The Identification of Dynamic Structural Shocks

Kenza Benhima and Jean-Paul Renne

2025-10-31

Contents

The Identification of Dynamic Structural Shocks

The identification and estimation of dynamic responses to structural shocks is one of the principal goals of macroeconometrics. These responses correspond to the effect, over time, of an exogenous intervention that propagates through the economy, as modeled by a system of simultaneous equations.

Over the last decades, several methodologies have been proposed so as to estimate these responses. The objective of this course, developed by Kenza Benhima and Jean-Paul Renne, is to provide an exhaustive view of these methodologies and to provide students with tools enabling them to implement them in various contexts.

Codes associated with this course are part of the `IdSS` package (**Id**entification of **S**tructural **S**hocks), which is available on GitHub. To load a package from GitHub, you need to use function `install_github` from the `devtools` package:

```
install.packages("devtools") # devtools allows to use "install_github"
library(devtools)
install_github("jrenne/IdSS")
library(IdSS)
```

Useful (R) links:

- Download R:
  - R software: https://cran.r-project.org (the basic R software)
  - RStudio: https://www.rstudio.com (a convenient R editor)
- Tutorials:
  - Rstudio: https://dss.princeton.edu/training/RStudio101.pdf (by Oscar Torres-Reyna)
  - R: https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf (by Emmanuel Paradis)
  - Our own tutorial: https://jrenne.shinyapps.io/Rtuto_publiShiny/

Chapter 1

VARs and IRFs: the basics

Often, impulse response functions (IRFs) are generated in the context of vectorial autoregressive (VAR) models. This section presents these models and show how they can be used to compute IRFs.

1.1 Definition of VARs (and SVARMA) models

Definition 1.1 ((S)VAR model)**.** Let y_t denote a $n \times 1$ vector of (endogenous) random variables. Process y_t follows a p^{th}-order (S)VAR if, for all t, we have

$$\begin{array}{rrcl} VAR: & y_t & = & c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \varepsilon_t, \\ SVAR: & y_t & = & c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + B\eta_t, \end{array} \tag{1.1}$$

with $\varepsilon_t = B\eta_t$, where $\{\eta_t\}$ is a white noise sequence whose components are mutually and serially independent, and that satisfies $\mathbb{V}ar(\eta_t) = Id$.

The first line of Eq. (1.1) corresponds to the **reduced-form** of the VAR model (**structural form** for the second line). While the **structural shocks** (the components of η_t) are mutually uncorrelated, this is not the case of the **innovations**, that are the components of ε_t. However, in both cases, vectors η_t and ε_t are serially correlated (through time).

Notice that $\mathbb{V}ar(\eta_t) = Id$ and $\varepsilon_t = B\eta_t$ jointly imply that

$$\Omega := \mathbb{V}ar(\varepsilon_t) = \mathbb{V}ar(B\eta_t) = B\mathbb{V}ar(\eta_t)B' = BB'.$$

Eq. (1.1) can also be rewritten using the lag operator L, which gives:[1]

$$y_t = c + \Phi(L)y_t + \varepsilon_t,$$

where $\Phi(L) = \Phi_1 L + \dots + \Phi_p L^p$.

As is the case for univariate models, VARs can be extended with MA terms in η_t, giving rise to VARMA models:

[1]The lag operator L transforms a process into a lagged process; e.g., $Ly_t = y_{t-1}$, and $L^k y_t = y_{t-k}$, where $k \in \mathbb{N}$.

Definition 1.2 ((S)VARMA model)**.** Let y_t denote a $n \times 1$ vector of random variables. Process y_t follows a VARMA model of order (p,q) if, for all t, we have

$$\begin{array}{rcl} VARMA: \; y_t & = & c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \\ & & \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \dots + \Theta_q \varepsilon_{t-q}, \\ SVARMA: \; y_t & = & c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \\ & & B_0 \eta_t + B_1 \eta_{t-1} + \dots + B_q \eta_{t-q}, \end{array} \tag{1.2}$$

with $\varepsilon_t = B_0\eta_t$, and $B_j = \Theta_j B_0$, for $j \geq 0$ (with $\Theta_0 = Id$), where $\{\eta_t\}$ is a white noise sequence whose components are are mutually and serially independent.

1.2 IRFs in SVARMA

One of the main objectives of macro-econometrics is to derive IRFs, that represent the dynamic effects of structural shocks (components of η_t) though the system of variables y_t.

Formally, an IRF is a difference in conditional expectations:

$$\boxed{\Psi_{i,j,h} = \mathbb{E}(y_{i,t+h}|\eta_{j,t} = 1) - \mathbb{E}(y_{i,t+h})} \tag{1.3}$$

(effect on $y_{i,t+h}$ of a one-unit shock on $\eta_{j,t}$).

IRFs closely relate to the Wold decomposition of y_t. Indeed, if the dynamics of process y_t can be described as a VARMA model, and if y_t is covariance stationary (see Def. 11.1), then y_t admits the following infinite MA representation (or Wold decomposition):

$$\boxed{y_t = \mu + \sum_{i=0}^{\infty} \Psi_i \eta_{t-i}.} \tag{1.4}$$

or, shifted by h periods:

$$\begin{array}{rcl} y_{t+h} & = & \mu + \sum_{i=0}^{\infty} \Psi_i \eta_{t+h-i} \\ & = & \mu + \underbrace{\sum_{i=0}^{h-1} \Psi_i \eta_{t+h-i}}_{\text{Effect of } \eta_{t+1},\dots,\eta_{t+h}} + \underbrace{\Psi_h \eta_t}_{\text{Effect of } \eta_t} + \underbrace{\sum_{i=h+1}^{\infty} \Psi_i \eta_{t+h-i}}_{\text{Effect of } \eta_{t-1},\eta_{t-2},\dots} \end{array}$$

With these notations, we get $\mathbb{E}(y_{i,t+h}|\eta_{j,t} = 1) = \mu_i + \Psi_{i,j,h}$, where $\Psi_{i,j,h}$ is the component (i,j) of matrix Ψ_h and μ_i is the i^{th} entry of vector μ. Since we also have $\mathbb{E}(y_{i,t+h}) = \mu_i$, we obtain Eq. (1.3).

Example 1.1 (IRF of a univariate AR model)**.** To fix ideas, consider a simple univariate AR(1) process:

$$y_t = \phi y_{t-1} + \eta_t, \quad \eta_t \sim \text{i.i.d. } (0, \sigma^2)$$

We assume $|\phi| < 1$ so that the process is covariance stationary.

Iterating forward:

$$y_{t+h} = \phi^h y_t + \sum_{i=0}^{h-1} \phi^i \eta_{t+h-i}.$$

A one-unit shock at time t means $\eta_t = 1$, while all other shocks have expectation zero. Therefore, the expected response at horizon h is:

$$\mathbb{E}(y_{t+h}|\eta_t = 1) - \mathbb{E}(y_{t+h}) = \phi^h.$$

Hence, the **impulse response function (IRF)** at horizon h is simply $\Psi_h = \phi^h$.

```
phi <- .8
plot(phi^(0:30),type="b",lwd=2,xlab="Periods after shock",ylab="",las=1)
```

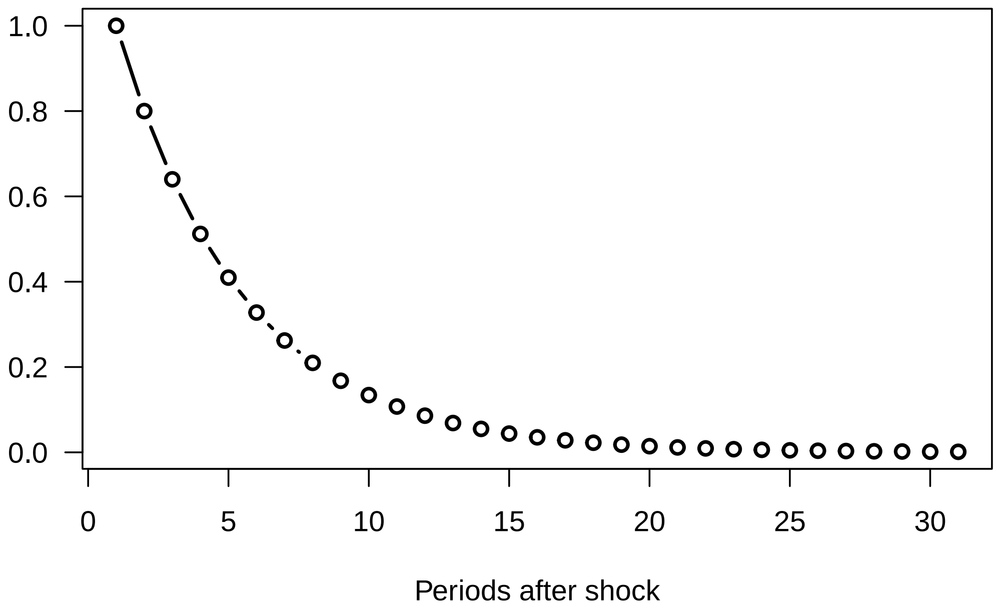

Figure 1.1: Impulse response functions for an AR(1) process.

Hence, estimating IRFs amounts to estimating the Ψ_h's. In general, there exist three main approaches for that:

- Calibrate and solve a (purely structural) Dynamic Stochastic General Equilibrium (DSGE) model at the first order (linearization). The solution takes the form of Eq. (1.4).

- Directly estimate the Ψ_h based on **projection approaches** (see Section 8).
- Approximate the infinite MA representation by estimating a parsimonious type of model, e.g. **VAR(MA) models** (see Section 1.4). Once a (Structural) VARMA representation is obtained, Eq. (1.4) is easily deduced using the following proposition:

Proposition 1.1 (IRF of an ARMA(p,q) process)**.** *If y_t follows the VARMA model described in Def. 1.2, then the matrices Ψ_h appearing in Eq. (1.4) can be computed recursively as follows:*

1. *Set $\Psi_{-1} = \dots = \Psi_{-p} = 0$.*
2. *For $h \geq 0$, (recursively) apply:*

$$\Psi_h = \Phi_1 \Psi_{h-1} + \dots + \Phi_p \Psi_{h-p} + \Theta_h B_0,$$

with $\Theta_0 = Id$ and $\Theta_h = 0$ for $h > q$.

Proof. This is obtained by applying the operator $\frac{\partial}{\partial \eta_t}$ on both sides of Eq. (1.2). □

Typically, consider the VAR(2) case. The first steps of the algorithm mentioned in the last bullet point are as follows:

$$\begin{aligned}
y_t &= \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + B\eta_t \\
&= \Phi_1(\Phi_1 y_{t-2} + \Phi_2 y_{t-3} + B\eta_{t-1}) + \Phi_2 y_{t-2} + B\eta_t \\
&= B\eta_t + \Phi_1 B\eta_{t-1} + (\Phi_2 + \Phi_1^2) y_{t-2} + \Phi_1\Phi_2 y_{t-3} \\
&= B\eta_t + \Phi_1 B\eta_{t-1} + (\Phi_2 + \Phi_1^2)(\Phi_1 y_{t-3} + \Phi_2 y_{t-4} + B\eta_{t-2}) + \Phi_1\Phi_2 y_{t-3} \\
&= \underbrace{B}_{=\Psi_0}\eta_t + \underbrace{\Phi_1 B}_{=\Psi_1}\eta_{t-1} + \underbrace{(\Phi_2 + \Phi_1^2)B}_{=\Psi_2}\eta_{t-2} + f(y_{t-3}, y_{t-4}).
\end{aligned}$$

In particular, we have $B = \Psi_0$. Matrix B indeed captures the contemporaneous impact of η_t on y_t. That is why matrix B is sometimes called **impulse matrix**.

Example 1.2 (IRFs of an SVAR model)**.** Consider the following VAR(2) model:

$$y_t = \underbrace{\left[\begin{array}{cc} 0.6 & 0.2 \\ 0 & 0.5 \end{array}\right]}_{=\Phi_1} y_{t-1} + \underbrace{\left[\begin{array}{cc} -0.1 & 0.1 \\ 0.2 & 0.3 \end{array}\right]}_{=\Phi_2} y_{t-2} + \underbrace{\left[\begin{array}{cc} 0.5 & 1.5 \\ -1 & 0.8 \end{array}\right]}_{=B} \eta_t. \tag{1.5}$$

We can use function `simul.VARMA` of package `IdSS` to produce IRFs (using `indic.IRF=1` in the list of arguments):

```r
library(IdSS)
# ---- Specify model: ----
Phi <- array(NaN,c(2,2,2)) # (2,2,2) for (n,n,p)
Phi[,,1] <- matrix(c(.6,0,.2,.5),2,2)
```

```
Phi[,,2] <- matrix(c(-.1,.2,.1,.3),2,2)
c <- c(1,2)
B <- matrix(c(.5,-1,1.5,.8),2,2)
Model <- list(c = c,Phi = Phi,B = B)
# ---- Define first shock: ----
eta0 <- c(1,0)
res.sim.1 <- simul.VARMA(Model,nb.sim=30,eta0=eta0,indic.IRF=1)
# ---- Define second shock: ----
eta0 <- c(0,1)
res.sim.2 <- simul.VARMA(Model,nb.sim=30,eta0=eta0,indic.IRF=1)
# ---- Prepare plots: ----
par(plt=c(.15,.95,.25,.8)) # define margins
par(mfrow=c(2,2)) # 2 rows and 2 columns of plots
for(i in 1:2){
  if(i == 1){res.sim <- res.sim.1}else{res.sim <- res.sim.2}
  for(j in 1:2){
    plot(res.sim$Y[j,],las=1,
         type="l",lwd=3,xlab="",ylab="",
         main=paste("Resp. of y",j,
                    " to a 1-unit increase in ",i,sep=""))
    abline(h=0,col="grey",lty=3)
  }}
```

The same type of output would be obtained by using the function `simul.VAR` of package `IdSS`:[2]

```
res.sim <- simul.VAR(c=c,Phi=Phi,B=B,nb.sim=30,indic.IRF=1,u.shock=eta0)
```

Example 1.3 (IRFs of an SVARMA model)**.** Consider the following VARMA(1,1) model:

$$y_t = \underbrace{\begin{bmatrix} 0.5 & 0.3 \\ -0.4 & 0.7 \end{bmatrix}}_{\Phi_1} y_{t-1} + \tag{1.6}$$

$$\underbrace{\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}}_{B} \eta_t + \underbrace{\begin{bmatrix} 0.4 & 0 \\ -1 & -0.5 \end{bmatrix}}_{\Theta_1} \underbrace{\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}}_{B} \eta_{t-1}. \tag{1.7}$$

We can use function `simul.VARMA` of package `IdSS` to produce IRFs (using `indic.IRF=1` in the list of arguments):

[2] In that case, `res.sim` is of dimension `nb.sim`$\times n$.

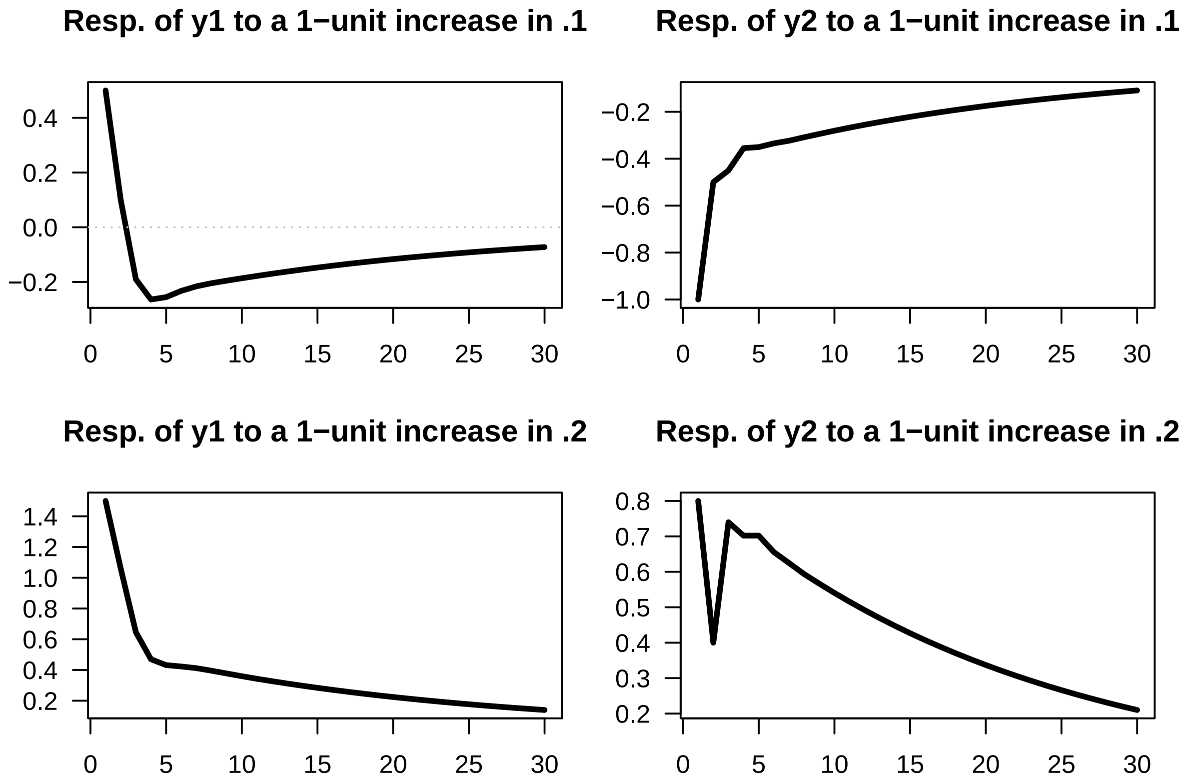

Figure 1.2: Impulse response functions (VAR(2) specified above).

```r
library(IdSS)
# ---- Specify model: ----
Phi <- array(NaN,c(2,2,1)) # (2,2,1) for (n,n,p)
Phi[,,1] <- matrix(c(.5,-.4,.3,.7),2,2)
p <- dim(Phi)[3]
Theta <- array(NaN,c(2,2,1))
Theta[,,1] <- matrix(c(.4,-1,0,-.5),2,2)
q <- dim(Theta)[3]
c <- rep(0,2)
B <- matrix(c(1,-1,2,1),2,2)
Model <- list(c = c,Phi = Phi,Theta = Theta,B = B)
# ---- Define first shock: ----
eta0 <- c(1,0)
res.sim.1 <- simul.VARMA(Model,nb.sim=30,eta0=eta0,indic.IRF=1)
# ---- Define second shock: ----
eta0 <- c(0,1)
res.sim.2 <- simul.VARMA(Model,nb.sim=30,eta0=eta0,indic.IRF=1)
par(plt=c(.15,.95,.25,.8))
par(mfrow=c(2,2))
for(i in 1:2){
  if(i == 1){res.sim <- res.sim.1}else{res.sim <- res.sim.2}
  for(j in 1:2){
    plot(res.sim$Y[j,],las=1,
         type="l",lwd=3,xlab="",ylab="",
         main=paste("Resp. of y",j,
                    " to a 1-unit increase in ",i,sep=""))
    abline(h=0,col="grey",lty=3)
  }}
```

1.3 Covariance-stationary VARMA models

Let's come back to the infinite MA case (Eq. (1.4)):

$$y_t = \mu + \sum_{h=0}^{\infty} \Psi_h \eta_{t-h}.$$

For y_t to be covariance-stationary (and ergodic for the mean), it has to be the case that

$$\sum_{i=0}^{\infty} \|\Psi_i\| < \infty, \tag{1.8}$$

where $\|A\|$ denotes a norm of the matrix A (e.g. $\|A\| = \sqrt{tr(AA')}$). This notably implies that if y_t is stationary (and ergodic for the mean), then $\|\Psi_h\| \to 0$ when h gets large.

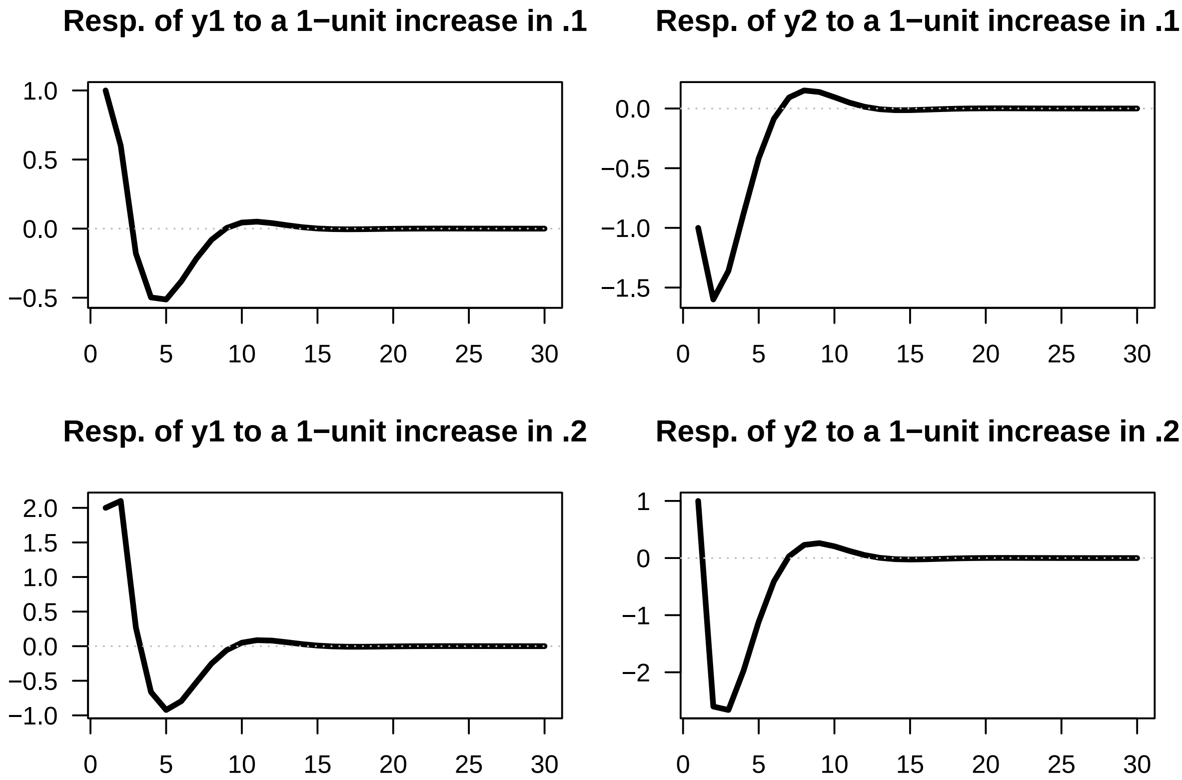

Figure 1.3: Impulse response functions (SVARMA(1,1) specified above).

What should be satisfied by Φ_k's and Θ_k's for a VARMA-based process (Eq. (1.2)) to be stationary? The conditions will be similar to that we have in the univariate case. Let us introduce the following notations:

$$\begin{aligned} y_t &= c + \underbrace{\Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p}}_{\text{AR component}} + \\ &\quad \underbrace{B\eta_t - \Theta_1 B\eta_{t-1} - \dots - \Theta_q B\eta_{t-q}}_{\text{MA component}} \\ \Leftrightarrow &\quad \underbrace{(I - \Phi_1 L - \dots - \Phi_p L^p)}_{=\Phi(L)} y_t = c + \underbrace{(I + \Theta_1 L + \dots + \Theta_q L^q)}_{=\Theta(L)} B\eta_t. \end{aligned} \tag{1.9}$$

Process y_t is stationary iff the roots of $\det(\Phi(z)) = 0$ are strictly outside the unit circle or, equivalently, iff the eigenvalues of

$$\underbrace{\Phi}_{np\times np} = \left[\begin{array}{cccc} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ I & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I & 0 \end{array}\right] \tag{1.10}$$

lie strictly within the unit circle. Hence, as is the case for univariate processes, the covariance-stationarity of a VARMA model depends only on the specification of its AR part.

Let's derive the first two unconditional moments of a (covariance-stationary) VARMA process.

Eq. (1.10) gives $\mathbb{E}(\Phi(L)y_t) = c$, therefore $\Phi(1)\mathbb{E}(y_t) = c$, or

$$\mathbb{E}(y_t) = (I - \Phi_1 - \dots - \Phi_p)^{-1} c.$$

The autocovariances of y_t can be deduced from the infinite MA representation (Eq. (1.4)). We have:

$$\gamma_j \equiv \mathbb{C}ov(y_t, y_{t-j}) = \sum_{i=j}^{\infty} \Psi_i \Psi'_{i-j}.$$

Indeed:

$$\begin{aligned} \mathbb{C}ov(y_t, y_{t-j}) &= \mathbb{E}\left(\left[\sum_{h=0}^{\infty} \Psi_h \eta_{t-h}\right] \left[\sum_{h=0}^{\infty} \Psi_h \eta_{t-j-h}\right]' \right) \\ &= \mathbb{E}\left(\left[\sum_{i=0}^{j-1} \Psi_i \eta_{t-i} + \sum_{i=j}^{\infty} \Psi_i \eta_{t-i}\right] \left[\sum_{i=j}^{\infty} \Psi_{i-j} \eta_{t-i}\right]' \right) \\ &= \mathbb{E}\left(\sum_{i=j}^{\infty} \Psi_i \eta_{t-i} \eta'_{t-i} \Psi'_{i-j} \right) = \sum_{i=j}^{\infty} \Psi_i \Psi'_{i-j}. \end{aligned}$$

(Remark: This infinite sum exists as soon as Eq. (1.8) is satisfied.)

Conditional means and autocovariances can also be deduced from Eq. (1.4). For $0 \le h$ and $0 \le h_1 \le h_2$:

$$\begin{aligned}
\mathbb{E}_t(y_{t+h}) &= \mu + \sum_{k=0}^{\infty} \Psi_{k+h}\eta_{t-k} \\
\mathbb{C}ov_t(y_{t+1+h_1}, y_{t+1+h_2}) &= \sum_{k=0}^{h_1} \Psi_k \Psi'_{k+h_2-h_1}.
\end{aligned}$$

The previous formula implies in particular that the forecasting error $y_{t+h} - \mathbb{E}_t(y_{t+h})$ has a variance equal to:

$$\boxed{\mathbb{V}ar_t(y_{t+h}) = \sum_{k=0}^{h-1} \Psi_k \Psi'_k.}$$

Because the components of η_t are mutually and serially independent (and therefore uncorrelated), we have:

$$\mathbb{V}ar(\Psi_k \eta_{t-k}) = \mathbb{V}ar\left(\sum_{i=1}^{n} \psi_{k,i}\eta_{i,t-k}\right) = \sum_{i=1}^{n} \psi_{k,i}\psi'_{k,i},$$

where $\psi_{k,i}$ denotes the i^{th} column of Ψ_k. This suggests the following decomposition of the variance of the forecast error (called **variance decomposition**):

$$\mathbb{V}ar_t(y_{t+h}) = \sum_{i=1}^{n} \underbrace{\sum_{k=0}^{h-1} \psi_{k,i}\psi'_{k,i}}_{\text{Contribution of } \eta_{i,t}}.$$

To illustrate, let us use the VAR(2) model proposed above, in Example 1.2. As explained above, the variance decomposition directly results from the knowledge of IRFs (shown in Figure 1.2):

Example 1.4 (Variance decomposition in the context of a VAR(2) defined.)**.** The chunk below computes, for different horizons, the shares of the variances of the two variables that are accounted for by each of the two components of η_t in the context of the model proposed in Example 1.2.

```
Phi <- array(NaN,c(2,2,2)) # (2,2,2) for (n,n,p)
Phi[,,1] <- matrix(c(.6,0,.2,.5),2,2)
Phi[,,2] <- matrix(c(-.1,.2,.1,.3),2,2)
B <- matrix(c(.5,-1,1.5,.8),2,2)
make_variance_decompo(Phi,B,maxHorizon=30)
```

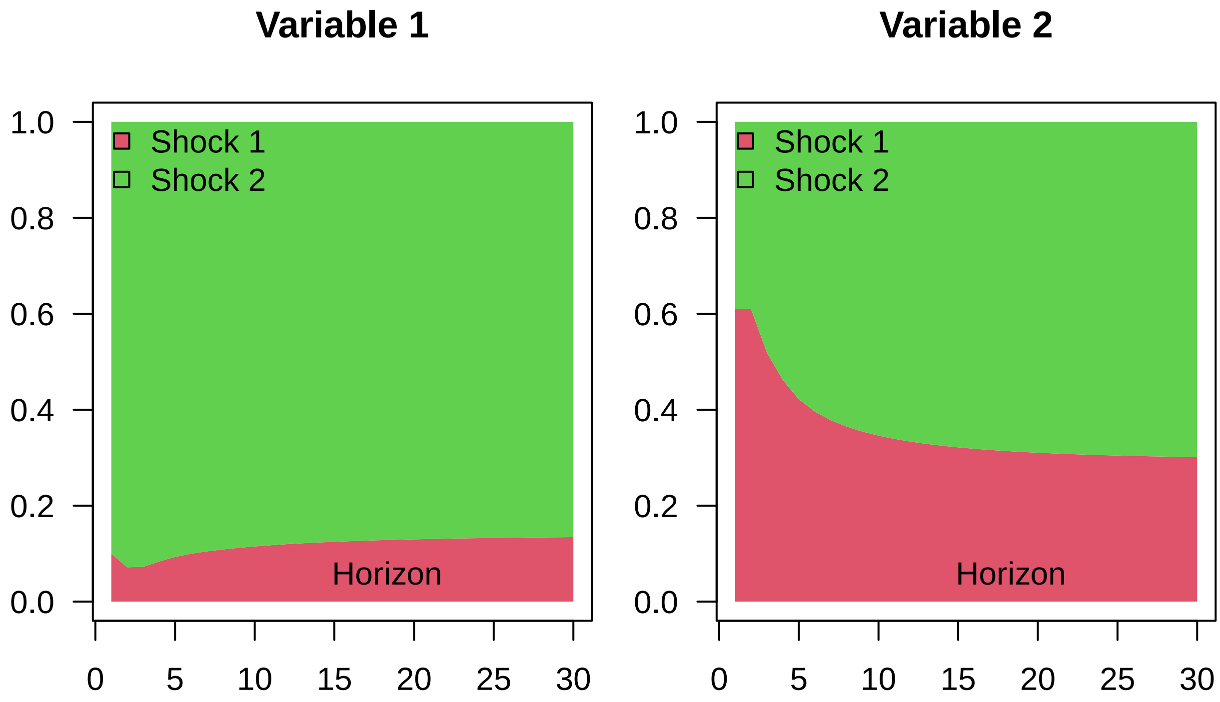

Figure 1.4: Variance decomposition of a VAR(2) model.

1.4 VAR estimation

This section discusses the estimation of VAR models.[3] Eq. (1.1) can be written:

$$y_t = c + \Phi(L)y_{t-1} + \varepsilon_t,$$

with $\Phi(L) = \Phi_1 + \Phi_2 L + \dots + \Phi_p L^{p-1}$.

Consequently:

$$y_t \mid y_{t-1}, y_{t-2}, \dots, y_{-p+1} \sim \mathcal{N}(c + \Phi_1 y_{t-1} + \dots \Phi_p y_{t-p}, \Omega).$$

Using Hamilton (1994)'s notations, denote with Π the matrix $\left[\begin{array}{ccccc} c & \Phi_1 & \Phi_2 & \dots & \Phi_p \end{array}\right]'$ and with x_t the vector $\left[\begin{array}{ccccc} 1 & y'_{t-1} & y'_{t-2} & \dots & y'_{t-p} \end{array}\right]'$, we have:

$$y_t = \Pi' x_t + \varepsilon_t. \tag{1.11}$$

The previous representation is convenient to discuss the estimation of the VAR model, as parameters are gathered in two matrices only: Π and Ω.

Let us start with the case where the shocks are Gaussian.

Proposition 1.2 (MLE of a Gaussian VAR)**.** *If y_t follows a VAR(p) (see Definition 1.1), and if $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \Omega)$, then the ML estimate of Π, denoted by $\hat{\Pi}$ (see Eq.* (1.11)*), is given by*

$$\hat{\Pi} = \left[\sum_{t=1}^{T} x_t x_t'\right]^{-1} \left[\sum_{t=1}^{T} y_t' x_t\right] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{1.12}$$

where $\mathbf{X}$ is the $T \times (1 + np)$ matrix whose t^{th} row is x_t and where $\mathbf{y}$ is the $T \times n$ matrix whose t^{th} row is y_t'.

That is, the i^{th} column of $\hat{\Pi}$ (b_i, say) is the OLS estimate of β_i, where:

$$y_{i,t} = \beta_i' x_t + \varepsilon_{i,t}, \tag{1.13}$$

(i.e., $\beta_i' = [c_i, \phi_{i,1}', \dots, \phi_{i,p}']'$).

The ML estimate of Ω, denoted by $\hat{\Omega}$, coincides with the sample covariance matrix of the n series of the OLS residuals in Eq. (1.13)*, i.e.:*

$$\hat{\Omega} = \frac{1}{T}\sum_{i=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t', \quad \textit{with } \hat{\varepsilon}_t = y_t - \hat{\Pi}' x_t. \tag{1.14}$$

The asymptotic distributions of these estimators are the ones resulting from standard OLS formula.

[3]The estimation of VARMA models is introduced in Appendix 11.2.

Proof. See Appendix 11.3. □

As stated by Proposition 1.3, when the shocks are not Gaussian, then the OLS regressions still provide consistent estimates of the model parameters. However, since x_t correlates to ε_s for $s < t$, the OLS estimator $\mathbf{b}_i$ of β_i is biased in small sample. (That is also the case for the ML estimator.) Indeed, denoting by ε_i the $T \times 1$ vector of $\varepsilon_{i,t}$'s, and using the notations of b_i and β_i introduced in Proposition 1.2, we have:

$$\mathbf{b}_i = \beta_i + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon_i. \tag{1.15}$$

We have non-zero correlation between x_t and $\varepsilon_{i,s}$ for $s < t$ and, therefore, $\mathbb{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon_i] \neq 0$.

However, when y_t is covariance stationary, then $\frac{1}{n}\mathbf{X}'\mathbf{X}$ converges to a positive definite matrix $\mathbf{Q}$, and $\frac{1}{n}X'\varepsilon_i$ converges to 0. Hence $\mathbf{b}_i \overset{p}{\to} \beta_i$. More precisely:

Proposition 1.3 (Asymptotic distribution of the OLS estimate of VAR coefficients (for one variable))**.** *If y_t follows a VAR model, as defined in Definition 1.1, we have:*

$$\sqrt{T}(\mathbf{b}_i - \beta_i) = \underbrace{\left[\frac{1}{T}\sum_{t=p}^{T} x_t x_t'\right]^{-1}}_{\overset{p}{\to}\mathbf{Q}^{-1}} \underbrace{\sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_{i,t}\right]}_{\overset{d}{\to}\mathcal{N}(0,\sigma_i^2\mathbf{Q})},$$

where $\sigma_i = \mathbb{V}ar(\varepsilon_{i,t})$ and where $\mathbf{Q} = plim\ \frac{1}{T}\sum_{t=p}^{T} x_t x_t'$ is given by:

$$\mathbf{Q} = \left[\begin{array}{ccccc} 1 & \mu' & \mu' & \dots & \mu' \\ \mu & \gamma_0 + \mu\mu' & \gamma_1 + \mu\mu' & \dots & \gamma_{p-1} + \mu\mu' \\ \mu & \gamma_1 + \mu\mu' & \gamma_0 + \mu\mu' & \dots & \gamma_{p-2} + \mu\mu' \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mu & \gamma_{p-1} + \mu\mu' & \gamma_{p-2} + \mu\mu' & \dots & \gamma_0 + \mu\mu' \end{array}\right], \tag{1.16}$$

where γ_j is the unconditional autocovariance matrix of order j of y_t, i.e., $\gamma_i = \mathbb{C}ov(y_t, y_{t-j})$.

Proof. See Appendix 11.3. □

The following proposition extends the previous proposition and includes covariances between different β_i's as well as the asymptotic distribution of the ML estimates of Ω.

Proposition 1.4 (Asymptotic distribution of the OLS estimates)**.** *If y_t follows a VAR model, as defined in Definition 1.1, we have:*

$$\sqrt{T}\left[\begin{array}{c} vec(\hat{\Pi} - \Pi) \\ vec(\hat{\Omega} - \Omega) \end{array}\right] \sim \mathcal{N}\left(0, \left[\begin{array}{cc} \Omega \otimes \mathbf{Q}^{-1} & 0 \\ 0 & \Sigma_{22} \end{array}\right]\right), \tag{1.17}$$

where the component of Σ_{22} corresponding to the covariance between $\hat{\sigma}_{i,j}$ and $\hat{\sigma}_{k,l}$ (for $i, j, l, m \in \{1, \dots, n\}^4$) is equal to $\sigma_{i,l}\sigma_{j,m} + \sigma_{i,m}\sigma_{j,l}$.

Proof. See, e.g., Hamilton (1994), Appendix of Chapter 11. □

In practice, to use the previous proposition (for instance to implement Monte-Carlo simulations, see Section 3.1), Ω is replaced with $\hat{\Omega}$, $\mathbf{Q}$ is replaced with $\hat{\mathbf{Q}} = \frac{1}{T}\sum_{t=p}^{T} x_t x_t'$ and Σ with the matrix whose components are of the form $\hat{\sigma}_{i,l}\hat{\sigma}_{j,m} + \hat{\sigma}_{i,m}\hat{\sigma}_{j,l}$, where the $\hat{\sigma}_{i,l}$'s are the components of $\hat{\Omega}$.

The simplicity of the VAR framework and the tractability of its MLE open the way to convenient econometric testing. Let's illustrate this with the likelihood ratio test (see Def. 11.2). The maximum value achieved by the MLE is

$$\log \mathcal{L}(Y_T; \hat{\Pi}, \hat{\Omega}) = -\frac{Tn}{2}\log(2\pi) + \frac{T}{2}\log\left|\hat{\Omega}^{-1}\right| - \frac{1}{2}\sum_{t=1}^{T}\left[\hat{\varepsilon}_t'\hat{\Omega}^{-1}\hat{\varepsilon}_t\right].$$

The last term is:

$$\begin{aligned}\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}^{-1}\hat{\varepsilon}_t &= \mathrm{Tr}\left[\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}^{-1}\hat{\varepsilon}_t\right] = \mathrm{Tr}\left[\sum_{t=1}^{T}\hat{\Omega}^{-1}\hat{\varepsilon}_t\hat{\varepsilon}_t'\right]\\ &= \mathrm{Tr}\left[\hat{\Omega}^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'\right] = \mathrm{Tr}\left[\hat{\Omega}^{-1}\left(T\hat{\Omega}\right)\right] = Tn.\end{aligned}$$

Therefore, the optimized log-likelihood is simply obtained by:

$$\log \mathcal{L}(Y_T; \hat{\Pi}, \hat{\Omega}) = -(Tn/2)\log(2\pi) + (T/2)\log\left|\hat{\Omega}^{-1}\right| - Tn/2. \tag{1.18}$$

Having the optimized likelihood in analytical form allows for flexible applications of the likelihood ratio (LR) test (Def. 11.2).

Example 1.5 (Ad-hoc lag selection based on the LR test)**.** Suppose we want to test whether a system of variables follows a VAR(p_0) process (the null hypothesis) against an alternative specification with p_1 lags, where $p_1 > p_0$. Let $\hat{L}_0$ and $\hat{L}_1$ denote the maximum log-likelihoods obtained under the two specifications, respectively. Under the null hypothesis (H_0: $p = p_0$), the LR statistic is given by

$$\xi^{LR} := 2(\hat{L}_1 - \hat{L}_0) = T\left(\log\left|\hat{\Omega}_1^{-1}\right| - \log\left|\hat{\Omega}_0^{-1}\right|\right) \sim \chi^2\left(n^2(p_1 - p_0)\right). \tag{1.19}$$

This statistic can be used to determine whether adding lags significantly improves the model fit. However, including too many lags quickly exhausts degrees of freedom: with p lags, each of the n equations in the VAR includes $n \times p$ coefficients plus an intercept term. While adding lags always improves the in-sample fit, it can lead to over-parameterization and deteriorate the model's out-of-sample predictive performance.

The approach above provides an ad-hoc way to select lag length using LR tests (Def. 11.2). More systematic procedures rely on information criteria, which balance goodness of fit and model parsimony. In the Gaussian case (see Eq. (1.18)), common criteria include:

$$
\begin{aligned}
AIC &= cst + \log\left|\widehat{\Omega}\right| + \frac{2}{T}N \\
BIC &= cst + \log\left|\widehat{\Omega}\right| + \frac{\log T}{T}N,
\end{aligned}
$$

where $N = p \times n^2$.

Example 1.6 (Three-variable VAR model)**.** The following example illustrates how to estimate and analyze a three-variable VAR model using U.S. quarterly data on inflation, the output gap, and the short-term nominal interest rate.

We begin by using the `VARselect` function from the `vars` package to determine the optimal lag length for the VAR specification.

```
library(vars)   # provides 'VARselect' function
library(IdSS)   # dataset is included here

First.date <- "1959-04-01"
Last.date  <- "2015-01-01"

data <- US3var
data <- data[(data$Date >= First.date) & (data$Date <= Last.date), ]
Y <- as.matrix(data[c("infl", "y.gdp.gap", "r")])

VARselect(Y, lag.max = 12)
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##     11      3      2     11
##
## $criteria
##                  1            2           3          4          5          6
## AIC(n) -0.14340897 -0.323002241 -0.39796730 -0.3888482 -0.4110019 -0.5229686
## HQ(n)  -0.06661725 -0.188616726 -0.20598799 -0.1392751 -0.1038350 -0.1582079
## SC(n)   0.04658647  0.009489796  0.07702133  0.2286370  0.3489799  0.3795098
## FPE(n)  0.86641130  0.724024392  0.67182519  0.6781505  0.6635579  0.5936165
##                  7           8           9         10         11         12
## AIC(n) -0.48840029 -0.50157851 -0.44613801 -0.4950938 -0.5372138 -0.4980589
## HQ(n)  -0.06604582 -0.02163024  0.09140405  0.1000420  0.1155158  0.2122645
## SC(n)   0.55657468  0.68589305  0.88383014  0.9773709  1.0777475  1.2593990
## FPE(n)  0.61498914  0.60758020  0.64308309  0.6133871  0.5892917  0.6143270
```

Next, we estimate the VAR model including an exogenous variable — the commodity price index.

```r
p <- 6
exogen <- matrix(data$commo, ncol = 1); colnames(exogen) <- "commo"
estVAR <- VAR(Y, p = p, exogen = exogen) # estimate the VAR model
Phi <- Acoef(estVAR)
eps <- residuals(estVAR)
Omega <- var(eps) # covariance matrix of OLS residuals
B <- t(chol(Omega)) # Cholesky decomposition of Omega (lower triangular)
```

We can now verify whether the estimated VAR is stationary:

```r
PHI <- make.PHI(Phi) # autoregressive matrix in companion form
print(abs(eigen(PHI)$values)) # eigenvalues must be < 1 for stationarity
```

```
##  [1] 0.9485855 0.9485855 0.9180083 0.8539724 0.8539724 0.7683116 0.7683116
##  [8] 0.7509592 0.7509592 0.7465920 0.7465920 0.7392400 0.7392400 0.6720905
## [15] 0.6720905 0.4474043 0.4474043 0.3904400
```

All eigenvalues have moduli smaller than one, confirming that the system is stationary.

We then compute and display the impulse response functions (IRFs) using the Cholesky identification approach (see Section 2.3), where the impact matrix B comes from the Cholesky factorization of $\Omega = \mathbb{V}ar(\varepsilon_t)$.

```r
n <- dim(Y)[2] # number of endogenous variables
par(mfrow = c(n, n))
par(plt = c(.3, .95, .2, .8))
Model <- list(c = c, Phi = Phi, B = B)
names.var <- c("Inflation", "Output gap", "Interest rate")

for (variable in 1:n) {
  for (shock in 1:n) {
    eta0 <- rep(0, n)
    eta0[shock] <- 1
    res.sim <- simul.VAR(c = NaN, Phi, B, nb.sim = 30, indic.IRF = 1,
                         u.shock = eta0)
    plot(res.sim[, variable], type = "l", lwd = 2, las = 1,
         xlab = "", ylab = ifelse(shock==1,
                                  paste("Resp. of ",
                                        names.var[variable],"...",sep=""),""),
         main = ifelse(variable==1,paste("... to shock ", shock, sep = ""),"")
         )}
}
```

Finally, we compute the variance decomposition associated with the same model.

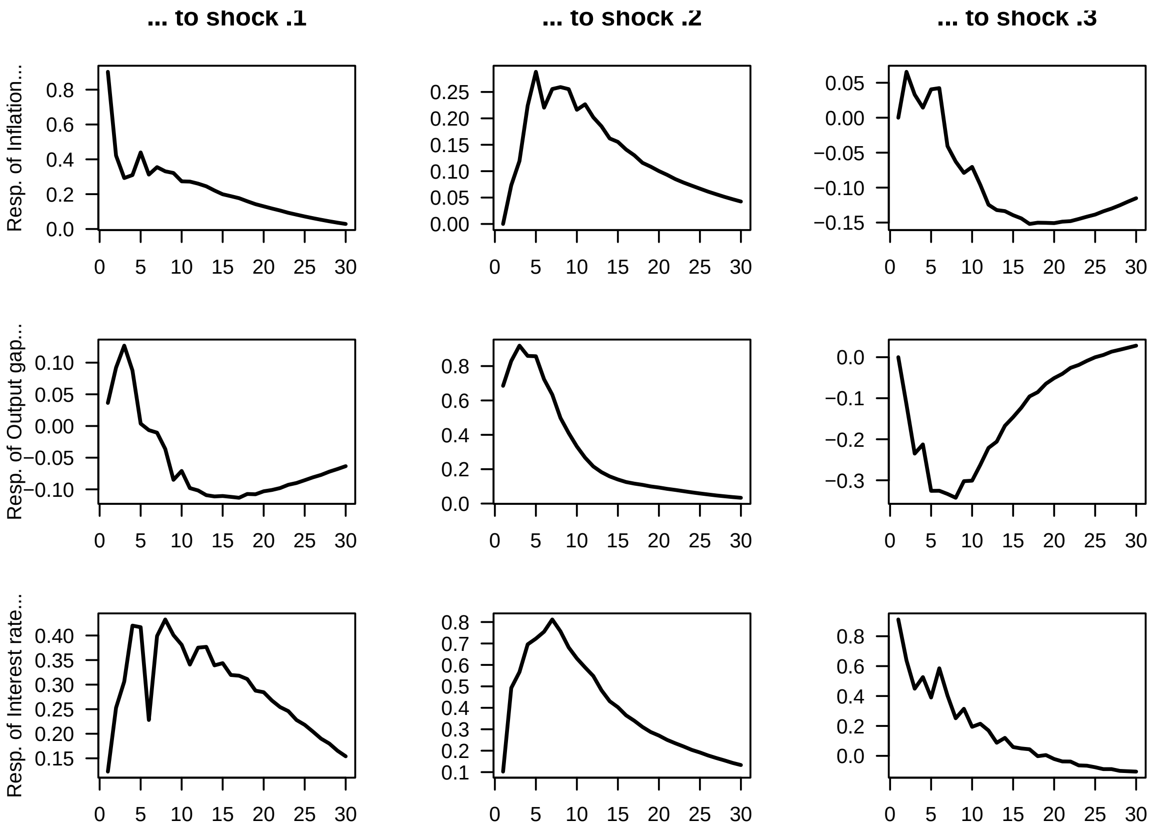

Figure 1.5: Impulse response functions for a 3-variable VAR model estimated on U.S. quarterly data.

```
make_variance_decompo(Phi, B, maxHorizon = 50,
                      names.var = c("Inflation", "Real activity",
                                    "Short-term rate"))
```

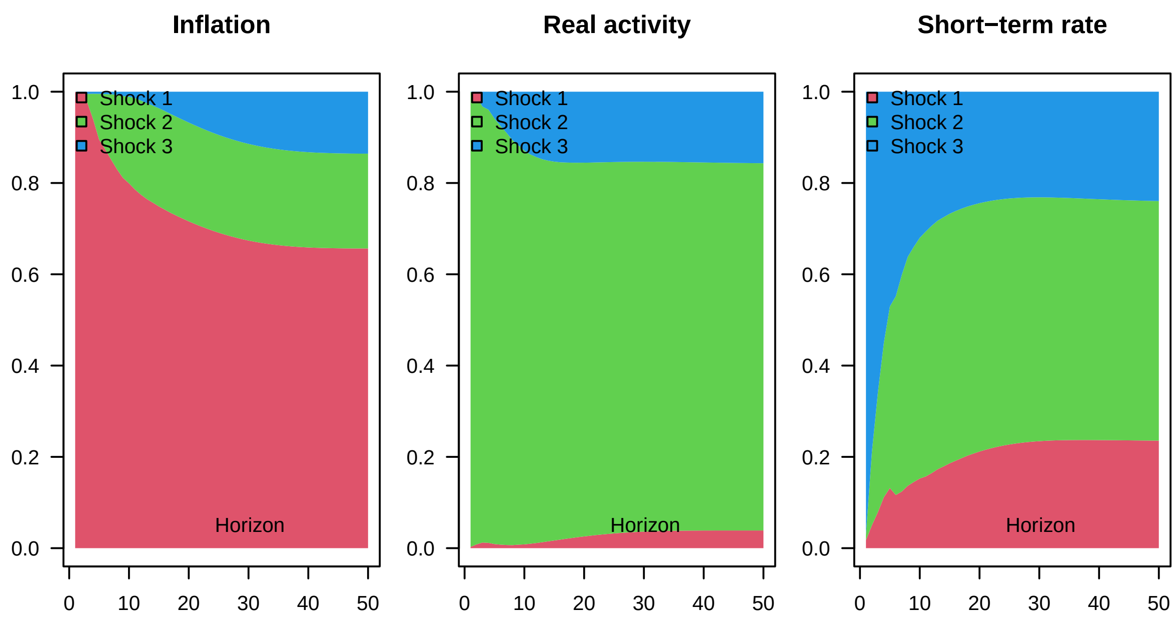

Figure 1.6: Variance decomposition for a 3-variable VAR model estimated on U.S. quarterly data.

1.5 Block exogeneity and Granger causality

1.5.1 Block exogeneity

In an unconstrained VAR model, where the coefficient matrices Φ_i are full, all lags of all endogenous variables may affect all current endogenous variables. However, in some situations one may suspect causal structures in which a subset of variables influences the others, but not the reverse. In such cases, the VAR can be formulated with **block exogeneity** restrictions.

Let us partition y_t into two subvectors $y_t^{(1)}$ $(n_1 \times 1)$ and $y_t^{(2)}$ $(n_2 \times 1)$, such that $y_t' =$

$[y_t^{(1)'}, y_t^{(2)'}]$ and $n = n_1 + n_2$. Then the VAR(1) representation can be written as

$$\begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi^{(1,1)} & \Phi^{(1,2)} \\ \Phi^{(2,1)} & \Phi^{(2,2)} \end{bmatrix} \begin{bmatrix} y_{t-1}^{(1)} \\ y_{t-1}^{(2)} \end{bmatrix} + \varepsilon_t.$$

Block exogeneity of $y_t^{(2)}$ (i.e., $y_t^{(2)}$ does not respond to past values of $y_t^{(1)}$) corresponds to the null hypothesis

$$\Phi^{(2,1)} = 0.$$

This restriction can be tested using a likelihood ratio test (see Def. 11.2). The same logic extends straightforwardly to a VAR(p) model by imposing $\Phi_i^{(2,1)} = 0$ for all $i = 1, \dots, p$.

Consider a model of the joint dynamics of US and Korean GDP growth and inflation rates. Figure 1.7 shows the data used in the estimation.

```
library(IdSS) # this package contains the data
library(vars) # this package contains standard VAR estimation tools
names.var <- c("date","GDP_USA","CPI_USA","GDP_KOR","CPI_KOR")
data <- international[,names.var]
data <- data[complete.cases(data),]
Y <- data[,c("GDP_USA","CPI_USA","GDP_KOR","CPI_KOR")]
par(mfrow=c(1,2))
par(plot=c(.1,.95,.15,.8))
```

```
## Warning in par(plot = c(0.1, 0.95, 0.15, 0.8)): "plot" is not a graphical
## parameter
```

```
plot(data$date,data$GDP_KOR,type="l",lwd=2,
     main="GDP growth rate (y-o-y)",xlab="",ylab="")
lines(data$date,data$GDP_USA,col="grey",lwd=2,las=1)
plot(data$date,data$CPI_KOR,type="l",lwd=2,
     main="Inflation (y-o-y)",xlab="",ylab="")
lines(data$date,data$CPI_USA,col="grey",lwd=2,las=1)
```

Let us first estimate an unrestricted VAR(p) model using these data.

```
p <- 4
estVAR <- VAR(Y, p = p) # estimate the VAR model
L0 <- logLik(estVAR)
```

The model log-likelihood is stored in `L0`.

Let us now estimate a retricted model:

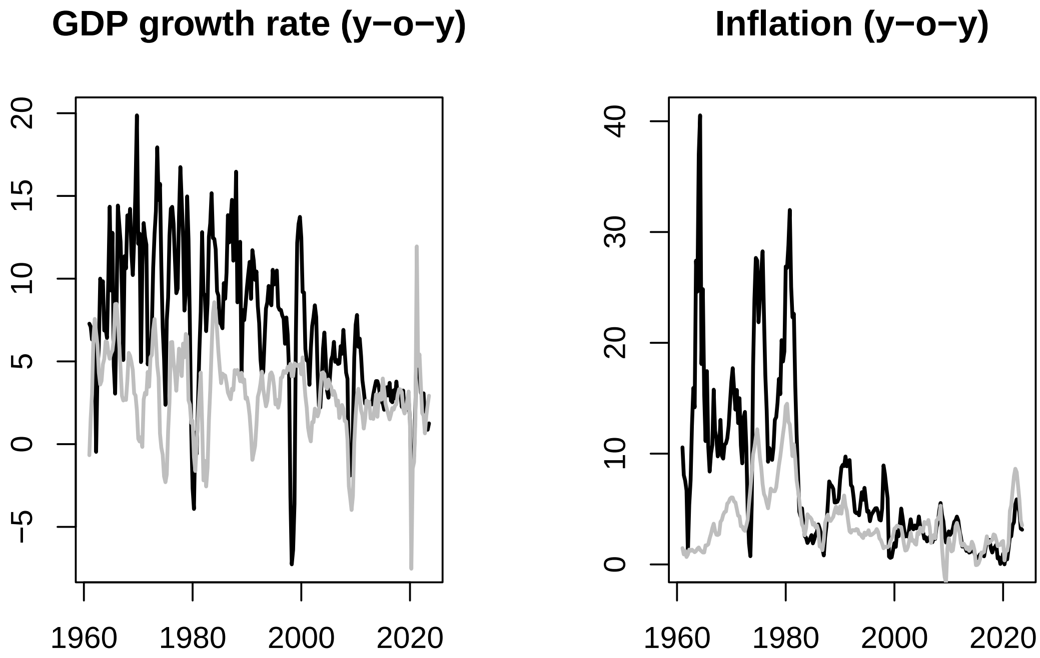

Figure 1.7: GDP log growth rates and inflation rates for the US and Korea. Black lines are for Korea; grey lines are for the US. All series are year-on-year growth rates, expressed in percent.

```r
n <- dim(Y)[2]
restrict <- matrix(1,n,1+p*n) # 'restrict' specifies restrictions; last column = 'c'.
for(i in 1:p){
  restrict[1:(n/2),(n/2+n*(i-1)+1):(i*n)] <- 0
}
estVARrestr <- restrict(estVAR, method = "man", resmat = restrict)
Lrestr <- logLik(estVARrestr)
```

The model log-likelihood is stored in `Lrestr`. We can perform a Likelihood Ratio (LR) test to see whether the data call for the exogeneity of the US variables. According to its definition (Def. 11.2), the LR test statistics is given by:

$$\xi^{LR} = 2 \times (\hat{L} - \hat{L}^*),$$

where $\hat{L}$ and $\hat{L}^*$ are, respectively, the unrestricted and restricted maximum log-likelihoods. Under the null hypothesis that the block exogeneity is satisfied, we have:

$$\xi^{LR} \sim \chi^2 \left[p \left(\frac{n}{2} \right)^2 \right].$$

In our context, we can therefore apply the test as follows:

```r
LRstat <- 2*(L0 - Lrestr)
df <- length(restrict)-sum(restrict)
pvalue <- 1 - pchisq(LRstat,df=df)
print(c(pvalue))
```

```
## [1] 0.2044077
```

Hence, the test cannot reject the null hypothesis of block exogeneity since the p-value is larger than conventional thresholds. In terms or coding, we could have used, equivalently, the `lrtest` function to perform this test:

```r
lrtest(estVAR,estVARrestr)
```

```
## Likelihood ratio test
##
## Model 1: VAR(y = Y, p = p)
## Model 2: VAR(y = Y, p = p)
##   #Df  LogLik  Df  Chisq Pr(>Chisq)
## 1  68 -1793.2
## 2  52 -1803.4 -16 20.361     0.2044
```

Let us compare the IRFs in both models (w/ and w/o block exogeneity). In Figure 1.8, the grey lines correspond to a contrained model, where Korean variables do not affect US ones. (In other words, in the restricted model, US variables are exogenous.) Notice that, in the restricted model, the last two shocks ($\eta_{3,t}$ and $\eta_{4,t}$) do not affect US variables.

```
# Unrestricted model:
Phi <- Acoef(estVAR)
eps <- residuals(estVAR)
Omega <- var(eps) # covariance matrix of OLS residuals
B <- t(chol(Omega)) # Cholesky decomposition of Omega (lower triangular)
# Restricted model:
Phi_restr <- Acoef(estVARrestr)
eps_restr <- residuals(estVARrestr)
Omega_restr <- var(eps_restr)
B_restr <- t(chol(Omega_restr))

par(mfrow = c(n, n))
par(plt = c(.35, .97, .25, .68))
Model <- list(c = c, Phi = Phi, B = B)
for (shock in 1:n) {
  eta0 <- rep(0, n)
  eta0[shock] <- 1
  res.sim <- simul.VAR(c = NaN, Phi, B, nb.sim = 30, indic.IRF = 1,
                       u.shock = eta0)
  res.sim_restr <- simul.VAR(c = NaN, Phi_restr, B_restr, nb.sim = 30,
                             indic.IRF = 1,u.shock = eta0)
  for (variable in 1:n) {
    plot(res.sim[, variable], type = "l", lwd = 2, las = 1,
         ylab = ifelse(variable==1,
                       paste("Shock  ", shock, sep = ""),""),xlab="",
         main = ifelse(shock==1,
                       paste("Resp. of ", names.var[variable],sep = ""),""))
    lines(res.sim_restr[, variable], type = "l", lwd = 2,col="grey")
  }
}
```

Finally, let us compute the variance decomposition in both models. Let us start with the unrestricted model:

```
make_variance_decompo(Phi, B, maxHorizon = 50,
                      names.var = names(Y),mfrow = c(2,2))
```

Figure 1.10 shows the variance decomposition in the context of the restricted model, that entails block exogeneity:

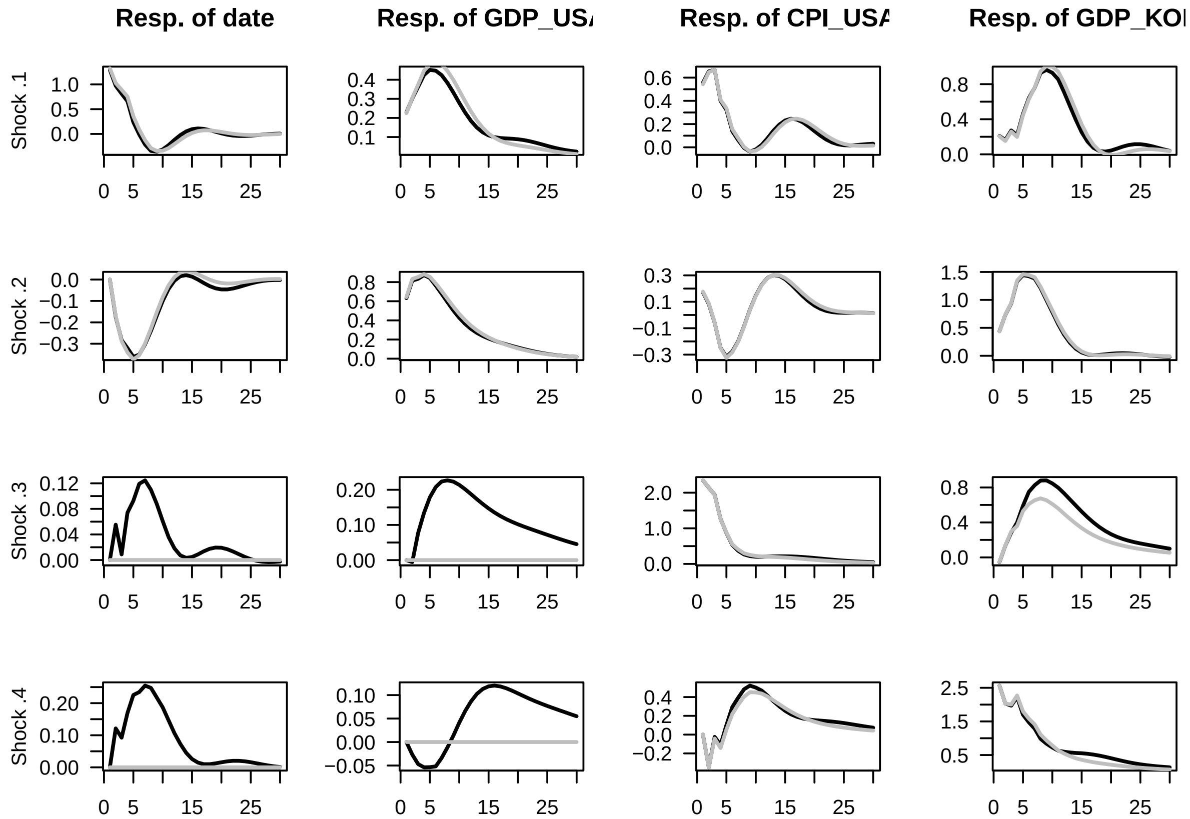

Figure 1.8: Impulse response functions in the context of VAR models depicting the joint dynamics of US and Korean GDP growth and inflation rates. Structural shocks are identified by using the Cholesky decomposition. The grey lines correspond to a contrained model, where Korean variables do not affect US ones. (In other words, in the restricted model, US variables are exogenous.)

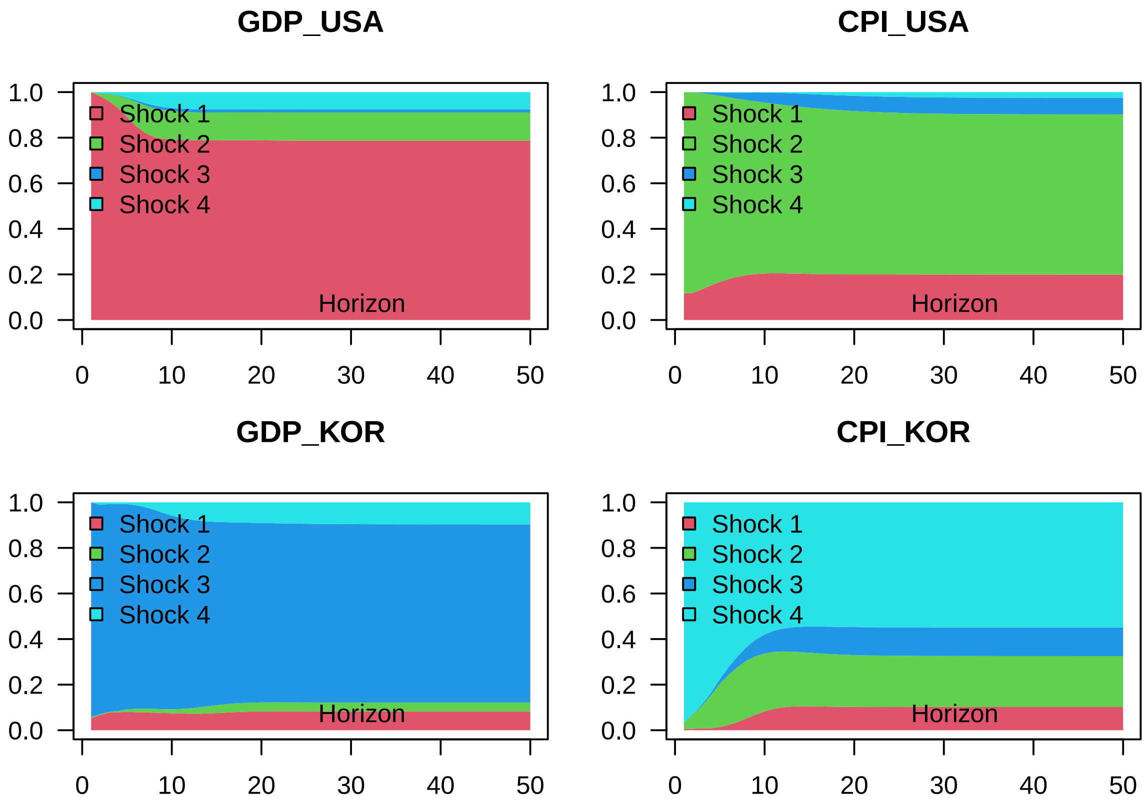

Figure 1.9: Variance decomposition in the context of a VAR model depicting the joint dynamics of US and Korean GDP growth and inflation rates. Structural shocks are identified by using the Cholesky decomposition. The grey lines correspond to a contrained model, where Korean variables do not affect US ones. (In other words, in the restricted model, US variables are exogenous.)

```
make_variance_decompo(Phi_restr, B_restr, maxHorizon = 50,
                      names.var = names(Y),mfrow = c(2,2))
```

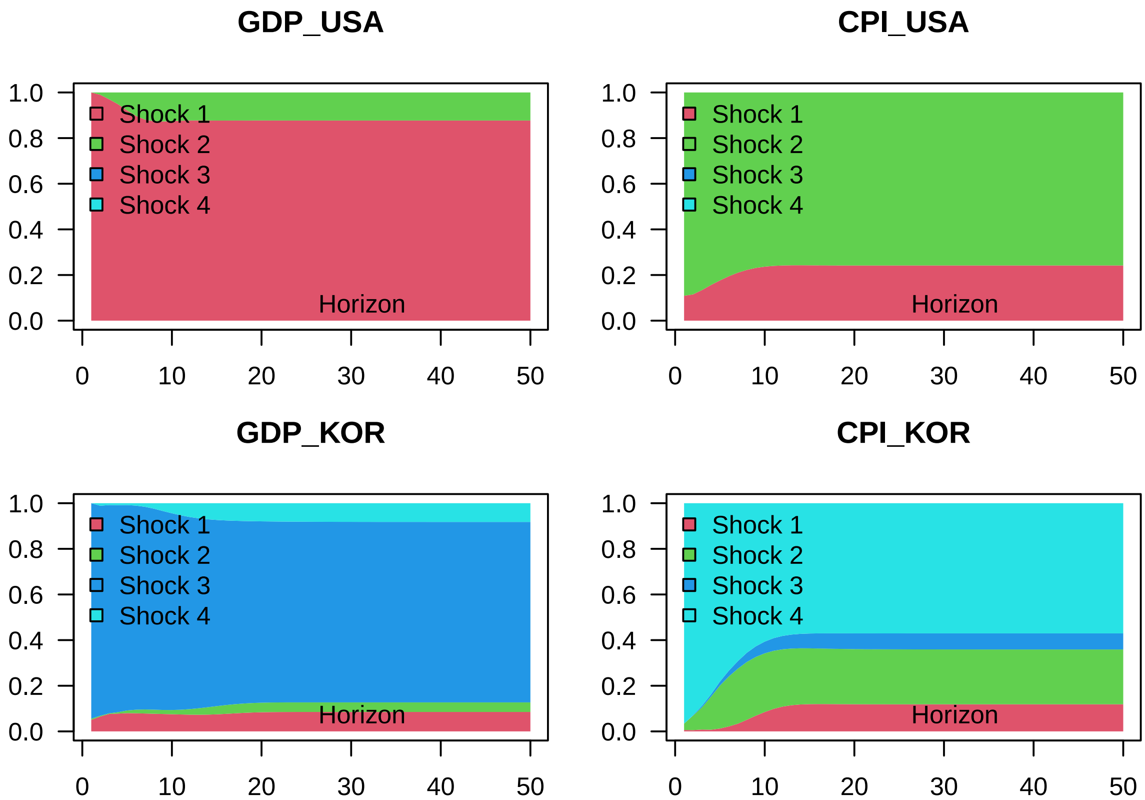

Figure 1.10: Variance decomposition in the context of a VAR model depicting the joint dynamics of US and Korean GDP growth and inflation rates. Structural shocks are identified by using the Cholesky decomposition. The model is contrained in such a way that Korean variables do not affect US ones. (In other words, in this restricted model, US variables are exogenous.)

1.5.2 Granger Causality

Granger (1969) developed a method to explore **causal relationships** among variables. The approach consists in determining whether the past values of $y_{1,t}$ can help explain the current $y_{2,t}$ (beyond the information already included in the past values of $y_{2,t}$).

Formally, let us denote three information sets:

$$
\begin{aligned}
\mathcal{I}_{1,t} &= \{y_{1,t}, y_{1,t-1}, \dots\} \\
\mathcal{I}_{2,t} &= \{y_{2,t}, y_{2,t-1}, \dots\} \\
\mathcal{I}_{t} &= \{y_{1,t}, y_{1,t-1}, \dots y_{2,t}, y_{2,t-1}, \dots\}.
\end{aligned}
$$

We say that $y_{1,t}$ Granger-causes $y_{2,t}$ if

$$\mathbb{E}\left[y_{2,t} \mid \mathcal{I}_{2,t-1}\right] \neq \mathbb{E}\left[y_{2,t} \mid \mathcal{I}_{t-1}\right].$$

To get the intuition behind the testing procedure, consider the following bivariate VAR(p) process:

$$\begin{aligned} y_{1,t} &= c_1 + \Sigma_{i=1}^{p} \Phi_i^{(11)} y_{1,t-i} + \Sigma_{i=1}^{p} \Phi_i^{(12)} y_{2,t-i} + \varepsilon_{1,t} \\ y_{2,t} &= c_2 + \Sigma_{i=1}^{p} \Phi_i^{(21)} y_{1,t-i} + \Sigma_{i=1}^{p} \Phi_i^{(22)} y_{2,t-i} + \varepsilon_{2,t}, \end{aligned}$$

where $\Phi_k^{(ij)}$ denotes the element (i,j) of Φ_k. Then, $y_{1,t}$ is said not to Granger-cause $y_{2,t}$ if

$$\Phi_1^{(21)} = \Phi_2^{(21)} = \dots = \Phi_p^{(21)} = 0.$$

The null and alternative hypotheses therefore are:

$$\begin{cases} H_0: & \Phi_1^{(21)} = \Phi_2^{(21)} = \dots = \Phi_p^{(21)} = 0 \\ H_1: & \Phi_1^{(21)} \neq 0 \text{ or } \Phi_2^{(21)} \neq 0 \text{ or} \dots \Phi_p^{(21)} \neq 0. \end{cases}$$

Loosely speaking, we reject H_0 if some of the coefficients on the lagged $y_{1,t}$'s are statistically significant. Formally, this can be tested using the F-test or asymptotic chi-square test. The F-statistic is

$$F = \frac{(RSS - USS)/p}{USS/(T-2p-1)},$$

where RSS is the Restricted sum of squared residuals and USS is the Unrestricted sum of squared residuals. Under H_0, the F-statistic is distributed as $\mathcal{F}(p, T-2p-1)$ (See Table 11.4).[4]

According to the following lines of code, the output gap Granger-causes inflation, but the reverse is not true:

```
grangertest(US3var[,c("y.gdp.gap","infl")],order=3)
```

```
## Granger causality test
##
## Model 1: infl ~ Lags(infl, 1:3) + Lags(y.gdp.gap, 1:3)
## Model 2: infl ~ Lags(infl, 1:3)
##   Res.Df Df      F   Pr(>F)
## 1    214
## 2    217 -3 3.9761 0.008745 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[4] We have $pF \underset{T\to\infty}{\to} \chi^2(p)$.

```
grangertest(US3var[,c("infl","y.gdp.gap")],order=3)
```

```
## Granger causality test
##
## Model 1: y.gdp.gap ~ Lags(y.gdp.gap, 1:3) + Lags(infl, 1:3)
## Model 2: y.gdp.gap ~ Lags(y.gdp.gap, 1:3)
##   Res.Df Df      F Pr(>F)
## 1    214
## 2    217 -3 1.5451 0.2038
```

Chapter 2

Identification problem and standard identification techniques

2.1 The identification problem

In Section 1.4, we have seen how to estimate $\mathbb{V}ar(\varepsilon_t) = \Omega$ and the Φ_k matrices in the context of a VAR model. But the IRFs are functions of B and of the Φ_k's, not of Ω the Φ_k's (see Section 1.2). We have $\Omega = BB'$, which provides some restrictions on the components of B, but this is not sufficient to fully identify B. Indeed, seen a system of equations whose unknowns are the $b_{i,j}$'s (components of B), the system $\Omega = BB'$ contains only $n(n+1)/2$ linearly independent equations. For instance, for $n = 2$:

$$
\begin{aligned}
&\left[\begin{array}{cc}\omega_{11} & \omega_{12}\\ \omega_{12} & \omega_{22}\end{array}\right] = \left[\begin{array}{cc}b_{11} & b_{12}\\ b_{21} & b_{22}\end{array}\right]\left[\begin{array}{cc}b_{11} & b_{21}\\ b_{12} & b_{22}\end{array}\right]\\
\Leftrightarrow &\left[\begin{array}{cc}\omega_{11} & \omega_{12}\\ \omega_{12} & \omega_{22}\end{array}\right] = \left[\begin{array}{cc}b_{11}^2+b_{12}^2 & b_{11}b_{21}+b_{12}b_{22}\\ b_{11}b_{21}+b_{12}b_{22} & b_{22}^2+b_{21}^2\end{array}\right].
\end{aligned}
$$

We then have 3 linearly independent equations but 4 unknowns. Therefore, B is not identified based on second-order moments. Additional restrictions are required to identify B. This section covers two standard identification schemes: **short-run** and **long-run** restrictions.

1. A **short-run restriction (SRR)** prevents a structural shock from affecting an endogenous variable contemporaneously.

- It is easy to implement: the appropriate entries of B are set to 0.
- A particular (popular) case is that of the **Cholesky, or recursive approach**.
- Examples include Bernanke (1986), Sims (1986), Galí (1992), Ruibio-Ramírez et al. (2010).

2. A **long-run restriction (LRR)** prevents a structural shock from having a cumulative impact on one of the endogenous variables.

- Additional computations are required to implement this. One needs to compute the cumulative effect of one of the structural shocks u_t on one of the endogenous variable.
- Examples include Blanchard and Quah (1989), Faust and Leeper (1997), Galí (1999), Erceg et al. (2005), Christiano et al. (2007).

The two approaches can be combined (see, e.g., Gerlach and Smets (1995)).

2.2 A stylized example motivating short-run restrictions

Let us consider a simple example that could motivate short-run restrictions. Consider the following stylized macro model:

$$\begin{array}{rcll} g_t &=& \bar{g} - \lambda(i_{t-1} - \mathbb{E}_{t-1}\pi_t) + \underbrace{\sigma_d \eta_{d,t}}_{\text{demand shock}} & \text{(IS curve)} \\ \Delta\pi_t &=& \beta(g_t - \bar{g}) + \underbrace{\sigma_\pi \eta_{\pi,t}}_{\text{cost push shock}} & \text{(Phillips curve)} \\ i_t &=& \rho i_{t-1} + \left[\gamma_\pi \mathbb{E}_t \pi_{t+1} + \gamma_g (g_t - \bar{g})\right] & \\ && + \underbrace{\sigma_{mp} \eta_{mp,t}}_{\text{Mon. Pol. shock}} & \text{(Taylor rule),} \end{array} \tag{2.1}$$

where:

$$\eta_t = \begin{bmatrix} \eta_{\pi,t} \\ \eta_{d,t} \\ \eta_{mp,t} \end{bmatrix} \sim i.i.d.\, \mathcal{N}(0, I). \tag{2.2}$$

Vector η_t is assumed to be a vector of structural shocks, mutually and serially independent. On date t:

- g_t is contemporaneously affected by $\eta_{d,t}$ only;
- π_t is contemporaneously affected by $\eta_{\pi,t}$ and $\eta_{d,t}$;
- i_t is contemporaneously affected by $\eta_{mp,t}$, $\eta_{\pi,t}$ and $\eta_{d,t}$.

System (2.1) could be rewritten as follows:

$$\begin{bmatrix} g_t \\ \pi_t \\ i_t \end{bmatrix} = \Phi(L) \begin{bmatrix} g_{t-1} \\ \pi_{t-1} \\ i_{t-1}+ \end{bmatrix} + \underbrace{\underbrace{\begin{bmatrix} 0 & \bullet & 0 \\ \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet \end{bmatrix}}_{=B} \eta_t.}_{=\varepsilon_t} \tag{2.3}$$

This is the **reduced-form** of the model. This representation suggests three additional restrictions on the entries of B; the latter matrix is therefore identified as soon as $\Omega = BB'$ is known (up to the signs of its columns).

2.3 Cholesky: a specific short-run-restriction situation

There are particular cases in which some well-known matrix decomposition of $\Omega = \mathbb{V}ar(\varepsilon_t)$ can be used to easily estimate some specific SVAR. This is the case for the so-called Cholesky decomposition. Consider the following context:

- A first shock (say, $\eta_{n_1,t}$) can affect instantaneously (i.e., on date t) only one of the endogenous variable (say, $y_{n_1,t}$);
- A second shock (say, $\eta_{n_2,t}$) can affect instantaneously (i.e., on date t) two endogenous variables, $y_{n_1,t}$ (the same as before) and $y_{n_2,t}$;
- ...

This implies

1. that column n_1 of B has only 1 non-zero entry (this is the n_1^{th} entry),
2. that column n_2 of B has 2 non-zero entries (the n_1^{th} and the n_2^{th} ones), etc.

Without loss of generality, we can set $n_1 = n$, $n_2 = n-1$, etc. In this context, matrix B is lower triangular. The Cholesky decomposition of Ω_ε then provides an appropriate estimate of B, since this matrix decomposition yields to a lower triangular matrix satisfying:

$$\Omega_\varepsilon = BB'.$$

For instance, Dedola and Lippi (2005) estimate 5 structural VAR models for the US, the UK, Germany, France and Italy to analyse the monetary-policy transmission mechanisms. They estimate SVAR(5) models over the period 1975-1997. The shock-identification scheme is based on Cholesky decompositions, the ordering of the endogenous variables being: (1) the industrial production, (2) the consumer price index, (3) a commodity price index, (4) the short-term rate, (5) monetary aggregate and (6) the effective exchange rate (except for the US). This ordering implies that monetary policy (i.e., the short-term rate) reacts to the shocks affecting the first three variables but that the latter react to monetary policy shocks with a one-period lag only.

Consider again the small structural example discussed above, whose reduced-form VAR representation is given in Eq. (2.3). If we reorder the structural shocks as $\eta_t = (\eta_{d,t}, \eta_{\pi,t}, \eta_{mp,t})'$ (the order being arbitrary), the impact matrix B can be written as

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

that is, B corresponds to the Cholesky factor of Ω.

However, it is not always possible to obtain a lower-triangular B matrix by reordering the structural shocks. For instance, if for a given ordering we have

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

then no reordering of the shocks will produce a lower-triangular structure.

The Cholesky approach can be employed when one is interested in one specific structural shock. This is the case, e.g., of Christiano et al. (1996). Their identification is based on the following relationship between ε_t and η_t:

$$\left[\begin{array}{c}\varepsilon_{S,t}\\ \varepsilon_{r,t}\\ \varepsilon_{F,t}\end{array}\right] = \left[\begin{array}{ccc}B_{SS} & 0 & 0\\ B_{rS} & B_{rr} & 0\\ B_{FS} & B_{Fr} & B_{FF}\end{array}\right]\left[\begin{array}{c}\eta_{S,t}\\ \eta_{r,t}\\ \eta_{F,t}\end{array}\right],$$

where S, r and F respectively correspond to *slow-moving variables*, the policy variable (short-term rate) and *fast-moving variables*. While $\eta_{r,t}$ is scalar, $\eta_{S,t}$ and $\eta_{F,t}$ may be vectors. The space spanned by $\varepsilon_{S,t}$ is the same as that spanned by $\eta_{S,t}$. As a result, because $\varepsilon_{r,t}$ is a linear combination of $\eta_{r,t}$ and $\eta_{S,t}$ (which are $\perp$), it comes that the $B_{rr}\eta_{r,t}$'s are the (population) residuals in the regression of $\varepsilon_{r,t}$ on $\varepsilon_{S,t}$. Because $\mathbb{V}ar(\eta_{r,t}) = 1$, B_{rr} is given by the square root of the variance of $B_{rr}\eta_{r,t}$. $B_{F,r}$ is finally obtained by regressing the components of $\varepsilon_{F,t}$ on the estimates of $\eta_{r,t}$.

An equivalent approach consists in computing the Cholesky decomposition of BB' and the contemporaneous impacts of the monetary policy shock (on the n endogenous variables) are the components of the column of B corresponding to the policy variable.

```
library(IdSS)
library(vars)
data("USmonthly")
# Select sample period:
First.date <- "1965-01-01";Last.date <- "1995-06-01"
indic.first <- which(USmonthly$DATES==First.date)
indic.last  <- which(USmonthly$DATES==Last.date)
USmonthly   <- USmonthly[indic.first:indic.last,]
considered.variables <- c("LIP","UNEMP","LCPI","LPCOM","FFR","NBR","TTR","M1")
y <- as.matrix(USmonthly[considered.variables])
res.svar.ordering <- svar.ordering(y,p=3,
                                   posit.of.shock = 5,
                                   nb.periods.IRF = 20,
                                   nb.bootstrap.replications = 100,
                                   confidence.interval = 0.90, # expressed in pp.
                                   indic.plot = 1 # Plots are displayed if = 1.
                                   )
```

2.4 Long-run restrictions

Let us now turn to **long-run restrictions**. Such a restriction concerns the long-run influence of a shock on an endogenous variable. Let us consider for instance a structural shock that is

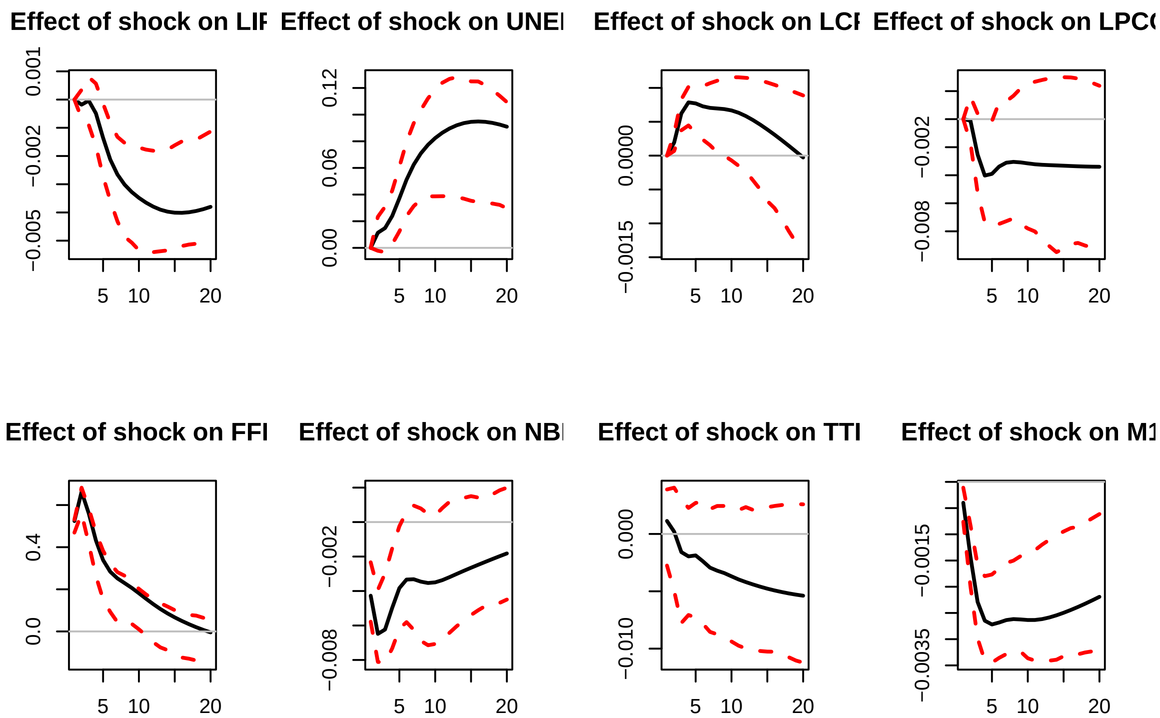

Figure 2.1: Response to a monetary-policy shock. Identification approach of Christiano, Eichenbaum and Evans (1996). Confidence intervals are obtained by boostrapping the estimated VAR model (see inference section).

assumed to have no “long-run influence” on GDP. How to express this? The long-run change in GDP can be expressed as $GDP_{t+h} - GDP_t$, with h large. Note further that:

$$GDP_{t+h} - GDP_t = \Delta GDP_{t+h} + \Delta GDP_{t+h-1} + \dots + \Delta GDP_{t+1}.$$

Hence, the fact that a given structural shock ($\eta_{i,t}$, say) has no long-run influence on GDP means that

$$\lim_{h\rightarrow\infty} \frac{\partial GDP_{t+h}}{\partial \eta_{i,t}} = \lim_{h\rightarrow\infty} \frac{\partial}{\partial \eta_{i,t}} \left(\sum_{k=1}^{h} \Delta GDP_{t+k} \right) = 0.$$

This long-run effect can be formulated as a function of B and of the matrices Φ_i when y_t (including ΔGDP_t) follows a VAR process.

2.4.1 The VAR(1) case

Let us start with the VAR(1) case. We have:

$$\begin{aligned}
y_t &= c + \Phi y_{t-1} + \varepsilon_t \\
&= c + \varepsilon_t + \Phi(c + \varepsilon_{t-1}) + \dots + \Phi^k(c + \varepsilon_{t-k}) + \dots \\
&= \mu + \varepsilon_t + \Phi \varepsilon_{t-1} + \dots + \Phi^k \varepsilon_{t-k} + \dots \\
&= \mu + B\eta_t + \Phi B \eta_{t-1} + \dots + \Phi^k B \eta_{t-k} + \dots,
\end{aligned}$$

which is the Wold representation of y_t.

The sequence of shocks $\{\eta_t\}$ determines the sequence $\{y_t\}$. What if $\{\eta_t\}$ is replaced with $\{\tilde{\eta}_t\}$, where $\tilde{\eta}_t = \eta_t$ if $t \neq s$ and $\tilde{\eta}_s = \eta_s + \gamma$? Assume $\{\tilde{y}_t\}$ is the associated “perturbated” sequence. We have $\tilde{y}_t = y_t$ if $t < s$. For $t \geq s$, the Wold decomposition of $\{\tilde{y}_t\}$ implies:

$$\tilde{y}_t = y_t + \Phi^{t-s} B \gamma.$$

Therefore, the cumulative impact of γ on $\tilde{y}_t$ will be (for $t \geq s$):

$$\begin{aligned}
(\tilde{y}_t - y_t) + (\tilde{y}_{t-1} - y_{t-1}) + \dots + (\tilde{y}_s - y_s) &= \\
(Id + \Phi + \Phi^2 + \dots + \Phi^{t-s}) B \gamma. & \qquad (2.4)
\end{aligned}$$

Consider a shock on $\eta_{1,t}$, with a magnitude of 1. This shock corresponds to $\gamma = [1, 0, \dots, 0]'$. Given Eq. (2.4), the long-run cumulative effect of this shock on the endogenous variables is given by:

$$\underbrace{\underbrace{(Id + \Phi + \dots + \Phi^k + \dots)}_{=(Id-\Phi)^{-1}} B}_{=:\Theta} \left[\begin{array}{c} 1 \\ 0 \\ \vdots \\ 0 \end{array}\right],$$

which is the first column of the $n \times n$ matrix $\Theta := (Id - \Phi)^{-1} B$.

In this context, consider the following long-run restriction: *"the j^{th} structural shock has no cumulative impact on the i^{th} endogenous variable"*. It is equivalent to

$$\Theta_{ij} = 0,$$

where Θ_{ij} is the element (i,j) of Θ.

If $n(n-1)/2$ restrictions of this type can be made, then B is identified. In particular, in the case case where Θ is lower-triangular, the problem admits an analytical solution. Indeed, let $J = (I - \Phi)^{-1}$ (assumed to be invertible). We want $\Theta = JB$ to be lower triangular and $\Omega = BB'$. Since $B = J^{-1}\Theta$, we have

$$\Omega = BB' = J^{-1}\Theta\Theta'(J^{-1})' \;\Rightarrow\; J\Omega J' = \Theta\Theta'.$$

Since Ω is positive definite and J is invertible, $\Sigma = J\Omega J'$ is positive definite. Take the (unique, with positive diagonal) Cholesky factorization $\Sigma = \Theta\Theta'$ with Θ lower triangular. Then set $B = J^{-1}\Theta$. This B satisfies $\Omega = BB'$ and $(I-\Phi)^{-1}B = \Theta$ is lower triangular.

2.4.2 The VAR(p) case

Several of the developments made above are still valid in the VAR(p) case since a VAR(p) process can be rewritten as a VAR(1) process by augmenting the state vector. More specifically, stack the last p values of vector y_t in vector $y_t^* = [y_t', \dots, y_{t-p+1}']'$; Eq. (1.1) can then be rewritten in its **companion form**:

$$y_t^* = \underbrace{\left[\begin{array}{c} c \\ 0_{n\times 1} \\ \vdots \\ 0_{n\times 1} \end{array}\right]}_{=c^*} + \underbrace{\left[\begin{array}{cccc} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ I & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I & 0 \end{array}\right]}_{=\Phi} y_{t-1}^* + \underbrace{\left[\begin{array}{c} \varepsilon_t \\ 0_{n\times 1} \\ \vdots \\ 0_{n\times 1} \end{array}\right]}_{\varepsilon_t^*}, \tag{2.5}$$

where matrices Φ and $\Omega^* := \mathbb{V}ar(\varepsilon_t^*)$ are of dimension $np \times np$. Matrix Φ had been introduced in Eq. (1.10), and

$$\Omega^* := \mathbb{V}ar(\varepsilon_t^*) = \left[\begin{array}{c} B \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{array}\right] \left[\begin{array}{cccc} B & 0_{n\times n} & \dots & 0_{n\times n} \end{array}\right] = \left[\begin{array}{cccc} \Omega & 0_{n\times n} & \dots & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & & \\ \vdots & & \ddots & \\ 0_{n\times n} & \dots & & 0_{n\times n} \end{array}\right].$$

In that context, the long-run effect on $y_t^* + y_{t+1}^* + y_{t+2}^* + \dots$ of a change in η_t by γ (that happens on date t) is given by:

$$\underbrace{(Id + \Phi + \dots + \Phi^k + \dots)}_{=(Id-\Phi)^{-1}} \left[\begin{array}{c} B \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{array}\right] \gamma.$$

In particular, the cumulated effect of this shock on y_t (and not y_t^* anymore) is $JB\gamma$, where J is the upper-left $n \times n$ submatrix of $(I-\Phi)^{-1}$. Let us denote the matrix JB by Θ. With these notations, the long-run restriction: *the j^{th} structural shock has no cumulative impact on the i^{th} endogenous variable* is equivalent to the fact that the component (i,j) of Θ is equal to zero.

If $n(n-1)/2$ restrictions of this type can be made, then B is identified. In particular, in the case where Θ is lower triangular, the problem admits an analytical solution. Using $B = J^{-1}\Theta$, we get

$$\Omega = BB' = J^{-1}\Theta\Theta'(J^{-1})' \;\Rightarrow\; J\Omega J' = \Theta\Theta'.$$

Since Ω is positive definite, $\Sigma = J\Omega J'$ is positive definite. Take the (unique, with positive diagonal) Cholesky factorization $\Sigma = \Theta\Theta'$ with Θ lower triangular. Then set $B = J^{-1}\Theta$. This B satisfies $\Omega = BB'$ and $JB = \Theta$ is lower triangular.

2.4.3 Blanchard and Quah (1989)

Blanchard and Quah (1989) have implemented such long-run restrictions in a small-scale VAR. Two variables are considered: GDP and unemployment. Consequently, the VAR is affected by two types of shocks. Specifically, authors want to identify **supply shocks** (that can have a permanent effect on output) and **demand shocks** (that cannot have a permanent effect on output).[1]

Blanchard and Quah (1989)'s dataset is quarterly, spanning the period from 1950:2 to 1987:4. Their VAR features 8 lags. Here are the data they use:

```
library(IdSS)
data(BQ)
par(mfrow=c(1,2))
plot(BQ$Date,BQ$Dgdp,type="l",main="GDP quarterly growth rate",
     xlab="",ylab="",lwd=2)
plot(BQ$Date,BQ$unemp,type="l",ylim=c(-3,6),main="Unemployment rate (gap)",
     xlab="",ylab="",lwd=2)
```

Estimate a reduced-form VAR(8) model:

```
library(vars)
y <- BQ[,2:3]
n <- dim(y)[2]
p <- 8
est.VAR <- VAR(y,p=p)
Omega <- var(residuals(est.VAR))
```

[1]The motivation of the authors regarding their long-run restrictions can be obtained from a traditional Keynesian view of fluctuations. The authors propose a variant of a model from Fischer (1977).

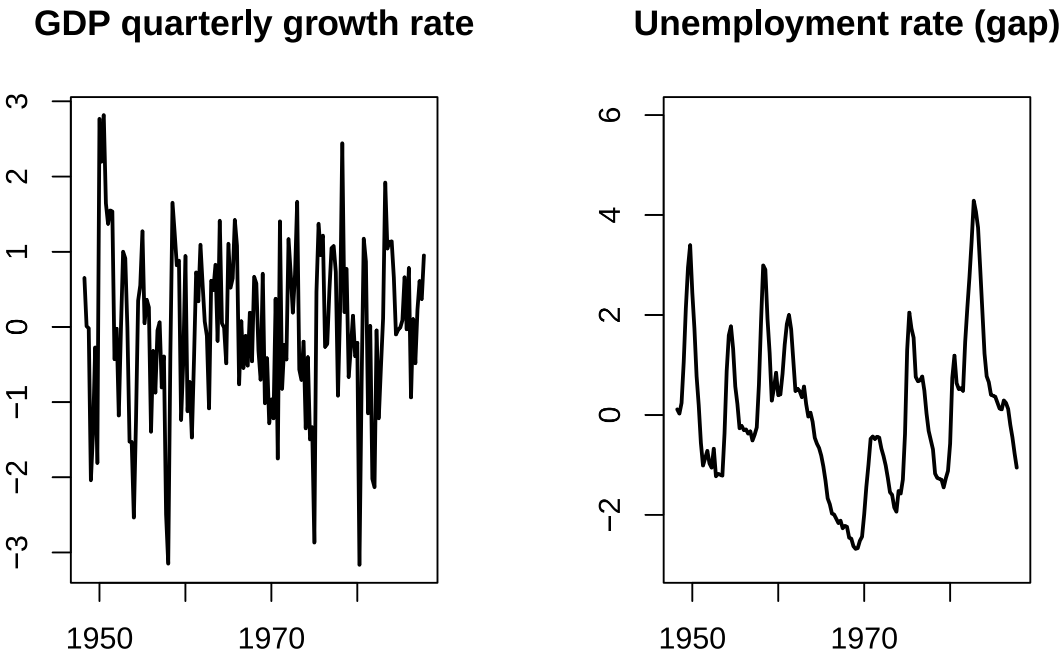

Figure 2.2: These data come from Blanchard and Quah (1989). GDP growth rates are calculated as the first differences of the logarithm of GDP.

Let us employ the approach described above:

```
Phi <- Acoef(est.VAR)
PHI <- make.PHI(Phi)
A <- diag(n*p) - PHI
J <- solve(A)[1:n,1:n]
Sigma <- J %*% Omega %*% t(J)
L <- t(chol(Sigma))
B <- solve(J) %*% L
```

Now, let us define a loss function (`loss`) that is equal to zero if (a) $BB' = \Omega$ and (b) the element (1,1) of $\Theta = (Id - \Phi)^{-1}B$ is equal to zero:

```
# Compute (Id - Phi)^{-1}:
Phi <- Acoef(est.VAR)
PHI <- make.PHI(Phi)
sum.PHI.k <- solve(diag(dim(PHI)[1]) - PHI)[1:2,1:2]
loss <- function(param){
  B <- matrix(param,2,2)
  X <- Omega - B %*% t(B)
  Theta <- sum.PHI.k[1:2,1:2] %*% B
  loss <- 10000 * ( X[1,1]^2 + X[2,1]^2 + X[2,2]^2 + Theta[1,1]^2 )
  return(loss)
}
res.opt <- optim(c(1,0,0,1),loss,method="BFGS",hessian=FALSE)
print(res.opt$par)
```

```
## [1]  0.8570358 -0.2396345  0.1541395  0.1921221
```

(Note: one can use that type of approach, based on a loss function, to mix short- and long-run restrictions.)

Figure 2.3 displays the resulting IRFs. Note that, for GDP, we cumulate the GDP growth IRF, so as to have the response of the GDP in level.

```
B.hat <- matrix(res.opt$par,2,2)
print(cbind(Omega,B.hat %*% t(B.hat)))
```

```
##             Dgdp       unemp
## Dgdp   0.7582704 -0.17576173  0.7582694 -0.17576173
## unemp -0.1757617  0.09433658 -0.1757617  0.09433558
```

```
nb.sim <- 40
par(mfrow=c(2,2));par(plt=c(.15,.95,.15,.8))
Y <- simul.VAR(c=matrix(0,2,1),Phi,B.hat,nb.sim,y0.star=rep(0,2*8),
               indic.IRF = 1,u.shock = c(1,0))
plot(cumsum(Y[,1]),type="l",lwd=2,xlab="",ylab="",main="Demand shock on GDP")
plot(Y[,2],type="l",lwd=2,xlab="",ylab="",main="Demand shock on UNEMP")
Y <- simul.VAR(c=matrix(0,2,1),Phi,B.hat,nb.sim,y0.star=rep(0,2*8),
               indic.IRF = 1,u.shock = c(0,1))
plot(cumsum(Y[,1]),type="l",lwd=2,xlab="",ylab="",main="Supply shock on GDP")
plot(Y[,2],type="l",lwd=2,xlab="",ylab="",main="Supply shock on UNEMP")
```

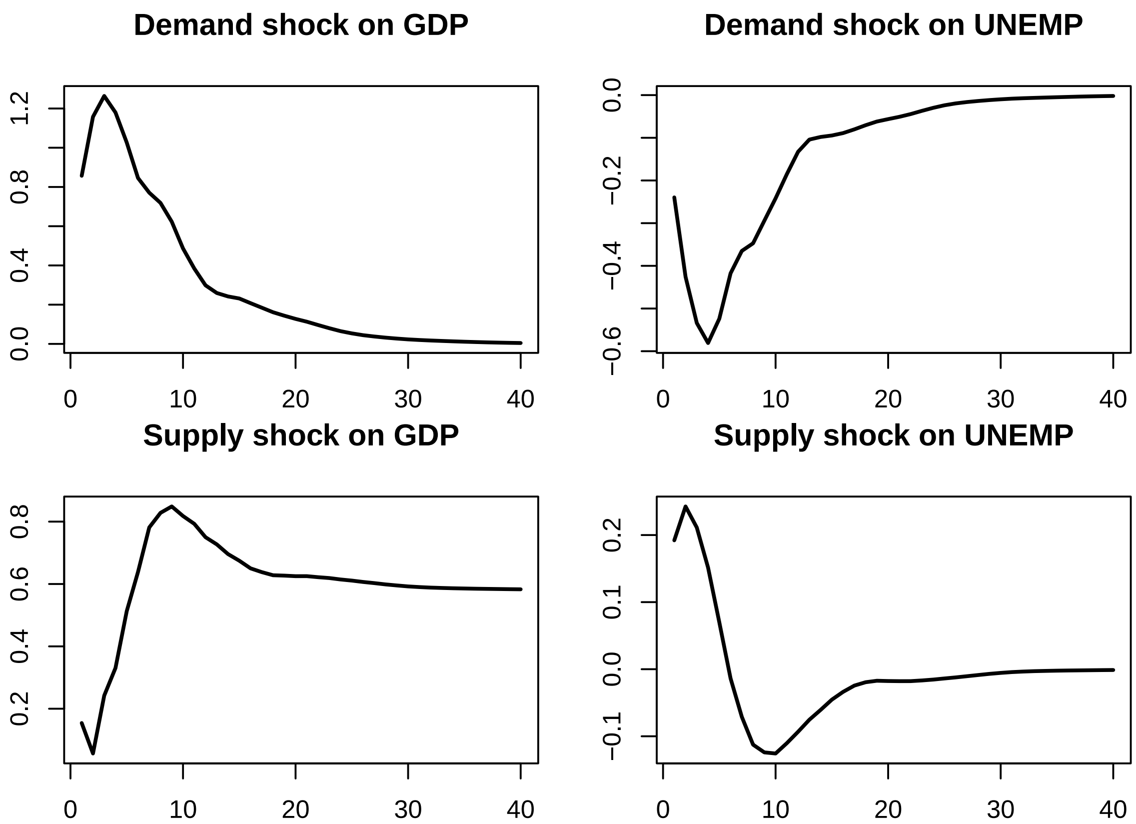

Figure 2.3: IRF of GDP and unemployment to demand and supply shocks.

Chapter 3

Inference

Consider the following SVAR model:

$$y_t = \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \varepsilon_t$$

with $\varepsilon_t = B\eta_t$, $\Omega_\varepsilon = BB'$.

The corresponding infinite MA representation (Eq. (1.4), or Wold representation, is:

$$y_t = \sum_{h=0}^{\infty} \Psi_h \eta_{t-h},$$

where $\Psi_0 = B$ and for $h = 1, 2, \dots$:

$$\Psi_h = \sum_{j=1}^{h} \Psi_{h-j}\Phi_j,$$

with $\Phi_j = 0$ for $j > p$ (see Prop. 1.1 for this recursive computation of the Ψ_j's).

Inference on the VAR coefficients $\{\Phi_j\}_{j=1,\dots,p}$ is straightforward (standard OLS inference). But inference is more complicated regarding IRFs. Indeed, as shown by the previous equation, the (infinite) MA coefficients $\{\Psi_j\}_{j=1,\dots}$ are non-linear functions of the $\{\Phi_j\}_{j=1,\dots,p}$ and of Ω_ε. An other issue pertain to small sample bias: typically, for persistent processes, auto-regressive parameters are known to be downward biased.

The main inference methods are the following:

- Monte Carlo method (Hamilton (1994))
- Asymptotic normal approximation (Lütkepohl (1990)), or Delta method
- Bootstrap method (Kilian (1998))

3.1 Monte Carlo method

We use Monte Carlo when we need to approximate the distribution of a variable whose distribution is unknown (here: the Ψ_j's) but which is a function of another variable whose distribution is known (here, the Φ_j's).

For instance, suppose we know the distribution of a random variable X, which takes values in $\mathbb{R}$, with density function p. Assume we want to compute the mean of $\varphi(X)$. We have:

$$\mathbb{E}(\varphi(X)) = \int_{-\infty}^{+\infty} \varphi(x)p(x)dx$$

Suppose that the above integral does not have a simple expression. We cannot compute $\mathbb{E}(\varphi(X))$ but, by virtue of the law of large numbers, we can approximate it as follows:

$$\mathbb{E}(\varphi(X)) \approx \frac{1}{N}\sum_{i=1}^{N} \varphi(X^{(i)}),$$

where $\{X^{(i)}\}_{i=1,\dots,N}$ are N independent draws of X. More generally, the distribution of $\varphi(X)$ can be approximated by the empirical distribution of the $\varphi(X^{(i)})$'s. Typically, if 10'000 values of $\varphi(X^{(i)})$ are drawn, the 5^{th} percentile of the p.d.f. of $\varphi(X)$ can be approximated by the 500^{th} value of the 10'000 draws of $\varphi(X^{(i)})$ (after arranging these values in ascending order).

As regards the computation of confidence intervals around IRFs, one has to think of $\{\widehat{\Phi}_j\}_{j=1,\dots,p}$, and of $\widehat{\Omega}$ as X and $\{\widehat{\Psi}_j\}_{j=1,\dots}$ as $\varphi(X)$. (Proposition 1.4 provides us with the asymptotic distribution of the "X.")

To summarize, here are the steps one can implement to derive confidence intervals for the IRFs using the Monte-Carlo approach: For each iteration k,

1. Draw $\{\widehat{\Phi}_j^{(k)}\}_{j=1,\dots,p}$ and $\widehat{\Omega}^{(k)}$ from their asymptotic distribution (using Proposition 1.4).
2. Compute the matrix $B^{(k)}$ so that $\widehat{\Omega}^{(k)} = B^{(k)}B^{(k)\prime}$, according to your identification strategy.
3. Compute the associated IRFs $\{\widehat{\Psi}_j\}^{(k)}$.

Perform N replications and report the median impulse response (and its confidence intervals).

The following code implements the Monte Carlo method.

```r
library(IdSS);library(vars);library(Matrix)
data("USmonthly")
First.date <- "1965-01-01"
Last.date <- "1995-06-01"
indic.first <- which(USmonthly$DATES==First.date)
indic.last  <- which(USmonthly$DATES==Last.date)
```

```
USmonthly   <- USmonthly[indic.first:indic.last,]
considered.variables<-c("LIP","UNEMP","LCPI","LPCOM","FFR","NBR","TTR","M1")
y <- as.matrix(USmonthly[considered.variables])
# ===================================
# CEE with Monte Carlo
# ===================================
res.svar.ordering <-
  svar.ordering.2(y,p=3,
                  posit.of.shock = 5,
                  nb.periods.IRF = 20,
                  inference = 3,# 0 -> no inference, 1 -> parametric bootst.,
                  # 2 <- non-parametric bootstrap, 3 <- Monte Carlo,
                  # 4 <- bootstrap-after-bootstrap
                  nb.draws = 200,
                  confidence.interval = 0.90, # expressed in pp.
                  indic.plot = 1 # Plots are displayed if = 1.
  )
```

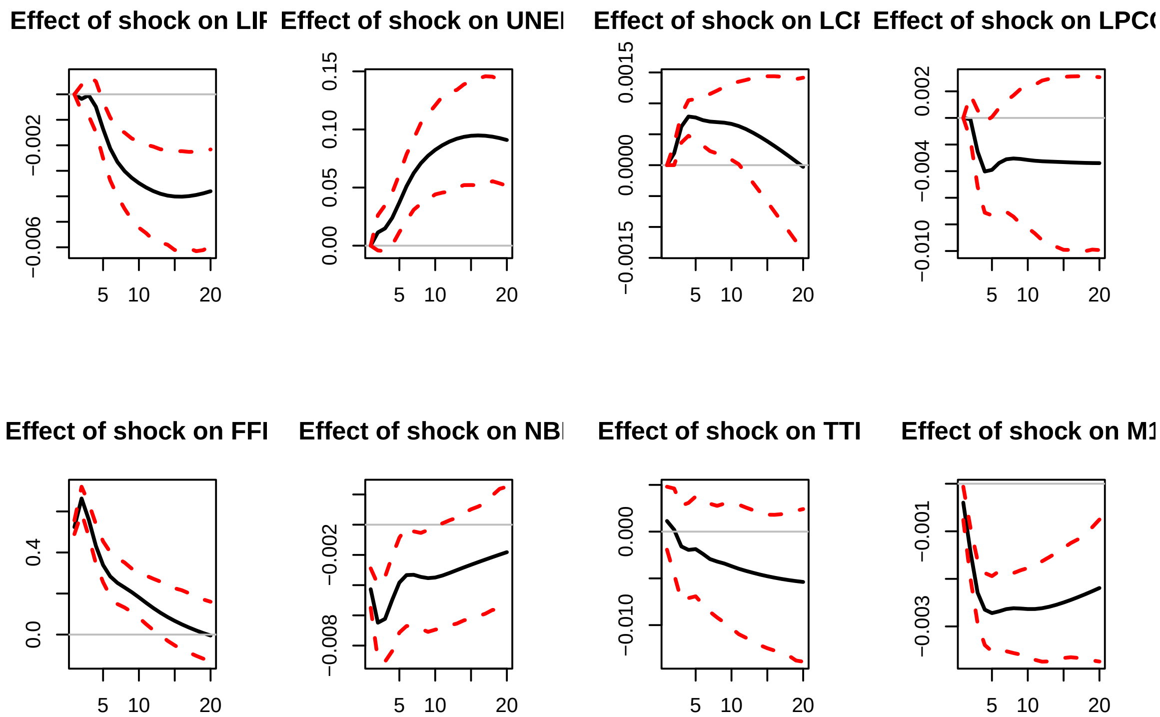

Figure 3.1: IRF associated with a monetary policy shock; Monte Carlo method.

```
IRFs.ordering <- res.svar.ordering$IRFs
median.IRFs.ordering <- res.svar.ordering$all.CI.median
simulated.IRFs.ordering <- res.svar.ordering$simulated.IRFs
```

Below we plot the estimated IRFs to a mom^netary policy shock, together with a sample of 50 IRFs simulated with the Monte Carlo method, and the median of all simulated IRFs.

```
par(mfrow=c(1,1))
plot(IRFs.ordering[,5],type="l")
for(i in 1:50){
  lines(simulated.IRFs.ordering[,5,i],col="red",type="l")
}
lines(IRFs.ordering[,5],col="black",type="l")
lines(median.IRFs.ordering[,5],col="blue",type="l")
```

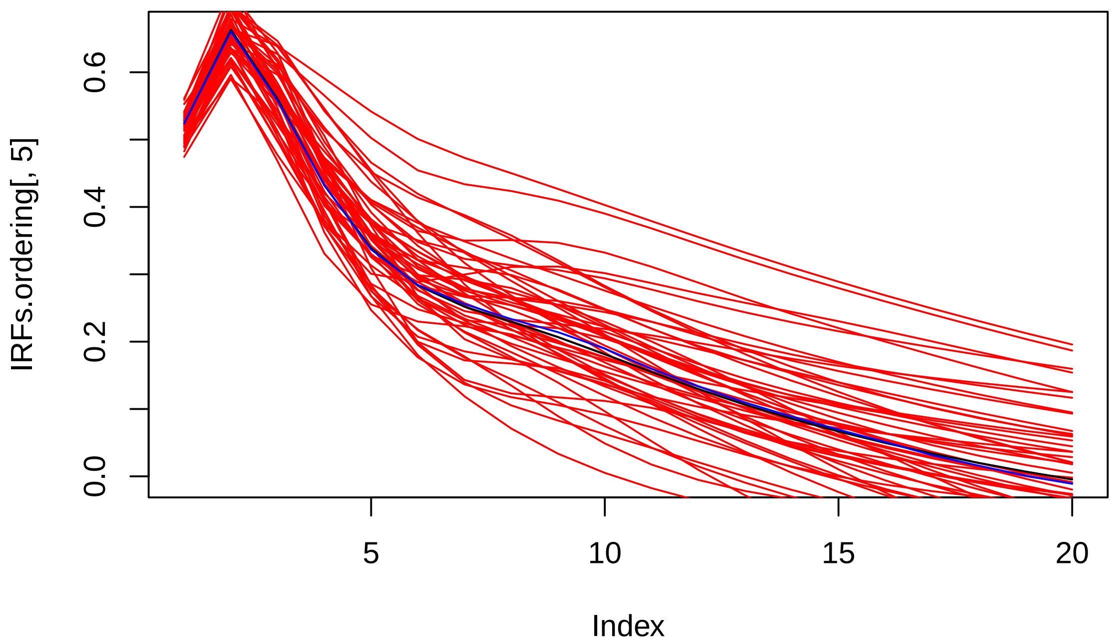

Figure 3.2: Estimated IRF associated with a monetary policy shock (black line) with simulated IRFs (red lines) and median IRF (blue line); Monte Carlo method.

3.2 Delta method

Suppose β is a vector of parameters and β is an estimator such that

$$\sqrt{T}(\hat{\beta}-\beta) \overset{d}{\to} \mathcal{N}(0, \Sigma_\beta),$$

where d denotes convergence in distribution, $N(0, \Sigma_\beta)$ denotes the multivariate normal distribution with mean vector 0 and covariance matrix Σ_β and T is the size of the sample used for estimation.

Let $\varphi(\beta) = (\varphi_l(\beta), ..., \varphi_m(\beta))'$ be a continuously differentiable function with values in $\mathbb{R}^m$, and assume that $\partial \varphi_i / \partial \beta' = (\partial \varphi_i / \partial \beta_j)$ is nonzero at β for $i = 1, \dots, m$. Then

$$\sqrt{T}(\varphi(\hat{\beta}) - \varphi(\beta)) \overset{d}{\to} \mathcal{N}\left(0, \frac{\partial \varphi}{\partial \beta'} \Sigma_\beta \frac{\partial \varphi'}{\partial \beta}\right).$$

Using this property, Lütkepohl (1990) provides the asymptotic distributions of the Ψ_j's.

A limit of the last two approaches (Monte Carlo and the Delta method) is that they rely on asymptotic results and the normality assumption. Boostrapping approaches are more robust in small-sample and non-normal situations.

3.3 Bootstrap

IRFs' confidence intervals are intervals where 90% (or 95%, 75%, …) of the IRFs would lie, if we were to repeat the estimation a large number of times in similar conditions (T observations). We obviously cannot do this, because we have only one sample: $\{y_t\}_{t=1,..,T}$. But we can try to *construct* such samples.

Bootstrapping consists in:

- re-sampling N times, i.e., constructing N samples of T observations, using the estimated VAR coefficients and

a. a sample of residuals from the distribution $N(0, BB')$ (**parametric approach**), or
b. a sample of residuals drawn randomly from the set of the actual estimated residuals $\{\hat{\varepsilon}_t\}_{t=1,..,T}$. (**non-parametric approach**).

- re-estimating the SVAR N times.

Here is the algorithm for the non-parametric approach:

1. Construct a sample
$$y_t^{(k)} = \widehat{\Phi}_1 y_{t-1}^{(k)} + \dots + \widehat{\Phi}_p y_{t-p}^{(k)} + \hat{\varepsilon}_t^{(k)},$$
with $\hat{\varepsilon}_t^{(k)} = \hat{\varepsilon}_{s_t^{(k)}}$, where $\{s_1^{(k)}, .., s_T^{(k)}\}$ is a random set from $\{1, .., T\}^T$. (Note: in the parametric approach, we would draw $\hat{\varepsilon}_t^{(k)}$ from the $N(0, BB')$ distribution)

2. Re-estimate the SVAR and compute the IRFs $\{\widehat{\Psi}_j\}^{(k)}$.

Perform N replications and report the median impulse response (and its confidence intervals).

The following code implements the bootstrap method.

```
library(IdSS);library(vars);library(Matrix)
data("USmonthly")
First.date <- "1965-01-01"
Last.date <- "1995-06-01"
indic.first <- which(USmonthly$DATES==First.date)
indic.last  <- which(USmonthly$DATES==Last.date)
USmonthly   <- USmonthly[indic.first:indic.last,]
considered.variables<-c("LIP","UNEMP","LCPI","LPCOM","FFR","NBR","TTR","M1")
y <- as.matrix(USmonthly[considered.variables])
# ====================================
# CEE with bootstrap
# ====================================
res.svar.ordering <-
  svar.ordering.2(y,p=3,
                  posit.of.shock = 5,
                  nb.periods.IRF = 20,
                  inference = 2,# 0 -> no inference, 1 -> parametric bootstr.,
                  # 2 <- non-parametric bootstrap, 3 <- monte carlo,
                  # 4 <- bootstrap-after-bootstrap
                  nb.draws = 200,
                  confidence.interval = 0.90, # expressed in pp.
                  indic.plot = 1 # Plots are displayed if = 1.
  )
```

```
IRFs.ordering.bootstrap <- res.svar.ordering$IRFs
median.IRFs.ordering.bootstrap <- res.svar.ordering$all.CI.median
simulated.IRFs.ordering.bootstrap <- res.svar.ordering$simulated.IRFs
```

3.4 Bootstrap-after-bootstrap

The previous simple bootstrapping procedure deals with non-normality and small sample distribution, since we use the actual residuals. However, it does not deal with the *small sample bias*, stemming, in particular, from small-sample bias associated with OLS coefficient estimates $\{\widehat{\Phi}_j\}_{j=1,..,p}$.

The code below illustrates the small sample bias.

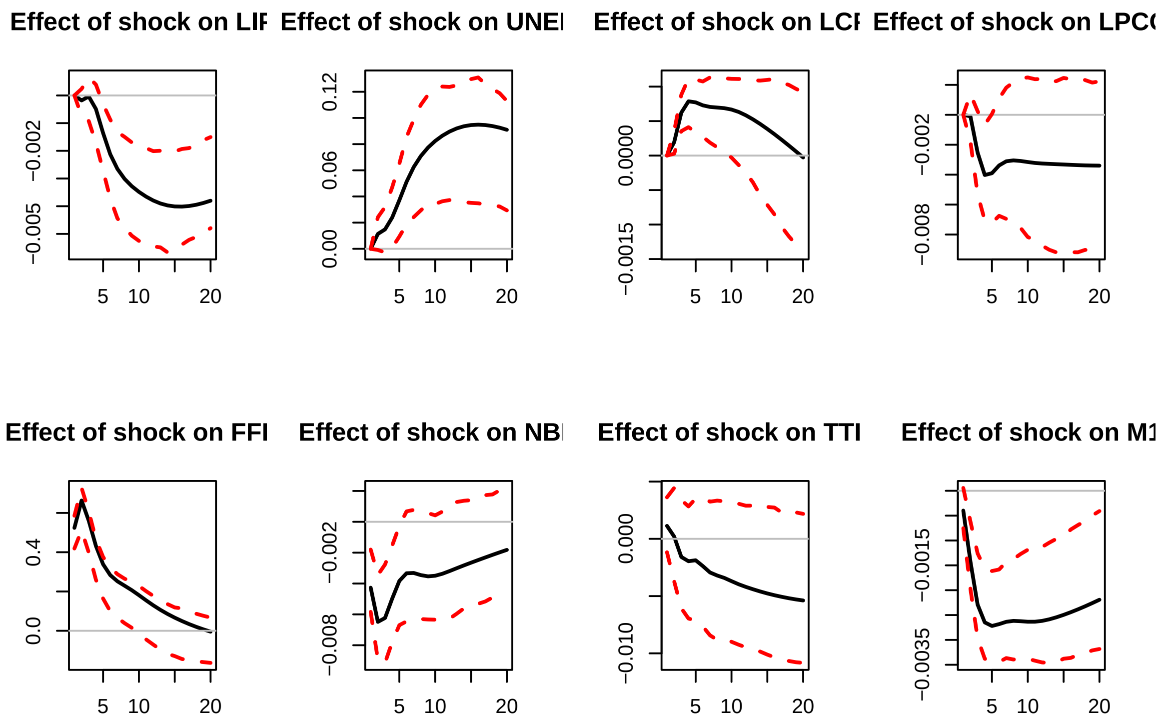

Figure 3.3: IRF associated with a monetary policy shock; bootstrap method.

```
# Distribution of coefficients stemming from non-parametric bootstrap
n <- length(considered.variables)
h <- 5
par(mfrow=c(2,ifelse(round(n/2)==n/2,n/2,(n+1)/2)))
for (i in 1:n){
  hist(simulated.IRFs.ordering.bootstrap[h,i,],xlab="",ylab="",
       main=paste("Effect at h = ",h," on ",
                  considered.variables[i],sep=""),cex.main=.9)
  lines(array(c(IRFs.ordering.bootstrap[h,i],
                IRFs.ordering.bootstrap[h,i],0,100),c(2,2)),col="red")
  lines(array(c(median.IRFs.ordering.bootstrap[h,i],
                median.IRFs.ordering.bootstrap[h,i],0,100),c(2,2)),col="blue")
  text(IRFs.ordering.bootstrap[h,i],25,label="Estimated coef.",col="red")
}
```

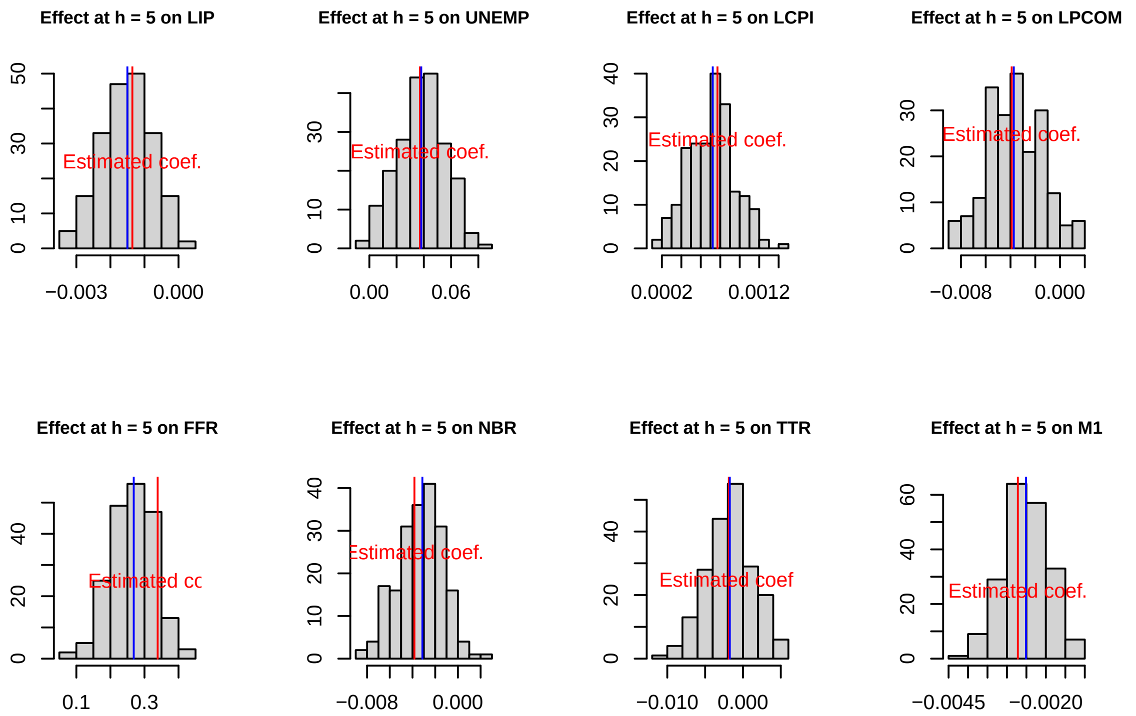

Figure 3.4: Estimated and bootstrapped coefficients.

The distribution of the bootstrapped coefficients is not centered around the estimated coefficients.

In the following code, we perform the VAR estimation and bootstrap inference after generating artificial data. We can then compare the IRFs and confidence intervals to the "true'' parameters used to generate the data.

```
# Simulate a small sample
est.VAR <- VAR(y,p=3)
Phi     <- Acoef(est.VAR)
cst     <- Bcoef(est.VAR)[,3*n+1]
resids  <- residuals(est.VAR)
Omega   <- var(resids)
B.hat   <- t(chol(Omega))
y0.star <- NULL
for(k in 3:1){
  y0.star <- c(y0.star,y[k,])
}
small.sample <- simul.VAR(c=rep(0,dim(y)[2]),
                          Phi,
                          B.hat,
                          nb.sim = 100,
                          y0.star,
                          indic.IRF = 0)
colnames(small.sample)  <- considered.variables
# Estimate the VAR with the small sample
res.svar.small.sample <-
  svar.ordering.2(small.sample,p=3,
                  posit.of.shock = 5,
                  nb.periods.IRF = 20,
                  inference = 2,# 0 -> no inference, 1 -> parametric bootstr.,
                  # 2 <- non-parametric bootstrap, 3 <- monte carlo
                  nb.draws = 200,
                  confidence.interval = 0.90, # expressed in pp.
                  indic.plot = 1 # Plots are displayed if = 1.
  )
```

```
IRFs.small.sample <- res.svar.small.sample$IRFs
median.IRFs.small.sample <- res.svar.small.sample$all.CI.median
simulated.IRFs.small.sample <- res.svar.small.sample$simulated.IRFs
```

```
# True IRFs
res.svar.ordering <-
  svar.ordering.2(y,p=3,
                  posit.of.shock = 5,
                  nb.periods.IRF = 20,
                  inference = 0,# 0 -> no inference, 1 -> parametric bootstr.,
                  # 2 <- non-parametric bootstrap, 3 <- monte carlo,
                  # 4 <- bootstrap-after-bootstrap
                  indic.plot = 0 # Plots are displayed if = 1.
```

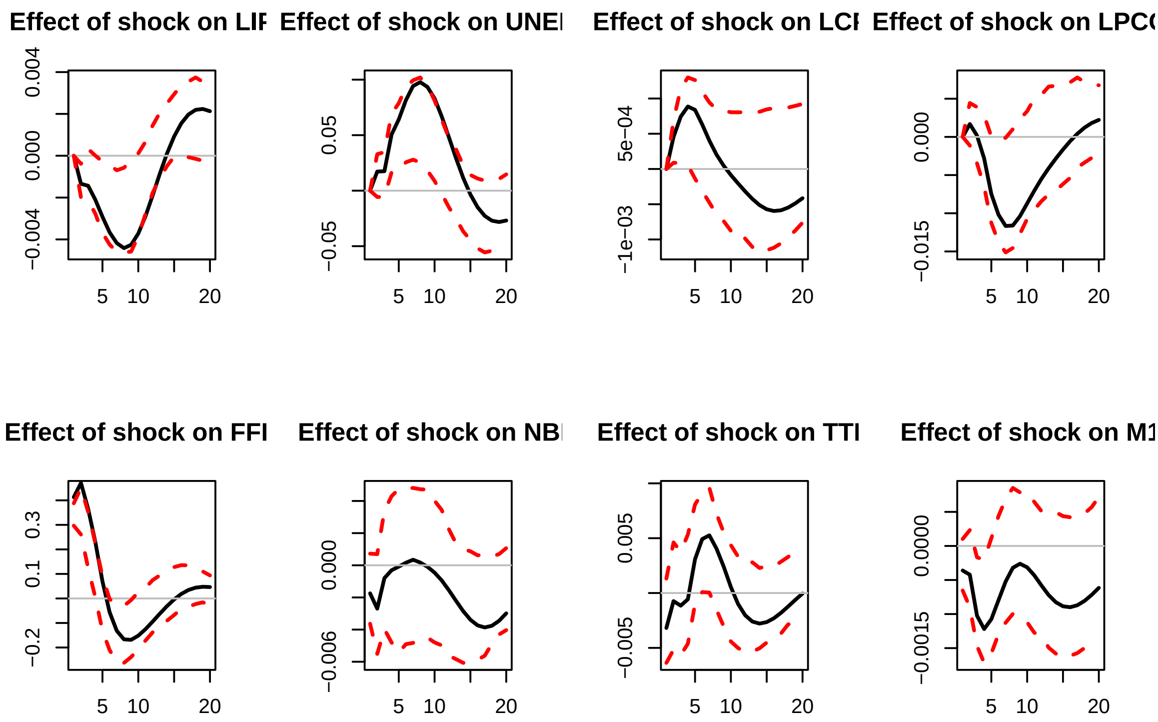

Figure 3.5: Simulated IRF associated with a monetary policy shock.

```
  )
IRFs.ordering.true <- res.svar.ordering$IRFs

# Distribution of coefficients resulting from the small sample VAR
h <- 5
par(mfrow=c(2,ifelse(round(n/2)==n/2,n/2,(n+1)/2)))
for (i in 1:n){
  hist(simulated.IRFs.small.sample[h,i,],xlab="",ylab="",
       main=paste("Effect at h = ",h," on ",
                  considered.variables[i],sep=""),cex.main=.9)
  lines(array(c(IRFs.small.sample[h,i],
                IRFs.small.sample[h,i],0,100),c(2,2)),col="red")
  lines(array(c(median.IRFs.small.sample[h,i],
                median.IRFs.small.sample[h,i],0,100),c(2,2)),col="blue")
  lines(array(c(IRFs.ordering.true[h,i],
                IRFs.ordering.true[h,i],0,100),c(2,2)),col="black")
  text(IRFs.small.sample[h,i],25,label="Estimated coef.",col="red")
  text(IRFs.ordering.true[h,i],30,label="True coef.",col="black")
}
```

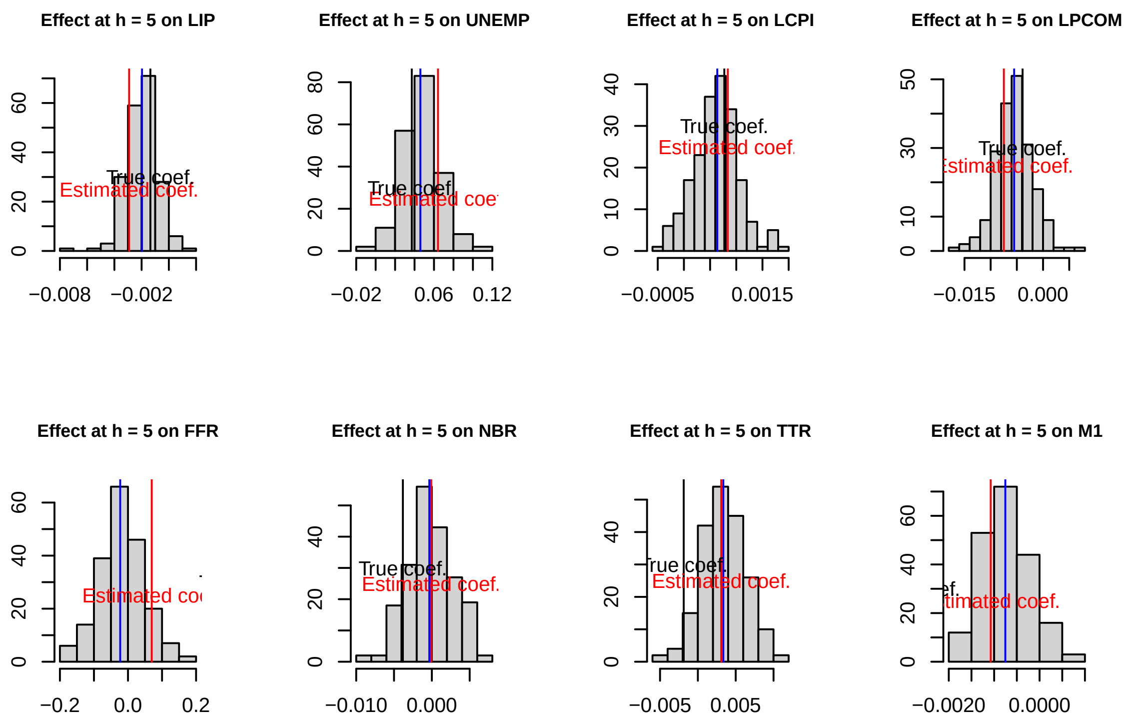

Figure 3.6: IRF associated with a monetary policy shock; sign-restriction approach.

The main idea of the bootstrap-after-bootstrap of Kilian (1998) is to run two consecutive boostraps: the objective of the first is to compute the bias, which can further be used to correct the initial estimates of the Φ_i's. Further, these corrected estimates are used —in the second boostrap— to compute a set of IRFs (as in the standard boostrap).

More formally, the algorithm is as follows:

1. Estimate the SVAR coefficients $\{\widehat{\Phi}_j\}_{j=1,..,p}$ and $\widehat{\Omega}$
2. **First bootstrap.** For each iteration k:

a. Construct a sample

$$y_t^{(k)} = \widehat{\Phi}_1 y_{t-1}^{(k)} + \dots + \widehat{\Phi}_p y_{t-p}^{(k)} + \hat{\varepsilon}_t^{(k)},$$

with $\hat{\varepsilon}_t^{(k)} = \hat{\varepsilon}_{s_t^{(k)}}$, where $\{s_1^{(k)}, .., s_T^{(k)}\}$ is a random set from $\{1, .., T\}^T$.

b. Re-estimate the VAR and compute the coefficients $\{\widehat{\Phi}_j\}_{j=1,..,p}^{(k)}$.

3. Perform N replications and compute the median coefficients $\{\widehat{\Phi}_j\}_{j=1,..,p}^{*}$.
4. Approximate the bias terms by $\widehat{\Theta}_j = \widehat{\Phi}_j^* - \widehat{\Phi}_j$.
5. Construct the bias-corrected terms $\widetilde{\Phi}_j = \widehat{\Phi}_j - \widehat{\Theta}_j$.
6. **Second bootstrap.** For each iteration k:

a. Construct a sample now from

$$y_t^{(k)} = \widetilde{\Phi}_1 y_{t-1}^{(k)} + \dots + \widetilde{\Phi}_p y_{t-p}^{(k)} + \hat{\varepsilon}_t^{(k)}.$$

b. Re-estimate the VAR and compute the coefficients $\{\widehat{\Phi}_j^*\}_{j=1,..,p}^{(k)}$.

c. Construct the bias-corrected estimates $\widetilde{\Phi}_j^{*(k)} = \widehat{\Phi}_j^{*(k)} - \widehat{\Theta}_j$.

d. Compute the associated IRFs $\{\widetilde{\Psi}_j^{*(k)}\}_{j\geq 1}$.

7. Perform N replications and compute the median and the confidence interval of the set of IRFs.

It should be noted that correcting for the bias can generate non-stationary results ($\tilde{\Phi}$ with eigenvalue with modulus > 1). Solution (Kilian (1998)):

In step 5, check if the largest eigenvalue of $\tilde{\Phi}$ is of modulus <1. If not, shrink the bias: for all js, set $\widehat{\Theta}_j^{(i+1)} = \delta_{i+1}\widehat{\Theta}_j^{(i)}$, with $\delta_{i+1} = \delta_i - 0.01$, starting with $\delta_1 = 1$ and $\widehat{\Theta}_j^{(1)} = \widehat{\Theta}_j$, and compute $\widetilde{\Phi}_j^{(i+1)} = \widehat{\Phi}_j - \widehat{\Theta}_j^{(i+1)}$ until the largest eigenvalue of $\tilde{\Phi}^{(i+1)}$ has modulus <1.

The following code implements the bootstrap-after-bootrap method.

```r
library(IdSS);library(vars);library(Matrix)
data("USmonthly")
First.date <- "1965-01-01"
Last.date <- "1995-06-01"
indic.first <- which(USmonthly$DATES==First.date)
indic.last  <- which(USmonthly$DATES==Last.date)
USmonthly   <- USmonthly[indic.first:indic.last,]
considered.variables<-c("LIP","UNEMP","LCPI","LPCOM","FFR","NBR","TTR","M1")
y <- as.matrix(USmonthly[considered.variables])
# ====================================
# CEE with bootstrap-after-bootstrap
# ====================================
res.svar.ordering <-
  svar.ordering.2(y,p=3,
                  posit.of.shock = 5,
                  nb.periods.IRF = 20,
                  inference = 4,# 0 -> no inference, 1 -> parametric bootstr.,
                  # 2 <- non-parametric bootstrap, 3 <- monte carlo,
                  # 4 <- bootstrap-after-bootstrap
                  nb.draws = 200,
                  confidence.interval = 0.90, # expressed in pp.
                  indic.plot = 1 # Plots are displayed if = 1.
  )
```

```r
IRFs.ordering <- res.svar.ordering$IRFs
median.IRFs.ordering <- res.svar.ordering$all.CI.median
simulated.IRFs.ordering <- res.svar.ordering$simulated.IRFs
```

As an alternative, function `VAR.Boot` of package `VAR.etp` (Kim (2022)) can be used to operate the bias-correction approach of Kilian (1998):

```r
library(VAR.etp)
library(vars) #standard VAR models
data(dat) # part of VAR.etp package
corrected <- VAR.Boot(dat,p=2,nb=200,type="const")
noncorrec <- VAR(dat,p=2)
rbind(corrected$coef[1,],
      (corrected$coef+corrected$Bias)[1,],
      noncorrec$varresult$inv$coefficients)
```

```
##         inv(-1)   inc(-1)   con(-1)    inv(-2)   inc(-2)   con(-2)       const
## [1,] -0.3216256 0.1805332 0.9627248 -0.1372523 0.1388058 0.9419710 -0.01856408
## [2,] -0.3196310 0.1459888 0.9612190 -0.1605511 0.1146050 0.9343938 -0.01672199
## [3,] -0.3196310 0.1459888 0.9612190 -0.1605511 0.1146050 0.9343938 -0.01672199
```

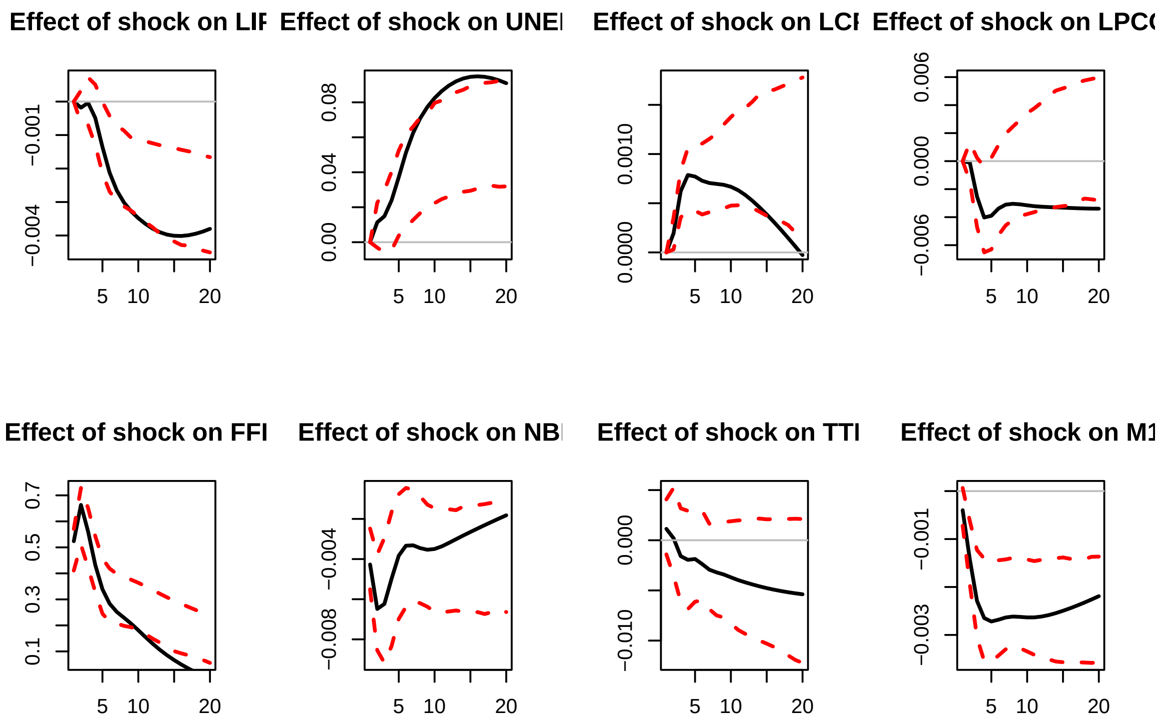

Figure 3.7: IRF associated with a monetary policy shock; sign-restriction approach.

Chapter 4

Sign restrictions

To identifiy the structural shocks, we need to find a matrix B that satisfies $\Omega = BB'$ (with $\Omega = \mathbb{V}ar(\varepsilon_t)$) and other restrictions. Indeed, as explained above, $\Omega = BB'$ is not sufficient to identify B since, if we take any orthogonal matrix Q (see Def. 4.1), then $\mathcal{P} = BQ$ also satisfies $\Omega = \mathcal{P}\mathcal{P}'$.

Definition 4.1 (Orthogonal matrix)**.** An orthogonal matrix Q is a matrix such that $QQ' = I$, i.e., all columns (rows) of Q are are orthogonal and unit vectors:

$$q_i'q_j = 0 \text{ if } i \neq j \text{ and } q_i'q_j = 1 \text{ if } i = j,$$

where q_i is the i^{th} column of Q.

4.1 The approach

The idea behind the sign-restriction approach is to "draw" random matrices $\mathcal{P}$ that satisfy $\Omega = \mathcal{P}\mathcal{P}'$, and then to constitute a set of admissible matrices, keeping in this set only the simulated $\mathcal{P}$ matrices that satisfy some predefined sign-based restriction. An example of restriction is "*after one year, a contractionary monetary-policy shocks has a negative impact on inflation*".

As suggested above, if B is any matrix that satisfies $\Omega = BB'$ (for instance, B can be based on the Cholesky decomposition of Ω), then we also have $\Omega = \mathcal{P}\mathcal{P}'$ as soon as $\mathcal{P} = BQ$, where Q is an orthogonal matrix. Therefore, to draw $\mathcal{P}$ matrices, it suffices to draw in the set of orthogonal matrices.

To fix ideas, consider dimension 2. In that case, the orthogonal matrices are rotation matrices, and the set of orthogonal matrices can be parameterized by the angle x, with:

$$Q_x = \begin{pmatrix} \cos(x) & \cos\left(x+\frac{\pi}{2}\right) \\ \sin(x) & \sin\left(x+\frac{\pi}{2}\right) \end{pmatrix} = \begin{pmatrix} \cos(x) & -\sin(x) \\ \sin(x) & \cos(x) \end{pmatrix}.$$

(This is an angle-x counter-clockwise rotation.) Hence, in that case, by drawing x randomly from $[0, 2\pi]$, we draw randomly from the set of 2×2 rotation matrices. For high-dimensional

VAR, we lose this simple geometrical representation, though. It is not always possible to parametrize a rotation matrix (high-dimensional VARs).

How to proceed, then? Arias et al. (2018) provide a procedure. Their approach is based on the so-called QR decomposition: any square matrix X may be decomposed as $X = QR$ where Q is an orthogonal matrix and R is an upper diagonal matrix. With this in mind, they propose a two-step approach:

i. Draw a random matrix X by drawing each element from independent standard normal distribution.
ii. Let $X = QR$ be the QR decomposition of X with the diagonal of R normalized to be positive. The random matrix Q is orthogonal and is a draw from the uniform distribution over the set of orthogonal matrices.

Equipped with this procedure, the sign-restriction is based on the following algorithm:

1. Draw a random orthogonal matrix Q (using step i. and ii. described above).
2. Compute $B = PQ$ where P is the Cholesky decomposition of the reduced form residuals Ω_ε.
3. Compute the impulse response associated with B $y_{t,t+k} = \Phi^k B$ or the cumulated response $\bar{y}_{t,t+k} = \sum_{j=0}^{k} \Phi^j B$.
4. Are the sign restrictions satisfied?

a. **Yes**. Store the impulse response in the set of admissible response.
b. **No**. Discard the impulse response.

5. Perform N replications and report the median impulse response (and its "confidence" intervals).

Note: to take into account the uncertainty in B and Φ, you can draw B and Φ in Steps 2 and 3 using an inference method (see Section 3).

The sign-restriction approach method has the advantage of being relatively agnostic. Moreover, it is fairly flexible, as one can impose sign restrictions on any variable, at any horizon.

4.2 An example

A prominent example is Uhlig (2005). Using US monthly data from 1965.I to 2003.XII, he employs sign restrictions to estimate the effect of monetary policy shocks.

According to conventional wisdom, monetary contractions should:[1]

[1]Standard identification schemes often fail to achieve these 4 points. Two puzzles regularly arise: *liquidity puzzle*: when identifying monetary policy shocks as surprise increases in the stock of money, interest rates tend to go up, not down; *price puzzle*: after a contractionary monetary policy shock, even with interest rates going up and money supply going down, inflation goes up rather than down.

- Raise the federal funds rate,
- Lower prices,
- Decrease non-borrowed reserves,
- Reduce real output.

The restrictions considered by Uhlig (2005) are as follows: an expansionary monetary policy shock leads to:

- Increases in prices
- Increase in nonborrowed reserves
- Decreases in the federal funds rate

What about output? Since is the response of interest, we leave it un-restricted.

```
library(IdSS);library(vars);library(Matrix)
data("USmonthly")
First.date <- "1965-01-01"
Last.date <- "1995-06-01"
indic.first <- which(USmonthly$DATES==First.date)
indic.last  <- which(USmonthly$DATES==Last.date)
USmonthly   <- USmonthly[indic.first:indic.last,]
considered.variables<-c("LIP","UNEMP","LCPI","LPCOM","FFR","NBR","TTR","M1")
n <- length(considered.variables)
y <- as.matrix(USmonthly[considered.variables])
sign.restrictions <- list()
horizon <- list()
#Define sign restrictions and horizon for restrictions
for(i in 1:n){
  sign.restrictions[[i]] <- matrix(0,n,n)
  horizon[[i]] <- 1
}
# Sign restrictions on shock 1 (monetary shock)
sign.restrictions[[1]][1,3] <- 1  # positive impact on price level
sign.restrictions[[1]][2,5] <- -1 # negative impact on interest rate
sign.restrictions[[1]][3,6] <- 1 # positive impact on non-borrowed reserves
horizon[[1]] <- 1:5 # from horizon 1 to 5
res.svar.signs <-
  svar.signs(y,p=3,
             nb.shocks = 1, #number of identified shocks
             nb.periods.IRF = 20,
             bootstrap.replications = 1, # = 1 if no bootstrap, = N if bootstrap
             confidence.interval = 0.90, # expressed in pp.
             indic.plot = 1, # Plots are displayed if = 1.
             nb.draws = 10000, # number of draws
```

```
                sign.restrictions,
                horizon,
                recursive =1 #  =0 <- draw Q directly, =1 <- draw q recursively
  )
```

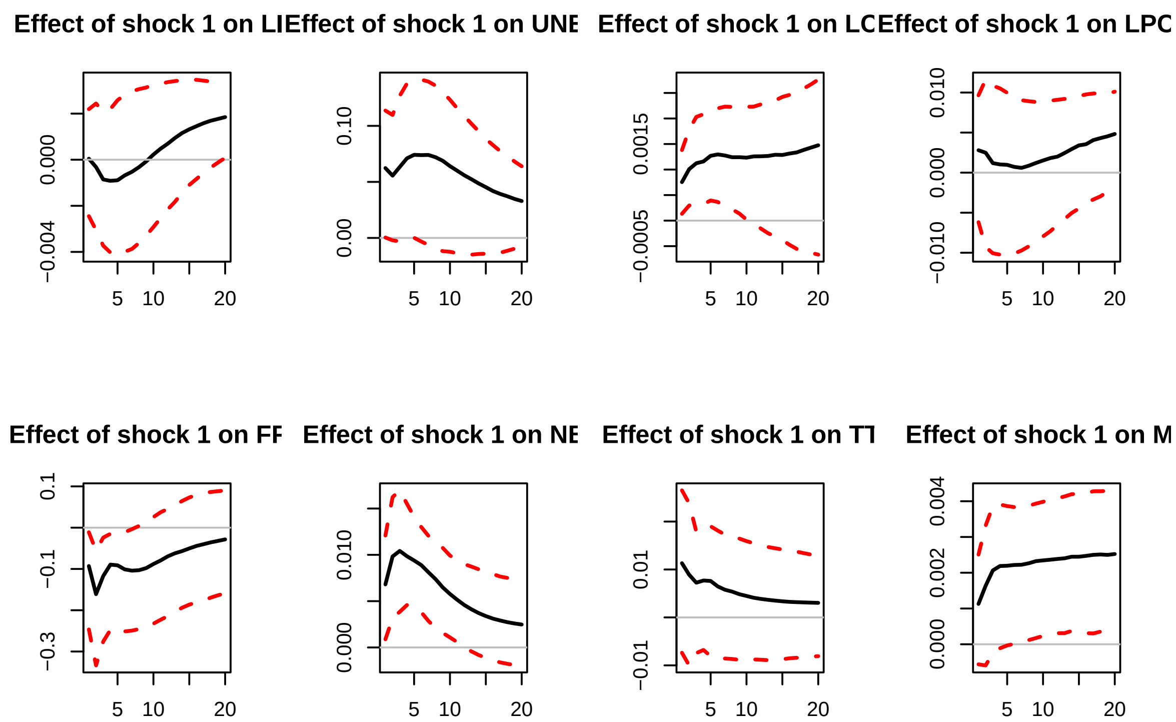

Figure 4.1: IRF associated with a monetary policy shock; sign-restriction approach.

```
# Output
IRFs.signs <- res.svar.signs$IRFs.signs # all the simulated IRFs
nb.rotations <- res.svar.signs$xx # total number of rotations
all.CI.median <- res.svar.signs$all.CI.median # median IRFs for the selected shocks
all.CI.lower.bounds <- res.svar.signs$all.CI.lower.bounds # lower-bound IRFs for the
all.CI.upper.bounds <- res.svar.signs$all.CI.upper.bounds # upper-bound IRFs for the
```

It has to be stressed that the sign restriction approach does not lead to a unique IRF, but to a set of admissible IRFs. Accordingly, we say that this approach is set-identified, not point-identified.

4.3 The penalty-function approach (PFA)

An alternative approach is the so-called **penalty-function approach** (PFA, Uhlig (2005), present in Danne (2015)'s package). This approach relies on a *penalty function*:

$$\begin{array}{rcll} f(x) & = & x & \text{if } x \leq 0 \\ & & 100.x & \text{if } x > 0 \end{array}$$

which penalizes positive responses and rewards negative responses.

Let $\psi_k^j(q)$ be the impulse response of variable j. The $\psi_k^j(q)$'s are the elements of $\psi_k(q) = \Psi_k q$.

Let σ_j be the standard deviation of variable j. Let $\iota_{j,k} = 1$ if we restrict the response of variable j at the k^th horizon to be negative, $\iota_{j,k} = -1$ if we restrict it to be positive, and $\iota_{j,k} = 0$ if there is no restriction. The total penalty is given by

$$\mathbf{P}(q) = \sum_{j=1}^{m}\sum_{k=0}^{K} f\left(\iota_{j,k}\frac{\psi_k^j(q)}{\sigma_j}\right).$$

We are looking for a solution to

$$\begin{array}{rl} & \min_q \mathbf{P}(q) \\ \text{s.t.} & q'q = 1. \end{array}$$

The problem is solved numerically.

4.4 Narrative sign restrictions

A related approach, introduced by Antolín-Díaz and Rubio-Ramírez (2018), consists in imposing that, on some specific dates (based on narrative evidence), the signs of some shocks are positive (or negative).[2] For instance, Antolín-Díaz and Rubio-Ramírez (2018) argue that one should rule out structural parameters that disagree with the view that "a negative oil supply shock occurred at the outbreak of the Gulf War in August 1990."

Suppose we want to impose the restriction that, at dates $\{t_1, \dots, t_J\}$, the signs of the j^{th} shock are all positive. Then, the narrative sign restrictions are simply imposed by:

$$\hat{\eta}_{j,t}(B) = e_j'\hat{\eta}_t(B) > 0,$$

where $\hat{\eta}_t(B)$ is the vector of structural shock associated with a given matrix B (and where e_j is the j^{th} column of the $n \times n$ identity matrix).

[2]See also Section 6.2.

Chapter 5

Combining sign and zero restrictions

Sometimes we need to combine different types of restrictions. For instance:

- One shock satisfies both zero and sign restrictions.
- Some shocks can be identified with zero restrictions (SR or LR), others with sign restrictions.
- Some shocks satisfy the same zero restrictions (e.g. no LR effect on output) but can be distinguished from each other through sign restrictions.

In such instances, we must make independent draws from the set of all structural parameters satisfying the zero restrictions. How to do that? Arias et al. (2018) propose to impose the zero restrictions on B, and then check signs. Remember, $\mathcal{P} = BQ$ is a candidate impact IRF. For each structural shock j, define the m-column matrices Z_j (zero restrictions) and S_j (sign restrictions). Each row of Z_j (resp. S_j) defines a zero (resp. sign) restriction. Z_j has $m - j$ rows at most (i.e., $m - j$ zero restriction at most).

Example 5.1. In a 4-variable VAR, we want to impose that the first structural shock has no effect on variable 1, affects positively variable 2 and negatively variable 3 on impact:

$$Z_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix},$$

$$S_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

For both zero and sign restrictions to be satisfied, we must have

$$Z_j b_j = 0 \quad \text{and} \quad S_j b_j > 0,$$

where b_j is the j^{th} column of B, i.e. the impact effect of the j^{th} structural shock.

The algorithm is as follows:

1. For $1 \leq j \leq m$, draw $u_j \in \mathbb{R}^{m+1-j-z_j}$ from a standard normal distribution (z_j is the number of zero restrictions imposed on the j^{th} shock) and set $w_j = u_j/||u_j||$.

2. Define $Q = (q_1 \quad \dots \quad q_m)$ recursively by $q_j = K_j w_j$ for any matrix K_j whose columns form an orthogonal basis for the null space of the matrix
$$M_j = \begin{pmatrix} q_1 & \dots & q_{j-1} & (Z_j P)' \end{pmatrix}'.$$
(Vector q_j will then be orthogonal to $\begin{pmatrix} q_1 & \dots & q_{j-1} \end{pmatrix}$ and satisfy the zero restriction.)
3. Set $B = PQ$.
4. Check sign restrictions ($S_j b_j > 0$ for all j?).

Perform N replications and report the median impulse response (and its confidence intervals).

Function `svar.signs` can run this algorithm. It is called as follows:[1]

```
library(IdSS);library(vars);library(Matrix)
data("USmonthly")
First.date <- "1965-01-01"
Last.date <- "1995-06-01"
indic.first <- which(USmonthly$DATES==First.date)
indic.last  <- which(USmonthly$DATES==Last.date)
USmonthly   <- USmonthly[indic.first:indic.last,]
considered.variables<-c("LIP","UNEMP","LCPI","LPCOM","FFR","NBR","TTR","M1")
n <- length(considered.variables)
y <- as.matrix(USmonthly[considered.variables])
sign.restrictions <- list()
SR.restrictions <- list()
horizon <- list()

#Define sign restrictions and horizon for restrictions
for(i in 1:n){
  sign.restrictions[[i]] <- matrix(0,n,n)
  horizon[[i]] <- 1
}
# 2 shocks on the demand for reserves
sign.restrictions[[1]][1,6] <- 1
sign.restrictions[[2]][1,7] <- 1
# 3 shocks that drive an endogenous response of the interest rate
sign.restrictions[[3]][1,1] <- 1
sign.restrictions[[3]][2,5] <- 1
sign.restrictions[[4]][1,2] <- -1
sign.restrictions[[4]][2,5] <- 1
sign.restrictions[[5]][1,3] <- 1
sign.restrictions[[5]][2,5] <- 1
# monetary policy shock
sign.restrictions[[6]][1,5] <- -1
```

[1] Note that outputs are not reported here; just copy-paste in Rstudio to get the 6 sets of IRFs.

```
sign.restrictions[[6]][2,3] <- 1
sign.restrictions[[6]][3,6] <- 1
# horizon for sign restrictions on monetary policy shock
horizon[[6]] <- 1:5

#Define zero restrictions
# 2 shocks on the demand for reserves
SR.restrictions[[1]] <- array(0,c(1,n))
SR.restrictions[[1]][1,5] <- 1 # no impact on the interest rate
SR.restrictions[[2]] <- array(0,c(1,n))
SR.restrictions[[2]][1,5] <- 1 # no impact on the interest rate
for(i in 3:n){
  SR.restrictions[[i]] <- array(0,c(0,n))
}

res.svar.signs.zeros <- svar.signs(y,p=3,
                                  nb.shocks = 6, #number of identified shocks
                                  nb.periods.IRF = 20,
                                  bootstrap.replications = 100, # = 1 if no bootstrap, =
                                  confidence.interval = 0.90, # expressed in pp.
                                  indic.plot = 1, # Plots are displayed if = 1.
                                  nb.draws = 10000, # number of draws
                                  sign.restrictions,
                                  horizon,
                                  recursive =0,
                                  SR.restrictions
)
```

```
# Output
IRFs.signs <- res.svar.signs.zeros$IRFs.signs # all the simulated IRFs
nb.rotations <- res.svar.signs.zeros$xx # total number of rotations
all.CI.median <- res.svar.signs.zeros$all.CI.median # median IRFs for the selected shock
all.CI.lower.bounds <- res.svar.signs.zeros$all.CI.lower.bounds # lower-bound IRFs for t
all.CI.upper.bounds <- res.svar.signs.zeros$all.CI.upper.bounds # upper-bound IRFs for t
```

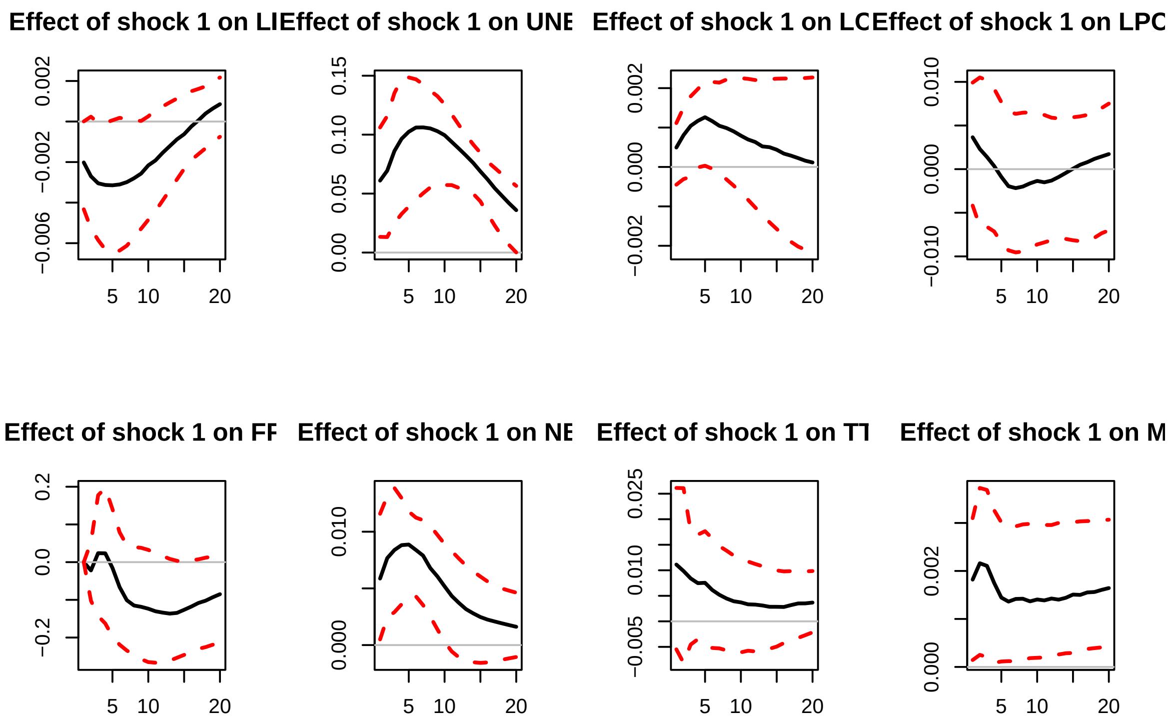

Figure 5.1: IRFs; sign-restriction approach.

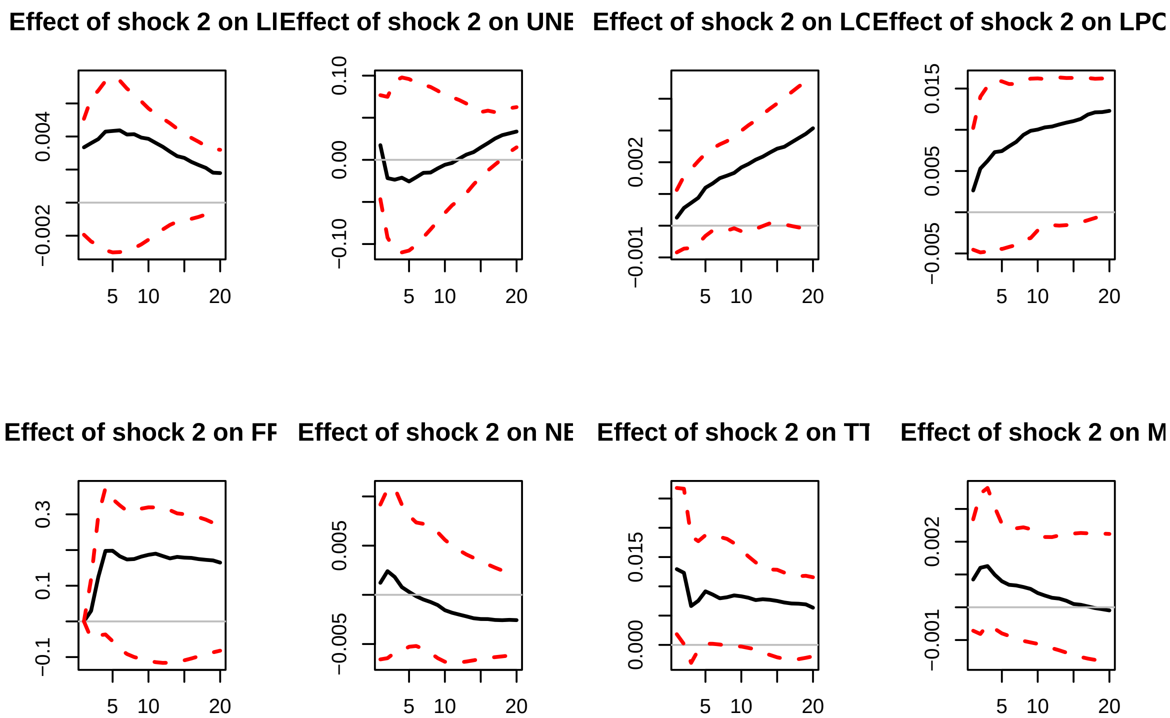

Figure 5.2: IRFs; sign-restriction approach.

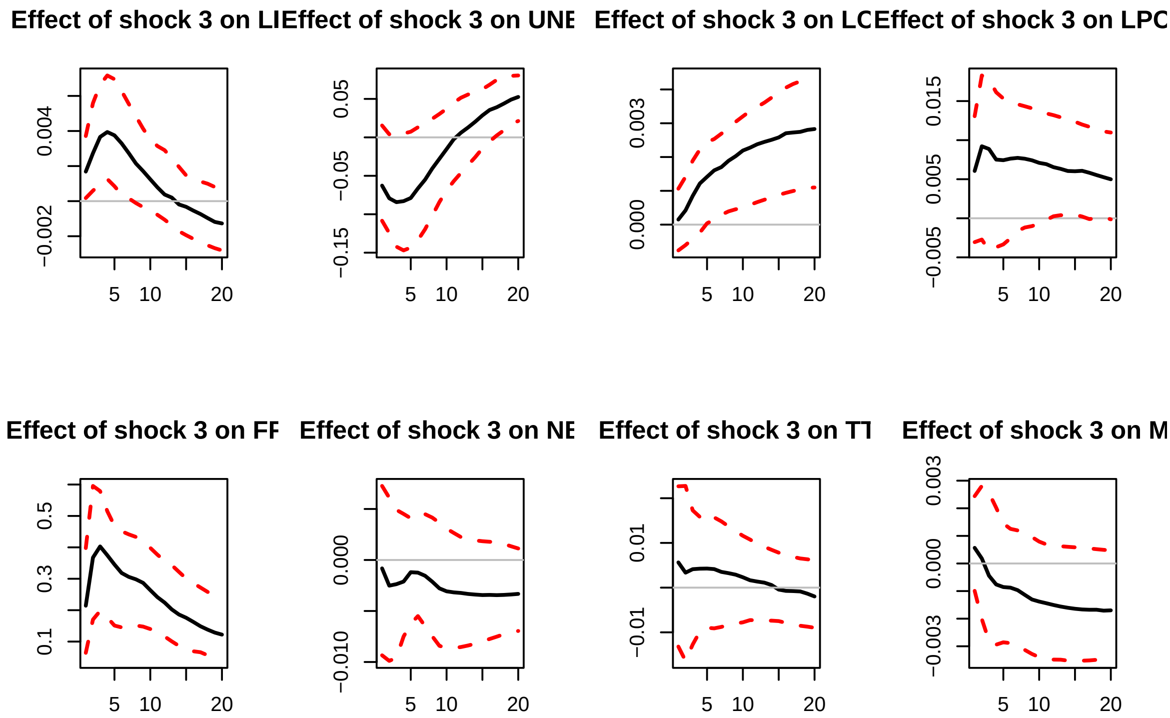

Figure 5.3: IRFs; sign-restriction approach.

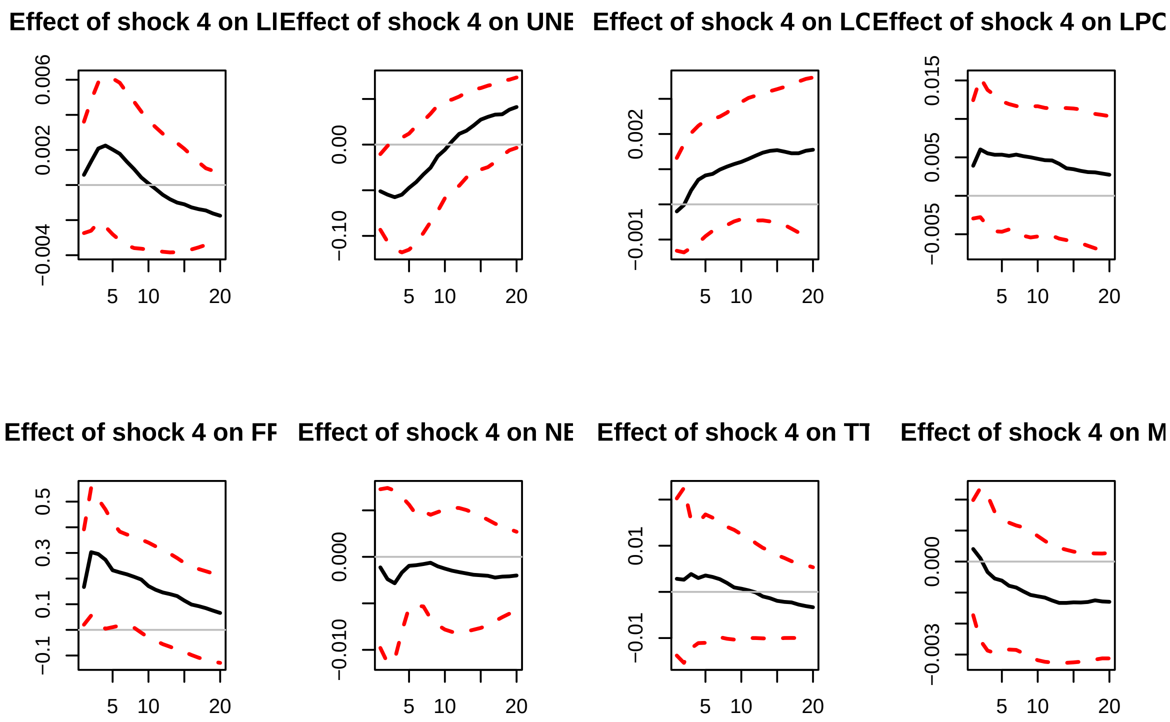

Figure 5.4: IRFs; sign-restriction approach.

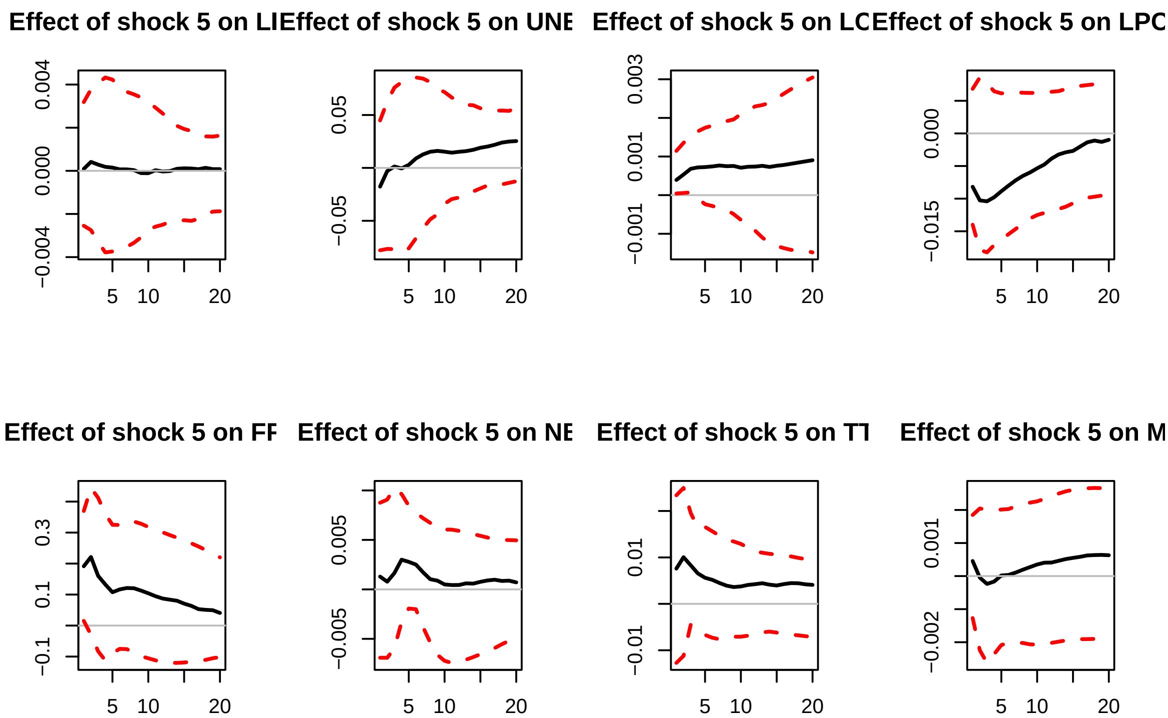

Figure 5.5: IRFs; sign-restriction approach.

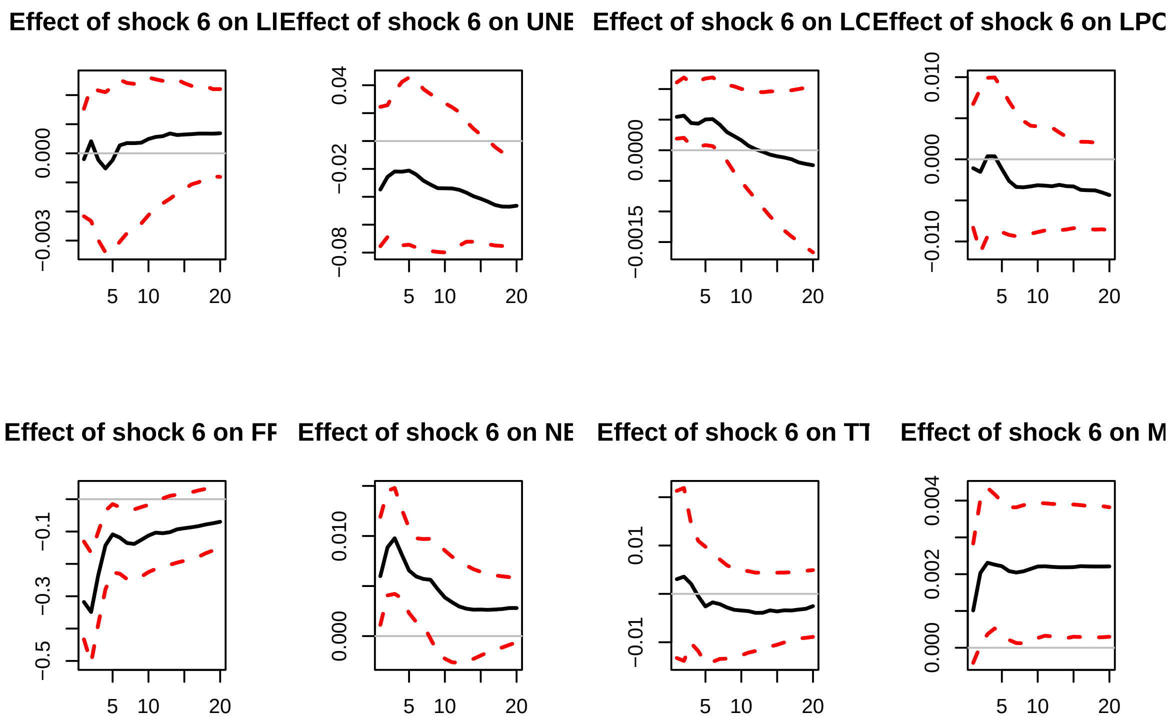

Figure 5.6: IRFs; sign-restriction approach.

Chapter 6

Forecast error variance maximization

6.1 The main (unconditional) approach

The approach presented in this section exploits the derivations of Uhlig (2004).

Consider a process $\{y_t\}$ that admits the infinite MA representation of Eq. (1.4). Let Q be an orthogonal matrix, an alternative decomposition of y_t is:

$$y_t = \sum_{h=0}^{+\infty} \Psi_h \underbrace{\eta_{t-h}}_{Q\tilde{\eta}_{t-h}} = \sum_{h=0}^{+\infty} \underbrace{\Psi_h Q}_{\tilde{\Psi}_h} \tilde{\eta}_{t-h} = \sum_{h=0}^{+\infty} \tilde{\Psi}_h \tilde{\eta}_{t-h}, \tag{6.1}$$

where $\tilde{\eta}_{t-h} = Q'\eta_{t-h}$ are the white-noise shocks associated with the new MA representation, Q being an orthgonal matrix. (They also satisfy $\mathbb{V}ar(\tilde{\eta}_t) = Id$.)

The h-step ahead prediction error of y_{t+h}, given all the data up to, and including, $t-1$ is given by

$$e_{t+h}(h) = y_{t+h} - \mathbb{E}_{t-1}(y_{t+h}) = \sum_{j=0}^{h} \tilde{\Psi}_h \tilde{\eta}_{t+h-j}.$$

The variance-covariance matrix of $e_{t+h}(h)$ is

$$\Omega^{(h)} = \sum_{j=0}^{h} \tilde{\Psi}_j \tilde{\Psi}_j' = \sum_{j=0}^{h} \Psi_j \Psi_j'.$$

We can decompose $\Omega^{(h)}$ into the contribution of each shock l (l^{th} component of $\tilde{\eta}_t$):

$$\Omega^{(h)} = \sum_{l=1}^{n} \Omega_l^{(h)}(Q)$$

with

$$\Omega_l^{(h)}(Q) = \sum_{j=0}^{h} (\Psi_j q_l)(\Psi_j q_l)',$$

where q_l is the l^{th} column of Q.

This decomposition can be used with the objective of finding the **impulse vector** b that is s.t. that it explains as much as possible of the sum of the h-step ahead prediction error variance of some variable i, say, for prediction horizons $h \in [\underline{h}, \overline{h}]$.

Formally, the task is to explain as much as possible of the variance

$$\sigma^2(\underline{h}, \overline{h}, q_1) = \sum_{h=\underline{h}}^{\overline{h}} \sum_{j=0}^{h} \left[(\Psi_j q_1)(\Psi_j q_1)'\right]_{i,i}$$

with a single impulse vector q_1.

Denote by E_{ii} the matrix that is filled with zeros, except for its (i, i) entry, set to 1. We have:

$$\begin{aligned}\sigma^2(\underline{h}, \overline{h}, q_1) &= \sum_{h=\underline{h}}^{\overline{h}} \sum_{j=0}^{h} \left[(\Psi_j q_1)(\Psi_j q_1)'\right]_{i,i} = \sum_{h=\underline{h}}^{\overline{h}} \sum_{j=0}^{h} Tr\left[E_{ii}(\Psi_j q_1)(\Psi_j q_1)'\right] \\ &= \sum_{h=\underline{h}}^{\overline{h}} \sum_{j=0}^{h} Tr\left[q_1' \Psi_j' E_{ii} \Psi_j q_1\right] \\ &= q_1' S q_1,\end{aligned}$$

where

$$\begin{aligned}S &= \textstyle\sum_{h=\underline{h}}^{\overline{h}} \sum_{j=0}^{h} \Psi_j' E_{ii} \Psi_j \\ &= \textstyle\sum_{j=0}^{\overline{h}} (\overline{h} + 1 - max(\underline{h}, j)) \Psi_j' E_{ii} \Psi_j \\ &= \textstyle\sum_{j=0}^{\overline{h}} (\overline{h} + 1 - max(\underline{h}, j)) \Psi_{j,i}' \Psi_{j,i}\end{aligned}$$

where $\Psi_{j,i}$ denotes row i of Ψ_j, i.e., the response of variable i at horizon j (when $Q = Id$).

The maximization problem subject to the side constraint $q_1' q_1 = 1$ can be written as a Lagrangian:

$$L = q_1' S q_1 - \lambda(q_1' q_1 - 1),$$

with the first-order condition $Sq_1 = \lambda q_1$ (the side constraint is $q_1' q_1 = 1$). From this equation, we see that the solution q_1 is an eigenvector of S, the one associated with eigenvalue λ. We also see that $\sigma^2(\underline{h}, \overline{h}, q_1) = \lambda$. Thus, to maximize this variance, we need to find the eigenvector of S that is associated with the maximal eigenvalue λ. That defines the first principal component (see Section 10.1). That is, if S admits the following spectral decomposition:

$$S = \mathcal{P} D \mathcal{P}',$$

where D is diagonal matrix whose entries are the (ordered) eigenvalues: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$, then $\sigma^2(\underline{h}, \overline{h}, q_1)$ is maximized for $q_1 = p_1$, where p_1 is the first column of $\mathcal{P}$.

The following code identifies a “main GDP shock’ ’ using Uhlig’s method.

```
library(IdSS)
library(readxl)
library(vars)
library(Matrix)
# Declare data:
TFP   <- levpan$tfp_lev
GDP   <- levpan$lngdpcap
E12   <- levpan$e12m
CONS  <- levpan$lnconcap
HOURS <- levpan$lnhrscap
y <- cbind(TFP,GDP,E12,CONS,HOURS)
names.of.variables <- c("TFP","GDP","E12","Consumption","Hours")
colnames(y)  <- names.of.variables
names.of.shocks <- c("Main GDP shock")
#define horizons for FEVM
H1 <- 1
H2 <- 20
variable <- 2 # variables for which we want to maximize FEVD
norm <- 15 # horizon at which the impact of the shock
# is normalized to be positive
res.svar.fevmax <- svar.fevmax(y,p=2,
                               nb.shocks=1,
                               names.of.shocks,
                               H1,
                               H2,
                               variable,
                               norm,
                               nb.periods.IRF = 20,
                               bootstrap.replications = 100, # This is used in
                               #the parametric bootstrap only
                               confidence.interval = 0.90, # expressed in pp.
                               indic.plot = 1 # Plots are displayed if = 1.
)
# Compute variance decomposition:
Variance.decomp <- variance.decomp(res.svar.fevmax$simulated.IRFs)
vardecomp <- Variance.decomp$vardecomp
mean(vardecomp[2,2,20,,1]) # mean contribution (across all simulated IRFs)
```

```
## [1] 0.809319
```

```
# of 1st shock to variance of second variable, horizon 20.
mean(vardecomp[1,2,10,,1]) # mean contribution of 1st shock
```

```
## [1] 0.8297299
```

```
# to covariance between 1st and 2nd variable, horizon 10.
```

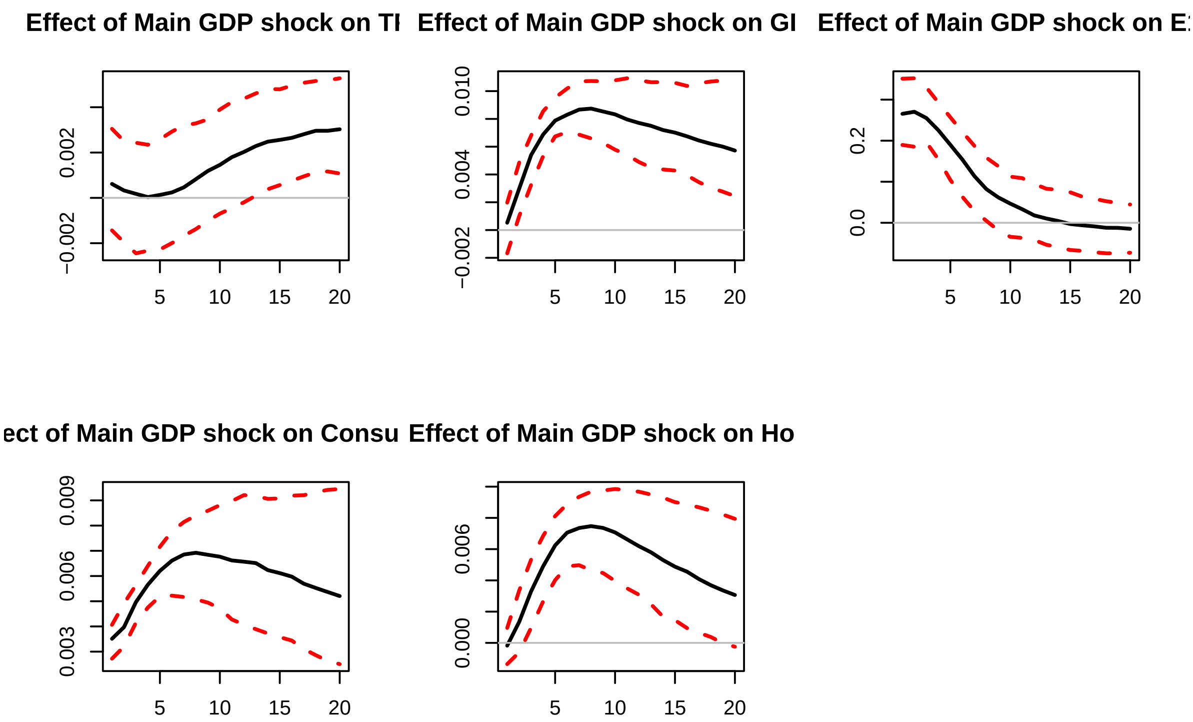

Figure 6.1: Main GDP shock.

The "main GDP shock' ' explains 81% of the variance of GDP at horizon 20.

Barsky and Sims (2011) exploit this approach to identify a "TFP news shock", that they define as the shock (a) that is orthogonal to the innovation in current utilization-adjusted TFP and (b) that best explains variation in future TFP. Levchenko and Pandalai-Nayar (2018) add a "sentiment shock" that they define as the shock (a) that is orthogonal to both the innovation in current utilization-adjusted TFP and to the TFP news shock, and (b) that best explains variation in consumer sentiment. The following code replicates Levchenko and Pandalai-Nayar (2018). They use a mix of zero and FEVD to identify TFP surprises, TFP news, and sentiment shocks. We adopt a different approach by using FEVD to capture zero restrictions.

```
names.of.shocks <- c("TFP surprise","TFP news","Sentiment")
#define horizons for FEVM
H1 <- array(0,c(1,3))
H2 <- array(0,c(1,3))
H1[1,1] <- 1
H1[1,2] <- 1
```

```
H1[1,3] <- 1
H2[1,1] <- 1
H2[1,2] <- 40
H2[1,3] <- 2
variable <- c(1,1,3) # variables for which we want to maximize FEVD
norm <- c(1,20,2) # horizon at which the impact of the
# shock is normalized to be positive
res.svar.fevmax <-
  svar.fevmax(y,p=2,
              nb.shocks=3,
              names.of.shocks,
              H1,
              H2,
              variable,
              norm,
              nb.periods.IRF = 40,
              bootstrap.replications = 100, # This is used in the
              #parametric bootstrap only
              confidence.interval = 0.90, # expressed in pp.
              indic.plot = 1 # Plots are displayed if = 1.
  )
```

```
Variance.decomp <- variance.decomp(res.svar.fevmax$simulated.IRFs)
vardecomp <- Variance.decomp$vardecomp
mean(vardecomp[2,2,40,,3])# mean contribution (across all simulated IRFs)
```

```
## [1] 0.1226119
```

```
# of 3rd shock to variance of second variable, horizon 40.
```

Sentiment shocks explain 12% of the variance of GDP, against 77% for TFP shocks (including 65% for TFP news shocks).

6.2 Restrictions based on narrative historical decomposition

A related approach, introduced by Antolín-Díaz and Rubio-Ramírez (2018), consists in imposing that, on some specific dates (based on narrative information), a particular shock was the most important contributor to the unexpected movement of some variable during a particular period.[1] This can be formalized in two different ways (respectively called Type A and Type B by Antolín-Díaz and Rubio-Ramírez (2018)):

[1] See also Section 4.4.

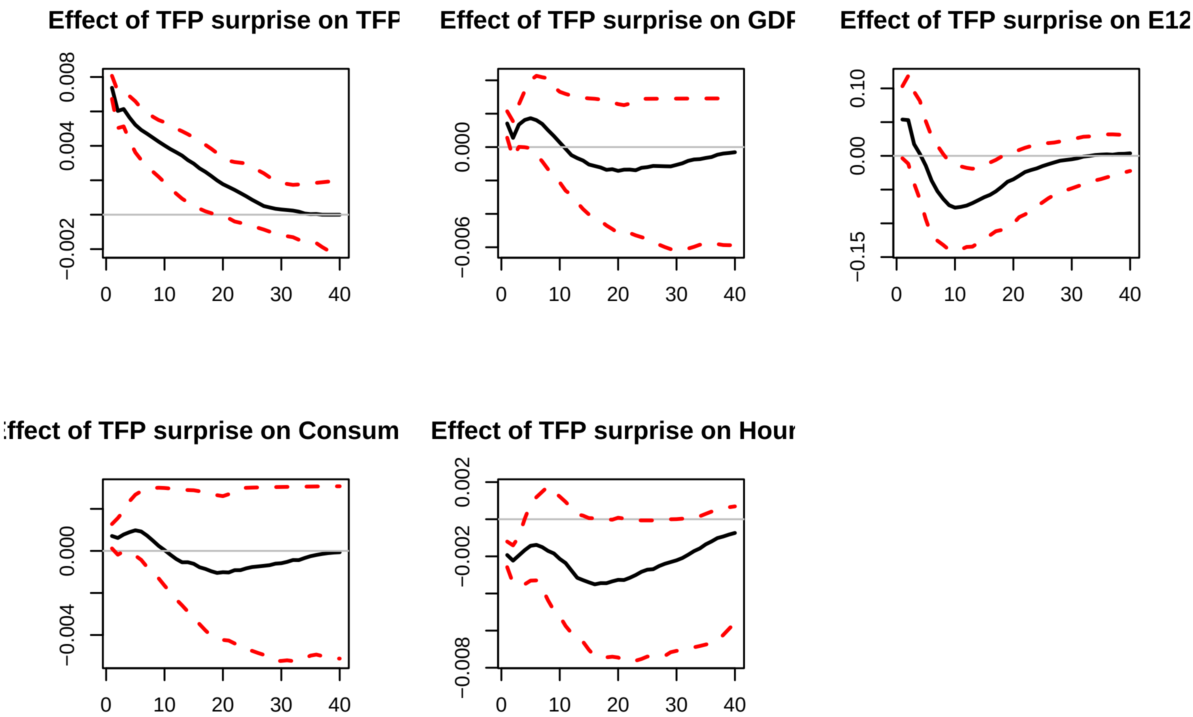

Figure 6.2: Replication of Levchenko and Pandalai-Nayar (2020). FEVD and zero restrictions.

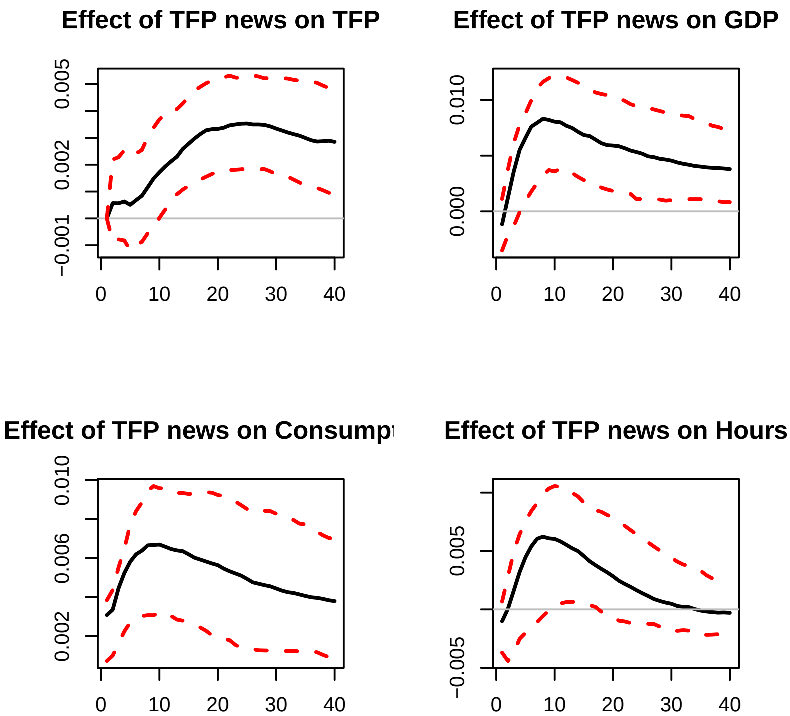

Figure 6.3: Replication of Levchenko and Pandalai-Nayar (2020). FEVD and zero restrictions.

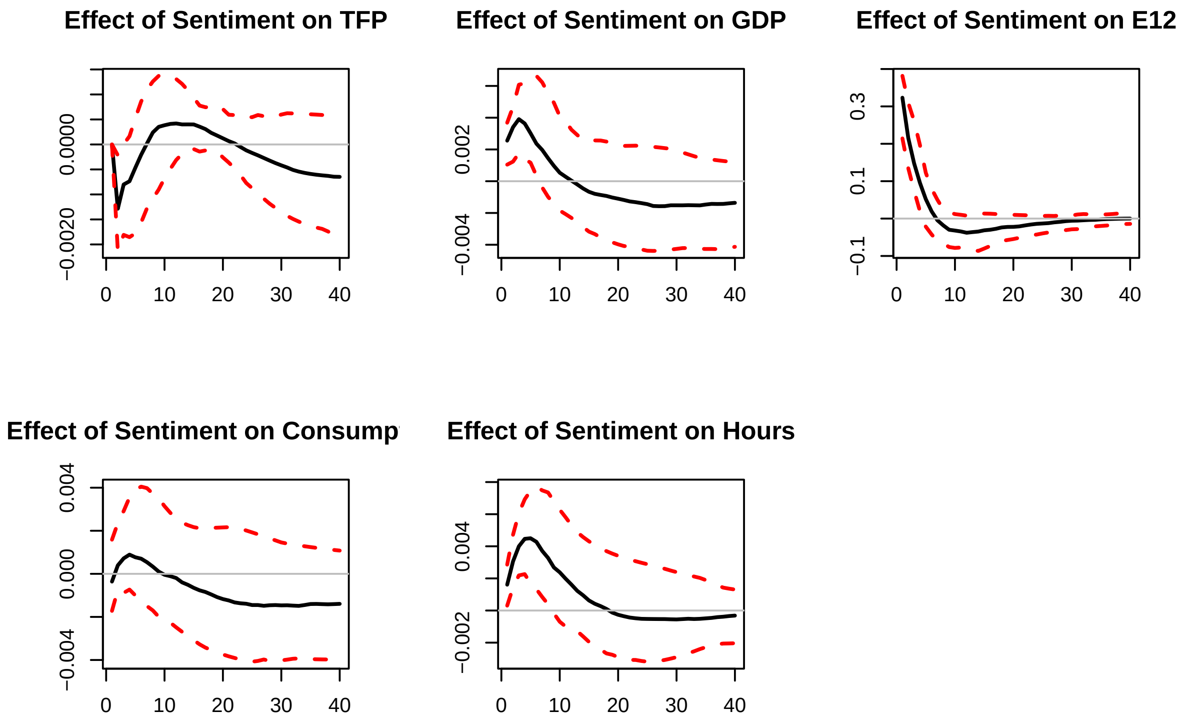

Figure 6.4: Replication of Levchenko and Pandalai-Nayar (2020). FEVD and zero restrictions.

- Type A: A given shock is the most important (least important) driver of the unexpected change in a variable during some periods. For these periods, the absolute value of its contribution to the unexpected change in a variable is larger (smaller) than the absolute value of the contribution of any other structural shock.
- Type B: A given shock is the overwhelming (negligible) driver of the unexpected change in a given variable during the period. For these periods, the absolute value of its contribution to the unexpected change in a variable is larger (smaller) than the sum of the absolute value of the contributions of all other structural shocks.

Chapter 7

Identification based on non-normality of the shocks

7.1 Intuition

In this section, we show that the non-identification of the structural shocks (η_t) is specific to the Gaussian case. We propose consistent estimation approaches for SVAR in the context of non-Gaussian shocks.

We have seen in what precedes that we cannot identify B based on first and second moments only. Since a Gaussian distribution is perfectly determined by the first two moments, it comes that one cannot achieve identification when the structural shocks are Gaussian. That is, even if we observe an infinite number of i.i.d. $B\eta_t$, we cannot recover B is the η_t's are Gaussian.

Indeed, if $\eta_t \sim \mathcal{N}(0, Id)$, then the distribution of $\varepsilon_t \equiv B\eta_t$ is $\mathcal{N}(0, BB')$. Hence $\Omega = BB'$ is observed (in the population), but for any orthogonal matrix Q (i.e. $QQ' = Id$), we also have $BQ\eta_t \sim \mathcal{N}(0, \Omega)$.

To illustrate, consider the following bivariate Gaussian situations, with $\Theta_1 = 0$):

$$\left[\begin{array}{c}\eta_{1,t}\\ \eta_{2,t}\end{array}\right] \sim \mathcal{N}(0, Id), \text{ with } B = \left[\begin{array}{cc}1 & 2\\ -1 & 1\end{array}\right] \text{ and } Q = \left[\begin{array}{cc}\cos(\pi/3) & -\sin(\pi/3)\\ \sin(\pi/3) & \cos(\pi/3)\end{array}\right] \text{ (rotation).}$$

Figure 7.1 shows that the distributions of $B\eta_t$ and of $BQ\eta_t$ are identical. However, the impulse response functions associated with one of the other impulse matrix (B or BQ) are different. This is illustrated by Figure 7.2, that shows the IRFs associated with two identical models (defined by Eq. (1.7)), the only difference being the impulse matrix (B or BQ).

Hence, in the Gaussian case, external restrictions (economic hypotheses) are needed to identify B (see previous sections). But such restrictions may not be necessary if the structural shocks are not Gaussian. That is, the identification problem is very specific to normally-distributed η_t's (Rigobon (2003), Normandin and Phaneuf (2004), Lanne and Lütkepohl (2008)).

To better see why this can be the case, consider again a bivariate vector of independent structural shocks ($\eta_{1,t}$ and $\eta_{2,t}$) but, now, assume that one of them is not Gaussian any

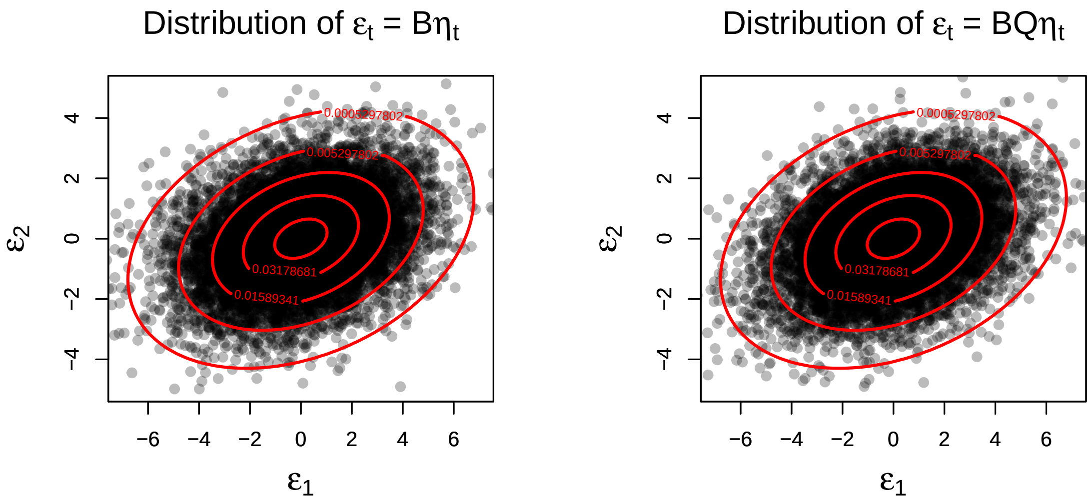

Figure 7.1: This figure compares the distributions of two Gaussian bivariate vectors, $B\eta_t$ and $BQ\eta_t$, where $\eta_t \sim \mathcal{N}(0, Id)$ (therefore $\eta_{1,t}$ and $\eta_{2,t}$ are independent), and Q is an orthogonal matrix.

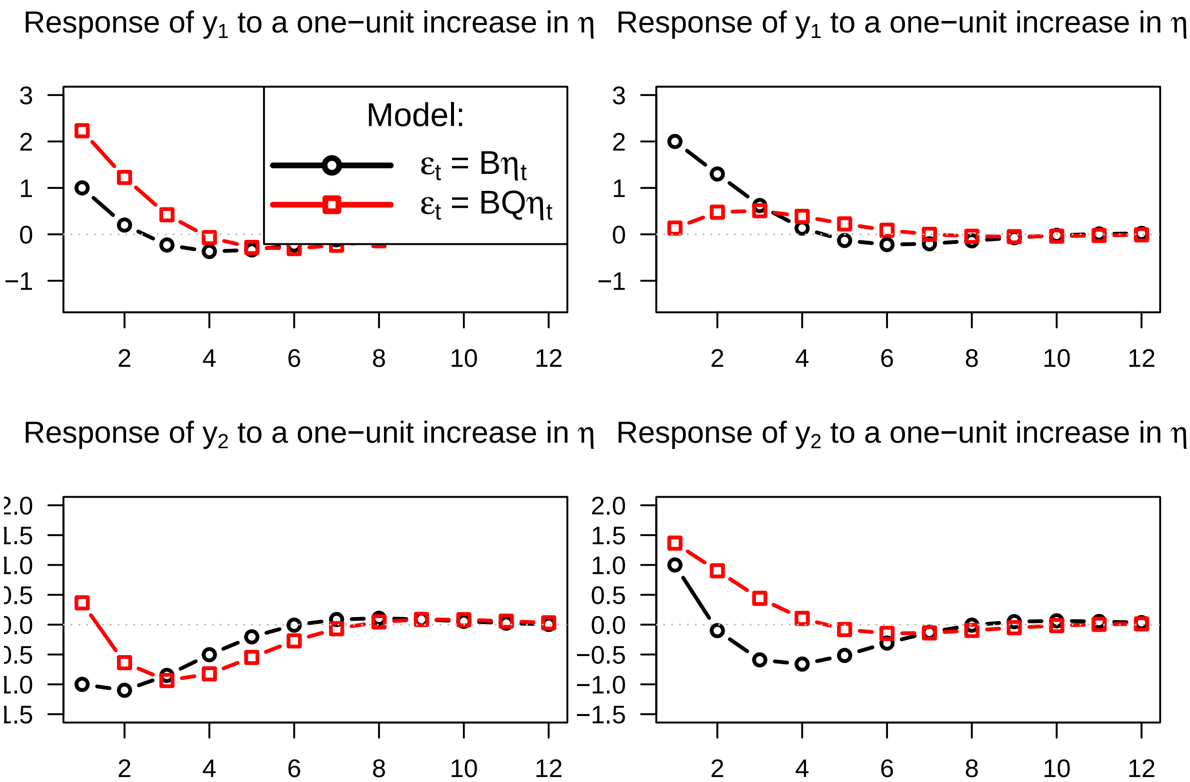

Figure 7.2: This figure shows that the impulse response functions associated with an impulse matrix equal to B (black line) or BQ (red line) are different (even if $BB' = BQ(BQ)'$).

more. Specifically, assume that $\eta_{2,t}$ is drawn from a Student distribution with 5 degrees of freedom: $\eta_{1,t} \sim \mathcal{N}(0,1)$, $\eta_{2,t} \sim t(5)$, $B = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$ and $Q = \begin{bmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{bmatrix}$.

Figure 7.3 shows that, in this case, $B\eta_t$ and $BQ\eta_t$ do not have the same distribution any more (in spite of the fact that, in both cases, we have $\mathbb{V}ar(\varepsilon_t) = BB'$). This opens the door to the identification of the impulse matrix (BQ) in the non-Gaussian case.

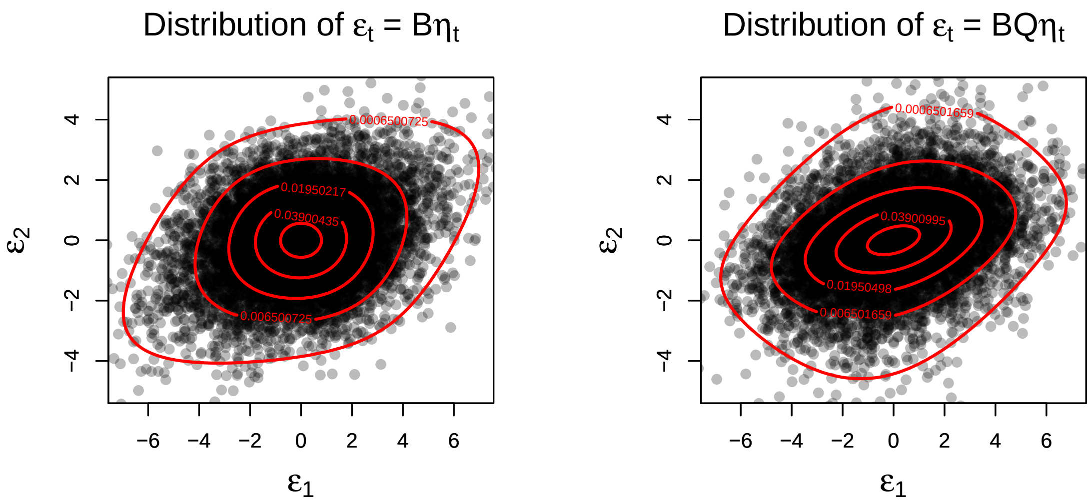

Figure 7.3: This figure compares the distributions of two Gaussian bivariate vectors, $B\eta_t$ and $BQ\eta_t$, where $\eta_t 1,t \sim \mathcal{N}(0,1)$, $\eta_t 2,t \sim t(5)$, and Q is an orthogonal matrix.

7.2 Independent Component Analysis (ICA)

The exercise that consists in identifying non-Gaussian independent shocks out of linear combinations of these shocks is a well-known problem of the signal-processing literature, called **independent component analysis (ICA)**. Let us denote by C the matrix that is such that $C = \Omega^{-1/2}B$, where $\Omega^{1/2}$ results from the Cholesky decomposition of $\Omega = BB'$ (implying, $\Omega^{1/2}\Omega^{1/2'} = \Omega$). It is easy to check that C is an orthogonal matrix (and we have $B = \Omega^{1/2}C$).

The classical ICA problem is as follows: Find C such that $\varepsilon_t = C\eta_t$ (or $\eta_t = C'\varepsilon_t$) given that:[1]

i. We observe the ε_t's,
ii. The components of η_t are independent,

[1]The ε_t's that we consider here are *standardized* VAR residuals, obtained by pre-multiplying the actual VAR residuals by $\Omega^{-1/2}$.

iii. $CC' = Id$ (i.e., C is orthogonal).

Figure 7.4 represents again some bivariate distributions. The black (red) lines correspond to the distributions of η_t ($C\eta_t$). It is important to note that the two components of vector $C\eta_t$ are not independent (contrary to those of η_t).

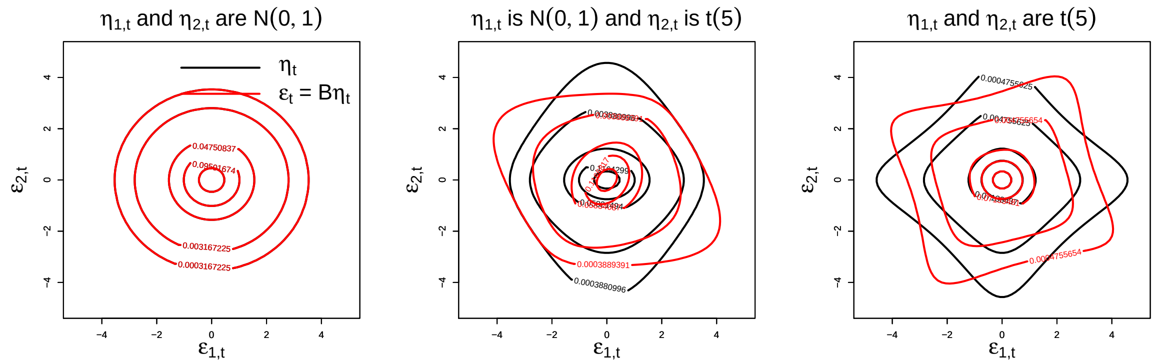

Figure 7.4: The three plots represent the bivariate distributions of η_t (black) and of $C\eta_t$ (red), where the two components of η_t are independent, of unit variance, and C is orthogonal. Hence, for each of the three plots, $\mathbb{V}ar(C\eta_t) = Id$.

In all cases, we have $\mathbb{V}ar(\varepsilon_t) = \mathbb{V}ar(\eta_t) = Id$. But the two components of ε_t are not independent. For instance, in the last two cases, we have $\mathbb{E}(\varepsilon_{2,t}|\varepsilon_{1,t} > 4) < 0$ (whereas $\mathbb{E}(\eta_{2,t}|\eta_{1,t} > 4) = 0$). The objective of ICA is to rotate ε_t to retrieve independent components (η_t).

Hypothesis 7.1. Process η_t satisfies:

i. The η_t's are i.i.d. (across time) with $\mathbb{E}(\eta_t) = 0$ and $\mathbb{V}ar(\eta_t) = Id$.
ii. The components $\eta_{1,t}, \dots, \eta_{n,t}$ are mutually independent. iii We have

$$\varepsilon_t = C_0\eta_t,$$

with $\mathbb{V}ar(\varepsilon_t) = Id$ (i.e., C_0 is orthogonal).

Theorem 7.1 (Eriksson, Koivunen (2004))**.** *If Hypothesis 7.1 is satisfied and if at most one of the components of η is Gaussian, then matrix C_0 is identifiable up to the post multiplication by DP, where P is a permutation matrix and D is a diagonal matrix whose diagonal entries are 1 or -1.}*

7.3 Pseudo-Maximum Likelihood (PML) approach

Hence, non-normal structural shocks are identifiable. But how to estimate them based on observations of the ε_t's? Gouriéroux et al. (2017) have proposed a **Pseudo-Maximum**

Likelihood (PML) approach. This approach consists in maximizing a so-called **pseudo log-likelihood function**, based on a set of p.d.f. $g_i(\eta_i), i = 1, \dots, n$ (that may be different from the true p.d.f. of the $\eta_{i,t}$'s):

$$\log \mathcal{L}_T(C) = \sum_{t=1}^{T} \sum_{i=1}^{n} \log g_i(c_i' \varepsilon_t), \tag{7.1}$$

where c_i is the i^{th} column of matrix C (or c_i' is the i^{th} row of C^{-1} since $C^{-1} = C'$).

The log-likelihood function (7.1) is computed as if the errors $\eta_{i,t}$ had the pdf $g_i(\eta_i)$. The PML estimator of matrix C maximizes the pseudo log-likelihood function:

$$\widehat{C_T} = \arg\max_{C} \sum_{t=1}^{T} \sum_{i=1}^{n} \log g_i(c_i' \varepsilon_t), \tag{7.2}$$

$$s.t.\ C'C = Id.$$

The restrictions $C'C = Id$ can be eliminated by parameterizing C in such a way that, whatever the considered parameters, C is orthogonal.[2] Gouriéroux et al. (2017) propose to use, for that, the Cayley's representation: any orthogonal matrix with no eigenvalue equal to -1 can be written as

$$C(A) = (Id + A)(Id - A)^{-1}, \tag{7.3}$$

where A is a skew symmetric (or antisymmetric) matrix, such that $A' = -A$. There is a one-to-one relationship with A, since:

$$A = (C(A) + Id)^{-1}(C(A) - Id). \tag{7.4}$$

Hence, the PML estimator of matrix C is obtained as $\widehat{C_T} = C(\hat{A}_T)$, where:

$$\hat{A}_T = \arg \max_{a_{i,j}, i>j} \sum_{t=1}^{T} \sum_{i=1}^{n} \log g_i[c_i(A)' \varepsilon_t]. \tag{7.5}$$

Under assumptions on the g_i functions (excluding the Gaussian distributions), Gouriéroux et al. (2017) derive the asymptotic properties of the PML estimator. Specifically, the PML estimator $\widehat{C_T}$ of C_0 is consistent (in $\mathcal{P}_0$, the set of matrices obtained by permutation and sign change of the columns of C_0) and asymptotically normal, with speed of convergence $1/\sqrt{T}$.

The asymptotic variance-covariance matrix of $vec\sqrt{T}(\widehat{C_T} - C_0)$ is $A^{-1} \left[\begin{array}{cc} \Gamma & 0 \\ 0 & 0 \end{array} \right] (A')^{-1}$, where matrices A and Γ are detailed in Gouriéroux et al. (2017).

Note that the potential misspecification of pseudo-distributions g_i has no effect on the consistency of these specific PML estimators.

[2] Jarocinski (2021) develops a ML approach that does not necessitates to parameterize the space of orthogonal matrices as he does not proceed under the assumption that $C'C$ is orthogonal.

Table 7.1 reports usual p.d.f. and their derivatives. (The latter are needed to compute the asymptotic variance-covariance matrix of $vec\sqrt{T}(\widehat{C_T} - C_0)$.)

Table 7.1: This table reports usual p.d.f. and their derivatives.

	$\log g(x)$	$\dfrac{d \log g(x)}{dx}$	$\dfrac{d^2 \log g(x)}{dx^2}$
Gaussian	$cst - x^2/2$	$-x$	-1
Student $t(\nu > 4)$	$-\dfrac{1-\nu}{2}\log\left(1+\dfrac{x^2}{\nu-2}\right)$	$-\dfrac{x(1+\nu)}{\nu-2+x^2}$	$-(1+\nu)\dfrac{\nu-2-x^2}{\nu-2+x^2}$
Hyperbolic secant	$cst - \log\left(\cosh\left\{\dfrac{\pi}{2}x\right\}\right)$	$-\dfrac{\pi}{2}anh\left(\dfrac{\pi}{2}x\right)$	$-\left(\dfrac{\pi}{2}\dfrac{1}{\cosh\left(\dfrac{\pi}{2}x\right)}\right)^2$
Subgaussian	$cst + \pi x^2 + \log\left(\cosh\left\{\dfrac{\pi}{2}x\right\}\right)$	$2\pi x + \dfrac{\pi}{2}\tanh\left(x\dfrac{\pi}{2}\right)$	$2\pi + \left(\dfrac{\pi}{2}\dfrac{1}{\cosh\left(\dfrac{\pi}{2}x\right)}\right)^2$

Example 7.1 (Non-Gaussian monetary-policy shocks)**.** We apply the PML-ICA approach on U.S. data coerving the period 1959:IV to 2015:I at the quarterly frequency ($T = 224$). We consider three dependent variables: inflation (π_t), economic activity (z_t, the output gap) and the nominal short-term interest rate (r_t). Changes in the log of oil prices are added as an exogenous variable (x_t).

```
library(IdSS)
First.date <- "1959-04-01"
Last.date  <- "2015-01-01"
data <- US3var
data <- data[(data$Date>=First.date)&(data$Date<=Last.date),]
Y <- as.matrix(data[c("infl","y.gdp.gap","r")])
names.var <- c("inflation","real activity","short-term rate")
T <- dim(Y)[1]
n <- dim(Y)[2]
```

Let us denote by W_t the set of information made of the past values of $y_t = [\pi_t, z_t, r_t]$, that is $\{y_{t-1}, y_{t-2}, \dots\}$, and of exogenous variables $\{x_t, x_{t-1}, \dots\}$. The reduced-form VAR model reads:

$$y_t = \underbrace{\mu + \sum_{i=1}^{p} \Phi_i y_{t-i} + \Theta x_t}_{a(W_t;\theta)} + u_t$$

where the u_t's are assumed to be serially independent, with zero mean and variance-covariance matrix Ω.

Matrices μ, Φ_i, Θ and Ω are consistently estimated by OLS. Jarque-Bera tests support the hypothesis of non-normality for all residuals.

```
nb.lags <- 6 # number of lags used in the VAR model
X <- NULL
for(i in 1:nb.lags){
  lagged.Y <- rbind(matrix(NaN,i,n),Y[1:(T-i),])
  X <- cbind(X,lagged.Y)}
X <- cbind(X,data$commo) # add exogenous variables
Phi <- matrix(0,n,n*nb.lags);mu <- rep(0,n)
effect.commo <- rep(0,n)
U <- NULL # Eta is the matrix of OLS residuals
for(i in 1:n){
  eq <- lm(Y[,i] ~ X)
  Phi[i,] <- eq$coef[2:(dim(Phi)[2]+1)]
  mu[i] <- eq$coef[1]
  U <- cbind(U,eq$residuals)
  effect.commo[i] <- eq$coef[length(eq$coef)]
}
Omega <- var(U) # Covariance matrix of the OLS residuals.
Omeg12 <- t(chol(Omega)) # Cholesky matrix associated with Omega (lower triang.)
Eps <- U %*% t(solve(Omeg12)) # Recover associated structural shocks
```

We want to estimate the orthogonal matrix C such that $u_t = \Omega^{1/2} C \eta_t$, where

- $\Omega^{1/2}$ results from the Cholesky decomposition of Ω and
- the components of η_t are independent, zero-mean with unit variance.

The PML approach is applied on standardized VAR residuals given by:

$$\hat{\varepsilon}_t = \hat{\Omega}_T^{-1/2} \underbrace{[y_t - a(W_t; \hat{\theta}_T)]}_{\text{VAR residuals}}.$$

By construction of $\hat{\Omega}_T^{-1/2}$, it comes that the covariance matrix of these residuals is Id.

The pseudo density functions are distinct and asymmetric mixtures of Gaussian distributions.

```
distri <- list(
  type=c("mixt.gaussian","mixt.gaussian","mixt.gaussian"),
  df=c(NaN,NaN,NaN),
  p=c(0.5,.5,.5),mu=c(.1,.1,.1),sigma=c(.5,.7,1.3))
AA.0 <- c(0,0,0)
res.optim <- optim(AA.0,func.2.minimize,
                   Y = Eps, distri = distri,
```

```
                    gr = d.func.2.minimize,
                    method="Nelder-Mead",
                    control=list(trace=FALSE,maxit=1000))
AA.0 <- res.optim$par
res.optim <- optim(AA.0,func.2.minimize,d.func.2.minimize,
                    Y = Eps, distri = distri,
                    method="BFGS",
                    control=list(trace=FALSE))
AA.est <- res.optim$par
n <- ncol(Y)
M <- make.M(n)
A.est <- matrix(M %*% AA.est,n,n)
C.PML <- (diag(n) + A.est) %*% solve(diag(n) - A.est)
eta.PML <- Eps %*% C.PML # eta.PML are the ICA-estimated structural shocks

# Compute asymptotic covariance matrix of C.PML:
V <- make.Asympt.Cov.delta(eta.PML,distri,C.PML)
param <- c(C.PML)
st.dev <- sqrt(diag(V))
t.stat <- c(C.PML)/sqrt(diag(V))
cbind(param,st.dev,t.stat) # print results of PML estimation
```

```
##            param      st.dev      t.stat
##  [1,]  0.94417705 0.040848382  23.1141845
##  [2,] -0.32711569 0.118802653  -2.7534376
##  [3,]  0.03905164 0.074172945   0.5264944
##  [4,]  0.32070293 0.119270893   2.6888616
##  [5,]  0.93977707 0.041629110  22.5749976
##  [6,]  0.11818924 0.060821400   1.9432179
##  [7,] -0.07536139 0.071980455  -1.0469702
##  [8,] -0.09906759 0.062185577  -1.5930959
##  [9,]  0.99222290 0.007785691 127.4418551
```

(Note: it is always useful to combine two optimization algorithms, such as `Nelder-Mead` and `BFGS`.)

We would obtain close results by neglecting commodity prices. In that case, one can simply use the function `estim.SVAR.ICA` of the `IdSS` package. Let us compare the C matrix obtained in the two cases (with or without commodity prices):

```
ICA.res.no.commo <- estim.SVAR.ICA(Y,distri = distri,p=6)
round(cbind(ICA.res.no.commo$C.PML,NaN,C.PML),3)
```

```
##          [,1]   [,2]    [,3]  [,4]    [,5]   [,6]    [,7]
```

```
## [1,]  0.956 0.287 -0.059  NaN  0.944 0.321 -0.075
## [2,] -0.292 0.950 -0.108  NaN -0.327 0.940 -0.099
## [3,]  0.025 0.121  0.992  NaN  0.039 0.118  0.992
```

Once C has been estimated, it remains to label the resulting structural shocks (components of η_t). Postulated shocks are monetary-policy, supply, and demand shocks. This labelling can be based on the following considerations:

- Contractionary **monetary-policy shocks** have a negative impact on real activity and on inflation.
- **Supply shock** have influences of opposite signs on economic activity and on inflation.
- **Demand shock** have influences of same signs on economic activity and on inflation.

Let us compute the IRFs associated with the three structural shocks. (For the sake of comparison, the first line of plots shows the IRFs to a monetary-policy shock obtained from a Cholesky-based approach where the short-term rate is ordered last.)

```
IRF.Chol <- array(NaN,c(n,41,n))
IRF.ICA  <- array(NaN,c(n,41,n))
PHI <- list();for(i in 1:nb.lags){PHI[[i]]<-array(Phi,c(3,3,nb.lags))[,,i]}
for(jjjj in 1:n){
  u.shock <- rep(0,n)
  u.shock[jjjj] <- 1
  IRF.Chol[,,jjjj] <-
    t(simul.VAR(c=rep(0,3),Phi=PHI,B=Omeg12,nb.sim=41,
                y0.star=rep(0,3*nb.lags),indic.IRF = 1,u.shock = u.shock))
  IRF.ICA[,,jjjj]  <-
    t(simul.VAR(c=rep(0,3),Phi=PHI,B=Omeg12%*%C.PML,nb.sim=41,
                y0.star=rep(0,3*nb.lags),indic.IRF = 1,u.shock = u.shock))
}
```

According to Figure 7.5, Shock 1 is a supply shock, Shock 2 is a demand shock, and Shock 3 is a monetary-policy shock. Note that Shock 3 is close to the one resulting from the Cholesky approach.

7.4 Relation with the Heteroskedasticity Identification

In some cases, where the ε_t's are heteroskedastic, the B matrix can be identified (Rigobon (2003), Lanne et al. (2010)).

Consider the case where we still have $\varepsilon_t = B\eta_t$ but where η_t's variance conditionally depends on a regime $s_t \in \{1, \dots, M\}$. That is:

$$\mathbb{V}ar(\eta_{k,t}|s_t) = \lambda_{s_t,k} \quad \text{for } k \in \{1, \dots, n\}$$

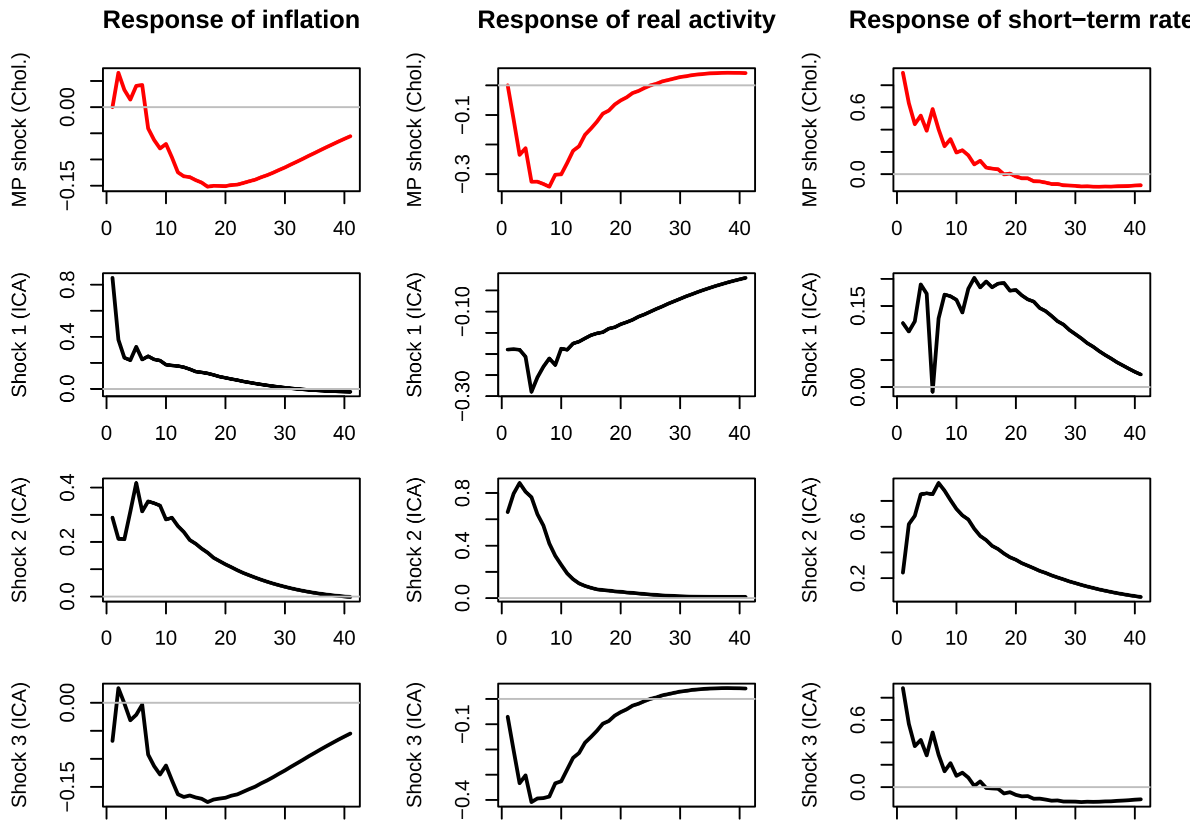

Figure 7.5: The first row of plots shows the responses of the three endogenous variables to the monetary policy shock in the context of a Cholesky-idendtified SVAR (ordering: inflation, output gap, interest rate). The next three rows of plots show the repsonses of the endogenous variables to the three structural shocks identified by ICA. The last one (Shock 3) is close to the Cholesky-identified monetary policy shock.

Denoting by Λ_i the diagonal matrix whose diagonal entries are the $\lambda_{i,k}$'s, it comes that:

$$\mathbb{V}ar(\eta_t|s_t) = \Lambda_{s_t}, \quad \text{and} \quad \mathbb{V}ar(\varepsilon_t|s_t) = B\Lambda_{s_t}B'.$$

Without loss of generality, it can be assumed that $\Lambda_1 = Id$.

In this context, B is identified, apart from sign reversal of its columns if for all $k \neq j \in \{1, \dots, n\}$, there is a regime i s.t. $\lambda_{i,k} \neq \lambda_{i,j}$. (Prop.1 in Lanne et al. (2010)).

Bivariate regime case ($M = 2$): B identified if the $\lambda_{2,k}$'s are all different. That is, identification is ensured if "there is sufficient heterogeneity in the volatility changes" (Lütkepohl and Netšunajev (2017)).

If the regimes s_t are exogenous and serially independent, then this situation is consistent with the "non-Gaussian" situation described above.

Chapter 8

Local projection methods

8.1 Overview of the approach

Consider the infinite MA representation of y_t (Eq. (1.4)):

$$y_t = \mu + \sum_{h=0}^{\infty} \Psi_h \eta_{t-h}.$$

As seen in Section 1.2, the entries (i, j) of the sequence of the Ψ_h matrices define the IRF of $\eta_{j,t}$ on $y_{i,t}$.

Assume that you observe $\eta_{j,t}$, then a consistent estimate of $\Psi_{i,j,h}$ is simply obtained by the OLS regression of $y_{i,t+h}$ on $\eta_{j,t}$:[1]

$$y_{i,t+h} = \mu_i + \Psi_{i,j,h}\eta_{j,t} + u_{i,j,t+h}. \tag{8.1}$$

Running that kind of regression (using instruments for $\eta_{j,t}$) is the core idea of the **local projection (LP) approach** proposed by Jordà (2005).

Now, how to proceed in the (usual) case where $\eta_{j,t}$ is not observed? We consider two situations. While the second requires some instruments, the first approach does not. This first approach (Section 8.2) is the original Jordà (2005)'s approach.

8.2 Situation A: Without IV

Assume that the structural shock of interest ($\eta_{1,t}$, say) can be consistently obtained as the residual of a regression of a variable x_t on a set of control variables w_t independent from $\eta_{1,t}$:

$$\eta_{1,t} = x_t - \mathbb{E}(x_t|w_t), \tag{8.2}$$

[1]Because the residuals $u_{i,j,t+h}$ are autocorrelated (for $h > 0$), estimates of the covariance of the OLS estimators of the $\Psi_{i,j,h}$ would then have to be based on robust estimators (e.g. Newey and West (1987)).

where $\mathbb{E}(x_t|w_t)$ is affine in w_t and where w_t is an affine transformation of $\eta_{2:n,t}$ and of past shocks $\eta_{t-1}, \eta_{t-2}, \dots$.

Eq. (8.2) implies that, conditional on w_t, the additional knowledge of x_t is useful only when it comes to forecast something that depends on $\eta_{1,t}$. Hence, given that $u_{i,1,t+h}$ (see Eq. (8.1)) is independent from $\eta_{1,t}$ (it depends on $\eta_{t+h}, \dots, \eta_{t+1}, \eta_{2:n,t}, \eta_{t-1}, \eta_{t-2}, \dots$), it comes that

$$\mathbb{E}(u_{i,1,t+h}|x_t, w_t) = \mathbb{E}(u_{i,1,t+h}|w_t). \tag{8.3}$$

This is the *conditional mean independence* case.

Using (8.2), one can rewrite Eq. (8.1) as follows:

$$\begin{aligned}
y_{i,t+h} &= \mu_i + \Psi_{i,1,h}\eta_{1,t} + u_{i,1,t+h} \\
&= \mu_i + \Psi_{i,1,h}x_t - \Psi_{i,1,h}\mathbb{E}(x_t|w_t) + u_{i,1,t+h},
\end{aligned}$$

Given Eq. (8.3), it comes that, conditional on x_t and w_t, the expectation of the blue term is a function of w_t. Assuming this expectation is linear, standard results in the conditional mean independence case imply that the OLS estimator in the regression of $y_{i,t+h}$ on x_t, controlling for w_t, provides a consistent estimate of $\Psi_{i,1,h}$:

$$y_{i,t+h} = \alpha_i + \Psi_{i,1,h}x_t + \beta' w_t + v_{i,t+h}. \tag{8.4}$$

This is for instance consistent with the case where $[\Delta GDP_t, \pi_t, i_t]'$ follows a VAR(1) and the monetary-policy shock does not contemporaneously affect ΔGDP_t and π_t. The IRFs can then be estimated by LP, taking $x_t = i_t$ and $w_t = [\Delta GDP_t, \pi_t, \Delta GDP_{t-1}, \pi_{t-1}, i_{t-1}]'$.

This approach closely relates to the SVAR Cholesky-based identification approach. Specifically, if $w_t = [y_{1,t}, \dots, y_{k-1,t}, y'_{t-1}, \dots, y'_{t-p}]'$, with $k \le n$, and $x_t = y_{k,t}$, then this approach corresponds, for $h = 0$, to the SVAR(p) Cholesky-based IRF (focusing on the responses to the k^{th} structural shock). However, the two approaches differ for $h > 0$, because the LP methodology does not assumes a VAR dynamics for y_t.[2]

In the following lines of code, we employ the Jordà (2005)'s approach on the same dataset as the one used in Section 2.3. (We were then illustrating Christiano et al. (1996)'s methodology.)

```
library(IdSS);library(vars)
data("USmonthly")
# Select sample period:
First.date <- "1965-01-01";Last.date <- "1995-06-01"
indic.first <- which(USmonthly$DATES==First.date)
indic.last  <- which(USmonthly$DATES==Last.date)
USmonthly   <- USmonthly[indic.first:indic.last,]
```

[2] This is reminiscent of the distinction betweem direct forecasting—based on regressions of y_{t+h} on $\{y_t, y_{t-1}, \dots\}$—and iterated forecasting—based on a recursive model where $y_{t+1} = g(y_t, y_{t-1}, \dots) + \varepsilon_{t+1}$, see Marcellino et al. (2006).

```
considered.variables <- c("LIP","UNEMP","LCPI","LPCOM","FFR","NBR","TTR","M1")
y <- as.matrix(USmonthly[considered.variables])
res.jorda <- make.jorda.irf(y,posit.of.shock = 5,
                            nb.periods.IRF = 12,
                            nb.lags.endog.var.4.control=3,
                            indic.plot = 1, # Plots are displayed if = 1.
                            confidence.interval = 0.90)
```

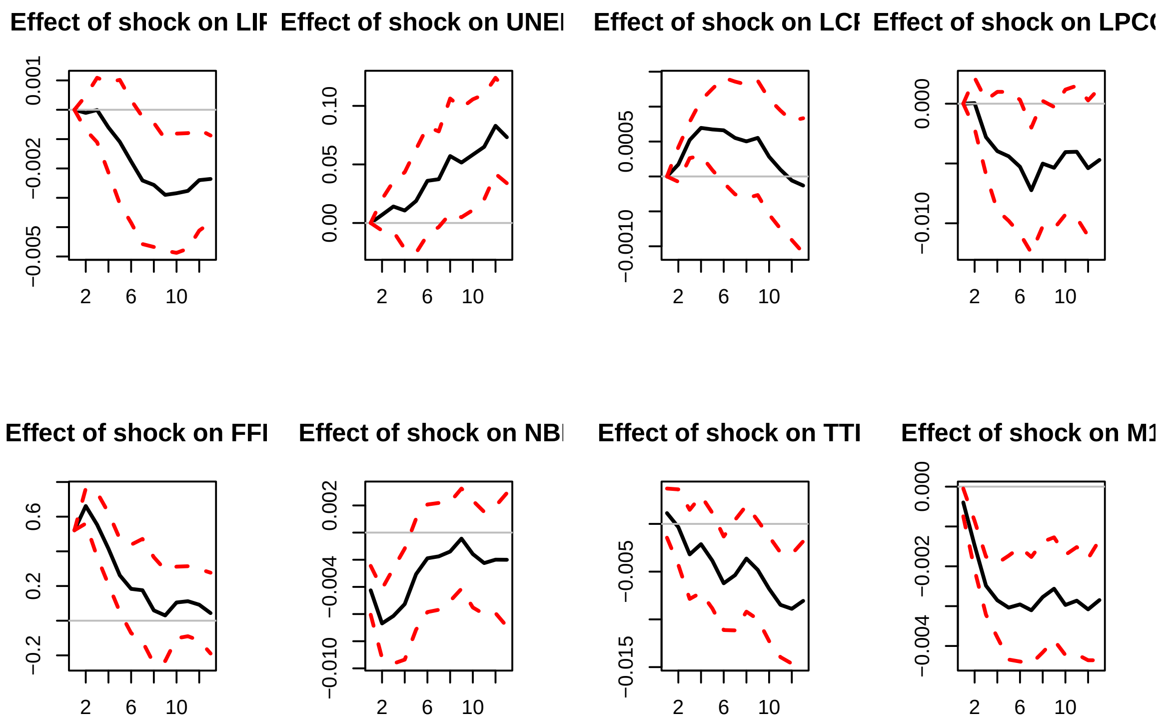

Figure 8.1: Response to a monetary-policy shock. Identification approach of Jorda (2005).

8.3 Situation B: IV approach

8.3.1 Instruments (proxies for structural shocks)

Consider now that we have a valid instrument z_t for $\eta_{1,t}$ (with $\mathbb{E}(z_t) = 0$). That is:

$$\left\{\begin{array}{llll}(IV.i) & \mathbb{E}(z_t\eta_{1,t}) & \neq 0 & & \text{(relevance condition)}\\ (IV.ii) & \mathbb{E}(z_t\eta_{j,t}) & = 0 & \text{for } j > 1 & \text{(exogeneity condition).}\end{array}\right. \tag{8.5}$$

The instrument z_t can be used to identify the structural shock. Eq. (8.5) implies that there exist $\rho \neq 0$ and a mean-zero variable ξ_t such that:

$$\eta_{1,t} = \rho z_t + \xi_t,$$

where ξ_t is correlated neither to z_t, nor to $\eta_{j,t}$, $j \geq 2$.

Proof. Define $\rho = \frac{\mathbb{E}(\eta_{1,t} z_t)}{\mathbb{V}ar(z_t)}$ and $\xi_t = \eta_{1,t} - \rho z_t$. It is easily seen that ξ_t satisfies the moment restrictions given above. □

Ramey (2016) reviews the different approaches employed to construct monetary policy-shocks (the two main approaches are presented in 8.1 and 8.2 below). She has also collected time series of such shocks, see her website. Several of these shocks are included in the `Ramey` dataset of package `IdSS`.

Example 8.1 (Identification of Monetary-Policy Shocks Based on High-Frequency Data)**.** Instruments for monetary-policy shocks can be extracted from high-frequency market data associated with interest-rate products.

The quotes of all interest-rate-related financial products are sensitive to monetary-policy announcements. That is because these quotes mainly depends on investors' expectations regarding future short-term rates: $\mathbb{E}_t(i_{t+s})$. Typically, if agents were risk-neutral, the maturity-h interest rate would approximatively be given by:

$$i_{t,h} \approx \mathbb{E}_t \left(\frac{1}{h} \int_0^h i_{t+s} ds \right) = \frac{1}{h} \int_0^h \mathbb{E}_t \left(i_{t+s} \right) ds.$$

In general, changes in $\mathbb{E}_t(i_{t+s})$, for $s > 0$, can be affected by all types of shocks that may trigger a reaction by the central bank.

However, if a MP announcement takes place between t and $t + \epsilon$, then most of $\mathbb{E}_{t+\epsilon}(i_{t+s}) - \mathbb{E}_t(i_{t+s})$ is to be attributed to the MP shock (see Figure 8.2, from Gürkaynak et al. (2005)). Hence, a monthly time series of MP shocks can be obtained by summing, over each month, the changes $i_{t+\epsilon,h} - i_{t,h}$ associated with a given interest rate (T-bills, futures, swaps) and a given maturity h.

See among others: Kuttner (2001), Cochrane and Piazzesi (2002), Gürkaynak et al. (2005), Piazzesi and Swanson (2008), Gertler and Karadi (2015). The time series named `FF4_TC`, `ED2_TC`, `ED3_TC`, `ED4_TC`, `GS1`, `ff1_vr`, `ff4_vr`, `ed2_vr`, `ff1_gkgreen`, `ff4_gkgreen`, `ed2_gkgreen` in the data frame `Ramey` of package `IdSS` are time series of shocks based on this approach (see Ramey's website for details).

Example 8.2 (Identification of Monetary-Policy Shocks Based on the Narrative Approach)**.** Romer and Romer (2004) propose a two-step approach:

a. derive a series for Federal Reserve intentions for the federal funds rate (the explicit target of the Fed) around FOMC meetings,

Figure 8.2: Source: Gurkaynak, Sack and Swanson (2005). Transaction rates of Federal funds futures on June 25, 2003, day on which a regularly scheduled FOMC meeting was scheduled. At 2:15 p.m., the FOMC announced that it was lowering its target for the federal funds rate from 1.25% to 1%, while many market participants were expecting a 50 bp cut. This shows that (i) financial markets seem to fully adjust to the policy action within just a few minutes and (ii) the federal funds rate surprise is not necessarily in the same direction as the federal funds rate action itself.

b. control for Federal Reserve forecasts.

This gives a measure of intended monetary policy actions not driven by information about future economic developments.

a. "intentions" are measured as a combination of narrative and quantitative evidence. Sources: (among others) Minutes of FOMC and "Blue Books".
b. Controls = variables spanning the information the Federal Reserve has about future developments. Data: Federal Reserve's internal forecasts (inflation, real output and unemployment), "Greenbook's forecasts" – usually issued 6 days before the FOMC meeting.

The shock measure is the residual series in the linear regression of (a) on (b). The time series `Ramey$rrshock83` and `Ramey$rrshock83b` (where `Ramey` is a data frame included in package `IdSS`) contain such shocks for the period 1983-2007. (`Ramey$rrshock83b` uses long-horizon Greenbook forecasts.)

To create a measure of *news* about future government spending, Ramey (2011) uses newspaper articles to construct a time series of (unexpected) fiscal shocks:[3]

[3] Data and replication codes can be found on her website.

Example 8.3
measure aims
to foreign poli
neoclassical fr
the public's ex

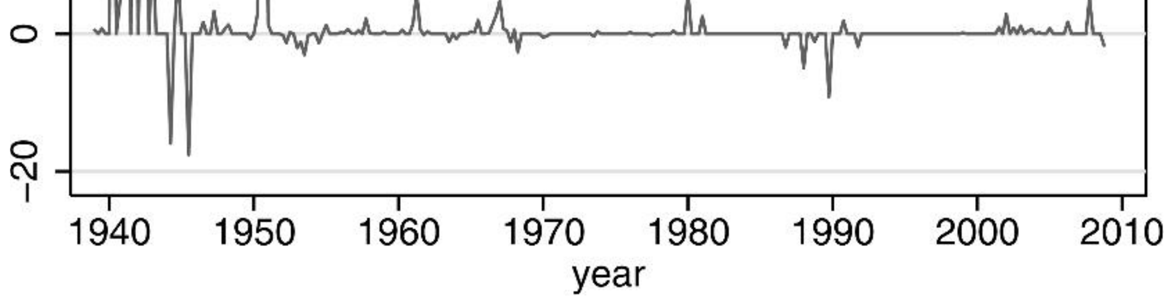

FIGURE IX
Defense News: PDV of Change in Spending as a Percent of GDP

According to Ramey (2011), government sources could not be used because (a) they were either not released in a timely manner or (b) were known to underestimate the costs of certain actions.

Figure 8.3 shows the resulting time series of shocks. Figure 8.4 shows the IRF of macro variables to the shock on expected government spending.

Figure 8.3: Source: Ramey (2011). Defense News: PDV of Change in Spending as a Percent of GDP.

There are two main IV approaches to estimate IRFs see Stock and Watson (2018):

a. The SVAR-IV approach (Subsection 8.3.2),

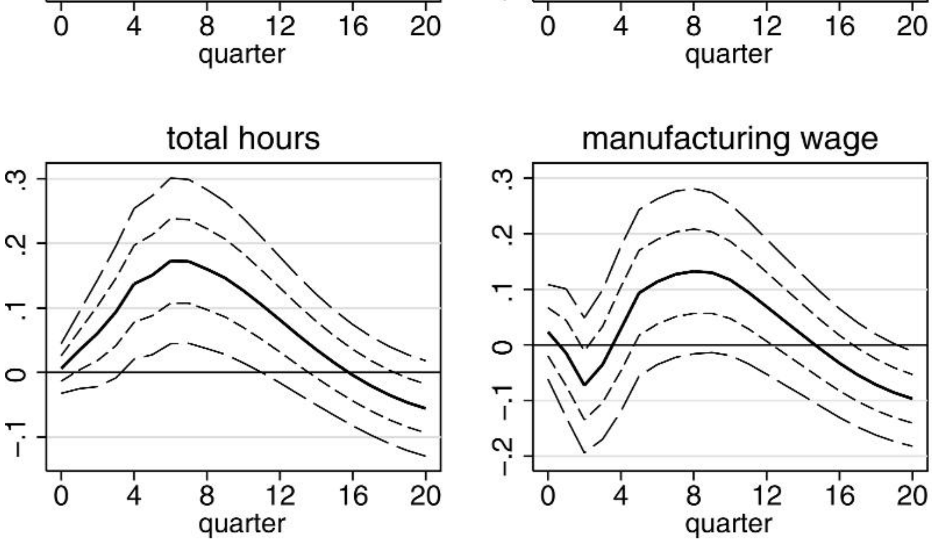

Figure 8.4: Source: Ramey (2011) [Figure X of the paper]. Responses of macro variables to a shock on expected government spending.

b. The LP-IV approach, where y_t's DGP is left unspecified (Subsection 8.3.3).

The LP-IV approach is based on a set of IV regressions (for each variable of interest, one for each forecast horizon). The SVAR-IV approach is based on IV regressions of VAR innovations only (one for each series of VAR innovations).

If the VAR adequately captures the DGP, then the IV-SVAR is optimal for all horizons. However, if the VAR is misspecified, then specification errors are compounded at each horizon and a local projection method would lead to better results.

8.3.2 Situation B.1: SVAR-IV approach

Assume you have consistent estimates of $\varepsilon_t = B\eta_t$, these estimates ($\hat{\varepsilon}_t$) coming from the estimation of a VAR model. We have, for $i \in \{1, \dots, n\}$:

$$\begin{aligned}\varepsilon_{i,t} &= b_{i,1}\eta_{1,t} + u_{i,t} \\ &= b_{i,1}\rho z_t + \underbrace{b_{i,1}\xi_t + u_{i,t}}_{\perp z_t}.\end{aligned} \tag{8.6}$$

($u_{i,t}$ is a linear combination of the $\eta_{j,t}$'s, $j \geq 2$).

Hence, up to a multiplicative factor (ρ), the (OLS) regressions of the $\hat{\varepsilon}_{i,t}$'s on z_t (that are consistent of the true $\varepsilon_{i,t}$'s) provide consistent estimates of the $b_{i,1}$'s.

Combined with the estimated VAR (the Φ_k matrices), this provides consistent estimates of the IRFs of $\eta_{1,t}$ on y_t, though up to a multiplicative factor. This scale ambiguity can be solved by rescaling the structural shock ("unit-effect normalisation", see Stock and Watson (2018)). Let us consider $\tilde{\eta}_{1,t} = b_{1,1}\eta_{1,t}$; by construction, $\tilde{\eta}_{1,t}$ has a unit contemporaneous effect on $y_{1,t}$. Denoting by $\tilde{B}_{i,1}$ the contemporaneous impact of $\tilde{\eta}_{1,t}$ on the i^{th} endogenous variable, we get:

$$\tilde{B}_1 = \frac{1}{b_{1,1}} B_1,$$

where B_1 denotes the 1^{st} column of B and $\tilde{B}_1 = [1, \tilde{B}_{2,1}, \dots, \tilde{B}_{n,1}]'$.

Eq. (8.6) gives:

$$\begin{aligned}\varepsilon_{1,t} &= \tilde{\eta}_{1,t} + u_{1,t} \\ \varepsilon_{i,t} &= \tilde{B}_{i,1}\tilde{\eta}_{1,t} + u_{i,t}.\end{aligned}$$

This suggests that $\tilde{B}_{i,1}$ can be estimated by regressing $\varepsilon_{i,t}$ on $\varepsilon_{1,t}$ (or $\hat{\varepsilon}_{i,t}$ on $\hat{\varepsilon}_{1,t}$ in practice), using z_t as an instrument.

What about inference? Once cannot use the usual TSLS standard deviations because the $\varepsilon_{i,t}$'s are not directly observed. Bootstrap procedures can be resorted to. Stock and Watson (2018) propose, in particular, a Gaussian parametric bootstrap:

Assume you have estimated $\{\widehat{\Phi}_1, \dots, \widehat{\Phi}_p, \widehat{B}_1\}$ using the SVAR-IV approach based on a size-T sample. Generate N (where N is large) size-T samples from the following VAR:

$$\left[\begin{array}{cc}\widehat{\Phi}(L) & 0\\ 0 & \widehat{\rho}(L)\end{array}\right]\left[\begin{array}{c}y_t\\ z_t\end{array}\right] = \left[\begin{array}{c}\varepsilon_t\\ e_t\end{array}\right],$$

$$\text{where}\quad \left[\begin{array}{c}\varepsilon_t\\ e_t\end{array}\right] \sim i.i.d.\,\mathcal{N}\left(\left[\begin{array}{c}0\\ 0\end{array}\right],\left[\begin{array}{cc}\Omega & S'_{\varepsilon,e}\\ S_{\varepsilon,e} & \sigma^2_e\end{array}\right]\right),$$

where $\widehat{\rho}(L)$ and σ^2_e result from the estimation of an AR process for z_t, and where Ω and $S_{\varepsilon,e}$ are sample covariances for the VAR/AR residuals.

For each simulated sample (of $\tilde{y}_t$ and $\tilde{z}_t$, say), estimate $\{\widetilde{\widehat{\Phi}}_1, \dots, \widetilde{\widehat{\Phi}}_p, \widetilde{\widehat{B}}_1\}$ and associated $\widetilde{\Psi}_{i,1,h}$. This provides e.g. a sequence of N estimates of $\Psi_{i,1,h}$, from which quantiles and conf. intervals can be deduced.

In the following lines of code, we use this approach to estimate the response of macroeconomic variables to a monetary policy shock. The instrument is `FF4_TC` from the Ramsey's database; they are base on the Gertler and Karadi (2015) approach, that use 3-month fed funds futures.

```r
library(vars);library(IdSS)
data("USmonthly")
First.date <- "1990-05-01";Last.date <- "2012-6-01"
indic.first <- which(USmonthly$DATES==First.date)
indic.last  <- which(USmonthly$DATES==Last.date)
USmonthly   <- USmonthly[indic.first:indic.last,]
shock.name <- c("FF4_TC","ED2_TC") # "ff1_vr", "rrshock83b"
indic.shock.name <- which(names(USmonthly)%in%shock.name)
Z <- as.matrix(USmonthly[,indic.shock.name])
par(plt=c(.1,.95,.1,.95))
plot(USmonthly$DATES,Z[,1],type="l",xlab="",ylab="",lwd=2)
lines(USmonthly$DATES,Z[,2],col="red",lwd=2,pch=3,lty=2)
```

```r
considered.variables <- c("GS1","LIP","LCPI","EBP")
Y <- as.matrix(USmonthly[,considered.variables])
n <- length(considered.variables)
colnames(Y) <- considered.variables
par(plt=c(.15,.95,.15,.8))
res.svar.iv <-
  svar.iv(Y,Z,p = 4,names.of.variables=considered.variables,
          nb.periods.IRF = 20,
          z.AR.order=1,
          nb.bootstrap.replications = 100,
          confidence.interval = 0.90,
          indic.plot=1)
```

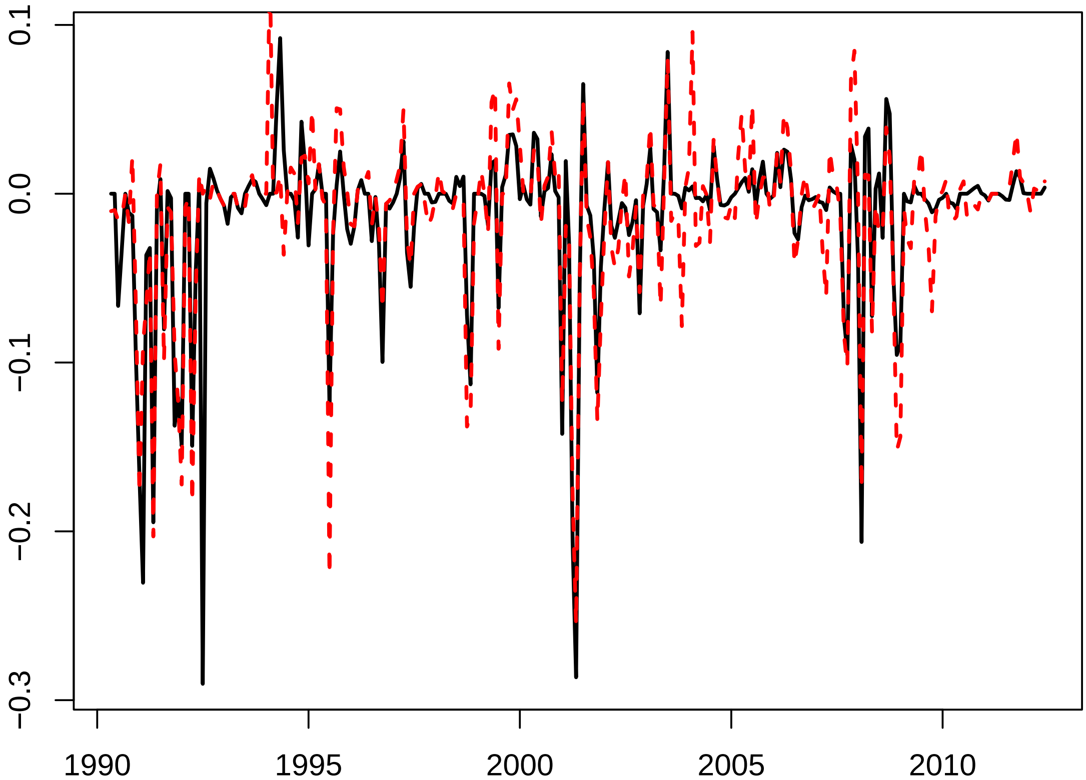

Figure 8.5: Gertler-Karadi monthly shocks, fed funds futures 3 months (resp. 6 months) in black (resp. in red).

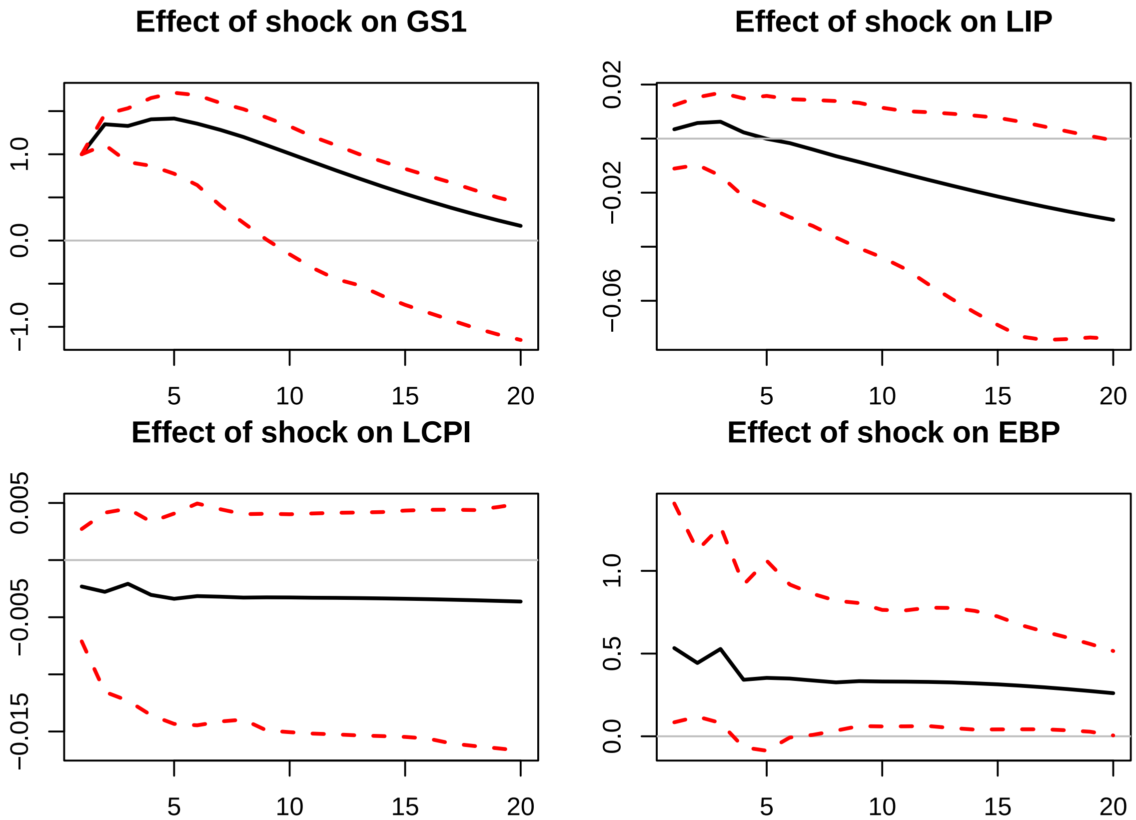

Figure 8.6: Reponses to a monetary-policy shock, SVAR-IV approach.

8.3.3 Situation B.2: LP-IV

If you do not want to posit a VAR-type dynamics for y_t –e.g., because you suspect that the true generating model may be a non-invertible VARMA model– you can directly proceed by IV-projection methods to obtain the $\tilde{\Psi}_{i,1,h} \equiv \Psi_{i,1,h}/b_{1,1}$ (that are the IRFs of $\tilde{\eta}_{1,t}$ on $y_{i,t}$).

However, Assumptions (IV.i) and (IV.ii) (Eq. (8.5)) have to be complemented with (IV.iii):

$$(IV.iii) \quad \mathbb{E}(z_t\eta_{j,t+h}) \quad = 0 \ \text{ for } h \neq 0 \quad \text{(lead-lag exogeneity)}$$

When (IV.i), (IV.ii) and (IV.iii) are satisfied, $\tilde{\Psi}_{i,1,h}$ can be estimated by regressing $y_{i,t+h}$ on $y_{1,t}$, using z_t as an instrument, i.e. by considering the TSLS estimation of:

$$y_{i,t+h} = \alpha_i + \tilde{\Psi}_{i,1,h} y_{1,t} + \nu_{i,t+h}, \tag{8.7}$$

where $\nu_{i,t+h}$ is correlated to $y_{1,t}$, but not to z_t.

We have indeed:

$$\begin{aligned} y_{1,t} &= \alpha_1 + \tilde{\eta}_{1,t} + v_{1,t} \\ y_{i,t+h} &= \alpha_i + \tilde{\Psi}_{i,1,h}\tilde{\eta}_{1,t} + v_{i,t+h}, \end{aligned}$$

where the $v_{i,t+h}$'s are uncorrelated to z_t under (IV.i), (IV.ii) and (IV.iii).

Note again that, for $h > 0$, the $v_{i,t+h}$ (and $\nu_{i,t+h}$) are auto-correlated. Newey-West corrections therefore have to be used to compute std errors of the $\tilde{\Psi}_{i,1,h}$'s estimates.

Consider the linear regression:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon,$$

where $\mathbb{E}(\varepsilon) = 0$, but where the explicative variables $\mathbf{X}$ can be correlated to the residuals ε. Moreover, the ε's may feature heteroskedasticity and be auto-correlated. We denote by $\mathbf{Z}$ the matrix of instruments, with $\mathbb{E}(\mathbf{X}'\mathbf{Z}) \neq 0$ but $\mathbb{E}(\varepsilon'\mathbf{Z}) = 0$.

The IV estimator of β is obtained by regressing $\hat{\mathbf{Y}}$ on $\hat{\mathbf{X}}$, where $\hat{\mathbf{Y}}$ and $\hat{\mathbf{X}}$ are the respective residuals of the regressions of $\mathbf{Y}$ and $\mathbf{X}$ on $\mathbf{Z}$.

$$\begin{aligned} \mathbf{b}_{iv} &= [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \\ \mathbf{b}_{iv} &= \beta + \frac{1}{\sqrt{T}} \underbrace{T[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}}_{=Q(\mathbf{X},\mathbf{Z}) \overset{p}{\to} \mathbf{Q}_{xz}} \underbrace{\sqrt{T}\left(\frac{1}{T}\mathbf{Z}'\varepsilon\right)}_{\overset{d}{\to}\mathcal{N}(0,S)}, \end{aligned}$$

where $\mathbf{S}$ is the long-run variance of $\mathbf{z}_t\varepsilon_t$.[4] The asymptotic covariance matrix of $\sqrt{T}\mathbf{b}_{iv}$ is $\mathbf{Q}_{xz}\mathbf{S}\mathbf{Q}_{xz}'$. Therefore, the covariance matrix of $\mathbf{b}_{iv}$ can be approximated by $\frac{1}{T}Q(\mathbf{X},\mathbf{Z})\hat{\mathbf{S}}Q(\mathbf{X},\mathbf{Z})'$ where $\hat{\mathbf{S}}$ is the Newey-West estimator of $\mathbf{S}$.[5]

---

[4]That is the sum of the covariance matrices of $\mathbf{z}_t\varepsilon_t$, for orders from $-\infty$ to $+\infty$.

[5]That is: $\hat{\mathbf{S}} = \hat{\gamma}_0 + 2\sum_{\nu=1}^{q}\left(1 - \frac{\nu}{q+1}\right)\hat{\gamma}_\nu$, where the $\hat{\gamma}_j$'s are sample auto-covariances of $\mathbf{z}_t\varepsilon_t$, see Newey and West (1987)

Assumption (IV.iii) is usually not restrictive for $h > 0$ (z_t is usually not affected by future shocks). By contrast, it may be restrictive for $h < 0$. This can be solved by adding controls in Regression (8.7). These controls should span the space of $\{\eta_{t-1}, \eta_{t-2}, \dots\}$.

If z_t is suspected to be correlated to past values of $\eta_{1,t}$ but not to the $\eta_{j,t}$'s, $j > 1$, then one can add lags of z_t as controls (method e.g. advocated by Ramey, 2016, p.108, considering the instrument by Gertler and Karadi (2015)).

In the general case, one can use lags of y_t as controls. Note that, even if (IV.iii) holds, adding controls may reduce the variance of the regression error.

```
res.LP.IV <- make.LPIV.irf(Y,Z,
                           nb.periods.IRF = 20,
                           nb.lags.Y.4.control=4,
                           nb.lags.Z.4.control=4,
                           indic.plot = 1, # Plots are displayed if = 1.
                           confidence.interval = 0.90)
```

Chapter 9

Panel VARs

Panel VARs have the same structure as VAR models, in the sense that all variables are assumed to be endogenous and interdependent, but a *cross sectional dimension* is added to the representation. There are N units indexed by $i \in \{1, ..., N\}$. The index i is generic and could indicate countries, sectors, markets… Then a panel VAR is

$$y_{it} = c_i + \Phi_i(L) y_{t-1} + \varepsilon_{it}.$$

where y_t is the stacked version of y_{it} and ε_t is i.i.d., with variance-covariance matrix Ω. Vector c_i and the lag polynomial $\Phi_i(L)$ may depend on the unit. Canova and Ciccarelli (2013) provide a survey of panel estimation methods.

Contrary to standard VARs, panel VARs may help study * similarities/differences in the transmission of shocks; * Spillovers, contagion.

But panel VARs are subject tothe *curse of dimensionality*. Indeed, they can be characterized by * *Dynamic interdependence*: potentially, the lags of *all* endogenous variables of *all* units can enter the model for unit i. * *Static interdependence*: ε_{it} are generally correlated across i. * *Cross sectional heterogeneity*: the intercept, the slope and the variance of the shocks may be unit-specific.

9.1 Without Dynamic interdependence

A panel VAR, assuming no dynamic interdependence, is of the form:

$$y_{it} = c_i + \Phi_i(L) {\color{red} y_{it-1} + \varepsilon_{it}}.$$

As a comparison, consider *micro panel data*, in the univariate cae (AR(1) case):

$$y_{it} = c_i + {\color{red}\phi} y_{it-1} + \varepsilon_{it}.$$

In that kind of context, we usually have no cross-sectional heterogeneity as $\phi_i = \phi$ for all i. Typically, we have a large cross-sectional dimension N, and a small time dimension T. If one uses a *"Fixed-effect'' regression*:

$$y_{it} - \frac{1}{T}\sum_{s=1}^{T} y_{is} = \phi(y_{it-1} - \frac{1}{T}\sum_{s=1}^{T} y_{is}) + \varepsilon_{it} - \frac{1}{T}\sum_{s=1}^{T} \varepsilon_{is},$$

then one faces the *Nickell bias*: with a lagged dependent variable, the estimator is biased, with a bias of size $\sim 1/T$. One can then use GMM regressions (Arellano and Bond (1991)) so as to get unbiased estimates.

Macro panel data have a different structure, with typically a moderate cross-sectional dimension N and a large time dimension T, so that the Nickell bias is negligible ($\to 0$ as $T \to \infty$).

9.1.1 Mean Group Estimator

When we etimate

$$y_{it} = c_i + \Phi_i(L)y_{it-1} + \varepsilon_{it},$$

we need to take into account the cross-sectional heterogeneity in the coefficients, i.e., different Φ_i's across i's. Pooled estimators (assuming cross-sectional homogeneity, i.e. identical Φ_i's across is) are *not consistent* (biased) if the underlying dynamics are actually heterogeneous. By contrast, the *Mean Group (MG) estimator*, which consists in estimating N separate regressions and calculating the coefficient means, is *consistent*.

9.1.2 Shock identification

Shock identification can be performed with standard methods (zeros, signs, FEVM, etc.). We make the assumption that Ω is block-diagonal (no interdependence on impact), which is consistent with the assumption of no cross-sectional dependence.

9.2 With Dynamic Interdependencies

A panel VAR that accommodates dynamic interdependence is of the form:

$$y_{it} = c_i + \Phi_i(L)y_{t-1} + \varepsilon_{it}.$$

We face a serious *curse of dimensionality* here: there are $NGp + 1$ coefficients to estimate in each equation.

A solution is to select some eligible dynamic links (See for instance del Negro (2011)). Another alternative is to use a factor model. This consists in capturing the dynamic interdependencies by a set of unobservable factors (See Canova and Ciccarelli (2004) and Canova and Ciccarelli (2009)). See Section 10 for more details on FAVAR models.

Chapter 10

Factor-Augmented VAR

VAR models are subject to the curse of dimensionality: If n, is large, then the number of parameters (in n^2) explodes.

In the case where one suspects that the $y_{i,t}$'s are mainly driven by a small number of random sources, a factor structure may be imposed, and **principal component analysis** (PCA, see Appendix 10.1) can be employed to estimate the relevant factors (Bernanke et al. (2005)).

10.1 Principal component analysis (PCA)

Principal component analysis (PCA) is a classical and easy-to-use statistical method to reduce the dimension of large datasets containing variables that are linearly driven by a relatively small number of factors. This approach is widely used in data analysis and image compression.

Suppose that we have T observations of a n-dimensional random vector x, denoted by $x_1, x_2, \dots, x_T$. We suppose that each component of x is of mean zero.

Let denote with X the matrix given by $\left[\begin{array}{cccc} x_1 & x_2 & \dots & x_T \end{array}\right]'$. Denote the j^{th} column of X by X_j.

We want to find the linear combination of the x_i's ($x.u$), with $\|u\| = 1$, with "maximum variance." That is, we want to solve:

$$\begin{array}{cl} \underset{u}{\arg\max} & u'X'Xu. \\ \text{s.t.} & \|u\| = 1 \end{array} \tag{10.1}$$

Since $X'X$ is a positive definite matrix, it admits the following decomposition:

$$\begin{aligned} X'X &= PDP' \\ &= P\left[\begin{array}{ccc} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{array}\right]P', \end{aligned}$$

where P is an orthogonal matrix whose columns are the eigenvectors of $X'X$.

We can order the eigenvalues such that $\lambda_1 \geq ... \geq \lambda_n$. (Since $X'X$ is positive definite, all these eigenvalues are positive.)

Since P is orthogonal, we have $u'X'Xu = u'PDP'u = y'Dy$ where $\|y\| = 1$. Therefore, we have $y_i^2 \leq 1$ for any $i \leq n$.

As a consequence:

$$y'Dy = \sum_{i=1}^{n} y_i^2 \lambda_i \leq \lambda_1 \sum_{i=1}^{n} y_i^2 = \lambda_1.$$

It is easily seen that the maximum is reached for $y = [1, 0, \cdots, 0]'$. Therefore, the maximum of the optimization program (Eq. (10.1)) is obtained for $u = P[1, 0, \cdots, 0]'$. That is, u is the eigenvector of $X'X$ that is associated with its larger eigenvalue (first column of P).

Let us denote with F the vector that is given by the matrix product XP. The columns of F, denoted by F_j, are called **factors**. We have:

$$F'F = P'X'XP = D.$$

Therefore, in particular, the F_j's are orthogonal.

Since $X = FP'$, the X_j's are linear combinations of the factors. Let us then denote with $\hat{X}_{i,j}$ the part of X_i that is explained by factor F_j, we have:

$$\begin{aligned} \hat{X}_{i,j} &= p_{ij} F_j \\ X_i &= \sum_j \hat{X}_{i,j} = \sum_j p_{ij} F_j. \end{aligned}$$

Consider the share of variance that is explained—through the n variables $(X_1, \dots, X_n)$—by the first factor F_1:

$$\frac{\sum_i \hat{X}'_{i,1} \hat{X}_{i,1}}{\sum_i X'_i X_i} = \frac{\sum_i p_{i1} F'_1 F_1 p_{i1}}{tr(X'X)} = \frac{\sum_i p_{i1}^2 \lambda_1}{tr(X'X)} = \frac{\lambda_1}{\sum_i \lambda_i}.$$

Intuitively, if the first eigenvalue is large, it means that the first factor captures a large share of the fluctutaions of the n X_i's.

By the same token, it is easily seen that the fraction of the variance of the n variables that is explained by factor j is given by:

$$\frac{\sum_i \hat{X}'_{i,j} \hat{X}_{i,j}}{\sum_i X'_i X_i} = \frac{\lambda_j}{\sum_i \lambda_i}.$$

Let us illustrate PCA on the term structure of yields. The term strucutre of yields (or yield curve) is know to be driven by only a small number of factors (e.g., Litterman and Scheinkman (1991)). One can typically employ PCA to recover such factors. The data used in the example below are taken from the Fred database (tickers: “DGS6MO”,“DGS1”, ...). The second plot shows the factor loardings, that indicate that the first factor is a level factor (loadings = black line), the second factor is a slope factor (loadings = blue line), the third factor is a curvature factor (loadings = red line).

To run a PCA, one simply has to apply function `prcomp` to a matrix of data:

```
library(IdSS)
USyields <- USyields[complete.cases(USyields),]
yds <- USyields[c("Y1","Y2","Y3","Y5","Y7","Y10","Y20","Y30")]
PCA.yds <- prcomp(yds,center=TRUE,scale. = TRUE)
```

Let us know visualize some results. The first plot of Figure 10.1 shows the share of total variance explained by the different principal components (PCs). The second plot shows the facotr loadings. The two bottom plots show how yields (in black) are fitted by linear combinations of the first two PCs only.

```
par(mfrow=c(2,2))
par(plt=c(.1,.95,.2,.8))
barplot(PCA.yds$sdev^2/sum(PCA.yds$sdev^2),
        main="Share of variance expl. by PC's")
axis(1, at=1:dim(yds)[2], labels=colnames(PCA.yds$x))
nb.PC <- 2
plot(-PCA.yds$rotation[,1],type="l",lwd=2,ylim=c(-1,1),
     main="Factor loadings (1st 3 PCs)",xaxt="n",xlab="")
axis(1, at=1:dim(yds)[2], labels=colnames(yds))
lines(PCA.yds$rotation[,2],type="l",lwd=2,col="blue")
lines(PCA.yds$rotation[,3],type="l",lwd=2,col="red")
Y1.hat <- PCA.yds$x[,1:nb.PC] %*% PCA.yds$rotation["Y1",1:2]
Y1.hat <- mean(USyields$Y1) + sd(USyields$Y1) * Y1.hat
plot(USyields$date,USyields$Y1,type="l",lwd=2,
     main="Fit of 1-year yields (2 PCs)",
     ylab="Obs (black) / Fitted by 2PCs (dashed blue)")
lines(USyields$date,Y1.hat,col="blue",lty=2,lwd=2)
Y10.hat <- PCA.yds$x[,1:nb.PC] %*% PCA.yds$rotation["Y10",1:2]
Y10.hat <- mean(USyields$Y10) + sd(USyields$Y10) * Y10.hat
plot(USyields$date,USyields$Y10,type="l",lwd=2,
     main="Fit of 10-year yields (2 PCs)",
     ylab="Obs (black) / Fitted by 2PCs (dashed blue)")
lines(USyields$date,Y10.hat,col="blue",lty=2,lwd=2)
```

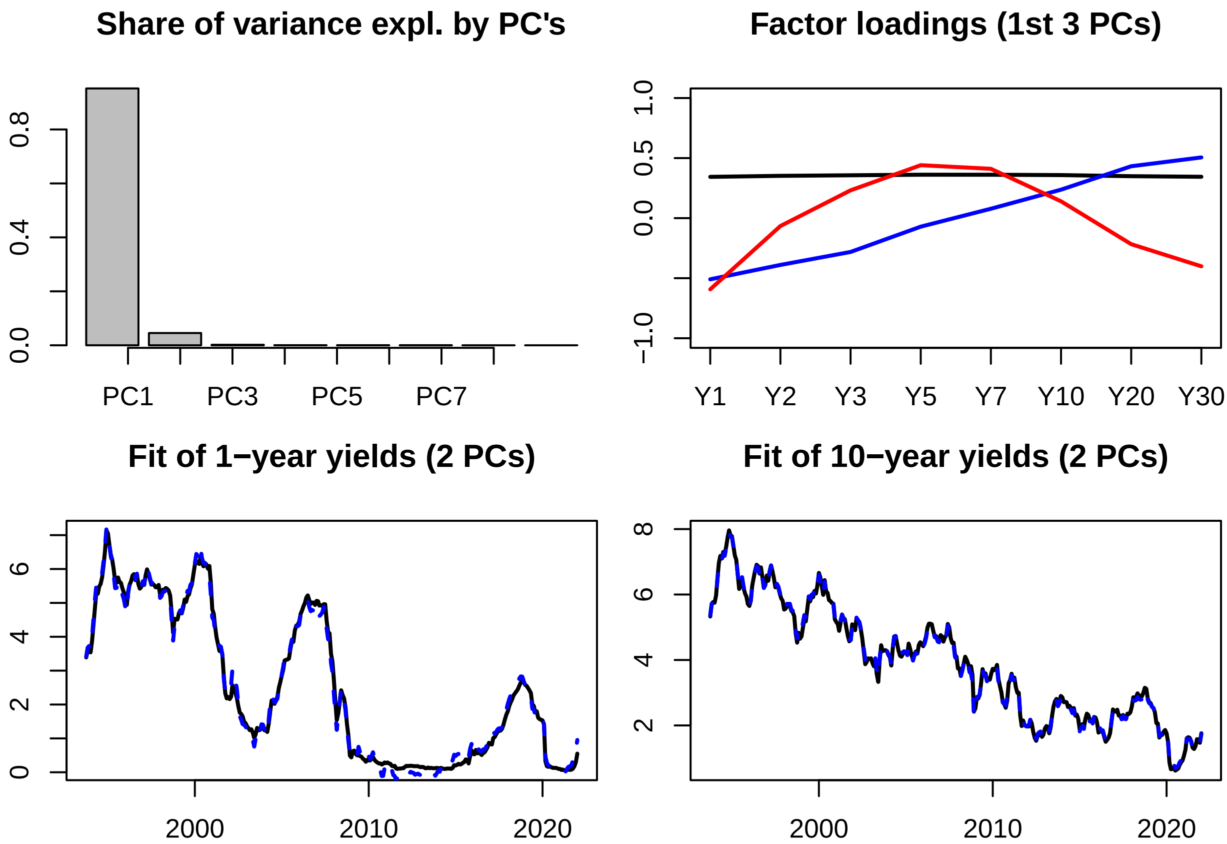

Figure 10.1: Some PCA results. The dataset contains 8 time series of U.S. interest rates of different maturities.

10.2 FAVAR models

Let us denote by F_t a k-dimensional vector of latent factors accounting for important shares of the variances of the $y_{i,t}$'s (with $K \ll n$) and by x_t is a small M-dimensional subset of y_t (with $M \ll n$). The following factor structure is posited:

$$y_t = \Lambda^f F_t + \Lambda^x x_t + e_t,$$

where the e_t are "small" serially and mutually i.i.d. error terms. That is F_t and x_t are supposed to drive most of the fluctuations of y_t's components.

The model is complemented by positing a VAR dynamics for $[F_t', x_t']'$:

$$\left[\begin{array}{c} F_t \\ x_t \end{array}\right] = \Phi(L) \left[\begin{array}{c} F_{t-1} \\ x_{t-1} \end{array}\right] + v_t. \tag{10.2}$$

Standard identification techniques of structural shocks can be employed in Eq. (10.2): Cholesky approach can be used for instance if the last component of x_t is the short-term interest rate and if it is assumed that a MP shock has no contemporaneous impact on other variables.

In their identification procedure, Bernanke et al. (2005) exploit the fact that macro-finance variables can be decomposed in two sets—fast-moving and slow-moving variables—and that only the former reacts contemporaneously to monetary-policy shocks. Now, how to estimate the (unobserved) factors F_t? Bernanke et al. (2005) note that the first $K + M$ PCA of the whole dataset (y_t), that they denote by $\hat{C}(F_t, x_t)$ should span the same space as F_t and x_t. To get an estimate of F_t, the dependence of $\hat{C}(F_t, x_t)$ in x_t has to be removed. This is done by regressing, by OLS, $\hat{C}(F_t, x_t)$ on x_t and on $\hat{C}^*(F_t)$, where the latter is an estimate of the common components other than x_t. To proxy for $\hat{C}^*(F_t)$, Bernanke et al. (2005) take principal components from the set of slow-moving variables, that are not comtemporaneously correlated to x_t. Vector $\hat{F}_t$ is then computed as $\hat{C}(F_t, x_t) - b_x x_t$, where b_x are the coefficients coming from the previous OLS regressions.

Note that this approach implies that the vectorial space spanned by $(\hat{F}_t, x_t)$ is the same as that spanned by $\hat{C}(F_t, x_t)$.

Below, we employ this method on the dataset built by McCracken and Ng (2016) —the FRED:MD database— that includes 119 time series.

```
library(BVAR)# contains the fred_md dataset
library(IdSS)
library(vars)
data <- fred_transform(fred_md,na.rm = FALSE, type = "fred_md")
data <- data[complete.cases(data),]
data.values <- scale(data, center = TRUE, scale = TRUE)
data_scaled <- data
data_scaled[1:dim(data)[1],1:dim(data)[2]] <- data.values
```

```
K <- 3
M <- 1
PCA <- prcomp(data_scaled) # implies that PCA$x %*% t(PCA$rotation) = data
C.hat <- PCA$x[,1:(K+M)]
fast_moving <- c("HOUST","HOUSTNE","HOUSTMW","HOUSTS","HOUSTW","HOUSTS","AMDMNOx",
                 "FEDFUNDS","CP3Mx","TB3MS","TB6MS","GS1","GS5","GS10",
                 "COMPAPFFx","TB3SMFFM","TB6SMFFM","T1YFFM","T5YFFM","T10YFFM",
                 "AAAFFM","EXSZUSx","EXJPUSx","EXUSUKx","EXCAUSx")
data.slow <- data_scaled[,-which(fast_moving %in% names(data))]
PCA.star <- prcomp(data.slow) # implies that PCA$x %*% t(PCA$rotation) = data
C.hat.star <- PCA.star$x[,1:K]
D <- cbind(data$FEDFUNDS,C.hat.star)
b.x <- solve(t(D)%*%D) %*% t(D) %*% C.hat
F.hat <- C.hat - data$FEDFUNDS %*% matrix(b.x[1,],nrow=1)
data_var <- data.frame(F.hat, FEDFUNDS = data$FEDFUNDS)
p <- 10
var <- VAR(data_var, p)
Omega <- var(residuals(var))
B <- t(chol(Omega))
D <- cbind(F.hat,data$FEDFUNDS)
loadings <- solve(t(D)%*%D) %*% t(D) %*% as.matrix(data_scaled)
irf <- simul.VAR(c=rep(0,(K+M)*p),Phi=Acoef(var),B,nb.sim=120,
                 y0.star=rep(0,(K+M)*p),indic.IRF = 1,
                 u.shock = c(rep(0,K+1),1))
irf.all <- irf %*% loadings
par(mfrow=c(2,2))
variables.2.plot <- c("FEDFUNDS","INDPRO","UNRATE","CPIAUCSL")
par(plt=c(.2,.95,.3,.95))
for(i in 1:length(variables.2.plot)){
  plot(cumsum(irf.all[,which(variables.2.plot[i]==names(data))]),lwd=2,
       type="l",xlab="months after shock",ylab=variables.2.plot[i])
}
```

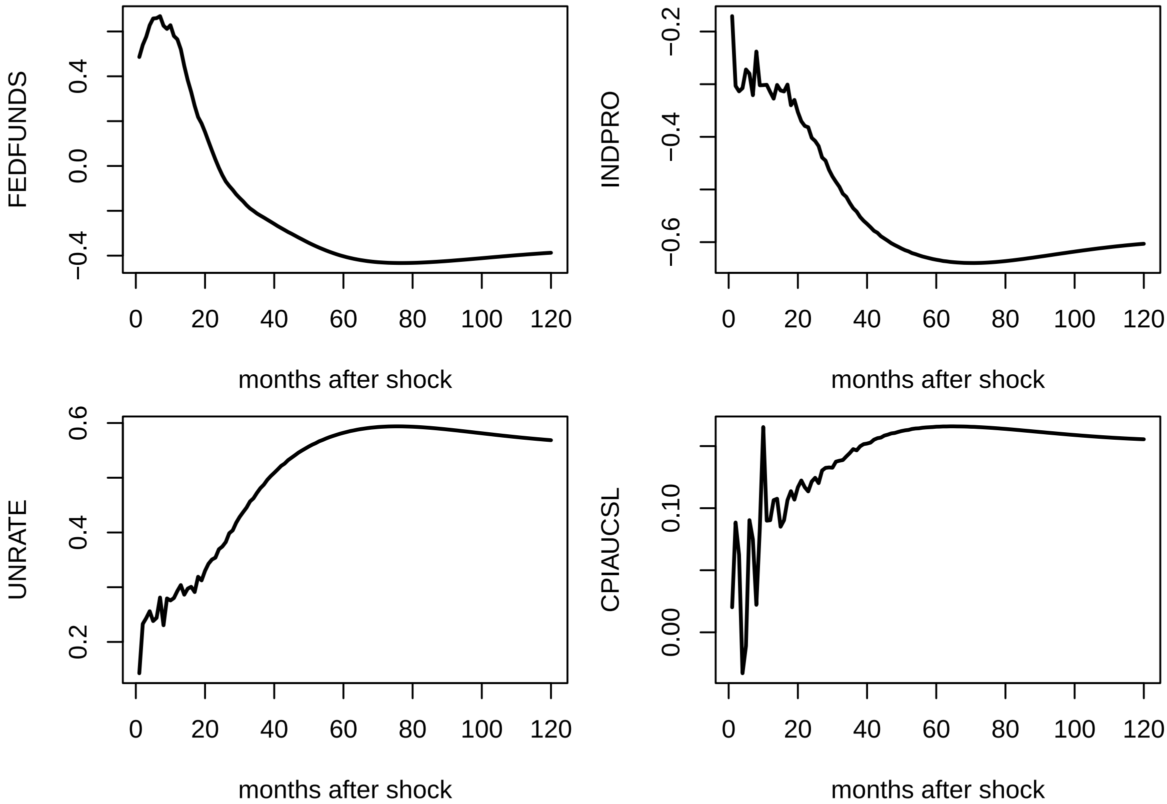

Figure 10.2: Responses of a monetary-policy shock. FAVAR approach of Bernanke, Boivin, and Eliasz (2005). FRED-MD dataset.

Chapter 11

Appendix

11.1 Definitions and statistical results

Definition 11.1 (Covariance stationarity)**.** The process y_t is covariance stationary —or weakly stationary— if, for all t and j,

$$\mathbb{E}(y_t) = \mu \quad \text{and} \quad \mathbb{E}\{(y_t - \mu)(y_{t-j} - \mu)\} = \gamma_j.$$

Definition 11.2 (Likelihood Ratio test statistics)**.** The likelihood ratio associated to a restriction of the form $H_0 : h(\theta) = 0$ (where $h(\theta)$ is a r-dimensional vector) is given by:

$$LR = \frac{\mathcal{L}_R(\theta; \mathbf{y})}{\mathcal{L}_U(\theta; \mathbf{y})} \quad (\in [0, 1]),$$

where $\mathcal{L}_R$ (respectively $\mathcal{L}_U$) is the likelihood function that imposes (resp. that does not impose) the restriction. The likelihood ratio test statistic is given by $-2\log(LR)$, that is:

$$\boxed{\xi^{LR} = 2(\log \mathcal{L}_U(\theta; \mathbf{y}) - \log \mathcal{L}_R(\theta; \mathbf{y})).}$$

Under regularity assumptions and under the null hypothesis, the test statistic follows a chi-square distribution with r degrees of freedom (see Table 11.3).

Proposition 11.1 (p.d.f. of a multivariate Gaussian variable)**.** *If $Y \sim \mathcal{N}(\mu, \Omega)$ and if Y is a n-dimensional vector, then the density function of Y is:*

$$\frac{1}{(2\pi)^{n/2}|\Omega|^{1/2}} \exp\left[-\frac{1}{2}\left(Y - \mu\right)' \Omega^{-1} \left(Y - \mu\right)\right].$$

11.2 Estimation of VARMA models

Section 1.4 discusses the estimation of VAR models and shows that standard VAR models can be esitmated by running OLS regressions.

If there is an MA component (i.e., if we consider a VARMA model), then OLS regressions yield biased estimates (even for asymptotically large samples). Assume for instance that y_t follows a VARMA(1,1) model:

$$y_{i,t} = \phi_i y_{t-1} + \varepsilon_{i,t},$$

where ϕ_i is the i^{th} row of Φ_1, and where $\varepsilon_{i,t}$ is a linear combination of η_t and η_{t-1}. Since y_{t-1} (the regressor) is correlated to η_{t-1}, it is also correlated to $\varepsilon_{i,t}$. The OLS regression of $y_{i,t}$ on y_{t-1} yields a biased estimator of ϕ_i (see Figure 11.1). Hence, SVARMA models cannot be consistently estimated by simple OLS regressions (contrary to VAR models, as we will see in the next section); instrumental-variable approaches can be employed to estimate SVARMA models (using past values of y_t as instruments, see, e.g., Gouriéroux et al. (2020)).

```
N <- 1000 # number of replications
T <- 100 # sample length
phi <- .8 # autoregressive parameter
sigma <- 1
par(mfrow=c(1,2))
for(theta in c(0,-0.4)){
  all.y <- matrix(0,1,N)
  y     <- all.y
  eta_1 <- rnorm(N)
  for(t in 1:(T+1)){
    eta <- rnorm(N)
    y <- phi * y + sigma * eta + theta * sigma * eta_1
    all.y <- rbind(all.y,y)
    eta_1 <- eta
  }
  all.y_1 <- all.y[1:T,]
  all.y   <- all.y[2:(T+1),]
  XX_1 <- 1/apply(all.y_1 * all.y_1,2,sum)
  XY   <- apply(all.y_1 * all.y,2,sum)
  phi.est.OLS <- XX_1 * XY
  plot(density(phi.est.OLS),xlab="OLS estimate of phi",ylab="",
       main=paste("theta = ",theta,sep=""))
  abline(v=phi,col="red",lwd=2)}
```

11.3 Proofs

Proof of Proposition 1.2

Proof. Using Proposition 11.1, we obtain that, conditionally on x_1, the log-likelihood is given

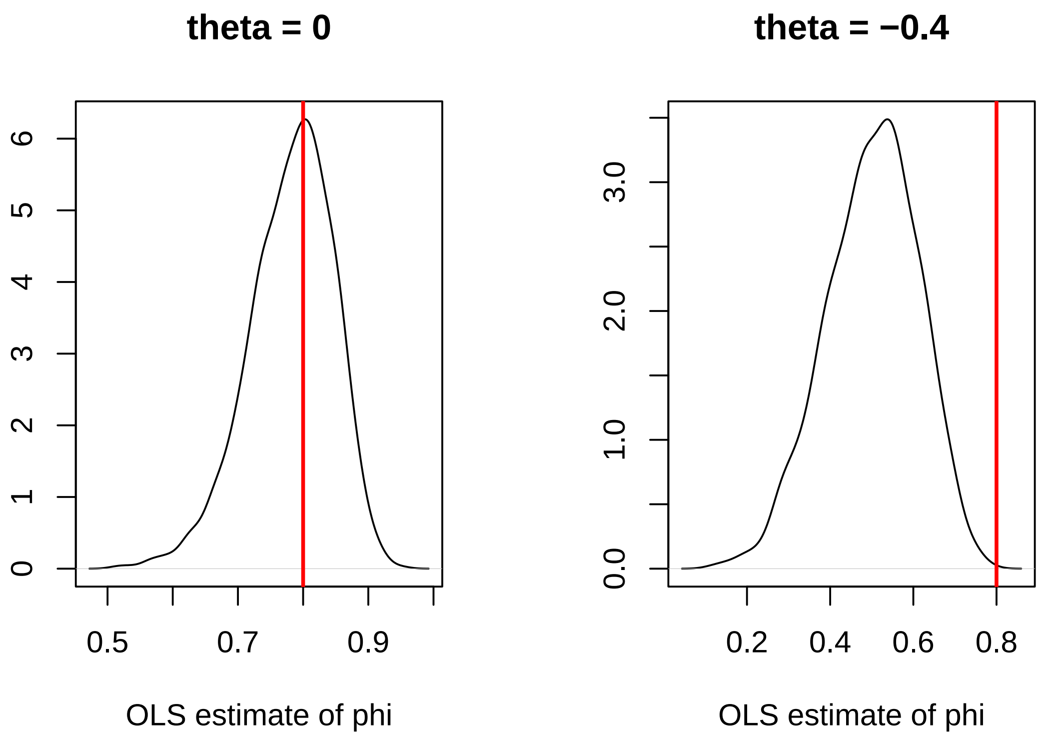

Figure 11.1: Illustration of the bias obtained when estimating the auto-regressive parameters of an ARMA process by (standard) OLS.

by

$$
\begin{aligned}
\log \mathcal{L}(Y_T;\theta) &= -(Tn/2)\log(2\pi) + (T/2)\log\left|\Omega^{-1}\right| \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}\left[\left(y_t-\Pi' x_t\right)'\Omega^{-1}\left(y_t-\Pi' x_t\right)\right].
\end{aligned}
$$

Let's rewrite the last term of the log-likelihood:

$$
\begin{aligned}
\sum_{t=1}^{T}\left[\left(y_t-\Pi' x_t\right)'\Omega^{-1}\left(y_t-\Pi' x_t\right)\right] &= \\
\sum_{t=1}^{T}\left[\left(y_t-\hat{\Pi}' x_t+\hat{\Pi}' x_t-\Pi' x_t\right)'\Omega^{-1}\left(y_t-\hat{\Pi}' x_t+\hat{\Pi}' x_t-\Pi' x_t\right)\right] &= \\
\sum_{t=1}^{T}\left[\left(\hat{\varepsilon}_t+(\hat{\Pi}-\Pi)' x_t\right)'\Omega^{-1}\left(\hat{\varepsilon}_t+(\hat{\Pi}-\Pi)' x_t\right)\right],&
\end{aligned}
$$

where the j^{th} element of the $(n \times 1)$ vector $\hat{\varepsilon}_t$ is the sample residual, for observation t, from an OLS regression of $y_{j,t}$ on x_t. Expanding the previous equation, we get:

$$
\begin{aligned}
\sum_{t=1}^{T}\left[\left(y_t-\Pi' x_t\right)'\Omega^{-1}\left(y_t-\Pi' x_t\right)\right] &= \sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega^{-1}\hat{\varepsilon}_t \\
+2\sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega^{-1}(\hat{\Pi}-\Pi)' x_t &+ \sum_{t=1}^{T}x_t'(\hat{\Pi}-\Pi)\Omega^{-1}(\hat{\Pi}-\Pi)' x_t.
\end{aligned}
$$

Let's apply the trace operator on the second term (that is a scalar):

$$
\begin{aligned}
\sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega^{-1}(\hat{\Pi}-\Pi)' x_t &= Tr\left(\sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega^{-1}(\hat{\Pi}-\Pi)' x_t\right) \\
= Tr\left(\sum_{t=1}^{T}\Omega^{-1}(\hat{\Pi}-\Pi)' x_t\hat{\varepsilon}_t'\right) &= Tr\left(\Omega^{-1}(\hat{\Pi}-\Pi)'\sum_{t=1}^{T}x_t\hat{\varepsilon}_t'\right).
\end{aligned}
$$

Given that, by construction (property of OLS estimates), the sample residuals are orthogonal to the explanatory variables, this term is zero. Introducing $\tilde{x}_t=(\hat{\Pi}-\Pi)' x_t$, we have

$$
\sum_{t=1}^{T}\left[\left(y_t-\Pi' x_t\right)'\Omega^{-1}\left(y_t-\Pi' x_t\right)\right] = \sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega^{-1}\hat{\varepsilon}_t + \sum_{t=1}^{T}\tilde{x}_t'\Omega^{-1}\tilde{x}_t.
$$

Since Ω is a positive definite matrix, Ω^{-1} is as well. Consequently, the smallest value that the last term can take is obtained for $\tilde{x}_t=0$, i.e. when $\Pi=\hat{\Pi}$.

The MLE of Ω is the matrix $\hat{\Omega}$ that maximizes $\Omega \overset{\ell}{\rightarrow} L(Y_T;\hat{\Pi},\Omega)$. We have:

$$
\log \mathcal{L}(Y_T;\hat{\Pi},\Omega) = -(Tn/2)\log(2\pi) + (T/2)\log\left|\Omega^{-1}\right| - \frac{1}{2}\sum_{t=1}^{T}\left[\hat{\varepsilon}_t'\Omega^{-1}\hat{\varepsilon}_t\right].
$$

Matrix $\hat{\Omega}$ is a symmetric positive definite. It is easily checked that the (unrestricted) matrix that maximizes the latter expression is symmetric positive definite matrix. Indeed:

$$\frac{\partial \log \mathcal{L}(Y_T; \hat{\Pi}, \Omega)}{\partial \Omega} = \frac{T}{2}\Omega' - \frac{1}{2}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}'_t \Rightarrow \hat{\Omega}' = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}'_t,$$

which leads to the result. □

Proof of Proposition 1.3

Proof. Let us drop the i subscript. Rearranging Eq. (1.15), we have:

$$\sqrt{T}(\mathbf{b} - \beta) = (X'X/T)^{-1}\sqrt{T}(X'\varepsilon/T).$$

Let us consider the autocovariances of $\mathbf{v}_t = x_t \varepsilon_t$, denoted by γ^v_j. Using the fact that x_t is a linear combination of past ε_ts and that ε_t is a white noise, we get that $\mathbb{E}(\varepsilon_t x_t) = 0$. Therefore

$$\gamma^v_j = \mathbb{E}(\varepsilon_t \varepsilon_{t-j} x_t x'_{t-j}).$$

If $j > 0$, we have $\mathbb{E}(\varepsilon_t \varepsilon_{t-j} x_t x'_{t-j}) = \mathbb{E}(\mathbb{E}[\varepsilon_t \varepsilon_{t-j} x_t x'_{t-j} | \varepsilon_{t-j}, x_t, x_{t-j}]) = \mathbb{E}(\varepsilon_{t-j} x_t x'_{t-j} \mathbb{E}[\varepsilon_t | \varepsilon_{t-j}, x_t, x_{t-j}]) = 0$. Note that we have $\mathbb{E}[\varepsilon_t | \varepsilon_{t-j}, x_t, x_{t-j}] = 0$ because $\{\varepsilon_t\}$ is an i.i.d. white noise sequence. If $j = 0$, we have:

$$\gamma^v_0 = \mathbb{E}(\varepsilon^2_t x_t x'_t) = \mathbb{E}(\varepsilon^2_t)\mathbb{E}(x_t x'_t) = \sigma^2 \mathbf{Q}.$$

The convergence in distribution of $\sqrt{T}(X'\varepsilon/T) = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} v_t$ results from the Central Limit Theorem for covariance-stationary processes, using the γ^v_j computed above. □

11.4 Statistical Tables

Table 11.1: Quantiles of the $\mathcal{N}(0,1)$ distribution. If a and b are respectively the row and column number; then the corresponding cell gives $\mathbb{P}(0 < X \leq a+b)$, where $X \sim \mathcal{N}(0,1)$.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.6179	0.7257	0.8159	0.8849	0.9332	0.9641	0.9821	0.9918	0.9965
0.1	0.5040	0.6217	0.7291	0.8186	0.8869	0.9345	0.9649	0.9826	0.9920	0.9966
0.2	0.5080	0.6255	0.7324	0.8212	0.8888	0.9357	0.9656	0.9830	0.9922	0.9967
0.3	0.5120	0.6293	0.7357	0.8238	0.8907	0.9370	0.9664	0.9834	0.9925	0.9968
0.4	0.5160	0.6331	0.7389	0.8264	0.8925	0.9382	0.9671	0.9838	0.9927	0.9969
0.5	0.5199	0.6368	0.7422	0.8289	0.8944	0.9394	0.9678	0.9842	0.9929	0.9970
0.6	0.5239	0.6406	0.7454	0.8315	0.8962	0.9406	0.9686	0.9846	0.9931	0.9971
0.7	0.5279	0.6443	0.7486	0.8340	0.8980	0.9418	0.9693	0.9850	0.9932	0.9972
0.8	0.5319	0.6480	0.7517	0.8365	0.8997	0.9429	0.9699	0.9854	0.9934	0.9973
0.9	0.5359	0.6517	0.7549	0.8389	0.9015	0.9441	0.9706	0.9857	0.9936	0.9974
1	0.5398	0.6554	0.7580	0.8413	0.9032	0.9452	0.9713	0.9861	0.9938	0.9974
1.1	0.5438	0.6591	0.7611	0.8438	0.9049	0.9463	0.9719	0.9864	0.9940	0.9975
1.2	0.5478	0.6628	0.7642	0.8461	0.9066	0.9474	0.9726	0.9868	0.9941	0.9976
1.3	0.5517	0.6664	0.7673	0.8485	0.9082	0.9484	0.9732	0.9871	0.9943	0.9977
1.4	0.5557	0.6700	0.7704	0.8508	0.9099	0.9495	0.9738	0.9875	0.9945	0.9977
1.5	0.5596	0.6736	0.7734	0.8531	0.9115	0.9505	0.9744	0.9878	0.9946	0.9978
1.6	0.5636	0.6772	0.7764	0.8554	0.9131	0.9515	0.9750	0.9881	0.9948	0.9979
1.7	0.5675	0.6808	0.7794	0.8577	0.9147	0.9525	0.9756	0.9884	0.9949	0.9979
1.8	0.5714	0.6844	0.7823	0.8599	0.9162	0.9535	0.9761	0.9887	0.9951	0.9980
1.9	0.5753	0.6879	0.7852	0.8621	0.9177	0.9545	0.9767	0.9890	0.9952	0.9981
2	0.5793	0.6915	0.7881	0.8643	0.9192	0.9554	0.9772	0.9893	0.9953	0.9981
2.1	0.5832	0.6950	0.7910	0.8665	0.9207	0.9564	0.9778	0.9896	0.9955	0.9982
2.2	0.5871	0.6985	0.7939	0.8686	0.9222	0.9573	0.9783	0.9898	0.9956	0.9982
2.3	0.5910	0.7019	0.7967	0.8708	0.9236	0.9582	0.9788	0.9901	0.9957	0.9983
2.4	0.5948	0.7054	0.7995	0.8729	0.9251	0.9591	0.9793	0.9904	0.9959	0.9984
2.5	0.5987	0.7088	0.8023	0.8749	0.9265	0.9599	0.9798	0.9906	0.9960	0.9984
2.6	0.6026	0.7123	0.8051	0.8770	0.9279	0.9608	0.9803	0.9909	0.9961	0.9985
2.7	0.6064	0.7157	0.8078	0.8790	0.9292	0.9616	0.9808	0.9911	0.9962	0.9985
2.8	0.6103	0.7190	0.8106	0.8810	0.9306	0.9625	0.9812	0.9913	0.9963	0.9986
2.9	0.6141	0.7224	0.8133	0.8830	0.9319	0.9633	0.9817	0.9916	0.9964	0.9986

Table 11.2: Quantiles of the Student-t distribution. The rows correspond to different degrees of freedom (ν, say); the columns correspond to different probabilities (z, say). The cell gives q that is s.t. $\mathbb{P}(-q < X < q) = z$, with $X \sim t(\nu)$.

	0.05	0.1	0.75	0.9	0.95	0.975	0.99	0.999
1	0.079	0.158	2.414	6.314	12.706	25.452	63.657	636.619
2	0.071	0.142	1.604	2.920	4.303	6.205	9.925	31.599
3	0.068	0.137	1.423	2.353	3.182	4.177	5.841	12.924
4	0.067	0.134	1.344	2.132	2.776	3.495	4.604	8.610
5	0.066	0.132	1.301	2.015	2.571	3.163	4.032	6.869
6	0.065	0.131	1.273	1.943	2.447	2.969	3.707	5.959
7	0.065	0.130	1.254	1.895	2.365	2.841	3.499	5.408
8	0.065	0.130	1.240	1.860	2.306	2.752	3.355	5.041
9	0.064	0.129	1.230	1.833	2.262	2.685	3.250	4.781
10	0.064	0.129	1.221	1.812	2.228	2.634	3.169	4.587
20	0.063	0.127	1.185	1.725	2.086	2.423	2.845	3.850
30	0.063	0.127	1.173	1.697	2.042	2.360	2.750	3.646
40	0.063	0.126	1.167	1.684	2.021	2.329	2.704	3.551
50	0.063	0.126	1.164	1.676	2.009	2.311	2.678	3.496
60	0.063	0.126	1.162	1.671	2.000	2.299	2.660	3.460
70	0.063	0.126	1.160	1.667	1.994	2.291	2.648	3.435
80	0.063	0.126	1.159	1.664	1.990	2.284	2.639	3.416
90	0.063	0.126	1.158	1.662	1.987	2.280	2.632	3.402
100	0.063	0.126	1.157	1.660	1.984	2.276	2.626	3.390
200	0.063	0.126	1.154	1.653	1.972	2.258	2.601	3.340
500	0.063	0.126	1.152	1.648	1.965	2.248	2.586	3.310

Table 11.3: Quantiles of the χ^2 distribution. The rows correspond to different degrees of freedom; the columns correspond to different probabilities.

	0.05	0.1	0.75	0.9	0.95	0.975	0.99	0.999
1	0.004	0.016	1.323	2.706	3.841	5.024	6.635	10.828
2	0.103	0.211	2.773	4.605	5.991	7.378	9.210	13.816
3	0.352	0.584	4.108	6.251	7.815	9.348	11.345	16.266
4	0.711	1.064	5.385	7.779	9.488	11.143	13.277	18.467
5	1.145	1.610	6.626	9.236	11.070	12.833	15.086	20.515
6	1.635	2.204	7.841	10.645	12.592	14.449	16.812	22.458
7	2.167	2.833	9.037	12.017	14.067	16.013	18.475	24.322
8	2.733	3.490	10.219	13.362	15.507	17.535	20.090	26.124
9	3.325	4.168	11.389	14.684	16.919	19.023	21.666	27.877
10	3.940	4.865	12.549	15.987	18.307	20.483	23.209	29.588
20	10.851	12.443	23.828	28.412	31.410	34.170	37.566	45.315
30	18.493	20.599	34.800	40.256	43.773	46.979	50.892	59.703
40	26.509	29.051	45.616	51.805	55.758	59.342	63.691	73.402
50	34.764	37.689	56.334	63.167	67.505	71.420	76.154	86.661
60	43.188	46.459	66.981	74.397	79.082	83.298	88.379	99.607
70	51.739	55.329	77.577	85.527	90.531	95.023	100.425	112.317
80	60.391	64.278	88.130	96.578	101.879	106.629	112.329	124.839
90	69.126	73.291	98.650	107.565	113.145	118.136	124.116	137.208
100	77.929	82.358	109.141	118.498	124.342	129.561	135.807	149.449
200	168.279	174.835	213.102	226.021	233.994	241.058	249.445	267.541
500	449.147	459.926	520.950	540.930	553.127	563.852	576.493	603.446

Table 11.4: Quantiles of the $\mathcal{F}$ distribution. The columns and rows correspond to different degrees of freedom (resp. n_1 and n_2). The different panels correspond to different probabilities (α) The corresponding cell gives z that is s.t. $\mathbb{P}(X \leq z) = \alpha$, with $X \sim \mathcal{F}(n_1, n_2)$.

	1	2	3	4	5	6	7	8	9	10
alpha = 0.9										
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937
50	2.809	2.412	2.197	2.061	1.966	1.895	1.840	1.796	1.760	1.729
100	2.756	2.356	2.139	2.002	1.906	1.834	1.778	1.732	1.695	1.663
500	2.716	2.313	2.095	1.956	1.859	1.786	1.729	1.683	1.644	1.612
alpha = 0.95										
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927
500	3.860	3.014	2.623	2.390	2.232	2.117	2.028	1.957	1.899	1.850
alpha = 0.99										
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503
500	6.686	4.648	3.821	3.357	3.054	2.838	2.675	2.547	2.443	2.356

Bibliography

Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for svars. *American Economic Review*, 108(10):2802–29.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58(2):277–297.

Arias, J. E., Rubio-Ramírez, J. F., and Waggoner, D. F. (2018). Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications. *Econometrica*, 86(2):685–720.

Barsky, R. B. and Sims, E. R. (2011). News shocks and business cycles. *Journal of Monetary Economics*, 58(3):273–289.

Bernanke, B. S. (1986). Alternative explanations of the money-income correlation. *Carnegie-Rochester Conference Series on Public Policy*, 25:49–99.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach*. *The Quarterly Journal of Economics*, 120(1):387–422.

Blanchard, O. J. and Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review*, 79(4):655–673.

Canova, F. and Ciccarelli, M. (2004). Forecasting and turning point predictions in a bayesian panel var model. *Journal of Econometrics*, 120(2):327–359.

Canova, F. and Ciccarelli, M. (2009). Estimating multi-country var models. *International Economic Review*, 50(3):929–959.

Canova, F. and Ciccarelli, M. (2013). Panel Vector Autoregressive Models: A Survey. CEPR Discussion Papers 9380, C.E.P.R. Discussion Papers.

Christiano, L. J., Eichenbaum, M., and Evans, C. (1996). The effects of monetary policy shocks: Evidence from the flow of funds. *The Review of Economics and Statistics*, 78(1):16–34.

Christiano, L. J., Eichenbaum, M., and Vigfusson, R. (2007). *Assessing Structural VARs*, pages 1–106. MIT Press.

Cochrane, J. H. and Piazzesi, M. (2002). The Fed and Interest Rates - A High-Frequency Identification. *American Economic Review*, 92(2):90–95.

Danne, C. (2015). VARsignR: Estimating VARs using sign restrictions in R. MPRA Paper 68429, University Library of Munich, Germany.

Dedola, L. and Lippi, F. (2005). The monetary transmission mechanism: Evidence from the industries of five oecd countries. *European Economic Review*, 49(6):1543–1569.

del Negro, M. (2011). 293 Bayesian Macroeconometrics. In *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press.

Erceg, C. J., Guerrieri, L., and Gust, C. (2005). Can Long-Run Restrictions Identify Technology Shocks? *Journal of the European Economic Association*, 3(6):1237–1278.

Faust, J. and Leeper, E. M. (1997). When do long-run identifying restrictions give reliable results? *Journal of Business & Economic Statistics*, 15(3):345–353.

Fischer, S. (1977). Long-term contracts, rational expectations, and the optimal money supply rule. *Journal of Political Economy*, 85(1):191–205.

Galí, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *American Economic Review*, 89(1):249–271.

Galí, J. (1992). How well does the is-lm model fit postwar u.s. data? *The Quarterly Journal of Economics*, 107(2):709–738.

Gerlach, S. and Smets, F. (1995). The Monetary Transmission Mechanism: Evidence from the G-7 Countries. CEPR Discussion Papers 1219, C.E.P.R. Discussion Papers.

Gertler, M. and Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2020). Identification and Estimation in Non-Fundamental Structural VARMA Models. *Review of Economic Studies*, 87(4):1915–1953.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017). Statistical inference for independent component analysis: Application to structural var models. *Journal of Econometrics*, 196(1):111–126.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

Gürkaynak, R. S., Sack, B., and Swanson, E. (2005). Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements. *International Journal of Central Banking*, 1(1).

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Jarocinski, M. (2021). Estimating the Fed's Unconventional Policy Shocks. Working Paper Series 20210, European Central Bank.

Jordà, O. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1):161–182.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *The Review of Economics and Statistics*, 80(2):218–230.

Kim, J. H. (2022). *VAR.etp: VAR Modelling: Estimation, Testing, and Prediction*. R package version 1.0.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of Monetary Economics*, 47(3):523–544.

Lanne, M. and Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility. *Journal of Money, Credit and Banking*, 40(6):1131–1149.

Lanne, M., Lütkepohl, H., and Maciejowska, K. (2010). Structural vector autoregressions with markov switching. *Journal of Economic Dynamics and Control*, 34(2):121–131.

Levchenko, A. A. and Pandalai-Nayar, N. (2018). Tfp, News, and "Sentiments": the International Transmission of Business Cycles. *Journal of the European Economic Association*, 18(1):302–341.

Litterman, R. and Scheinkman, J. (1991). Common Factors Affecting Bond Returns. *Journal of Fixed Income*, (1):54–61.

Lütkepohl, H. (1990). Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. *The Review of Economics and Statistics*, 72(1):116–25.

Lütkepohl, H. and Netšunajev, A. (2017). Structural vector autoregressions with heteroskedasticity: A review of different volatility models. *Econometrics and Statistics*, 1:2–18.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135(1-2):499–526.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Normandin, M. and Phaneuf, L. (2004). Monetary policy shocks:: Testing identification conditions under time-varying conditional volatility. *Journal of Monetary Economics*, 51(6):1217–1243.

Piazzesi, M. and Swanson, E. T. (2008). Futures prices as risk-adjusted forecasts of monetary policy. *Journal of Monetary Economics*, 55(4):677–691.

Ramey, V. A. (2011). Identifying government spending shocks: It's all tn the timing. *The Quarterly Journal of Economics*, 126(1):1–50.

Ramey, V. A. (2016). Macroeconomic Shocks and Their Propagation. NBER Working Papers 21978, National Bureau of Economic Research, Inc.

Rigobon, R. (2003). Identification Through Heteroskedasticity. *The Review of Economics and Statistics*, 85(4):777–792.

Romer, C. D. and Romer, D. H. (2004). A New Measure of Monetary Shocks: Derivation and Implications. *American Economic Review*, 94(4):1055–1084.

Ruibio-Ramírez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, 77(2):665–696.

Sims, C. A. (1986). Are forecasting models usable for policy analysis? *Quarterly Review*, 10(Win):2–16.

Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128(610):917–948.

Uhlig, H. (2004). What moves GNP? Econometric Society 2004 North American Winter Meetings 636, Econometric Society.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.