$\begin{aligned} \sigma_1^2 &= b_{11}^2 + b_{12}^2 \\ \sigma_{1,2} &= b_{11}b_{21} + b_{12}b_{22} \\ \sigma_2^2 &= b_{21}^2 + b_{22}^2 \end{aligned}$$

and it is clear that it is impossible to recover the four parameters b_{11}, b_{12}, b_{21}, b_{21} from the three parameters σ_1^2, $\sigma_{1,2}$, σ_2^2.

In this case we say that the model is not identified (is unidentified or underidentified) because there are infinitely many observational equivalent points in the parameters space.

We also say that **the shocks are not identified (unidentification)** because we are not able to assess, on the basis of the specified matrix B, what are the distinguishing features of the shock $\varepsilon_{1,t}$ (shock1) and those of the shock $\varepsilon_{2,t}$ (shock2).

However, imagine that, based on economic reasoning (or because we have in mind a precise theory), we can defend the restriction $b_{12} = 0$, which implies that the structural shock $\varepsilon_{2,t}$ does not affect contemporaneously $u_{1,t}$ ($W_{1,t}$) (while $\varepsilon_{1,t}$ affects contemporaneously both $u_{1,t}$ ($W_{1,t}$) and $u_{2,t}$ ($W_{2,t}$)).

We have:

$$B = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix}$$

$$\begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} b_{11}^2 & b_{11}b_{21} \\ b_{11}b_{21} & b_{21}^2 + b_{22}^2 \end{pmatrix}$$

so that

$$\begin{aligned} \sigma_1^2 &= b_{11}^2 \\ \sigma_{1,2} &= b_{11}b_{21} \\ \sigma_2^2 &= b_{21}^2 + b_{22}^2. \end{aligned}$$

By solving this system of three equations and three variables we can recover the non-zero elements of B as function of the elements in $\mathbf{\Sigma}_u$:

$$\begin{aligned} b_{11} &= \sigma_1 \\ b_{21} &= \sigma_{1,2}/\sigma_1 \\ b_{22} &= \left(\sigma_2^2 - (\sigma_{1,2})/\sigma_1^2\right)^{1/2}. \end{aligned}$$

In this case, we have **exact identification** because we can **uniquely** recover the three structural parameters b_{11}, b_{21} and b_{22} (recall that $b_{12} = 0$) from three reduced form parameters σ_1^2, $\sigma_{1,2}$, σ_2^2 and **the number of structural parameters is equal to the number of reduced form parameters.** Note that with $b_{12} = 0$, the specified matrix B **corresponds to the Choleski factor of Σ_u**!

Imagine further that we also know, from economic reasoning, that $b_{21} = (1/3)$, i.e. we know that $\varepsilon_{1,t}$ affects contemporaneously both $u_{1,t}$ $(W_{1,t})$ and $u_{2,t}$ $(W_{2,t})$ and that the instantaneous impact on $u_{2,t}$ $(W_{2,t})$ is equal to 1/3. In this case

$$B = \begin{pmatrix} b_{11} & 0 \\ 1/3 & b_{22} \end{pmatrix}$$

so that

$$\begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} b_{11}^2 & b_{11}(1/3) \\ b_{11}(1/3) & (1/3)^2 + b_{22}^2 \end{pmatrix}.$$

By solving the model we have

$$b_{11} = \sigma_1$$

and

$$b_{22} = \left(\sigma_2^2 - (1/9)\right)^{1/2}.$$

Here we have **overidentification** because we can still recover the two structural parameters b_{11} and b_{22} (recall b_{21} is fixed to $1/3$) from the three reduced form parameters σ_1^2, $\sigma_{1,2}$, σ_2^2, and **the number of structural parameters (2) is less than the number of reduced form parameters (3)**. In this case, we have a model which incorporates 3-2=1 testable overideintification restriction. To understand this fact note that in this case the reduced form parameters σ_1^2, $\sigma_{1,2}$, σ_2^2 are not free in the sense that the restrictions $b_{12} = 0$ and $b_{21} = (1/3)$ imply the constraint

$$\sigma_1 = 3\sigma_{1,2}$$

which involves Σ_u.

The overidentification constraint is testable from the data (e.g. by a Wald-type test) and allows us to establish empirically whether the two identification restrictions $b_{12} = 0$ and $b_{21} = (1/3)$ are supported by the data!

How do we check the identification of a SVAR model?

The necessary order condition can be easily checked: we simply count the number of restrictions placed on B and check whether $l \geq 1/2M(M-1)$.

But if only the necessary condition holds, we can not yet say that the model is identified.

In order to check the validity of the necessary and sufficient rank condition for identification, there are two possible routes.

One is to generate randomly many values $\tilde{\beta}$ of β (e.g. from the $U(0,1)$-distribution), and then check whether the conditions

$$rank\left\{2D_n^+(\tilde{B}\otimes I_M)S_B\right\} = b$$

is respected at the randomly generated points; notice that here $\tilde{B} = B(\tilde{\beta})$.

Another way is to estimate the SVAR by Maximum Likelihood and, given $\hat{B}_T = B(\hat{\beta}_T)$, one then checks whether

$$rank\left\{2D_n^+(\hat{B}_T\otimes I_M)S_B\right\} = b.$$

It is important to remark that the identification studied so far **holds up to sign normalization of the columns of** B.

What do we mean by that?

Consider the matrix

$$B = \begin{pmatrix} b_{11} & 0 \\ 1/3 & b_{22} \end{pmatrix}$$

taken from an example above. We know that the SVAR based on this matrix B is identified (is actually overidentified as $l = 2 > 1/2M(M-1) = 1$) respects the necessary and sufficient rank condition. In this case

$$\mathbf{\Sigma}_u = BB' = \begin{pmatrix} b_{11}^2 & \frac{1}{3}b_{11} \\ \frac{1}{3}b_{11} & b_{22}^2 + \frac{1}{9} \end{pmatrix}.$$

But now consider the matrix

$$B_s = \left(\begin{array}{cc} -b_{11} & 0 \\ -1/3 & b_{22} \end{array} \right)$$

which is obtained from B by simply changing the sign of the coefficients in the first column (notice that formally $B \neq B_s$).

Also B_s is identified (is overidentified) as it respects the necessary and sufficient rank condition.

But

$$\Sigma_u = B_s B_s' = \left(\begin{array}{cc} b_{11}^2 & \frac{1}{3} b_{11} \\ \frac{1}{3} b_{11} & b_{22}^2 + \frac{1}{9} \end{array} \right) = BB'.$$

The example above tells us that our identification rules allow to identify B **up to sign normalization.** Accordingly, also the IRFs will depend on the sign of the columns of B (e.g. recall that $IRF_{\bullet,\bullet}(0) = B$).

This means that once we get an estimate $\hat{B}_T = B(\hat{\beta}_T)$ (see below) we can "normalize" the sign of the columns without affecting the identification of the model

A typical normalization practitioners use in applied work **is to set the element on the diagonal of B ($\hat{B}_T$) to positive.**

Relation between identification and the theory of rotation matrices

Example Consider

$$B = \begin{pmatrix} 1/2 & 1 & 0 \\ 2 & 1/3 & 0 \\ 1/4 & -3/4 & 1 \end{pmatrix}$$

where we know we have unidentification.

Here, $\mathbf{\Sigma}_u = BB'$ is given by

$$\begin{aligned} \mathbf{\Sigma}_u &= \begin{pmatrix} 1/2 & 1 & 0 \\ 2 & 1/3 & 0 \\ 1/4 & -3/4 & 1 \end{pmatrix} \times \begin{pmatrix} \frac{1}{2} & 2 & \frac{1}{4} \\ 1 & \frac{1}{3} & -\frac{3}{4} \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{5}{4} & \frac{4}{3} & -\frac{5}{8} \\ \frac{4}{3} & \frac{37}{9} & \frac{1}{4} \\ -\frac{5}{8} & \frac{1}{4} & \frac{13}{8} \end{pmatrix}. \end{aligned}$$

Now consider the new matrix B^* selected as

$$
\begin{aligned}
B^* &= BO = \begin{pmatrix} 1/2 & 1 & 0 \\ 2 & 1/3 & 0 \\ 1/4 & -3/4 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{3} & 2 & 0 \\ -\frac{3}{4} & \frac{1}{4} & 1 \end{pmatrix} \neq B = \begin{pmatrix} 1/2 & 1 & 0 \\ 2 & 1/3 & 0 \\ 1/4 & -3/4 & 1 \end{pmatrix}.
\end{aligned}
$$

where O is **an orthogonal matrix,** i.e. such that $O'O = I_3 = OO'$.

Then

$$\boldsymbol{\Sigma}_u = B^* B^{*\prime}$$

$$\boldsymbol{\Sigma}_u = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{3} & 2 & 0 \\ -\frac{3}{4} & \frac{1}{4} & 1 \end{pmatrix} \times \begin{pmatrix} 1 & \frac{1}{3} & -\frac{3}{4} \\ \frac{1}{2} & 2 & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{5}{4} & \frac{4}{3} & -\frac{5}{8} \\ \frac{4}{3} & \frac{37}{9} & \frac{1}{4} \\ -\frac{5}{8} & \frac{1}{4} & \frac{13}{8} \end{pmatrix}$$

It turns out that B and B^* give rise to the same reduced form covariance matrix $\boldsymbol{\Sigma}_u$. This means that both "theories" for shocks identification based on B and B^* are consistent with the data. In other words, the shocks ε_t and ε_t^* obtained from

$$\varepsilon_t = B^{-1} u_t \quad , \quad \varepsilon_t^* = (B^*)^{-1} u_t^*$$

produce the same impact (the same outcomes) on the variables. Instead we are looking for a **unique** B!

How do we solve the problem? Imagine that we qualify the matrix B by imposing that its upper triangle must be exactly equal to the number 1, 0 and 0 (in bold):

$$B = \begin{pmatrix} 1/2 & \mathbf{1}_{fixed} & \mathbf{0}_{fixed} \\ 2 & 1/3 & \mathbf{0}_{fixed} \\ 1/4 & -3/4 & 1 \end{pmatrix}.$$

In this case, there exists only one orthogonal matrix O ($O'O = I_3 = OO'$) such that $\mathbf{\Sigma}_u = BB' = B^*B^{*\prime}$, and this matrix is the **identity matrix**! We obtain identification! **Thus we can achieved identification by putting restrictions on B that make B unique and directly recovered from $\mathbf{\Sigma}_u = BB'$.**

7 - AB-SVARs: specification, identification and estimation

The AB-model mixes features of both the A-model and the B-model.

The SVAR "AB-model" is given by the specification

$$AW_t = \tau + \Upsilon_1 W_{t-1} + ... + \Upsilon_p W_{t-p} + Au_t$$

$$\begin{aligned} Au_t &= B\varepsilon_t \\ E(\varepsilon_t) &= 0_{M\times 1},\ E(\varepsilon_t\varepsilon_t') = I_M \end{aligned}$$

where A is the $M \times M$ **invertible** matrix of structural parameters **which captures the simultaneous relationships between the variables** and B is the $M \times M$ **invertible** matrix **which captures the instantaneous impact of the structural shocks on** Au_t (not u_t!)

Observe that

$$u_t = A^{-1}B\varepsilon_t = C\varepsilon_t$$

so that $C = A^{-1}B$ captures the true instantaneous impact of the structural shocks on the variables W_t.

The structure of C depends on the structural specification of A and B, hence C might not contain zeros despite A and or B do.

In other words, a trade-off in the AB-model is that while practitioners must specify both A and B for identification (see below), the advantage is that the final impact of the structural shocks on the variables, as captured by the elements of C, tends to have 'very few' zeros, if any at all.

Implied moment conditions:

$$A\mathbf{\Sigma}_u A' = BB' \;\Rightarrow \mathbf{\Sigma}_u = A^{-1}BB'(A')^{-1}.$$

Now, the symmetry of $\mathbf{\Sigma}_u$ places $\frac{1}{2}M(M+1)$ restrictions on the $2M^2$ elements of A and B, so that there are still at least $2M^2 - \frac{1}{2}M(M+1)$ restrictions that must be imposed on these matrices to achieve identification.

We need identification restrictions on both A and B.

Note that with $A := I_M$ model collapses to the B-model

Note that with $B := I_M$ model collapses to the A-model.

Linear restrictions on A and B :

$$\begin{aligned} vec(A) &= S_A\alpha + s_\alpha \\ vec(B) &= S_B\beta + s_\beta \end{aligned}$$

where S_A is an $M^2 \times b_A$ selection matrix, α is the $b_A \times 1$ vector containing the unrestricted (free) elements in A and, finally, s_A is $M^2 \times 1$ and contains known elements; S_B is an $M^2 \times b_B$ selection matrix, β is the $b_B \times 1$ vector containing the unrestricted (free) elements in B and, finally, s_B is $M^2 \times 1$ and contains known elements.

Compactly, we can jointly represent identifying restrictions on A and B;

$$\begin{pmatrix} vec(A) \\ vec(B) \end{pmatrix} = \begin{pmatrix} S_A & 0 \\ 0 & S_B \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} s_A \\ s_B \end{pmatrix}$$

and this shows that the vector of (free) unrestricted structural parameters in the AB-model is given by the vector

$$\gamma = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{matrix} b_A \times 1 \\ b_B \times 1 \end{matrix} \quad b \times 1 \quad , \quad b = b_A + b_B.$$

We immediately understand that **necessary order condition** for identification is that $b = b_A + b_B \leq m = (1/2)M(M+1)$ or, alternatively, $l = l_A + l_B \geq 2M^2 - \frac{1}{2}M(M+1)$ where l_A is the number of restrictions on A, and l_B the number of restrictions on B.

To discuss **necessary and sufficient** rank condition for identification, we realize that

$$\mathbf{\Sigma}_u = \underbrace{A^{-1}B}_{C}\underbrace{B'(A')^{-1}}_{C'}$$

and the restrictions above imply the mapping:

$$\sigma_u^+ = g(\gamma) \equiv vech\left[A^{-1}BB'(A')^{-1}\right]$$

so that the focus will be on the $m \times b$ Jacobian matrix:

$$J_{A,B}(\gamma) = \frac{\partial \sigma_u^+}{\partial \gamma'} = \left(\begin{array}{ccc} \frac{\partial \sigma_u^+}{\partial \alpha'} & \vdots & \frac{\partial \sigma_u^+}{\partial \beta'} \end{array} \right)$$

Matrix computations show that:

$$\begin{aligned} J_{A,B}(\gamma) &= 2D_M^+ \Big(-(\boldsymbol{\Sigma}_u^{AB} \otimes A^{-1}) \ \vdots \ (A^{-1}B \otimes A^{-1}) \Big) \\ &\qquad \times \begin{pmatrix} S_A & 0_{M^2 \times b_B} \\ 0_{M^2 \times b_A} & S_B \end{pmatrix} \end{aligned}$$

where $\boldsymbol{\Sigma}_u^{AB} = A^{-1}BB'(A')^{-1}$ is the error covariance matrix subject to the restrictions.

We have the necessary and sufficient rank condition

$$rank\left[J_{A,B}(\gamma_0)\right] = b_A + b_B.$$

Summing up, we have the following **result on the identification** of the AB-model.

Consider the SVAR model based on

$$Au_t = B\varepsilon_t \quad , \quad \mathbf{\Sigma}_u = A^{-1}BB'(A')^{-1}$$

with $l = l_A + l_B$ ($l = M^2 - b$) linear restrictions on A and B of the form

$$\begin{pmatrix} vec(A) \\ vec(B) \end{pmatrix} = \begin{pmatrix} S_A & 0 \\ 0 & S_B \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} s_A \\ s_B \end{pmatrix}.$$

Then, **necessary order condition for identification** is

$$b = b_A + b_B \leq \frac{1}{2}M(M+1) \quad \text{or alternatively} \quad l = l_A + l_B \geq 2$$

Necessary and sufficient rank condition for **local** identification is

$$rank\left\{2D_M^+\left(-(\boldsymbol{\Sigma}_u^{A_0B_0}\otimes A_0^{-1}) \;\; \vdots\; (A_0^{-1}B_0\otimes A_0^{-1}) \right)\right.$$
$$\left.\times\left(\begin{array}{cc} S_A & 0_{n^2\times b_B} \\ 0_{n^2\times b_A} & S_B \end{array}\right)\right\} = b$$

where $A_0 = A(\alpha_0)$, $B_0 = B(\beta_0)$ and $\boldsymbol{\Sigma}_u^{A_0B_0} = A_0^{-1}B_0B_0'(A_0')^{-1}$.

Provided identification holds, the **Structural IRFs** implied by the AB-model are given by

$$IRF_{\bullet,\bullet}(h) = \Phi_h = C_h A^{-1} B = C_h C \quad , \quad h = 0, 1, 2, ...$$

$$= \left(R\,(\mathcal{C})^h\, R' \right) A^{-1} B \quad , \quad h$$

or, if we specifically focus on the dynamic response of the i-th variable to shocks in the j-th variable,

$$IRF_{i,j}(h) = \phi_{i,j,h} = e_i' C_h A^{-1} B e_j \ , \quad h = 0, 1, 2, ...$$

where recall that $A = A(\alpha)$ and $B = A(\beta)$.

We understand that in order to estimate Φ_h it is necessary to estimate both the VMA parameters C_h (and we know how to do this) and the structural parameters γ, i.e. those that enter the matrices $A = A(\alpha)$ and $B = B(\beta)$.

Estimation of A and B

Given

$$\ln L(\underbrace{g(\gamma)}_{\sigma_u^+}) = C - \frac{T}{2}\log[\det(A^{-1}BB'A'^{-1})] - \frac{T}{2}tr[A'B^{-1\prime}B^{-1}A\ \hat{\Sigma}_u]$$

The ML estimator of γ, denoted $\hat{\gamma}_{T,ML}$ is obtained by solving (numerically):

$$\max_{\gamma} \ln L_T(g(\gamma)).$$

The ML estimator $\hat{\gamma}_{T,ML} \equiv \hat{\gamma}_T$ has the typical properties of ML estimators in stationary VARs under standard regularity conditions, i.e.:

$$\hat{\gamma}_T \xrightarrow[T\to\infty]{p} \gamma_0$$

$$T^{1/2}(\hat{\gamma}_T - \gamma_0) \xrightarrow[T\to\infty]{d} N(0_{b\times 1}, V_\gamma) \quad , \quad V_\gamma = (I_\infty(\gamma_0))^{-1}$$

As we know from the theory of ML estimation

$$\hat{V}_\gamma = \left[-\frac{1}{T} H_T(\hat{\gamma}_T)\right]^{-1} \xrightarrow[T\to\infty]{p} (I_\infty(\gamma_0))^{-1} = V_\gamma$$

where

$$H_T(\hat{\gamma}_T) = \left.\frac{\partial^2 \ln L_T(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma=\hat{\gamma}_T}.$$

It turns out that reported standard errors for $\hat{\gamma}_T$ are the square roots of the diagonal elements of the matrix $[-H_T(\hat{\gamma}_T)]^{-1}$.

Any package that performs SVAR estimation typically provides **analytic standard errors** for the estimated coefficients in this way.

Alternatively, from the theory of identification of the AB-model, we have seen that

$$I_T(\gamma) = J_{A,B}(\gamma)' I_T(\sigma_u^+) J_{A,B}(\gamma)$$

where $I_T(\sigma_u^+)$ is the $m \times m$ information matrix associated with the reduced form parameters σ_u^+ of the VAR.

Thus,

$$I_\infty(\gamma) = \lim_{T\to\infty} \frac{1}{T} I_T(\gamma) = J_{A,B}(\gamma)' I_\infty(\sigma_u^+) J_{A,B}(\gamma)$$

and this means that a consistent estimator of $I_\infty(\gamma)$ may be obtained from the expression

$$\hat{I}_\infty(\hat{\gamma}_T) = J_{A,B}(\hat{\gamma}_T)' I_\infty(\hat{\sigma}_{u,T}^+) J_{A,B}(\hat{\gamma}_T)$$

and from this

$$\hat{V}_\gamma = \left(I_\infty(\hat{\gamma}_T)\right)^{-1}.$$

Finally, we have $\hat{\gamma}_T = (\hat{\alpha}'_T, \hat{\beta}'_T)'$ and

$$\begin{aligned} vec(\hat{A}_T) &= S_A \hat{\alpha}_T + s_A \\ vec(\hat{B}_T) &= S_B \hat{\beta}_T + s_B \end{aligned}$$

so that $\hat{A}_T = A(\hat{\alpha}_T)$ is obtained from the "unvec" of $vec(\hat{A}_T)$ and $\hat{B}_T = B(\hat{\beta}_T)$ is obtained from the "unvec" of $vec(\hat{B}_T)$.

Coming back to the structural IRFs:

$$IRF_{\bullet,\bullet}(h) = \Phi_h = C_h A^{-1} B \quad , \quad h = 0, 1, 2, ...$$

these are estimated by

$$\widehat{IRF_{\bullet,\bullet}}(h) = \hat{\Phi}_{h,T} = \hat{C}_{h,T} \left(\hat{A}_T\right)^{-1} \hat{B}_T \quad , \quad h = 0, 1, 2, ...$$

7.1-Case study: Blanchard and Perotti's (2002, QJE) fiscal model

Blanchard and Perotti (2002) (henceforth BP) use a SVAR-AB model to infer the size of the fiscal multipliers in the US.

To do so they estimate the dynamic response of output to exogenous tax and spending shocks.

BP is one of the few but important examples in the literature where an AB-SVAR is estimated.

BP problem is to identify exogenous tax and spending shocks from the estimation of a SVAR-AB model for the US economy.

BP consider the following variables ($M = 3$)

$$W_t = \begin{pmatrix} Tax_t \\ G_t \\ GDP_t \end{pmatrix} \quad \begin{array}{l} \text{log of total tax revenues} \\ \text{log of government spending} \\ \text{log of GDP} \end{array}$$

and in our dataset we have quarterly US data that cover the period 1950Q1-2006Q4 (T=228).

After the reduced form specification analysis of the VAR, the reduced form system is given by

$$\begin{aligned} W_t &= \Pi_1 W_{t-1} + \Pi_2 W_{t-2} + \Pi_3 W_{t-3} \\ &\quad + \Pi_4 W_{t-4} + \mu + \delta t + \lambda D_t^{75Q2} + u_t \end{aligned}$$

i.e. it includes $p = 4$ lags.

The deterministic part of the VAR is given by a constant (μ), a deterministic linear trend (δt) and a deterministic dummy variable D_t^{75Q2} which is equal to 1 in 1975Q2 and zero elsewhere in order to account for a large net tax cut episode, which is defined as a "well-identified, isolated, temporary, tax cut".

Given the reduced form, the AB-form reads

$$
\begin{aligned}
AW_t &= \Upsilon_1 W_{t-1} + \Upsilon_1 W_{t-2} + \Upsilon_1 W_{t-3} \\
&\qquad + \Upsilon_1 W_{t-4} + A\mu + A\delta t + A\lambda D_t^{75Q2} + Au_t
\end{aligned}
$$

$$Au_t = B\varepsilon_t.$$

Recall that since $M = 3$, in the AB-model **we have to place at least**

$$l = l_A + l_B \geq 2M^2 - \frac{1}{2}M(M+1) = 12$$

restrictions on the matrices A and B.

BP baseline specification:

$$\begin{pmatrix} 1 & 0 & -2.08 \\ 0 & 1 & 0 \\ a_{31} & a_{32} & 1 \end{pmatrix} \begin{pmatrix} u_t^{Tax} \\ u_t^{G} \\ u_t^{GDP} \end{pmatrix}$$

$$= \begin{pmatrix} b_{11} & b_{12} & 0 \\ b_{21} & b_{22} & 0 \\ 0 & 0 & b_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_t^{Tax} \\ \varepsilon_t^{G} \\ \varepsilon_t^{GDP} \end{pmatrix}$$

The value $a_{13} = -2.08$ is calibrated by Blanchard and Perotti (2002) from previous studies where it is estimated the elasticity of taxes to output (taxes respond contemporaneously to output because of the business cycle).

BP observe that $b_{12} = 0$ (the spending shock has no instantaneous impact on tax revenues) or $b_{21} = 0$ (the tax shock has no instantaneous impact on government spending) can be equally accepted.

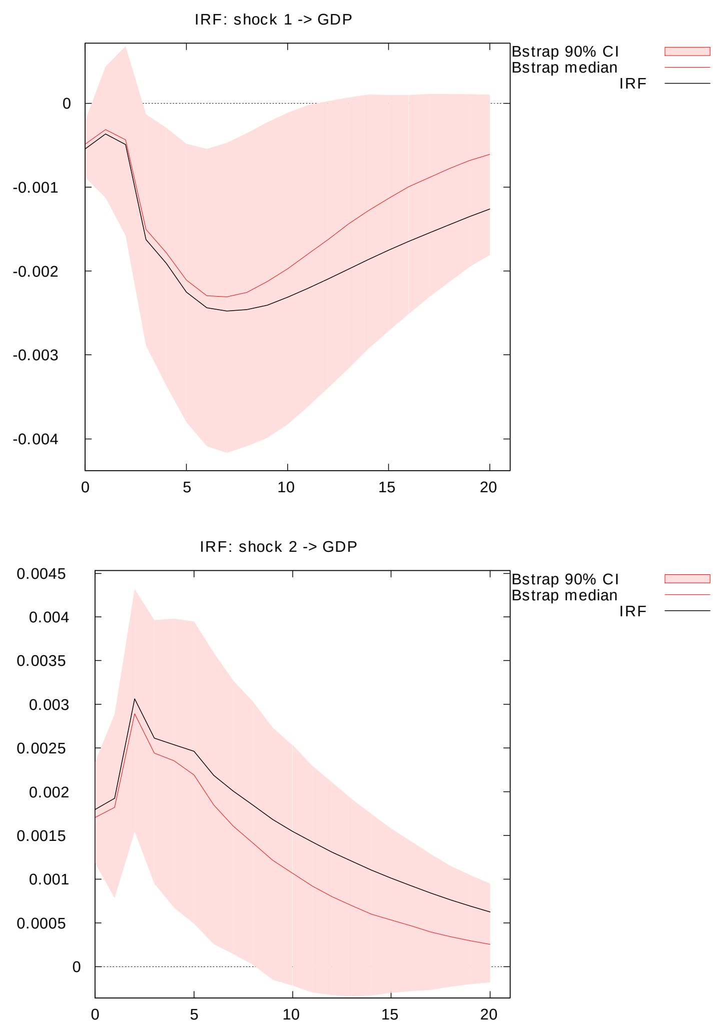
Let us assume that $b_{21} = 0$ and we leave b_{12} "free".

We have a total of $l = 12$ restrictions that identify exactly the SVAR.

One can check that also the necessary and sufficient rank condition is satisfied.

The estimated matrices A and B are obtained from the Gretl script code:

This code produces also the IRFs reported in the next slide.



Responses of GDP to the identified tax shock (upper panel) and government spending shock (lower panel) from Blanchard and Perotti (2002, QJE) specification, quarterly data 1950Q1-2006Q4.

8 - FEVDs: short account

Structural VMA representation ($u_t = K\varepsilon_t$):

$$W_t = \underbrace{K}_{\Phi_0}\varepsilon_t + \underbrace{C_1K}_{\Phi_1}\varepsilon_{t-1} + \underbrace{C_2K}_{\Phi_2}\varepsilon_{t-2} + ..+$$

where K can be:

$K \equiv P$ Choleski factor of Σ_u (Choleski-SVAR);

$K \equiv B$ in the B-SVAR;

$K \equiv A^{-1}$ in the A-SVAR;

$K \equiv A^{-1}B$ in the AB-SVAR.

Recall that $C_0 = I_{M.}$

Suppose for a moment we know all parameter values.

Let

$$K = (\kappa_1, \kappa_2, ..., \kappa_M),$$

with column κ_j capturing the instantaneous impact of the structural shock $\varepsilon_{j,t}$ on the variables.

Then

$$u_t = K\varepsilon_t \equiv \kappa_1 \varepsilon_{1,t} + \kappa_2 \varepsilon_{2,t} + ... + \kappa_M \varepsilon_{M,t}$$

so that

$$\mathbf{\Sigma}_u = KK' \equiv \kappa_1 \kappa_1' + \kappa_2 \kappa_2' + ... + \kappa_M \kappa_M'.$$

This implies, calling σ_i^2 the i-th diagonal elements of $\mathbf{\Sigma}_u$ (i.e. the variance of the innovation $u_{i,t}$):

$$\sigma_i^2 = e_i' \mathbf{\Sigma}_u e_i = \underbrace{e_i'(\kappa_1 \kappa_1')e_i}_{\text{contrib. shock 1}} + \underbrace{e_i'(\kappa_2 \kappa_2')e_i}_{\text{contrib. shock 2}}$$

$$+ ... + \underbrace{e_i'(\kappa_M \kappa_M')e_i}_{\text{contrib. shock } M}$$

$$= (e_i'\kappa_1)(e_i'\kappa_1)' + (e_i'\kappa_2)(e_i'\kappa_2)' + ... + (e_i'\kappa_M)(e_i'\kappa_M)'$$

$$= (\kappa_{i,1})^2 + (\kappa_{i,2})^2 + ... + (\kappa_{i,M})^2$$

Therefore, the quantity

$$\omega_{i,j} = \frac{e_i'(\kappa_j \kappa_j')e_i}{e_i'\Sigma_u e_i} = \frac{(\kappa_{i,j})^2}{e_i'\Sigma_u e_i}$$

captures the fraction of the variance σ_i^2 (the variability of the i-th variable once adjusted for the dynamics) **explained by the** j-th **identified structural shock** $\varepsilon_{j,t}$.

Obviously, by construction:

$$\underbrace{\omega_{i,1}}_{\text{fraction shock 1}} + \underbrace{\omega_{i,2}}_{\text{fraction shock 2}} + ... + \underbrace{\omega_{i,M}}_{\text{fraction shock } M} = 1$$

For $h = 1$, we have

$$W_{t+1} = \Phi_0 \varepsilon_{t+1} + \Phi_1 \varepsilon_t + \Phi_2 \varepsilon_{t-1} + ...$$

where $\Phi_0 = K$; $\Phi_l = C_l K$, $l = 1, 2, ...$

The implied conditional forecast at the forecast horizon $h = 1$ is:

$$\hat{W}_{t+1} = E\left(\hat{W}_{t+1} \mid \mathcal{F}_t\right) = \Phi_1 \varepsilon_t + \Phi_2 \varepsilon_{t-2} + \Phi_3 \varepsilon_{t-3} + ...$$

hence, the **forecast error** at the forecast horizon $h = 1$ is:

$$W_{t+1} - \hat{W}_{t+1} = \Phi_0 \varepsilon_{t+1}$$

$$= \phi_{0,1} \varepsilon_{1,t+1} + \phi_{0,2} \varepsilon_{2,t+1} + ... + \phi_{0,M} \varepsilon_{M,t+1}$$

where, again

$$\Phi_0 = (\phi_{0,1}, \phi_{0,2}, ..., \phi_{0,M}) \equiv K.$$

The associated MSE of the forecast error at the forecast horizon $h = 1$ is:

$$\underbrace{E\left[(W_{t+1} - \hat{W}_{t+1})(W_{t+1} - \hat{W}_{t+1})'\right]}_{MSE(1)} = \Phi_0 \Phi_0' \equiv \sum_{l=0}^{h-1} \Phi_l \Phi_l'$$

$$= \phi_{0,1}\phi_{0,1}' + ... + \phi_{0,M}\phi_{0,M}'$$

so that the fraction

$$\omega_{i,j,1} = \frac{e_i'(\phi_{0,j}\phi_{0,j}')e_i}{e_i'(\Phi_0\Phi_0')e_i} = \frac{(e_i'\phi_{0,j})(e_i'\phi_{0,j})'}{e_i'(\Phi_0\Phi_0')e_i} = \frac{(\phi_{0,i,j})^2}{e_i'(\Phi_0\Phi_0')e_i}$$

captures **the contribution of the identified structural shock j on the forecast error variance associated with variable i at the forecast horizon $h = 1$.**

Note that $\phi_{0,i,j}$ is the (i, j) element of Φ_0.

For $h = 2$, we have

$$W_{t+2} = \Phi_0 \varepsilon_{t+2} + \Phi_1 \varepsilon_{t+1} + \Phi_2 \varepsilon_t + \Phi_3 \varepsilon_{t-1} + ...$$

The implied conditional forecast at the forecast horizon $h = 2$ is

$$\hat{W}_{t+2} = E\left(\hat{W}_{t+2} \mid \mathcal{F}_t\right) = \Phi_2 \varepsilon_t + \Phi_3 \varepsilon_{t-1} + ...$$

hence, the **forecast error** at the forecast horizon $h = 2$ is:

$$W_{t+2} - \hat{W}_{t+2} = \Phi_0 \varepsilon_{t+2} + \Phi_1 \varepsilon_{t+1}$$

and the associated MSE is:

$$\underbrace{E\left[(W_{t+2} - \hat{W}_{t+2})(W_{t+2} - \hat{W}_{t+2})'\right]}_{MSE(2)}$$

$$= \Phi_0 \Phi_0' + \Phi_1 \Phi_1' \equiv \sum_{l=0}^{h-1} \Phi_l \Phi_l'$$

$$= (\phi_{0,1}\phi_{0,1}' + ... + \phi_{0,M}\phi_{0,M}') + (\phi_{1,1}\phi_{1,1}' + ... + \phi_{1,M}\phi_{1,M}').$$

Now,

$$\begin{aligned}\omega_{i,j,2} &= \frac{e_i'(\phi_{0,j}\phi_{0,j}' + \phi_{1,j}\phi_{1,j}')e_i}{e_i'(\Phi_0\Phi_0' + \Phi_1\Phi_1')e_i} \\ &= \frac{e_i'(\phi_{0,j}\phi_{0,j}')e_i + e_i'(\phi_{1,j}\phi_{1,j}')e_i}{e_i'(\Phi_0\Phi_0' + \Phi_1\Phi_1')e_i}\end{aligned}$$

$$= \frac{(\phi_{0,i,j})^2 + (\phi_{1,i,j})^2}{e_i'(\Phi_0\Phi_0' + \Phi_1\Phi_1')e_i} = \frac{\sum_{l=0}^{h-1}(\phi_{l,i,j})^2}{e_i'(\Phi_0\Phi_0' + \Phi_1\Phi_1')e_i}$$

captures **the contribution of the identified structural shock j on the forecast error variance associated with variable i at the forecast horizon $h = 2$.**

Note that $\phi_{0,i,j}$ is the (i,j) element of Φ_0 and $\phi_{1,i,j}$ is the (i,j) element of Φ_1.

Again, by construction:

$$\underbrace{\omega_{i,1}}_{\text{fraction shock 1}}$$

$$\sum_{j=1}^{M} \omega_{i,j,2} = \underbrace{\omega_{i,1,2}}_{\text{fraction shock 1 at } h=2} + \omega_{i,2,2} + \ldots + \omega_{i,M,2} = 1$$

In general, given the SVMA representation:

$$
\begin{aligned}
W_{t+h} &= \Phi_0 \varepsilon_{t+h} + \Phi_1 \varepsilon_{t+h-1} + \Phi_2 \varepsilon_{t+h-2} + \dots \\
& \quad .... + \Phi_{h-1} \varepsilon_{t+1} + \Phi_h \varepsilon_t + \Phi_{h+1} \varepsilon_{t-1} + ...
\end{aligned}
$$

the implied conditional h-step forecast is:

$$
\hat{W}_{t+h} = E\left(\hat{W}_{t+h} \mid \mathcal{F}_t\right) = \Phi_h \varepsilon_t + \Phi_{h+1} \varepsilon_{t-1} + ...
$$

and the h-**step forecast error** is:

$$
\begin{aligned}
& W_{t+h} - \hat{W}_{t+h} \\
= \; & \Phi_0 \varepsilon_{t+h} + \Phi_1 \varepsilon_{t+h-1} + \Phi_2 \varepsilon_{t+h-2} + \dots + \Phi_{h-1} \varepsilon_{t+1}
\end{aligned}
$$

$$
= \sum_{l=0}^{h-1} \Phi_l \varepsilon_{t+h-l}
$$

The MSE of h-**step forecast error** is

$$\underbrace{E\left[(W_{t+h}-\hat{W}_{t+h})(W_{t+h}-\hat{W}_{t+h})'\right]}_{MSE(h)}=\sum_{l=0}^{h-1}\Phi_l\Phi_l'$$

Now,

$$\omega_{i,j,h}=\frac{\sum_{l=0}^{h-1}(\phi_{l,i,j})^2}{e_i'(\sum_{l=0}^{h-1}\Phi_l\Phi_l')e_i}$$

captures **the contribution of the identified structural shock** j **on the forecast error variance associated with variable** i **at the forecast horizon** h.

So far we have treated parameters as known.

The estimation problem is the same as the ones we have studies for SVAR so far.

Hence,

$$\hat{\omega}_{i,j,h} = \frac{\sum_{l=0}^{h-1}(\hat{\phi}_{l,i,j})^2}{e_i'(\sum_{l=0}^{h-1}\hat{\Phi}_l\hat{\Phi}_l')e_i}$$

where

$\hat{\Phi}_0 = K(\hat{\gamma}_T)$; $\hat{\Phi}_l = \hat{C}_l K(\hat{\gamma}_T)$, $l = 1, 2, ...$

Standard errors for the $\hat{\omega}_{i,j,h}$, if necessary can be obtained by bootstrap methods.

So, for instance, in the simple small scale SVAR for output (gdp_t), inflation (π_t) and the short term interest rate (R_t), $W_t = (gdp_t, \pi_t, R_t)'$ whatever your identification method, imagine that you identify these three shocks:

$$\begin{pmatrix} \varepsilon_{gdp,t} \\ \varepsilon_{\pi,t} \\ \varepsilon_{MP,t} \end{pmatrix} \quad \begin{array}{l} \text{output shock} \\ \text{inflation shock} \\ \text{MP shock} \end{array}$$

Then you would like to assess the contribution of the monetary policy shock $\varepsilon_{MP,t}$ in explaining the variance of the forecast error associated with output (gdp_t) after 4/5 quarters (where MP shock is expected to exert its maximal effect).

So for $h = 5$:

$$\hat{\omega}_{gdp,\varepsilon_{MP},5} = \frac{\sum_{l=0}^{5-1}(\hat{\phi}_{l,dgp,\varepsilon_{MP}})^2}{\widehat{MSE}_{gdp}(5)} = \frac{\sum_{l=0}^{5-1}(\hat{\phi}_{l,dgp,\varepsilon_{MP}})^2}{e_1'(\sum_{l=0}^{5-1}\hat{\Phi}_l\hat{\Phi}_l')e_1}.$$

9 - Analytic confidence bands for IRFs (and the delta-method)

The IRFs (take most general case, see below):

$$IRF_{\bullet,\bullet}(h) = \Phi_h = C_h K \quad , \qquad h = 0, 1, 2, ...$$

track the identified dynamic causal effects, i.e. the impact of shocks of interest on the variables over time.

We are interested in the estimation of these quantities but also in quantifying the uncertainty surrounding these estimates.

In all reported graphs we have seen that confidence bands for the IRFs are obtained "by the bootstrap".

In the next slides we discuss the problem of making inference on the IRFs of interest.

We consider a general setup and the IRFs

$$IRF_{\bullet,\bullet}(h) = \Phi_h = C_h K \quad , \quad h = 0, 1, 2, ...$$

where:

- $K \equiv P$, with P Choleski factor of $\Sigma_u = PP'$ in the Choleski SVAR;

- $K \equiv B$ and $B = B(\beta)$ in the B-model;

- $K \equiv A^{-1}$and $A = A(\alpha)$ in the A-model;

- $K \equiv A^{-1}B$ and $A = A(\alpha)$, $B = B(\beta)$ in the AB-model.

We know that given the estimates of VMA matrices C_h via $C_h = \left(R\,(\mathcal{C})^h\,R'\right)$ and estimates of the structural parameters, the **point estimate** of $IRF_{\bullet,\bullet}(h)$ is obtained by the plug-in method:

$$\widehat{IRF_{\bullet,\bullet}}(h) = \hat{\Phi}_{h,T} = \hat{C}_{h,T}\hat{K}_T \ , \qquad h = 0, 1, 2, ...$$

where for $h = 0, 1, 2, ...$ $\hat{C}_{h,T}$ are the plug-in ML estimates of the VMA coefficients, and

- $\hat{K}_T \equiv \hat{P}_T$, in the Choleski SVAR;

- $\hat{K}_T \equiv \hat{B}_T = B(\hat{\beta}_T)$ in the B-model;

- $\hat{K}_T \equiv \hat{A}_T^{-1}$, $\hat{A}_T = A(\hat{\alpha}_T)$ in the A-model;

- $\hat{K}_T \equiv \hat{A}_T^{-1}\hat{B}_T$, $\hat{A}_T = A(\hat{\alpha}_T)$, $\hat{B}_T = B(\hat{\beta}_T)$ in the AB-model.

Thus, imagine that we are interested in the impact of the shock in the variable j on the variable i after h periods:

$$\phi_{i,j,h} = e_i' C_{h,T} K e_j$$

The estimated effect is given by

$$\hat{\phi}_{i,j,h} = e_i' \hat{C}_{h,T} \hat{K}_T e_j$$

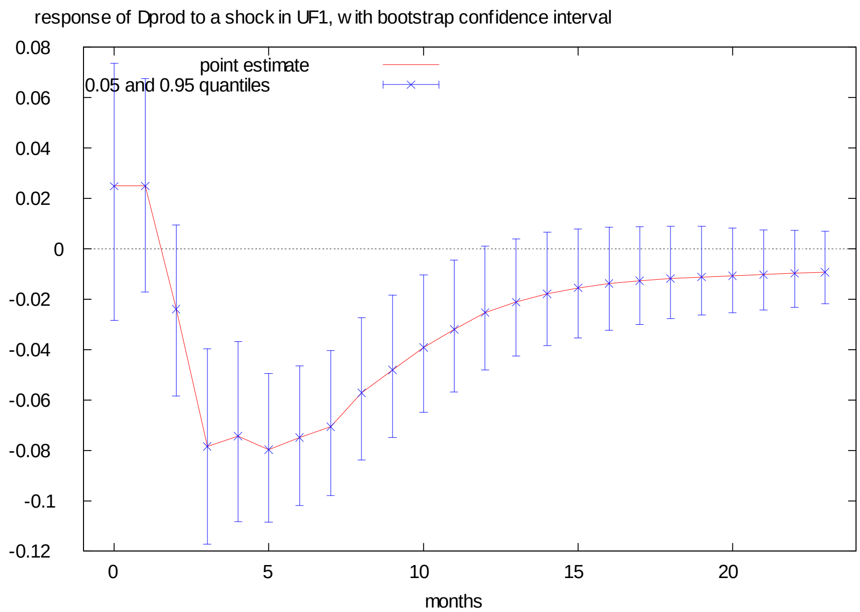
$$= e_i' \left(R \left(\widehat{\mathcal{C}} \right)^h R' \right) \hat{K}_T e_j$$

where recall from the VMA and companion form representations of the VAR, that R is a selection matrix, $\mathcal{C}$ is the VAR companion matrix that depends on the autoregressive parameters $\Pi_1, ..., \Pi_k$.

Above we have only computed **point estimates** of the dynamic causal effect but no measure of variability (uncertainty).

We also need a measure of variability (a standard error) associated with the estimated impact to assess its statistical significance.

Actually, other than IRFs we like to build a **confidence interval for the estimated impulse response** with pre-assigned coverage probability level, something like



$$(\hat{\phi}_{i,j,h,lower}, \hat{\phi}_{i,j,h,upper}), \quad h = 0, 1, 2, \ldots.$$

To obtain such a measure we need to study the distribution of the estimator of the IRFs in large samples

Asymptotic distribution of IRFs

To make the discussion as general as possible, we assume that $K = K(\gamma)$, where γ is the vector of identified structural parameters: $\gamma \equiv p$ in the Choleski-SVAR, $\gamma \equiv \beta$ in the B-model, $\gamma \equiv \alpha$ in the A-model and $\gamma \equiv (\alpha', \beta')'$ in the AB-model.

We have

$$\Phi_h = C_h K = \left(R\left(\mathcal{C}\right)^h R' \right) K(\gamma) \quad , \quad h = 0, 1, 2, ...$$

so that the target dynamic causal effect is given by the (i, j)-element of Φ_h, denoted $\phi_{i,j,h}$.

It is seen that $\phi_{i,j,h}$ is a continuos function of the parameters:

$$\rho = \left(\begin{array}{c} \pi \\ \gamma \end{array} \right) \begin{array}{ll} \text{autoregressive VAR param in } \mathcal{C}. & g \times 1 \\ \text{structural param in } K. & b \times 1 \end{array}$$

hence we have the mapping:

$$\phi_{i,j,h} = f(\rho)$$

where $f(\cdot)$ is a continuos nonlinear function.

It turns out that if ρ_0 is the true value of ρ, then $\phi^0_{i,j,h}$ $= f(\rho_0)$ is the true value of the dynamic causal effect we are interest in.

Let $\hat{\rho}_T$ be estimator:

$$\hat{\rho}_T = \begin{pmatrix} \hat{\pi}_T \\ \hat{\gamma}_T \end{pmatrix} \quad \begin{array}{l} \text{estimator of autoregressive VAR param.} \\ \text{estimator of structural param.} \end{array}$$

Given $\hat{K}_T = K(\hat{\gamma}_T)$, the estimator of $\phi_{i,j,h}$ is given by the **plug-in estimator**:

$$\begin{aligned} \hat{\phi}_{i,j,h} &= f(\hat{\rho}_T) \\ &= e_i' \hat{C}_{h,T} \hat{K}_T e_j = e_i' \left(R \left(\widehat{\mathcal{C}} \right)^h R' \right) K(\hat{\gamma}_T) e_j \end{aligned}$$

Recall that we know that under the assumptions $u_t \sim iidN(0, \Sigma_u)$, $t = 1, ..., T$ or $u_t \sim$MDS + conditional homoskedasticity, the ML estimator of the reduced form parameters of the VAR system are such that

$$
\begin{aligned}
T^{1/2}(\widehat{\delta}_T - \delta_0) &= T^{1/2} \begin{pmatrix} \hat{\pi}_T - \pi_0 \\ \hat{\sigma}^+_{u,T} - \sigma^+_{u,0} \end{pmatrix} \\
&\underset{T\to\infty}{\overset{d}{\longrightarrow}} N\left\{ \begin{pmatrix} 0_{g\times 1} \\ 0_{m\times 1} \end{pmatrix}, \begin{pmatrix} V_\pi & 0_{g\times m} \\ 0_{m\times g} & V_{\sigma^+_u} \end{pmatrix} \right\}
\end{aligned}
$$

where

$$
\begin{aligned}
V_\pi &= (\Sigma_u \otimes V_{XX}^{-1}) \quad , \quad V_{XX} = E(X_t X_t') \\
V_{\sigma^+_u} &= 2D^+_M(\Sigma_u \otimes \Sigma_u)(D^+_M)'.
\end{aligned}
$$

Hence, we have

$$T^{1/2}\,(\hat{\pi}_T - \pi_0) \xrightarrow[T\to\infty]{d} N(0_{g\times 1}, V_\pi)\ , \quad V_\pi = (\mathbf{\Sigma}_u \otimes V_{XX}^{-1})$$

and we also have, for instance, that in the most general case (AB-model):

$$T^{1/2}(\hat{\gamma}_T - \gamma_0) \xrightarrow[T\to\infty]{d} N(0_{b\times 1}, V_\gamma) \quad , \quad V_\gamma = (I_\infty(\gamma_0))^{-1}$$

$$I_\infty(\gamma) := J_{A,B}(\gamma)' I_\infty(\sigma_u^+) J_{A,B}(\gamma).$$

Since γ is obtained from σ_u^+ (identification problem) and $T^{1/2}\,(\hat{\pi}_T - \pi_0)$ and $T^{1/2}(\hat{\sigma}_{u,T}^+ - \sigma_{u,0}^+)$ are **asymptotic independent under stated conditions**, it follows that

$$T^{1/2}(\hat{\rho}_T - \rho_0) \xrightarrow[T\to\infty]{d} N(0_{(g+b)\times 1}, V_\rho) \quad , \; V_\rho = \begin{pmatrix} V_\pi & 0_{g\times b} \\ 0_{b\times g} & V_\gamma \end{pmatrix}$$

Delta-method, general formulation

The delta-method says that whenever we have a consistent and asymptotically Gaussian estimator $\hat{\theta}_T$ ($a \times 1$) like e.g.:

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(0_{a\times 1},\, V_\theta\right)$$

with V_θ ($a \times a$) positive definite asymptotic covariance matrix, and we have another estimator $\hat{\phi}_T$ ($d \times 1$) which can be expressed as

$$\hat{\phi}_T = f(\hat{\theta}_T) \xrightarrow{p} f(\theta_0) = \phi_0$$

$f(\cdot)$ being is a continuous function, then

$$T^{1/2}(\hat{\phi}_T - \phi_0) \xrightarrow{d} N\left(0_{d\times 1},\, V_\phi\right)$$

where the asymptotic covariance matrix V_ϕ is ($d \times d$) and given by the 'sandwitch form'

$$V_\phi = G_\theta V_\theta G_\theta'$$

where $G_\theta = \frac{\partial \phi}{\partial \theta'} \equiv \frac{\partial f(\theta)}{\partial \theta'}$ is a $d \times a$ Jacobian matrix, such that rank(V_ϕ)$= d$.

So, given the result

$$T^{1/2}(\hat{\rho}_T - \rho_0) \xrightarrow{d} N(0_{(g+b)\times 1}, V_\rho)$$

and the continuity of the function

$$\phi_{i,j,h} \ = f(\rho),$$

by the Delta-method:

$$T^{1/2}\left(\hat{\phi}_{i,j,h} - \phi^0_{i,j,h}\right) \xrightarrow{d} N\left(0,\ V_{\phi_{i,j,h}}\right),\ V_{\phi_{i,j,h}} = G_\rho V_\rho G'_\rho$$

where

$$G_\rho = \frac{\partial \phi_{i,j,h}}{\partial \rho'} \equiv \frac{\partial f(\rho)}{\partial \rho'} \quad \text{is} \ \ 1 \times (g+b).$$

In principle, it is possible to derive an analytic expression for the Jacobian matrix $G_\rho = G_\rho(\rho)$.

This step appears unnecessary for our purposes.

Indeed, G_ρ can be computed by numerical derivatives from packages.

For the moment, let us assume that we have a consistent estimator of G_ρ and of V_ρ, so that we have a consistent estimator of $V_{\phi_{i,j,h}}$, say $\hat{V}_{\phi_{i,j,h}} = \hat{G}_\rho \hat{V}_\rho \hat{G}'_\rho$.

The standard error associated with the estimated IRF $\hat{\phi}_{i,j,h}$ is

$$\text{s.e.}(\hat{\phi}_{i,j,h}) = \left(\frac{1}{T}\hat{V}_{\phi_{i,j,h}}\right)^{1/2}.$$

From the asymptotic normality result:

$$\Pr\left\{q_\varsigma \leq \frac{\left(\hat{\phi}_{i,j,h} - \phi^0_{i,j,h}\right)}{\text{s.e.}(\hat{\phi}_{i,j,h})} \leq q_{1-\varsigma}\right\} = 1 - \varsigma$$

where $q_{\varsigma/2}$ and $q_{1-\varsigma/2}$ are quantiles from the N(0,1) distribution.

Thus, a 90% confidence interval for $\phi^0_{i,j,h}$ is e.g. given by

$$[\hat{\phi}_{i,j,h} - 1.96\text{s.e.}(\hat{\phi}_{i,j,h})\ ,\ \ \hat{\phi}_{i,j,h} + 1.96\text{s.e.}(\hat{\phi}_{i,j,h})]$$

These are pointwise intervals, i.e. they are valid for each fixed $h = 0, 1, ...$

The Jacobian matrix

$$G_\rho = \frac{\partial \phi_{i,j,h}}{\partial \rho'}$$

is crucial to compute the asymptotic covariance matrix $V_{\phi_{i,j,h}} = G_\rho V_\rho G'_\rho$ and make inference on the IRFs.

Althoug it is in principle possible to derive an analytic expression for G_ρ or it is possible to compute it numerically, seldom analytic confidence bands are reported in empirical SVAR analysis.

This is so mainly for two reasons.

First, for finite T, the normal approximation

$$\frac{\left(\hat{\phi}_{i,j,h} - \phi^0_{i,j,h}\right)}{\left(\hat{V}_{\phi_{i,j,h}}\right)^{1/2}} \approx N\left(0,\ 1\right)$$

works poorly, in the sense that "analytic" confidence bands based on this result tend to provide empirical coverage (the frequency of times in which the true value $\phi^0_{i,j,h}$ is contained in the confidence interval) smaller than the expected nominal level, $1-\varsigma$.

Second, there exists omputationally convenient simulation methods which allow us to get a ‘better’ approximation in fine samples.

In applied research the typical confidence bands for IRFs are computed by bootstrap methods that we review next.

10 - Bootstrap confidence bands for IRFs (short account)

The idea is to treat the estimated VAR paramaters as pseudo-true population values. The model is then used to generate (simulate) pseudo-samples.

Consider the reduced form VAR:

$$W_t = \mu + \Pi_1 W_{t-1} + ... + \Pi_p W_{t-p} + u_t$$

and assume that we **know** that

$$u_t \sim iidN(0_{M\times 1}, \Sigma_u)$$

i.e. that the VAR innovations are iid and Gaussian (questionable assumption).

Given the ML (OLS) estimates $\hat{\mu}$, $\hat{\Pi}_1, ..., \hat{\Pi}_p$ and $\hat{\Sigma}_u$ obtained on the sample, one can think to the following algorithm called the **residual-based parametric iid bootstrap**:

1. generate u_t^* randomly (iid) from a multivariate distribution with zero mean and covariance matrix $\hat{\Sigma}_u$;

2. generate (reconstruct) the bootstrap sample $W_1^*, W_2^*, ...,$ by iterating, for $t = 1, ..., T$, the model

$$W_t = \hat{\mu} + \hat{\Pi}_1 W_{t-1} + ... + \hat{\Pi}_p W_{t-p} + u_t^*$$

   where initial conditions $W_0, ..., W_{1-p}$ are fixed at the original sample values;

3. estimate the SVAR model (e.g. B-model, or A-model or AB-model) by ML (OLS) on the bootstrap observations $W_1^*, W_2^*, ..., W_T^*$ generated in the previous step. This step produces the bootstrap estimates $\hat{\mu}^*$, $\hat{\Pi}_1^*, ..., \hat{\Pi}_p^*$, $\hat{\Sigma}_u^*$ of the parameters. From these compute the VMA coefficients $\hat{C}_h^*$ and the structural parameters $\hat{K}^*$ and the bootstrap estimate of the IRF of interest, $\hat{\phi}_{i,j,h}^*$ (recall that $\hat{\phi}_{i,j,h}^*$ is the (i,j)-element of $\hat{\Phi}_h^* = \hat{C}_h^* \hat{K}^*$);

4. repeat the steps 1-3 N times, obtaining the sequence of bootstrap estimates of the IRFs of interest:

$$\hat{\phi}^{*1}_{i,j,h}, \hat{\phi}^{*2}_{i,j,h}, ..., \hat{\phi}^{*N}_{i,j,h};$$

5. compute the **percentile bootstrap confidence itervals** for $\phi^0_{i,j,h}$ by $(q^{\phi *}_{\varsigma/2},\ q^{\phi *}_{1-\varsigma/2})$ where $q^{\phi *}_{\varsigma/2}$ and $q^{\phi *}_{1-\varsigma/2}$ are the $\varsigma/2$ and $1-\varsigma/2$ quantiles of the empirical distribution of $\hat{\phi}^{*1}_{i,j,h}, \hat{\phi}^{*2}_{i,j,h}, ..., \hat{\phi}^{*N}_{i,j,h}$. Alternatively, compute **the basic bootstrap confidence interval** for $\phi^0_{i,j,h}$ as $(2\hat{\phi}_{i,j,h} - q^{\phi *}_{1-\varsigma/2}, 2\hat{\phi}_{i,j,h} + q^{\phi *}_{\varsigma/2})$.

The residual-based bootstrap procedure described in the algorithm above is "parametric" in the sense that it is based on the knowlegde that u_t is Gaussian distributed other than iid.

The normality hypothesis is not the rule in applied work and generating u_t^* from the multivariate Gaussian can be misleading.

A possible solution to this problem is to consider a **nonparametric residual-based iid bootstrap** algorithm, which requires replacing the step 1 above with the following:

1' Given the VAR residuals

$$\hat{u}_1, \hat{u}_2 ..., \hat{u}_T$$

consider the re-centered sequence

$$\hat{u}_1^c, \hat{u}_2^c ..., \hat{u}_T^c$$

where

$$\hat{u}_t^c = \hat{u}_t - \frac{1}{T} \sum_{i=1}^{T} \hat{u}_i \quad , \quad t = 1, ..., T$$

and perform an iid-resampling on $\hat{u}_1^c, \hat{u}_2^c ..., \hat{u}_T^c$ obtaining

$$u_1^*, u_2^* ..., u_T^*$$

then continue with the steps 2-5 seen above.

Moving Block Bootstrap: accounts for Garch-type VAR innovations: see Brüggemann, R., Jentsch, C. and Trenkler, C. (2016), Inference in VARs with conditional volatility of unknown form, *Journal of Econometrics* 191, 69-85.

10.1 The notion of bootstrap consistency (short account)

Imagine we have the result

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, V_\theta)$$

so that

$$\Gamma_T := T^{1/2} V_\theta^{-1/2} (\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, I)$$

where, given spectral decomposition: $V_\theta = P D_\theta P'$, with D_θ diagonal having the eigenvalues of V_θ on the diagonal, and P such that $P'P = I = P^{-1}P$, then $V_\theta^{1/2} := P D_\theta^{1/2} P'$ ($\Rightarrow V_\theta^{-1/2} = P D_\theta^{-1/2} P'$).

Thus, if we have a consistent estimator of the asymptotic covariance matrix V_θ :

$$\hat{V}_\theta \xrightarrow{p} V_\theta$$

it holds:

$$\hat{\Gamma}_T := T^{1/2}\hat{V}_\theta^{-1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, I) \equiv \underbrace{\Phi_z(x)}_{\text{cdf}}$$

with $\hat{V}_\theta^{1/2} := \hat{P}\hat{D}_\theta^{1/2}\hat{P}'$.

Now, define

$$\hat{\Gamma}_T^* := T^{1/2}\hat{V}_\theta^{-1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$$

where $\hat{\theta}_T^*$ is the bootstrap counterpart of the estimator $\hat{\theta}_T$.

The question is:

$$\hat{\Gamma}_T^* := T^{1/2}\hat{V}_\theta^{-1/2}(\hat{\theta}_T^* - \hat{\theta}_T) \rightarrow \text{ ??}$$

Note that there are **two probability distributions** involved in the statistics defined above:

The standard probability distribution we have on the data (reflecting the fact that the observed data are realizations of stochastic processes);

The probability distribution **induced by the bootstrap** resampling.

Let $G_T^*(x) = \Pr(\hat{\Gamma}_T^* \leq x)$ the cdf of the bootstrap estimator $\hat{\Gamma}_T^*$ (its variabilty must be considered in the bootstrap probability world, conditional on the original data).

We say that the bootstrap is **consistent** if $\forall \epsilon > 0$ it holds;

$$\lim_{T\to\infty} \Pr\left(\sup_{x\in\mathbb{R}} |G_T^*(x) - \Phi_z(x)| > \epsilon\right) = 0$$

or,

$$\sup_{x\in\mathbb{R}} |G_T^*(x) - \Phi_z(x)| \xrightarrow{p} 0 \quad \text{as } T \to \infty$$

this is **a uniform convergence result**.

Under bootstrap consistency we can claim that:

$$\hat{\Gamma}^*_T := T^{1/2}\hat{V}_\theta^{-1/2}(\hat{\theta}^*_T - \hat{\theta}_T) \xrightarrow{d}_{p^*} N(0, I)$$

namely, **the bootstrap reproduces, in the bootstrap world, the same asymptotic distribution as the estimator** $\hat{\Gamma}_T := T^{1/2}\hat{V}_\theta^{-1/2}(\hat{\theta}_T - \theta_0)$ in the original probability data distribution.

The symbol "$\xrightarrow{d}_{p^*}$" indicates that convergence occurs "conditional on the data" in the boostrap world.

Observe that from the bootstrap algorithm, we can obtain N replications of $\hat{\theta}_T^*$:

$$\hat{\theta}_{T:1}^*, \hat{\theta}_{T:2}^*, ..., \hat{\theta}_{T:N}^*$$

where N can be arbitrarily large !

So, we have:

$$\hat{\Gamma}_{T:1}^*, \hat{\Gamma}_{T:2}^*, ..., \hat{\Gamma}_{T:N}^*$$

where

$$\hat{\Gamma}_{T:n}^* := T^{1/2} \hat{V}_\theta^{-1/2} (\hat{\theta}_{T:n}^* - \hat{\theta}_T) \quad , \quad n = 1, ..., N$$

which means that **we can always build the empirical bootstrap cdf:**

$$G_{T:N}^*(x) = \frac{1}{N} \sum_{n=1}^{N} 1(\hat{\Gamma}_{T:n}^* \leq x) \quad (1(\cdot) \text{ indicator function})$$

Glivenko-Cantelli theorem ensures that as $N \to \infty$ (not T):

$$\sup_{x \in \mathbb{R}} \left| G^*_{T,N}(x) - G^*_T(x) \right| \overset{a.s.}{\to} 0$$

11- Some current - novel - identification schemes:

So far we have been dealing with "conventional" SVARs.

The recent literature has moved towards **novel shock identification schemes** in SVARs aiming to minimize restrictions on the matrix B (A) as much as possible.

The rationale behind this approach is twofold: firstly, to impose as few controversial restrictions as possible;

secondly, to adopt a "partial" identification logic, wherein we concentrate exclusively on a subset of the structural shocks in the system (termed **target structural shocks**).

This allows us to refrain from taking a stance on the other structural shocks, which are not of primary interest (referred to as **non-target structural shocks**)."

For example: we want to identify a MP shock without taking a stand on a supply/demand shock that might be potentially embedded in the VAR.

Besides the monetary policy example, there are many other cases in which specifications based on zero identification restrictions are controversial/unsatisfactory.

There are several possible methods by which one can deal with a full" or "near-full" matrix B (A) in SVAR analysis.

To do so, we have to move to novel identification schemes.

In the current literature, popular approaches are:

- the **sign restrictions** (or **set-identification**) approach;

- the **Independent Component Analysis (ICA)** approach

- the **changes-in-volatility** (heteroskedasticity-) approach;

- the **external variables approach** known as **proxy-SVAR (SVAR-IV)** approach.

- the **local projections** (**LP**) approach (which may also make use of external variables).

11.1- The sign-restrictions approach: a frequentist view (short account);

The SVAR approach we have studied so far is based on the concept of **point-identification**, which means that we identify and estimate the elements of the matrices B or A or A and B.

The sign restrictions approach is based on the concept of **set-identification**, which means that the structural shocks not are identified by imposing zero restrictions, but by imposing theory-based restrictions about the sign of the responses contained in

$$IRF_{\bullet,\bullet}(h) = \Phi_h = C_h K \quad , \quad h = 0, 1, ...$$

These sign restrictions can be imposed both on impact ($h = 0$) or at the time horizons ($h = 1, 2, ..$).

The sign restrictions approach rests on the idea that in many cases we may have **a safe theoretical view about the signs** of the impact and propagation of the shocks.

Accordingly, there can be **many IRFs, not just one** that, at given horizons, respect the signs.

Each sign restriction corresponds to an a-priori idea of the impact of the j-th structural shock on the i-th variable.

This implies **set-identification**, as opposed to **point-identification** discussed so far.

Most methods that have been used to **construct pointwise coverage bands** for IRFs of sign-restricted SVARs are **justified from a Bayesian perspective**.

The paper by:

Granziera, E. L., Moon, H.R. and Shorfheide, F. (2018), Inference for VARs identified with sign restrictions, *Quantitative Economics* 9, 1087-1121.

demonstrates how to formulate the inference problem for sign-restricted SVARs within a **moment-inequality framework**.

In these slides we skip the issue of frequentist inference in sign-restricted SVARs and stick to a more "operational" approach.

Example: Kilian's (2012) global crude oil market model

Example taken from (recommended for students interested to the issue):

Kilian, L. and Murphy, D. P. (2012), Why agnostic sign restictions are not enough: Understanding the dynamics of oil market VAR models, *Journal of the European Economic Association*, 10, 1166-1188.

Kilian and Murphy (2012) consider a VAR model for

$$W_t = \begin{pmatrix} \Delta prod_t \\ rea_t \\ rpoil_t \end{pmatrix} \begin{matrix} \text{percentage change in global oil production} \\ \text{index of real economic activity} \\ \text{log of real price of oil} \end{matrix}$$

on the period 1973M1-2008M9 (monthly observations).

SVAR specification:

$$AW_t = \tau + \sum_{i=1}^{24} \Upsilon_i W_{t-i} + \varepsilon_t$$

In previous article, using the same VAR specification Kilian considers the following (triangular - Choleski) identification scheme ($B = A^{-1}$):

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{pmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_t^{\text{oil-supply}} \\ \varepsilon_t^{\text{aggregate-demamd}} \\ \varepsilon_t^{\text{oil-market-specific-dem}} \end{pmatrix}$$

where, obviously if B triangular also A is triangular.

In Kilian and Murphy (2012) instead, they consider the following table with baseline sign restrictions on the on-impact effects ($h = 0$) of the shocks:

Variables/Shocks	Oil supply disruption	Aggregate demand	Oil-specific demand
Oil production	-	$+$	$+$
Real activity	-	$+$	-
Real price of oil	$+$	$+$	$+$

Kilian and Murphy (2012, pp.1173-1174):

"*We postulate that a positive aggregate demand shock will tend to raise oil production, stimulate real activity and increase the real price of oil on impact*.

A positive oil-market-specific demand shock on impact will raise the real price of oil and stimulate oil production, but lower real activity.

An unexpected oil supply disruption will by construction lower oil production on impact. It also will lower real activity, while increasing the real price of oil".

Thus, with the sign restrictions approach, we want to consider a framework in which the coefficients of the IRFs $IRF_{\bullet,\bullet}(h) = \Phi_h = C_h K$ satisfy the desired signs at specified values of h.

In Kilian and Murphy's (2012) example, the sign restrictions in the Table refer to $IRF_{\bullet,\bullet}(0) = \Phi_0 = C_0 K$.

However, **one could also impose sign restrictions on lagged responses!**

In general, let $\phi_{i,j,h}$ be element of $\Phi_h = C_h K$.

The restrictions on the sign on the IRFs coefficients can take the form:

$$a_L \leq \phi_{i,j,h} \leq a_U$$

where

- the time horizon h "is chosen" by the practitioner according to his/her knowledge/theory;

- a_L could be finite or -∞; a_U could be finite or $+\infty$.

Obviously, for $h = 0$, the restrictions are put on the sign of the elements of the matrix $\Phi_0 = K$;

for $h = 1$ the sign restrictions are placed on the elements of $\Phi_1 = C_1 K$ which capture the response of the variables after one period, and so forth.

The simplest way to consider the sign restrictions approach **in the frequentist framework** is to consider the structural specification (B-form) $u_t = K\varepsilon_t$ with

$$K = PQ$$

where:

- P is the Choleski factor of $\mathbf{\Sigma}_u$;

- Q is an **orthogonal matrix**, i.e. any matrix such that $Q'Q = I_M$, $QQ' = I_M$;

There exist potentially infinite matrices Q that satisfy $QQ' = I_M$.

Hence, given Q_1 and Q_2 orthogonal matrices, one has

$$\begin{aligned} \mathbf{\Sigma}_u &= PP' = (PQ_1)(PQ_1)' \\ &= (PQ_2)(PQ_2)' \end{aligned}$$

which means that albeit PQ_1 and PQ_2 might have elements with different signs, both PQ_1 and PQ_2 are consistent with the (unique) covariance matrix $\mathbf{\Sigma}_u$ (i.e. with the data).

Moreover, although P is triangular, PQ is **no longer triangular**; in general, it will be a "full" matrix.

Importantly, the coefficients of the matrix PQ **may have the desired sign or not, depending on the particular selected matrix** Q **.**

Imagine now to write an algorithm that, given the estimated Choleski factor $\hat{P}$ of $\hat{\Sigma}_u$, (i) draws **may times** matrices Qs from the set of orthogonal matrices; (ii) computes $\hat{K} = \hat{P}Q$; (iii) **discards** the $\hat{K} = \hat{P}Q$ for which the implied IRFs $\hat{\Phi}_h = \hat{C}_h\hat{K} = \hat{C}_h\hat{P}Q$ do not satisfy the desired signs; (iv) **stores** the $\hat{K} = \hat{P}Q$ for which the implies IRFs $\hat{\Phi}_h = \hat{C}_h\hat{K} = \hat{C}_h\hat{P}Q$ satisfy the desired signs.

There are many possible algorithms one can use to generate matrices Qs randomly.

A simple, commonly used algorithm is as follows:

1. draw the $M \times M$ matrix G from NID(0,1) (which means that each element of G is generated from independent N(0,1) random variables);

2. compute the so-called QR-decomposition of G, i.e. $G = QR$, where Q is such that $QQ' = I_M$.

Whatever the algorithm one uses to generate randomly the Qs, at the end of the procedure we have stored a set of theoretically-admissible SVARs.

We understand that this algorithm **does not produce point estimates of the structural IRFs, but a family of IRFs**, where each menber of the family is compatible with the estimated reduced form covariance matrix $\hat{\Sigma}_u$.

It turns out that **without further information,** it is impossible to know which of the IRFs in the admissible set "represents" the true system dynamics.

Plotting all admissible response functions typically results in a bundle of lines.

Typically practitioners report some central tendency (mean, median), quantiles and the magnitude of the spread of responses.

For instante, it is typical to report percentiles such as the 5 percent, 50 percent and 95 percent of the IRFs stored in the admissible set.

It is important to stress that differently from what we have studied so far (with e.g. booootstrap methods), quantiles here **have nothing to do with sampling uncertainty**!!

Recommended reading: Fry and Pagan (2011, JEP), see below.

Suggested readings for students who want to elaborate on this topic:

Kilian, L. and Murphy, D. P. (2012), Why agnostic sign restrictions are not enougg: Understanding the dynamics of oil market VAR models, *Journal of the European Economic Association* 10, 1166-1188.

Fry, R. and Pagan, A. (2011), Sign restrictions in Structural Vector Autoregressions: A critical review, *Journal of Economic Perspectives* 49, 938-960.

11.2 The - "statistical"- Independent Component Aanalysis (ICA) approach

We focus on the link beteen structural shocks and VAR innovation in the B-model (imagine zero lags for simplicity):

$$u_t = B\varepsilon_t$$

Assume that no restriction is placed on B.

We compensate the lack of information on the elements of B with purely statistical information, in particular, with two joint pieces of information regarding the elements in ε_t:

- **statistical independence** (across elements) of ε_t;

- **non-normality (non-Gaussianity)** of the elements of ε_t (actually, at most only one component can be Gaussian but non-Gaussian all the rest).

More precisely, assume that we have these information on ε_t :

1. The components $\varepsilon_{j,t}$ of $\varepsilon_t := (\varepsilon_{1,t}, \varepsilon_{2,t} ..., \varepsilon_{M,t})'$ have zero mean and variance $0 < \sigma^2_{\varepsilon,j} < \infty$;

2. Any $\varepsilon_{j,t}$ is **independent** on $\varepsilon_{i,t}$, $j \neq i$ and **at most one** of the elements (only one) in $\varepsilon_t := (\varepsilon_{1,t}, \varepsilon_{2,t} ..., \varepsilon_{M,t})'$ **can have marginal Gaussian distribution (meaning that all the others are non-Gaussian)**;

3. $Cov(\varepsilon_{j,t}, \varepsilon_{j,t+k}) = 0$ for each $j = 1, ..., M$, i.e. all components in $\varepsilon_t := (\varepsilon_{1,t}, \varepsilon_{2,t} ..., \varepsilon_{M,t})$ **are serially uncorrelated (they need not necessarily be serially independent!).**

Let's turn on our initial definition (Definition 0) of structural shocks introduced in one of our first slides:

Definition 0. Shocks. *Shocks are primitive exogenous forces that are uncorrelated with each other and should be economically meaningful.*

Definition 0.1 They must be **exogenous** with respect to the other current and lagged endogenous variables in the model.

Definition 0.2 They must be **uncorrelated** with other exogenous shocks $\leftarrow$ **BIG DIFFERENCE HERE**

Definition 0.3 They must be **unanticipated**.

We notice that relative to Definition 0, the very novelty of the ICA is:

2. Any $\varepsilon_{j,t}$ is **independent** on $\varepsilon_{i,t}$, $j \neq i$ and **at most one** of the elements (only one) in $\varepsilon_t := (\varepsilon_{1,t}, \varepsilon_{2,t} \dots, \varepsilon_{M,t})'$ has **marginal Gaussian distribution**.

Let D be an $M \times M$ diagonal matrix **with positive elements on the diagonal.**

Let P be an $M \times M$ **permutation matrix**, where a permutation matrix is a square matrix if each row and each column of P contains a single element 1, and the remaining elements are zero.

An $M \times M$ permutation matrix P then contains M ones and $M(M-1)$ zeros.

It can be proved that any permutation matrix is **nonsingular and is orthogonal**, that is:

$$P^{-1} = P'.$$

Then, given **two arbitrary** matrices D and P as above, it always holds:

$$u_t = B\varepsilon_t$$

$$= BDPP^{-1}D^{-1}\varepsilon_t = \underbrace{BDP}_{B^*}\,\underbrace{P^{-1}D^{-1}\varepsilon_t}_{\varepsilon_t^*} = B^*\varepsilon_t^*$$

where:

$B^* = BDP \equiv \tilde{B}P$ is a **novel matrix of on-impact parameters** that differs from the original matrix B because of the different scaling of its columns (operated by the diagonal elements of D) **and** because of the permutation of its columns (operated by the permutation matrix P):

$\varepsilon_t^* = P^{-1}D^{-1}\varepsilon_t \equiv P^{-1}\breve{\varepsilon}_t$ is a **novel vector of structural shocks** that differ from the original vector ε_t because of the different variance of its elements **and** because of the different ordering of the elements in $P^{-1}\breve{\varepsilon}_t$:

if the variance of ε_t is I_M then the variance of $\breve{\varepsilon}_t := D^{-1}\varepsilon_t$ is D^{-2};

if the variance of ε_t is $\mathbf{\Sigma}_\varepsilon$ then the variance of $\breve{\varepsilon}_t := D^{-1}\varepsilon_t$ is $D^{-1}\mathbf{\Sigma}_\varepsilon D^{-1\prime} = D^{-2}\mathbf{\Sigma}_\varepsilon$

Note that the equivalence

$$u_t = B\varepsilon_t = \underbrace{BDP}_{B^*} \underbrace{P^{-1}D^{-1}\varepsilon_t}_{\varepsilon_t^*}$$

holds for **any** diagonal matrix D with positive elements on the diagonal and any permutation matrix P.

Independence and non-normality of ε_t have played no role so far.

But imagine now that by jointly exploiting the (cross-)independence and non-normality of the elements in ε_t one can estimate B^* consistently.

This means that one can consistently estimate IRFs of the type:

$$\Phi_h^* = C_h B^* = \left(R\left(\mathcal{C}\right)^h R'\right) B^* \quad , \quad h = 0, 1, 2, \ldots$$

So, by simply exploiting **two statistical properties** postulated for the structural shocks, we can potentially estimate the IRFs

$$\Phi_h^* = C_h B^* = \left(R\,(\mathcal{C})^h\,R'\right)B^* \quad , \quad h = 0, 1, 2, ...$$

with no use of economic/theoretical information.

The problem here is that we are not able to associate a-priori each column of B^* to a structural shock because P is identermined!

We recover B^* from the data (not B), which can be considered a version of the "true" B **up to scaling and column ordering**!

In other words, we may claim that we identify on-impact coefficients B **up to scaling and ordering**!

This means that we don't know ex-ante which response to which shock the matrix Φ_h^* collects !

We can explore the estimated IRFs and try to label **ex-post** the elements in ε_t^* using **some additional criteria**.

For instance, imagine that we apply ICA analysis to the VAR for $W_t = (\tilde{y}_t, \pi_t, R_t)'$.

We estimate the IRFs Φ_h^* and, given $\widehat{\Phi}_h^*$, plot them.

Imagine we observe that output $\tilde{y}_t$ and inflation π_t tend to decline in response to the shock $\varepsilon_{1,t}^*$, while R_t increases and then converges to equilibrium.

Then, we can label $\varepsilon_{1,t}^*$ as a MP-type shock.

Insights

For ML estimation of B^* (Maximum Likelihood, maintained of correct specification of the non-Gaussian distribution) refer to:

- Lanne, M., M. Meitz, and P. Saikkonen (2017). Identification and Estimation of Non-Gaussian Structural Vector Autoregressions, *Journal of Econometrics* 196, 288 – 304.

For likelihhod-based inference under potential distribution misspecification, refer to:

- Fiorentini, G. and Sentana, E. (2022), Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions, *Journal of Econometrics*

....

11.3 - Heteroskedasticity approach: constant IRFs

VAR practitioners often find that VAR disturbances display conditional heteroskedasticity (as seen through ARCH-type tests on the VAR residuals) but also **unconditional heteroskedasticity**, i.e. changes (breaks) in the covariance matrix Σ_u.

We can call breaks in the uncoditional covariance matrix Σ_u **changes in (unconditional) volatility**.

Example: A classical stylized fact of US macroeconomic data (but also of other developed countries) is a general drop of volatility of macroeconomic variables from the mid-eighties until the Global Financial Crisis, the so-called **Great Moderation** period.

The existence of **different volatility regimes in the data**, associated with recognizable historical events, provides an interesting opportunity to identify structural shocks in SVARs.

For instance, coming back to the financial/macro uncertainty example, we estimated a VAR model on the period 1960M7-2015M4 ($T = 651$ monthly observations).

Actually, one might argue that $\boldsymbol{\Sigma}_u$ can not constant over a such large span of data.

Angelini, Bacchiocchi, Caggiano and Fanelli (2019, JAE) argue that there are at least three volatilility regimes on the period 1960M7-2015M4 associated with the two break dates $T_{B_1} =$ 1984M3 and $T_{B_2} =$2007M12:

1960M8-1984M3 ($T = 280$)	Great Inflation, $\boldsymbol{\Sigma}_{u,1}$
1984M4-2007M12 ($T = 285$)	Great Moderation, $\boldsymbol{\Sigma}_{u,2}$
2008M1-2015M4 ($T = 88$)	Great Rec.+Slow Rec., $\boldsymbol{\Sigma}_{u,3}$

They test the null hypothesis that there are no volatility regimes in the data ($H_0 : \boldsymbol{\Sigma}_{u,1} = \boldsymbol{\Sigma}_{u,2} = \boldsymbol{\Sigma}_{u,3}$) and reject it.

Two volatility regimes

For simplicity we assume that there is a single exogenous structural break in the error covariance matrix which gives to two volatility regimes in the data.

The approaches we discuss below can be obviously extended (non-trivially) to more than two volatility regimes.

In particular, imagine that in our VAR it holds the condition

$$\boldsymbol{\Sigma}_u = \begin{cases} \boldsymbol{\Sigma}_{u,1} & \text{for } t = 1, ..., T_0 & \text{volatility regime1} \\ \boldsymbol{\Sigma}_{u,2} & \text{for } t = T_0 + 1, ..., T & \text{volatility regime2} \end{cases}$$

where $1 < T_0 < T$ is a break date which **is known or has been correctly inferred from the data**, and $\boldsymbol{\Sigma}_{u,1} \neq \boldsymbol{\Sigma}_{u,2}$.

The break date T_0 is "exogenous" relative to the shocks we want to identify.

$$\Sigma_u = \begin{cases} \Sigma_{u,1} & \text{for } t = 1, ..., T_0 \qquad \text{volatility regime1} \\ \Sigma_{u,2} & \text{for } t = T_0 + 1, ..., T \quad \text{volatility regime2} \end{cases}$$

Practitioners might be tempted to split the empirical analysis into two separate analyses: one "standard" SVAR analysis on data on the first volatility regime and one "standard" SVAR analysis on data on the second volatility regime.

This is not a mistake per se but doing so leads the practitioner to loose important identification information!

The fact that there are "two" covariance matrices, enhances the identification possibilities and allows us to relax undesired zero restrictions!

The identification schemes we analyze below **support non-triangular structures.**

But we have to deal with just one SVAR, not two!

There are two possible approaches one can consider when dealing with identification of SVARs based on heteroskedasticity:

- the Lanne and Lutkephol's (2010, JMCB) approach;

- the Bacchiocchi and Fanelli's (2015, OBES) approach.

Lanne and Lutkepohl (2010, JMCB)'s approach

A result of algebra says that if $\Sigma_{u,1} \neq \Sigma_{u,2}$, then it is always valid the simultaneous decomposition:

$$\begin{aligned} \Sigma_{u,1} &= BB' \\ \Sigma_{u,2} &= BVB' \end{aligned}$$

where:

- B is an $M \times M$ **nonsingular and "full"**;

- V is a **diagonal** matrix with **positive and distinct elements** on the diagonal, and distinct from '1'.

Thus, we have $\frac{1}{2}M(M+1) + \frac{1}{2}M(M+1) = M(M+1)$ reduced form parameters in $\Sigma_{u,1}$ and $\Sigma_{u,2}$, and $M^2 + M = M(M+1)$ parameters in B and V.

It is possible to show that the relationships

$$\begin{aligned}\boldsymbol{\Sigma}_{u,1} &= BB' \\ \boldsymbol{\Sigma}_{u,2} &= BVB'\end{aligned}$$

can be "inverted" in the sense that the elements in B and V (those on the diagonal) can be uniquely recovered as function of the elements in $vech(\boldsymbol{\Sigma}_{u,1})$ and $vech(\boldsymbol{\Sigma}_{u,2})$.

Hence, the IRFs are given by

$$IRF_{\bullet,\bullet}(h) = C_h B \ , \ \ h = 0, 1, ...$$

Notice that although there is a break in volatility at $t = T_0$, in this approach the parameters in B **do not change** in the move from the first to the second volatility regime.

In other words, **the IRFs are constant across the two volatility regimes.**

11.4 - Heteroskedasticity approach: regime-dependent IRFs

Bacchiocchi and Fanelli (2015, OBES) approach

If there is a break in unconditional volatility at $t = T_0$, then it can be the case that also structural parameters (and perhaps the autoregressive ones) do change!

Idea:

$$\begin{array}{ll} u_t = B\varepsilon_t & \text{for } t = 1, ..., T_0 \\ u_t = (B+G)\varepsilon_t & \text{for } t = T_0 + 1, ..., T \end{array}$$

where G is an $M \times M$ matrix that captures the changes in B in the move from the first to the second volatility regime.

Implied moment conditions

$$\begin{array}{ll} \Sigma_{u,1} = BB' & \text{volatility regime1} \\ \Sigma_{u,2} = (B+G)(B+G)' & \text{volatility regime2} \end{array}$$

Thus, we have $\frac{1}{2}M(M+1) + \frac{1}{2}M(M+1) = M(M+1)$ reduced form parameters in $\Sigma_{u,1}$ and $\Sigma_{u,2}$, and $M^2 + M^2$ parameters in B and G.

It turns out that we need at least $l \geq 2M^2 - M(M+1)$ joint identification restrictions on B and G!

$$
\begin{array}{ll}
u_t = B\varepsilon_t & \text{volatility regime1} \\
u_t = (B+G)\varepsilon_t & \text{volatility regime2}
\end{array}
$$

Joint identification restrictions on B and G:

$$
\begin{pmatrix} vec(B) \\ vec(G) \end{pmatrix} = \begin{pmatrix} S_B & 0 \\ 0 & S_G \end{pmatrix} \begin{pmatrix} \beta \\ g \end{pmatrix} + \begin{pmatrix} s_B \\ s_G \end{pmatrix}
$$

where S_B, S_G, s_B and s_G have the usual meaning.

The crucial fact is that the matrices B (first volatility regime) and $(B+G)$ (second volatility regime) can be "full" or almost full, depending on the restrictions, see examples below.

Now IRFs **change with volatility regimes**:

$$
\begin{aligned}
IRF_{\bullet,\bullet}(h) &= \begin{cases} C_h B & \text{first volatility regime} \\ C_h(B+G) & \text{second volatility regime.} \end{cases} \\
h &= 0, 1, \ldots
\end{aligned}
$$

In this case, necessary and sufficient rank condition for identification are derived in Bacchiocchi and Fanelli (2015, OBES).

The idea is always the same: given

$$\begin{array}{ll} \mathbf{\Sigma}_{u,1} = BB' & \text{volatility regime1} \\ \mathbf{\Sigma}_{u,2} = (B+G)(B+G)' & \text{volatility regime2} \end{array}$$

one must establish conditions such that the unrestricted "free" elements in B and G can be uniquely recovered from the $M(M+1)$ free elements in $\mathbf{\Sigma}_{u,1}$ and $\mathbf{\Sigma}_{u,2}$.

Important observation: this framework **nests** Lanne and Lutkepohl's (2010, JMCB) approach.

For Λ taken as in LL, consider $G = B(\Lambda^{1/2} - I_M)$ ...

Example 10 DSGE-consistent monetary SVAR Consider the three-variable ($M = 3$) monetary policy SVAR based on $W_t{=}(y_t, \pi_t, R_t)'$, a break in the error covariance matrix at T_0 :

$$\underbrace{\begin{pmatrix} u_t^y \\ u_t^\pi \\ u_t^R \end{pmatrix}}_{u_t} = \left\{ \underbrace{\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}}_{B} + \begin{pmatrix} g_{11} & 0 & 0 \\ 0 & g_{22} & 0 \\ 0 & 0 & g_{33} \end{pmatrix} \times 1\left(t > T_0\right) \right\} \underbrace{\begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^\pi \\ \varepsilon_t^R \end{pmatrix}}_{\varepsilon_t}$$

We interpret ε_t^R as the 'monetary policy shock', ε_t^y as the 'output shock' and ε_t^π as the 'inflation shock'. The matrices B (first volatility regime) and $(B + G)$ (second volatility regime) are both "full" and the break changes only the impact of the shocks to variable i on variable i, $i = 1, ..., M$.

Example 11 Monetary SVAR with changing policy reaction function. Consider the same three-variable monetary policy SVAR of the previous example, and the structural specification

$$\underbrace{\begin{pmatrix} u_t^y \\ u_t^\pi \\ u_t^R \end{pmatrix}}_{u_t} = \left\{ \underbrace{\begin{pmatrix} b_{11} & 0 & b_{13} \\ 0 & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}}_{B} \right.$$

$$\left. + \begin{pmatrix} g_{11} & 0 & 0 \\ 0 & g_{22} & 0 \\ g_{31} & g_{32} & g_{33} \end{pmatrix} \times 1\,(t > T_0) \right\} \underbrace{\begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^\pi \\ \varepsilon_t^R \end{pmatrix}}_{\varepsilon_t}$$

This model describes an heteroskedastic SVAR in which, in addition to the non-triangular (non-recursive) structure, the last row of the specified matrix G accommodates changes in the parameters governing the response of the policy reaction function to the structural shocks.

In particular, the response on impact of the nominal short term interest rate R_t to the shocks ε_t^y, ε_t^π and ε_t^R, is postulated to change from the levels b_{3j} in the 'pre-break' regime, to the levels $b_{3j} + b_{3j}$ in the 'post-break' regime, $j = 1, 2, 3$. Moreover, this system shares with the 'DSGE-consistent monetary SVAR' presented in the Example 10 the fact that the monetary policy shock may have an instantaneous impact on the macroeconomic variables y_t and π_t in both volatility regimes, as implied by the non-zero specification of the last column of the matrix B.

Suggested readings for students who want to elaborate on this topic

Angelini, G., Bacchiocchi, E., Caggiano, G. and Fanelli, L. (2019), Uncertainty across volatility regimes, Uncertainty across volatility regimes, *Journal of Applied Econometrics* 34(3), 437-455.

Bacchiocchi, E. and Fanelli, L. (2015). Identification in Structural Vector Autoregressive models with structural changes, with an application to U.S. monetary policy, *Oxford Bulletin of Economics and Statistics* 77, 761-779.

Bacchiocchi, E., Castelnuovo, E., and Fanelli, L. (2018). Give me a break! Identification and estimation of the macroeconomic effects of monetary policy shocks in the U.S., *Macroeconomic Dynamics* 22(6), 1613-1651.

Lanne, M., & Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* 40, 1131-1149.

Lütkepohl, H., & Netšunajev, A. (2017). Structural vector autoregressions with heteroskedasticity: A review of different volatility models. *Econometrics and Statistics* 1, 2-18.

A **very recent** -robust - approach to identification through heteroskedasticity may be found in:

Lewis, D. J. (2021), Identifying Shocks via Time-Varying Volatility, *Review of Economic Studies*, forthcoming.

11.5-The "external variables" approach

The external variables approach is motivated by the idea that in many applications we are interested in the identification of only a subset $k < M$ of the structural shocks, not all of them.

Very often (but not always), we are interested in just one shock, $k = 1$.

Many cases of interest are based on $k = 2$ structural shocks; see Fanelli and Marsi (2022, EER) for $k = 3$ (monetary policy in the Euro area).

This is a **partial identification** logic that also the sign-restrictions approach can address.

The two most popular approaches that make use of external instruments for the indentification of targest structural shocks are the **Proxy-SVAR approach** and the **LPs approach**.

11.5.1-Proxy-SVARs, identification and estimation

The proxy-SVAR approach is motivated by the idea that in many applications we are interested in the identification of only a subset $k < M$ of the structural shocks, not all of them.

Example:

$$\begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_t^{FU} \\ \varepsilon_t^{MU} \\ \varepsilon_t^{y} \end{pmatrix} \begin{array}{l} \text{financial unc. shock} \\ \text{macro uncer. shock} \\ \text{real econ. activity shoc} \end{array}$$

$$= (b_1 \vdots b_2 \vdots b_3) \begin{pmatrix} \varepsilon_t^{FU} \\ \varepsilon_t^{MU} \\ \varepsilon_t^{y} \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix}}_{b_1} \varepsilon_t^{FU} + \underbrace{\begin{pmatrix} b_{12} \\ b_{22} \\ b_{32} \end{pmatrix}}_{b_2} \varepsilon_t^{MU} + \underbrace{\begin{pmatrix} b_{13} \\ b_{23} \\ b_{33} \end{pmatrix}}_{b_3} \varepsilon_t^{y}.$$

$$\begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_{11} \\ \mathbf{b}_{21} \\ \mathbf{b}_{31} \end{pmatrix} \underset{\text{interest}}{\underset{\uparrow}{\varepsilon_t^{FU}}} + \begin{pmatrix} b_{12} & b_{13} \\ b_{22} & b_{23} \\ b_{32} & b_{33} \end{pmatrix} \underset{\text{not of int.}}{\underset{\uparrow}{\begin{pmatrix} \varepsilon_t^{MU} \\ \varepsilon_t^{y} \end{pmatrix}}}$$

Imagine we are interested in the identification of the financial uncertainty shock ε_t^{FU} alone, **regardless of the other structural shocks of the system** $(\varepsilon_t^{MU}, \varepsilon_t^{y})$. Thus, we are solely interested in the IRFs:

$$IRF_{\bullet,1}(h) = C_h b_1 \ , \ \ h = 0, 1, ...$$

where b_1 is the column in bold above.

In other words, we want to take a stand only **on a subset of the structural shocks of the system** and we do not want to think about the other strutural shocks which are not of primary interest.

This is called a **partial identification approach** (you want to get rid of $\varepsilon_{2,t} = (\varepsilon_t^{MU}, \varepsilon_t^{y})'$).

The idea of the proxy-SVAR approach is to find an **observable** variable, say Z_t, called **instrument or proxy**, "external" to the VAR (i.e. not included in the original VAR system), which satisfies these two conditions:

$$Cov(Z_t, \varepsilon_t^{FU}) = E(Z_t \varepsilon_t^{FU}) = \phi, \;\; |\phi| >> 0 \quad \text{relevance conditio}$$
$$E\left\{ Z_t \begin{pmatrix} \varepsilon_t^{MU} \\ \varepsilon_t^{y} \end{pmatrix} \right\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{exogeneity conditio}$$

The external instrument **must be correlated with the structural shock of interest (relevance condition)** and **must be uncorrelated with the non-instrumented shocks (exogeneity of orthogonality condition).**

Thus, the idea is the same as IV regressions, the main difference being that now **the instrumented variables is an unobservable shocks!**

A conveniet way to formalize the relevance and exogeneity conditions is to write the linear measurement equation:

$$Z_t = \phi\varepsilon_t^{FU} + \underbrace{\mathbf{0} \times \varepsilon_t^{MU}}_{\text{exogeneity}} + \underbrace{\mathbf{0} \times \varepsilon_t^{y}}_{\text{exogeneity}} + \underbrace{\sigma_\omega \omega_t^{\circ}}_{\omega_t}$$

$$= \phi\varepsilon_{1,t} + \omega_t$$

where ω_t° is a normalized **measurement error** indipendent on ε_t, $\sigma_\omega > 0$.

This equation shows that the proxy Z_t carries information on the latent structural shock of interest ε_t^{FU} (through the parameter ϕ), up to a measurement error.

In this example, we have considered a setup with **one shock of interest** ($k = 1$) and **one proxy alone** ($r = 1$), which is the case often considered in applied research.

The analysis can be generalized to **the multiple instrument-multiple shocks** framework, $r \geq k \geq 1$, where

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \begin{array}{cl} k & \text{structural shocks of interest} \\ M-k & \text{structural shocks not of interest} \end{array}$$

and the relavance and orthogonality conditions become:

$$\begin{array}{ll} E(Z_t \varepsilon_{1,t}') = \Phi \neq 0_{r\times k} \quad , \quad rank(\Phi) = k & \text{relevance} \\ E\left(Z_t \varepsilon_{2,t}'\right) = 0_{r\times(M-k)} & \text{exogeneity} \end{array}.$$

Here (V_ω pos. def.):

$$\underset{r\times 1}{Z_t} = \underset{r\times k}{\Phi}\ \underset{k\times 1}{\varepsilon_{1,t}} + \underset{r\times r}{V_\omega}\ \underset{r\times 1}{\omega_t^\circ}$$

To simplify exposition, we continue for the moment with the example based on $r = k = 1$.

By multiplying both sides of equation

$$Z_t = \phi \varepsilon_t^{FU} + \sigma_\omega \omega_t^\circ$$

by the VAR disturbances we obtain:

$$Z_t u_t' = \phi \varepsilon_t^{FU} u_t' + \sigma_\omega \omega_t^\circ u_t'$$

and taking expectations (and using the fact that $E(\omega_t u_t') = E(\omega_t^\circ \varepsilon_t' B') = 0_{1\times M}$):

$$E(Z_t u_t') = \phi E(\varepsilon_t^{FU} u_t')$$

$$= \phi E \left\{ \varepsilon_t^{FU} \left[\varepsilon_t^{FU} b_1' + \varepsilon_t^{FU} \varepsilon_t^{MU} b_2' + \varepsilon_t^{FU} \varepsilon_t^{y} b_3' \right] \right\}$$

$$= \phi b_1'.$$

Summing up, setting $E(u_t Z_t) = \mathbf{\Sigma}_{u,Z}$, we have the moment condition:

$$\mathbf{\Sigma}_{u,Z} = \phi b_1.$$

Given the relationship (obtained under the assumed conditions on the external instrument)

$$\mathbf{\Sigma}_{u,Z} = \phi b_1$$

the covariance $\mathbf{\Sigma}_{u,Z}$ can be easily estimated from the reduced form VAR innovations and the proxy.

Indeed, the "natural" estimator of the population moment, $E(u_t Z_t)$, is its sample counterpart:

$$\hat{\mathbf{\Sigma}}_{u,Z} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t Z_t$$

where $\hat{u}_t$ are the VAR residuals:

$$\hat{u}_t = W_t - \hat{\mathbf{\Pi}}_T X_t \quad , \quad t = 1, ..., T.$$

Under stationary processes and **fairly general conditions**:

$$\hat{\mathbf{\Sigma}}_{u,Z} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t Z_t \xrightarrow{p} \mathbf{\Sigma}_{u,Z} = E(u_t Z_t).$$

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T} vec(\hat{u}_t Z_t) \xrightarrow{d} N(vec(\mathbf{\Sigma}_{u,Z}) \ , \ V)$$

Thus, we have a consistent estimate of ϕb_1 given by

$$\hat{\mathbf{\Sigma}}_{u,Z} = \widehat{\phi b_1} \xrightarrow{p} \mathbf{\Sigma}_{u,Z} = \phi b_1$$

which corresponds to an estimate of b_1 "up-to-scale" ϕ, where the scale is here given by the relevance parameter ϕ.

This result holds regardless of the strength of the external instruments.

To get separate estimates of ϕ and b_1, there are two possible ways to proceed.

One is incorporate the information stemming from the "standard" VAR moment restrictions, i.e.: $\mathbf{\Sigma}_u = BB'$, in the analysis, obtaining **novel additional moment conditions**;

The other is is to impose a 'unit effect normalization'.

However, we look for a separate estimator of the on-impact effect b_1, **the strength of the proxy (its connection to the target shock) plays an important role in inference.**

I'll sqy something on this later on, using the blackboard.

Adding moment conditions

We exploi the information from $\mathbf{\Sigma}_u = BB'$.

We maintain that $Cov(Z_t, \varepsilon_{1,t}) = \phi$, $|\phi| >> 0$ (strong proxy, intuitive characterization).

Noting that $\mathbf{\Sigma}_{Z,u} = \mathbf{\Sigma}'_{u,Z} = \phi b'_{1.}$, it holds the relationship

$$\mathbf{\Sigma}_{Z,u} (\mathbf{\Sigma}_u)^{-1} \mathbf{\Sigma}_{u,Z} = \phi^2$$

$$= \phi b_1' (BB')^{-1} b_1 \phi = \phi b_1' (B')^{-1} B^{-1} b_1 \phi.$$

Now, since $B = (b_1 \vdots\, B_2)$ and

$$B^{-1}B = B^{-1}(b_1 \vdots B_2) = (B^{-1}b_1 \vdots B^{-1}B_2) = I_M,$$

it must hold:

$$B^{-1}b_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad b_1'(B')^{-1} = (1, 0, ..., 0)$$

so that

$$\Sigma_{Z,u} (\Sigma_u)^{-1} \Sigma_{u,Z} = \phi b_1'(B')^{-1}B^{-1}b_1\phi = \phi^2$$

which **provide additional moment conditions** (in this case an additional one) which can be used to estimate ϕ and b_1 separately.

Indeed, by using the expression

$$\mathbf{\Sigma}_{Z,u}(\mathbf{\Sigma}_u)^{-1}\mathbf{\Sigma}_{u,Z} = \phi^2$$

we obtain

$$\hat{\phi} = \mp \left\{ \hat{\mathbf{\Sigma}}_{Z,u} \left(\hat{\mathbf{\Sigma}}_u \right)^{-1} \hat{\mathbf{\Sigma}}_{u,Z} \right\}^{1/2}$$

where the sign of $\hat{\phi}$ is determined by the practitioner on the basis of his/her knowledge about the link between the proxy and the structural shocks of interest.

Given $\hat{\phi}$, we can solve $\mathbf{\Sigma}_{u,Z} = \phi b_1$ for b_1 and obtain

$$\hat{b}_1 = \frac{1}{\hat{\phi}} \hat{\mathbf{\Sigma}}_{u,Z}.$$

Unit effect normalization: estimation of relative responses

It is instructive to examine also the estimation method suggested by Stock and Watson (2018).

They only consider the moments $\Sigma_{u,Z} = \phi b_1$ **and an additional normalization condition.**

Consider the following partition of $\Sigma_{u,Z} = \phi b_1$:

$$\begin{array}{c} 1\times 1 \\ (M-1)\times 1 \end{array} \quad \left(\begin{array}{c} \Sigma_{u_1,Z} \\ \Sigma_{u_2,Z} \end{array}\right) = \phi \left(\begin{array}{c} b_{11} \\ b_{21} \end{array}\right) \quad \begin{array}{c} 1\times 1 \\ (M-1)\times 1 \end{array}$$

where, recall, b_{11} captures the instantaneous impact of the target shock (financial uncertanty shock ε_t^{FU}) on the first variable in Y_t, which is financial uncertainty UF_t.

If we want to normalize the **instantaneous impact** of ε_t^{FU} on UF_t to 1, we can write

$$\begin{pmatrix} \Sigma_{u_1,Z} \\ \Sigma_{u_2,Z} \end{pmatrix} = \begin{pmatrix} \phi b_{11} \\ \phi b_{21} \end{pmatrix} = \phi b_{11} \begin{pmatrix} 1 \\ b_{21}/b_{11} \end{pmatrix}$$

$$= \begin{pmatrix} \tilde{\phi} \\ \tilde{\phi}\tilde{b}_{21} \end{pmatrix} \quad , \qquad \begin{matrix} \tilde{\phi} = \phi b_{11} \\ \tilde{b}_{21} = b_{21}/b_{11} \end{matrix}$$

$$= \begin{pmatrix} 1 \\ \tilde{b}_{21} \end{pmatrix} \Sigma_{u_1,Z} \;\Rightarrow\; \Sigma_{u_2,Z} = \tilde{b}_{21}\Sigma_{u_1,Z}$$

from which we obtain the representation:

$$\tilde{b}_{21} = \frac{1}{\Sigma_{u_1,Z}}\Sigma_{u_2,Z}.$$

Notice that $\tilde{b}_{21}$ captures the **relative impact** of the shock ε_t^{FU} on the $(M-1)$ variables of the system, **given the 'unit effect' normalization.** It follows that, by replacing $\Sigma_{u_1,Z}$ and $\Sigma_{u_2,Z}$ by their consistent estimates:

$$\widehat{\tilde{b}_{21}} = \frac{1}{\hat{\Sigma}_{u_1,Z}} \times \hat{\Sigma}_{u_2,Z}$$

Intuition on the concept of strong and weak proxies and some implications for inference:

BLACKBOARD

Minimum distance approach

Angelini and Fanelli (2019, JAE) suggest addressing the estimation of the proxy-SVAR by a method alternative to IV methods, known "classical minimum distance" (CMD).

This method is particularly useful to understand what happens in the presence of **multiple target shocks:** $r \geq k \geq 2$.

We start by providing the main intuition for the case $r = k = 1$, already discussed above.

Consider the moment conditions (alredy discussed):

$$\begin{aligned}\Omega_Z &= \phi^2 \\ \Sigma_{u,Z} &= \phi b_1\end{aligned}$$

where $\Omega = \Sigma_{Z,u}(\Sigma_u)^{-1}\Sigma_{u,Z}$ is a scalar when $r = k = 1$.

(Ω_Z is a nonlinear function of the reduced form parameters $\Sigma_{Z,u}$ and Σ_u).

Again, we maintain that $Cov(Z_t, \varepsilon_{1,t}) = \phi$, $|\phi| >> 0$

The matrices Ω_Z and $\Sigma_{u,Z}$ in the left-hand side involve $1 + M$ reduced form parameters and the parameters ϕ and in b_1 in the right-hand-side are $1 + M$.

For simplicity, we collect the $1 + M$ parameters of the proxy-SVAR in the vector $\theta = (\phi, b_1')'$ and summarize the moment conditions above in the (nonlinear) function $\gamma = g(\theta)$, where

$$\gamma = vec\begin{pmatrix}\Omega_Z \\ \Sigma_{u,Z}\end{pmatrix} \quad ; \quad g(\theta) = vec\begin{pmatrix}\phi^2 \\ \phi b_1\end{pmatrix}.$$

The quantity

$$\gamma - g(\theta) = 0_{(M+1)\times 1}$$

defines a 'distance' between the reduced form parameters in γ and the 'structural and relevance' parameters in θ.

We can estimate the reduced form paramaters in γ consistently from the data because, under stationarity and fairly general conditions:

$$\hat{\Sigma}_{u,Z} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t Z_t \rightarrow_p E(u_t Z_t)$$

$$\hat{\Omega}_Z = \hat{\Sigma}_{Z,u}\left(\hat{\Sigma}_u\right)^{-1}\hat{\Sigma}_{u,Z} \rightarrow_p E(Z_t u_t)\left\{E(u_t u_t')\right\}^{-1} E(u_t Z_t)$$

$$= \Sigma_{Z,u}\left(\Sigma_u\right)^{-1}\Sigma_{u,Z} = \phi b_1'(B')^{-1}B^{-1}b_1\phi = \phi^2$$

which implies that the **plug-in estimator** $\hat{\gamma}_T$ is consistent:

$$\hat{\gamma}_T = vec\begin{pmatrix} \hat{\Omega}_Z \\ \hat{\Sigma}_{u,Z} \end{pmatrix} \xrightarrow{p} \gamma_0 = \begin{pmatrix} \phi_0^2 \\ \phi_0 b_{1,0} \end{pmatrix}.$$

Moreover, it is possible to show that under stationarity and the usual regularity conditions on the VAR, **plus the strong proxy condition**, it holds

$$T^{1/2}(\hat{\gamma}_T - \gamma_0) \xrightarrow{d} N(0_{(M+1)\times 1}, V_\gamma)$$

where the asymptotic covariance matrix V_γ can be estimated consistently from the data (details in Angelini and Fanelli, 2019, JAE).

Summing up, we have a 'distance'

$$\gamma - g(\theta) = 0_{(M+1)\times 1}$$

and an asymptotic Gaussian estimator of the reduced form parameters γ if a strong proxy condition holds.

Given a consistent estimate of the asymptotoc covariance matrix V_γ, say $\hat{V}_\gamma$, the CMD estimator of θ is obtained from the solution to the following problem

$$\min_\theta \left\{\hat{\gamma}_T - g(\theta)\right\}' \left(\hat{V}_\gamma\right)^{-1} \left\{\hat{\gamma}_T - g(\theta)\right\}.$$

Let $\hat{\theta}_T$ be the estimator obtained (numerically) as

$$\hat{\theta}_T = \arg\min_{\theta} \{\hat{\gamma}_T - g(\theta)\}' \left(\hat{V}_\gamma\right)^{-1} \{\hat{\gamma}_T - g(\theta)\}.$$

It is possible to show that if the $(M+1) \times (M+1)$ Jacobian matrix:

$$H(\theta) = \frac{\partial g(\theta)}{\partial \theta'} = \begin{pmatrix} 2\phi & 0_{1\times M} \\ b_1 & \phi I_n \end{pmatrix}$$

has **full rank** in a neighborhood of the true value θ_0, then:

$$\hat{\theta}_T \xrightarrow{p} \theta_0$$

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0_{(M+1)\times 1}, V_\theta)$$

where $V_\theta = \left\{H'(\theta_0)\,(V_\gamma)^{-1}\,H(\theta_0)\right\}^{-1}$.

Intuition on what happens to

$$H(\theta) = \frac{\partial g(\theta)}{\partial \theta'} = \begin{pmatrix} 2\phi & 0_{1\times M} \\ b_1 & \phi I_n \end{pmatrix}$$

when $\phi_0 = \frac{c}{T^{1/2}}$, $|c| < \infty$ (local-to-zero proxy):

$$rank[H(\theta_0)] < (M+1)$$

Standard asymptotic theory breaks down.

Minimum distance approach, multiple target shocks

In the presence **of multiple target shocks** (assume for simplicity $r = k \geq 1$), the situation is:

$$u_t = \underbrace{B_1}_{n\times k} \varepsilon_{1,t} + \underbrace{B_1}_{n\times(n-k)} \varepsilon_{2,t}$$

$$\underset{k\times 1}{Z_t} = \underset{k\times k}{\Phi} \underset{k\times 1}{\varepsilon_{1,t}} + \underset{k\times k}{V_\omega} \underset{k\times 1}{\omega_t^\circ}$$

so that the relevant moment conditions are obtained by taking expectations of:

$$u_t Z_t' = B_1 \varepsilon_{1,t} Z_t' + B_1 \varepsilon_{2,t} Z_t'$$

obtaining

$$\underbrace{\Sigma_{u,Z}}_{n\times k} = B_1 \Phi'$$

$$\underbrace{\Omega_Z}_{k\times k} = \Phi\Phi'$$

In the absence of restrictions on B_1 and Φ, there are $nk + \frac{1}{2}k(k+1)$ reduced form parameters in

$$\mathbf{\Sigma}_{u,Z} \text{ and } \Omega_Z$$

easily estimable from the data, and $nk+k^2$ parameters in

$$B_1 \text{ and } \Phi.$$

Thus, at least $\frac{1}{2}k(k-1)$ restrictions must be placed on

$$G = \begin{pmatrix} B_1 \\ \mathbf{\Phi} \end{pmatrix}$$

to identify the k target structural shocks.

Restrictions on G :

$$vec\begin{pmatrix} B_1 \\ \Phi \end{pmatrix} = vec(G) = S\theta + s$$

where:

θ vector $m \times 1$ of "free" (unconstrained) parameters in G;

S known selection matrix (selects) with column rank $= \dim(\theta) = m$;

s known vector: imagine we want to set one parameter to 2.08.

Thus,

$$\gamma = \underbrace{\begin{pmatrix} vec(\boldsymbol{\Sigma}_{u,Z}) \\ vech(\boldsymbol{\Omega}_Z) \end{pmatrix}}_{\left(nk+\frac{1}{2}k(k+1)\right)\times 1} \quad \text{easily estimable fron the data}$$

$$g(\theta) = \begin{pmatrix} vec(B_1\Phi') \\ vech(\Phi\Phi') \end{pmatrix} \quad \text{where } G \text{ restricted as above}$$

The proxy-SVAR is thus condensed in the distance function

$$\gamma - g(\theta) = 0_{\left(nk+\frac{1}{2}k(k+1)\right)\times 1}$$

hence

$$\hat{\theta}_T = \arg\min_{\theta} \left\{\hat{\gamma}_T - g(\theta)\right\}' \left(\hat{V}_\gamma\right)^{-1} \left\{\hat{\gamma}_T - g(\theta)\right\}.$$

where

$$\hat{\gamma}_T = \left(\begin{array}{c} vec(\hat{\Sigma}_{u,Z}) \\ vech(\hat{\Omega}_Z) \end{array} \right);\ \hat{V}_\gamma \text{ estimator of } V_\gamma.$$

The identification of θ in this case depends on the rank of the Jacobian matrix:

$$H(\theta) = \frac{\partial g(\theta)}{\partial \theta'} \qquad \left(nk + \frac{1}{2}k(k+1)\right) \times m$$

has **full rank** in a neighborhood of the true value θ_0, then:

$$\hat{\theta}_T \xrightarrow{p} \theta_0$$

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0_{m\times 1}, V_\theta)$$

where $V_\theta = \left\{H'(\theta_0)\,(V_\gamma)^{-1}\,H(\theta_0)\right\}$.

Suggested readings to complete this topic

Notes by Michele Piffer, available https://drive.google.com/fi

Angelini, G. and Fanelli, L. (2019), Exogenous uncertainty and the identification of Structural Vector autoregressions with external instruments, *Journal of Applied Econometrics*, forthcoming.

Gertler, M. and Karadi, P. (2015), Monetary Policy Surprises, Credit Costs,

and Economic Activity, *American Economic Journal: Macroeconomics* 2015, 7: 44–76.

Lakdawala, A. (2019), Decomposing the Effects of Monetary Policy Using an External Instruments SVAR, *Journal of Applied Econometrics*, forthcoming.

Mertens, K. and Ravn, M. (2013), The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* 103, 1212-1247.

Mertens, K. and Ravn, M. (2014), A reconciliation of SVAR and narrative estimates of tax multipliers, *Journal of Monetary Economics* 68, S1-S19.

Stock, J.H., & Watson, M.W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal* 128, 917-948.

11.5.2 Local Projections (LP) - work in progress part

IRFs are key object of interest in empirical analysis because they track the impact of identified shocks of interest on target variables over time.

It is becoming increasingly popular to estimate IRFs of interest using the **method of local projections (LPs).**

LPs are simple linear regressions of a future outcome on current covariates, conditional on some exogenous controls.

We introduce LPs in a very simple and intuitive way, which is the way LPs are typically used and interpred in the current empirical macroeconomic literature.

Let y_t be the **scalar target or response variable**, i.e. the variable whose dynamic response to a "shock" (broadly defined) is the object of interest (ex. GDP growth, unemployment, etc.);

Let x_t be an **observable scalar** variable that **directly measures a shock** at time t (e.g. a MP shock, a fiscal shock, an uncertainty shock), or that **approximates a shock at time t with measurement error**; or that represents and **indicator/instrument** of a policy, an indicator of uncertainty, etc.

Let f_t be **a vector of control variables;** (they can also include deterministic variables);

Let $q_t = (y_t, x_t, f_t')'$;

Finally, we denote with $\mathcal{F}_{t-1}^q$ the information set spanned by $q_1, ..., q_{t-1}$.

It is assumed that the the stochastic process $\{q_t\}$ is **covariance stationary**.

We call LPs the sequences of $H+1$ regressions:

$$y_{t+h} = \underbrace{c_h + \beta_h x_t + \gamma_h' f_t + \sum_{\ell=1}^{p} \delta_{h,\ell}' q_{t-\ell}}_{\Pr oj(y_{t+h}|x_t, f_t, \mathcal{F}_{t-1}^q)} + \underbrace{\eta_{t+h}}_{\equiv \eta_{h,t}}$$

for $h = 0, 1, ..., H$.

The term η_{t+h}, often indicated as $\eta_{h,t}$ or $\eta_{t,h}$, is called **projection error.**

For $h = 0$, $\eta_t \equiv \eta_{0,t} \equiv u_t \sim \text{WN}(0, \sigma_u^2)$.

The sequence β_0, $\beta_1, ..., \beta_H$ are the **population IRFs**.

More precisely, for $h = 0, 1, ..., H$:

$$\beta_h = E(y_{t+h} \mid x_t = 1, f_t, \mathcal{F}_{t-1}^q) - E(y_{t+h} \mid x_t = 0, f_t, \mathcal{F}_{t-1}^q)$$

tracks the impact of a unit change in x_t on y_{t+h}, **conditional on the controls and the past.**

A short way of writing the LPs is (Stock and Watson, 2018), for $h = 0, 1, ..., H$:

$$y_{t+h} = c_h + \beta_h x_t + \text{controls} + \eta_{t+h}$$

where

$$\text{controls}=\underbrace{\gamma_h' f_t + \sum_{\ell=1}^{p} \delta_{h,\ell}' q_{t-\ell}}_{\text{controls}}.$$

Note that, for $h = 1, ..., H$, the projection errors η_{t+h} in the model

$$y_{t+h} = c_h + \beta_h x_t + \text{controls} + \eta_{t+h}$$

have MA(h) structure.

Why?

Obviously, for $h = 0$, $\eta_t \equiv u_t$ ~WN$(0, \sigma_u^2)$ is MA(0)≡WN.

To simplify exposition, we provide the underlying intuition by considering the simplest DGP of a dynamic regression model:

$$y_t = \underbrace{\rho y_{t-1}}_{\beta_0 x_t} + u_t \quad , \quad u_t \sim WN(0, \sigma_u^2).$$

Consider the projection of y_{t+1} onto $\mathcal{F}_{t-1}$:

$$E\left(y_{t+1} \mid \mathcal{F}_{t-1}\right) = E\left(E\left(y_{t+1} \mid \mathcal{F}_t\right) \mid \mathcal{F}_{t-1}\right)$$

$$E\left(\rho y_t + u_{t+1} \mid \mathcal{F}_{t-1}\right) = \underbrace{\rho^2 y_{t-1}}_{\beta_1 x_t}$$

Then consider the decomposition

$$y_{t+1} = E\left(y_{t+1} \mid \mathcal{F}_{t-1}\right) + \eta_{t+1}.$$

The implied projection error is

$$\eta_{t+1} = y_{t+1} - E\left(y_{t+1} \mid \mathcal{F}_{t-1}\right)$$

$$= \underbrace{y_{t+1} - E\left(y_{t+1} \mid \mathcal{F}_t\right)}_{u_{t+1}} + \underbrace{E\left(y_{t+1} \mid \mathcal{F}_t\right)}_{\rho y_t} - \underbrace{E\left(y_{t+1} \mid \mathcal{F}_{t-1}\right)}_{\rho^2 y_{t-1}}$$

$$= u_{t+1} + \rho y_t - \rho^2 y_{t-1}$$

$$= u_{t+1} + \rho(y_t - \rho y_{t-1}) = u_{t+1} + \rho u_t \sim \mathrm{MA}(1)$$

The argument can be generalized recursively: when we consider the projection of y_{t+h} onto $\mathcal{F}_{t-1}$, we obtain the model:

$$y_{t+h} = \beta_h y_{t-1} + \eta_{t+h}$$

where $\beta_h = \rho^{h+1}$ ($\beta_0 \equiv \rho$) and η_{t+h} ~MA(h).

Estimation issues.

Assume we have a method to estimate, for $h = 0, 1, ..., H$ the coefficients β_h.

The method will be either OLS or IV; see below.

Let $\hat{\beta}_h$ be the point estimates, $h = 0, 1, ..., H$.

Then confidence intervals for the responses of interest are given by

$$\hat{\beta}_h \pm quan_{1-\alpha} \times \text{s.e.}(\hat{\beta}_h) \quad , \quad h = 0, 1, ..., H$$

where $quan_{1-\alpha}$ is the $1-\alpha$ percentile taken from the $N(0,1)$ distribution and s.e.$(\hat{\beta}_h)$ is typically an **HAC standard error** associated with $\hat{\beta}_h$.

Recall that $\eta_{t+h} \sim$MA(h), hence we need autocorrelation adjustment.

In many applications homoskedasticity is not obvious!

Then the relevant question for estimation is: when x_t is correlated with η_{t+h}?

To estimate the IRFs $\beta_0, \beta_1, ..., \beta_H$, it is crucial to know whether (recall $\eta_t = u_t$)

$$E(x_t \eta_t) = \begin{cases} 0 \\ \neq 0 \end{cases}$$

$$\Downarrow$$

$$E(x_t \eta_{t+h}) = \begin{cases} 0 \\ \varrho_h \neq 0 \end{cases}$$

When we can claim that $E(x_t \eta_t) \equiv E(x_t u_t) = 0$, then we have LP-OLS!

When we can claim that $E(x_t \eta_t) \equiv E(x_t u_t) \neq 0$, then we have LP-IV!

To understand the previous claim, consider, e.g., that

for $h = 1$, since η_{t+1} ~MA(1):

$$E(x_t \eta_{t+1}) = E(x_t[u_{t+1} + \rho_1 u_t])$$

$$= \underbrace{E(x_t u_{t+1})}_{0} + \theta E(x_t u_t) = \begin{cases} 0 & \text{if } E(x_t u_t) = 0 \\ \varrho_1 \neq 0 & \text{if } E(x_t u_t) \neq 0 \end{cases} \quad ;$$

For $h = 2$, since η_{t+1} ~MA(2):

$$E(x_t \eta_{t+2}) = E(x_t[u_{t+2} + \rho_{11} u_{t+1} + \rho_2 u_t])$$

$$= \underbrace{E(x_t u_{t+2})}_{0} + \underbrace{\theta_1 E(x_t u_{t+1})}_{0} + \theta_2 E(x_t u_t)$$

$$= \begin{cases} 0 & \text{if } E(x_t u_t) = 0 \\ \varrho_2 \neq 0 & \text{if } E(x_t u_t) \neq 0 \end{cases}$$

The argument can be generalized:

$$E(x_t \eta_t) = \begin{cases} 0 \\ \neq 0 \end{cases}$$

$$\Downarrow$$

$$E(x_t \eta_{t+h}) = \begin{cases} 0 \\ \varrho_h \neq 0 \end{cases}$$

Cases where external instruments are needed: measurement errors

In many cases, the resercher observes $\tilde{x}_t$, where

$$\tilde{x}_t = \underbrace{x_t}_{\text{true value}} + \underbrace{m_t}_{\text{measurement error}}$$

x_t **is the true (unobserved) shock** and m_t is a **measurement error** (assumed orthogonal to the true shock and all other shocks potentially playing a role).

A typical case is when one uses narrative counterpart of shocks; think e.g. about Romer & Romer MP narratives of fiscal narratives;

If this is the case, consider the LP model for $h = 0$:

$$y_t = c_0 + \beta_0 \tilde{x}_t + \text{controls} + \eta_t.$$

$$y_t = c_0 + \beta_0 (x_t + m_t) + \text{controls} + \eta_t.$$

The model can be re-written as

$$y_t = c_0 + \beta_0 \tilde{x}_t + \text{controls} + v_t \quad , \quad v_t = \eta_t + \beta_0 m_t$$

and, therefore,

$$Cov(\tilde{x}_t, v_t) = Cov(x_t + m_t, \eta_t + \beta_0 m_t)$$

$$= \underbrace{Cov(x_t, \eta_t)}_{0} + \beta_0 \underbrace{Cov(x_t, m_t)}_{0}$$

$$+ \underbrace{Cov(m_t, \eta_t)}_{0} + \beta_0 Var(m_t) \neq 0$$

.

This leads to biased OLS estimation of β_0.

The bias remains also for the LPs obtained for $h = 1, ..., H$, because, as we have already seen, the projection errors η_{t+h} depend on u_t via their MA-type dependence structure.

The common fix to this problem is to consider **an instrument (proxy)** z_t for $\tilde{x}_t$ such that:

$$E(\tilde{x}_t z_t) \neq \phi \ , \quad \mid \phi \mid >> 0 \quad \text{, relevance}$$

$$E(z_t \eta_t) = E(z_t u_t) = 0 \quad \text{exogeneity}$$

so that one can estimate the model

$$y_{t+h} = c_h + \beta_h \tilde{x}_t + \text{controls} + v_{t+h}$$

by using z_t as an instrument for $\tilde{x}_t$.

So one can apply 2SLS.

LP-IV as opposed to LP-OLS.

Summing up, we can estimate IRFs from LP either by OLS or IV.

In the first case we have to be sure that x_t represents a measure of the shock of interest which is not seriously affected by measurement errors.

There are also cases in which x_t represents an indicator of policy, or there might be issues about reverse causality.

In the second and third cases, we need an instrument for x_t.

In both cases, the efficient estimation standard errors (which are required to compute confidence bands for the IRFs) must take into account the correlation structure (MA-type) of the projection errors, otherwise we may obtain inflationated confidence intervals.

Implementing LP-OLS or LP-IV is very easy with every econometric package.

It is important to remark that the method of local projections provides an alternative to SVARs (proxy-SVARs) for the estimation of IRFs, with a small caveat.

From the VMA representation of the AR(1) we have:

$$\frac{\partial y_{t+h}}{\partial u_t} = E(y_{t+h} \mid u_t = 1, \mathcal{F}_{t-1}) - E(y_{t+h} \mid \mathcal{F}_{t-1})$$

$$\begin{aligned} = \quad & \rho^h(u_t = 1) + \rho^{h+1}u_{t-1} + \rho^{h+2}u_{t-2} + \ldots \\ & -(\rho^{h+1}u_{t-1} + \rho^{h+2}u_{t-2} + \ldots) \end{aligned}$$

$$= \rho^h$$

while from LPs we have that ρ^{h+1} is the coefficient associated with the projection of y_{t+h} onto $\mathcal{F}_{t-1}$.

Suggested readings to complete this topic

Jordà, O. (2005), Estimation and Inference of Impulse Responses by Local Projections, *American Economic Review* 95, 161-182.

Plagborg-Møller, M. and Christian K. Wolf. (2020), Local Projections and VARs Estimate the Same Impulse Responses, *Econometrica*, forthcoming.

Stock, J.H., & Watson, M.W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal* 128, 917-948.