

Advanced Econometrics

Jean-Paul Renne

2022-09-16

Contents

| | | |
|----------|---|-----------|
| 1 | Before starting | 5 |
| 2 | Time Series | 7 |
| 2.1 | Multivariate models | 7 |
| 3 | Appendix | 55 |
| 3.1 | Statistical Tables | 55 |
| 3.2 | Statistics: definitions and results | 55 |
| 3.3 | Some properties of Gaussian variables | 67 |
| 3.4 | Proofs | 68 |
| 3.5 | Additional codes | 77 |

Chapter 1

Before starting

This course covers various econometric topics, including linear regression models, discrete-choice models, and time series analysis. It provides examples or simulations based on R codes.

The R codes use various packages that can be obtained from CRAN. Several pieces of code also involve procedures and data from my **AEC** package. The latter is available on GitHub. To install it, one need to employ the **devtools** library:

```
library(devtools)
install_github("jrenne/AEC")
library(AEC)
```

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)
 - My own tutorial: https://jrenne.shinyapps.io/Rtuto_publiShiny/

Chapter 2

Time Series

2.1 Multivariate models

This section presents Vector Auto-Regressive Moving-Average (SVARMA) models. These models are widely used in macroeconomic analysis. While simple and easy to estimate, they make it possible to conveniently capture the dynamics of complex multivariate systems. VAR popularity is notably due to Sims (1980)'s influential work. A nice survey is proposed by Stock and Watson (2016).

In economics, VAR models are often employed in order to identify *structural* shocks, that are independent primitive exogenous forces that drive economic variables (Ramey (2016)). They are often given a specific economic meaning (e.g., demand and supply shocks).

Working with these models (VAR and VARMA models) often is often based on two steps: in a first step, the **reduced-form** version of the model is estimated; in a second step, structural shocks are identified and IRFs are produced.

Kilian (1998) See this page

Sign restrictions: package, Danne (2015).

Pfaff (2008)

Korobilis (2013): R package to estimate Bayesian VAR models.

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

```
##          inv(-1)   inc(-1)   con(-1)   inv(-2)   inc(-2)   con(-2)   const
## [1,] -0.3146496  0.2087952  0.9400172 -0.1499038  0.1253967  0.9521322 -0.01837906
## [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
```

2.1.1 Definition of VARs (and SVARMA) models

Definition 2.1 ((S)VAR model). Let y_t denote a $n \times 1$ vector of random variables. Process y_t follows a p^{th} -order VAR if, for all t , we have

$$\begin{aligned} VAR: y_t &= c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \varepsilon_t, \\ SVAR: y_t &= c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + B\eta_t, \end{aligned} \quad (2.1)$$

with $\varepsilon_t = B\eta_t$. We assume that $\{\eta_t\}$ is a white noise sequence whose components are mutually and serially independent.

The first line of Eq. (2.1) 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).

As was the case for univariate models, VARs can be extended with MA terms in η_t :

Definition 2.2 ((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

$$\begin{aligned} VARMA: y_t &= c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \dots + \Theta_q, \\ SVARMA: y_t &= c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + B_0 \eta_t + B_1 \eta_{t-1} + \dots + B_q \eta_{t-q}, \end{aligned} \quad (2.2)$$

with $\varepsilon_t = B_0 \eta_t$ (and $B_j = \Theta_j B_0$, for $j \geq 1$). We assume that $\{\eta_t\}$ is a white noise sequence whose components are mutually and serially independent.

2.1.2 IRFs 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) through the system of variables y_t .

Formally, an IRF is a difference in conditional expectations:

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

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

If the dynamics of process y_t can be described as a VARMA model, and if y_t is covariance stationary (see Def. ??), then y_t admits the following infinite MA representation (MA(∞)):

$$y_t = \mu + \sum_{h=0}^{\infty} \Psi_h \eta_{t-h}. \quad (2.3)$$

This is also the Wold decomposition of process $\{y_t\}$ (see Theorem ??).

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. (2.3).
- Directly estimate the Ψ_h based on **projection approaches** (see Section ??).
- Approximate the infinite MA representation by estimating a parsimonious type of model, e.g. **VAR(MA) models** (see Section 2.1.4). Once a (Structural) VARMA representation is obtained, Eq. (2.3) is easily deduced. For that, one can use the same recursive algorithm as for univariate processes (see Prop. ??).

Typically, consider the AR(2) case. The first steps of the algorithm 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 .

Let us consider the following VARMA(1,1) model:

$$y_t = \underbrace{\begin{bmatrix} 0.5 & 0.3 \\ -0.4 & 0.7 \end{bmatrix}}_{\Phi_1} y_{t-1} + \underbrace{\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}}_B \eta_t + \underbrace{\begin{bmatrix} 2 & 0 \\ 1 & 0.5 \end{bmatrix}}_{\Theta_1} \underbrace{\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}}_B \eta_{t-1} \quad (2.4)$$

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

```

library(AEC)
distri <- list(type=c("gaussian","gaussian"),df=c(4,4))
n <- length(distri$type) # dimension of y_t
nb.sim <- 30
eps <- simul.distri(distri,nb.sim)
Phi <- array(NaN,c(n,n,1))
Phi[, ,1] <- matrix(c(.5,-.4,.3,.7),2,2)
p <- dim(Phi)[3]
Theta <- array(NaN,c(n,n,1))
Theta[, ,1] <- -matrix(c(2,1,0,.5),2,2)
q <- dim(Theta)[3]
Mu <- rep(0,n)
C <- matrix(c(1,-1,2,1),2,2)
Model <- list(
  Mu = Mu,Phi = Phi,Theta = Theta,C = C,distri = distri
)
Y0 <- rep(0,n)
eta0 <- c(1,0)
res.sim.1 <- simul.VARMA(Model,nb.sim,Y0,eta0,indic.IRF=1)
eta0 <- c(0,1)
res.sim.2 <- simul.VARMA(Model,nb.sim,Y0,eta0,indic.IRF=1)

par(plt=c(.1,.95,.25,.8))
par(mfrow=c(2,2))
plot(res.sim.1$Y[1,],las=1,
      type="l",lwd=3,xlab="",ylab="",
      main=expression(paste("Response of ",y[1,"*","t"]," to a one-unit increase in ",e
abline(h=0,col="grey",lty=3)
plot(res.sim.2$Y[1,],las=1,
      type="l",lwd=3,xlab="",ylab="",
      main=expression(paste("Response of ",y[1,"*","t"]," to a one-unit increase in ",e
abline(h=0,col="grey",lty=3)
plot(res.sim.1$Y[2,],las=1,
      type="l",lwd=3,xlab="",ylab="",
      main=expression(paste("Response of ",y[2,"*","t"]," to a one-unit increase in ",e
abline(h=0,col="grey",lty=3)
plot(res.sim.2$Y[2,],las=1,
      type="l",lwd=3,xlab="",ylab="",
      main=expression(paste("Response of ",y[2,"*","t"]," to a one-unit increase in ",e
abline(h=0,col="grey",lty=3)

```

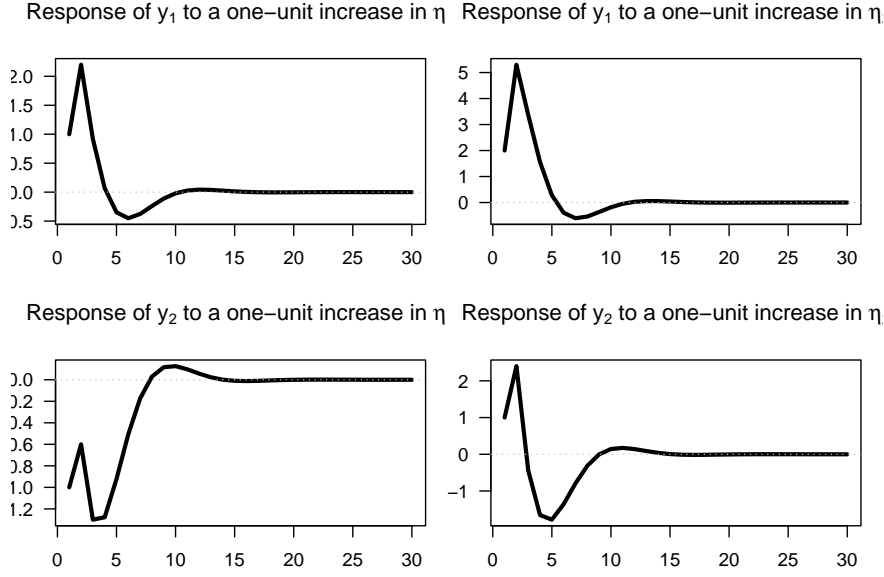


Figure 2.1: Impulse response functions

2.1.3 Covariance-stationary VARMA models

Let's come back to the infinite MA case (Eq. (2.3)):

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

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

$$\sum_{i=0}^{\infty} \|\Psi_i\| < \infty, \quad (2.5)$$

where $\|A\|$ denotes a norm of the matrix A (e.g. $\|A\| = \sqrt{\text{tr}(AA')}$).

This notably implies that if y_t is stationary (and ergodic for the mean), then $\|\Psi_h\| \rightarrow 0$ when h gets large.

What should be satisfied by Φ_k 's and Θ_k 's for a VARMA-based process (Eq. @ref{eq:VARMAstd}) to be stationary? The conditions will be similar to that we had in the univariate case (see Prop. ??). Let us introduce the following

notations:

$$\begin{aligned}
 y_t &= c + \underbrace{\Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p}}_{\text{AR component}} + \underbrace{B\eta_t + \Theta_1 B\eta_{t-1} + \dots + \Theta_q B\eta_{t-q}}_{\text{MA component}} \\
 &\Leftrightarrow \underbrace{(I - \Phi_1 L - \dots - \Phi_p L^p)}_{=\Phi(L)} y_t = c + \underbrace{(I - \Theta_1 L - \dots - \Theta_q L^q)}_{=\Theta(L)} B\eta_t.
 \end{aligned} \tag{2.6}$$

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

$$\Phi = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix} \tag{2.7}$$

lie strictly within the unit circle.

Hence, as was 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.

Based on Eq. (2.6), we have $\mathbb{E}(\Phi(L)y_t) = c$, which gives $\Phi(1)\mathbb{E}(y_t) = c$, 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. (2.3)). We have:

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

(Note that this infinite sum exists as soon as Eq. (2.5) is satisfied.)

Conditional means and autocovariances can also be deduced from Eq. (2.3). For $0 \leq h$ and $0 \leq h_1 \leq h_2$:

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

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

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

Because the 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**):

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

Let us now turn to the estimation of VAR(MA) models.

If there is a MA component, OLS regressions yield biased estimates (even for asymptotically large samples).

Assume y_t follows a VARMA(1,1) model. We have:

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

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 .

Estimation methods of VARMA models will be presented in Section ??.

2.1.4 VAR estimation

This section discusses the estimation of VAR models. (The estimation of SVARMA models is more challenging, see, e.g., Gouriéroux et al. (2020).) Eq. (2.1) 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:

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

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

$$y_t = \Pi' x_t + \varepsilon_t. \quad (2.8)$$

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 2.1 (MLE of a Gaussian VAR). *If y_t follows a $\text{VAR}(p)$ (see Definition 2.1), and if $\varepsilon_t \sim i.i.d. \mathcal{N}(0, \Omega)$, then the ML estimate of Π , denoted by $\hat{\Pi}$ (see Eq. (2.8)), is given by*

$$\hat{\Pi} = \left[\sum_{t=1}^T x_t x_t' \right]^{-1} \left[\sum_{t=1}^T y_t' x_t \right] = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}, \quad (2.9)$$

where \mathbf{X} is the $T \times (np)$ 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:

$$y_{i,t} = \beta_i' x_t + \varepsilon_{i,t}, \quad (2.10)$$

(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. (2.10), i.e.:

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t', \quad \text{with } \hat{\varepsilon}_t = y_t - \hat{\Pi}' x_t. \quad (2.11)$$

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

Proof. See Appendix 3.4.7. □

As stated by Proposition 3.4.8, 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 2.1, we have:

$$\mathbf{b}_i = \beta_i + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon_i. \quad (2.12)$$

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 $\frac{1}{n}\mathbf{X}'\varepsilon_i$ converges to 0. Hence $\mathbf{b}_i \xrightarrow{p} \beta_i$. More precisely:

Proposition 2.2 (Asymptotic distribution of the OLS estimate of β_i). *If y_t follows a VAR model, as defined in Definition 2.1, we have:*

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

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

$$\mathbf{Q} = \begin{bmatrix} 1 & \mu' & \mu' & \dots & \mu' \\ \mu & \gamma_0 + \mu\mu' & \gamma_1 + \mu\mu' & \dots & \gamma_{p-1} + \mu\mu' \\ \mu & \gamma_1 + \mu\mu' & \gamma_0 + \mu\mu' & \dots & \gamma_{p-2} + \mu\mu' \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mu & \gamma_{p-1} + \mu\mu' & \gamma_{p-2} + \mu\mu' & \dots & \gamma_0 + \mu\mu' \end{bmatrix}. \quad (2.13)$$

Proof. See Appendix 3.4.8. □

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 2.3 (Asymptotic distribution of the OLS estimates). *If y_t follows a VAR model, as defined in Definition 2.1, we have:*

$$\sqrt{T} \begin{bmatrix} \text{vec}(\hat{\Pi} - \Pi) \\ \text{vec}(\hat{\Omega} - \Omega) \end{bmatrix} \sim \mathcal{N} \left(0, \begin{bmatrix} \Omega \otimes \mathbf{Q}^{-1} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \right), \quad (2.14)$$

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 Hamilton (1994), Appendix of Chapter 11. □

Naturally, in practice, Ω is replaced by $\hat{\Omega}$, Q is replaced with $\hat{Q} = \frac{1}{T} \sum_{t=p}^T x_t x_t'$ 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. The maximum value achieved by the MLE is

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

The last term is:

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

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

$$\log \mathcal{L}(Y_T; \hat{\Pi}, \hat{\Omega}) = -(Tn/2) \log(2\pi) + (T/2) \log |\hat{\Omega}^{-1}| - Tn/2. \quad (2.15)$$

Assume that we want to test the null hypothesis that a set of variables follows a VAR(p_0) against the alternative specification of p_1 ($> p_0$).

Let us denote by \hat{L}_0 and \hat{L}_1 the maximum log-likelihoods obtained with p_0 and p_1 lags, respectively.

Under the null hypothesis ($H_0: p = p_0$), we have:

$$2(\hat{L}_1 - \hat{L}_0) = T(\log |\hat{\Omega}_1^{-1}| - \log |\hat{\Omega}_0^{-1}|) \sim \chi^2(n^2(p_1 - p_0)).$$

What precedes can be used to help determine the appropriate number of lags to use in the specification. In a VAR, using too many lags consumes numerous degrees of freedom: with p lags, each of the n equations in the VAR contains $n \times p$ coefficients plus the intercept term. Adding lags improve in-sample fit, but is likely to result in over-parameterization and affect the **out-of-sample** prediction performance.

To select appropriate lag length, **selection criteria** can be used (see Definition ??). In the context of VAR models, using Eq. (2.15), we have:

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

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

2.1.5 Block exogeneity and Granger causality

Block exogeneity

Let's decompose y_t into two subvectors $y_t^{(1)}$ ($n_1 \times 1$) and $y_t^{(2)}$ ($n_2 \times 1$), with $y_t' = [y_t^{(1)'} , y_t^{(2)'}]$ (and therefore $n = n_1 + n_2$), such that:

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

Using, e.g., a likelihood ratio test (see Def. ??), one can easily test for block exogeneity of $y_t^{(2)}$ (say). The null assumption can be expressed as $\Phi^{(2,1)} = 0$ and $\Sigma^{(2,1)} = 0$.

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:

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

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

$$\mathbb{E}[y_{2,t} | \mathcal{J}_{2,t-1}] \neq \mathbb{E}[y_{2,t} | \mathcal{J}_{t-1}].$$

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

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

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

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

Therefore the hypothesis testing is

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

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 $\mathcal{F}(p, T - 2p - 1)$.

Note that $pF \xrightarrow{T \rightarrow \infty} \chi^2(p)$. Therefore, for large samples and under H_0 :

$$F \sim \chi^2(p)/p.$$

2.1.6 Identification problem and standard identification techniques

The Identification Issue

In Section 2.1.4, 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 the Φ_k 's, not of Ω the Φ_k 's (see Section 2.1.2). We have $\Omega = BB'$, but this is not sufficient to recover 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 $n = 2$:

$$\begin{aligned} \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix} &= \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{bmatrix} \\ \Leftrightarrow \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix} &= \begin{bmatrix} b_{11}^2 + b_{12}^2 & b_{11}b_{21} + b_{12}b_{22} \\ b_{11}b_{21} + b_{12}b_{22} & b_{21}^2 + b_{22}^2 \end{bmatrix}. \end{aligned}$$

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.
 - Easy to implement: the appropriate entries of B are set to 0.
 - Particular case: **Cholesky, or recursive approach**.

- Examples: 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: 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)).

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

$$\begin{aligned}
 g_t &= \bar{g} - \lambda(i_{t-1} - \mathbb{E}_{t-1}\pi_t) + \underbrace{\sigma_d \eta_{d,t}}_{\text{demand shock}} && \text{(IS curve)} \\
 \Delta\pi_t &= \beta(g_t - \bar{g}) + \underbrace{\sigma_\pi \eta_{\pi,t}}_{\text{cost push shock}} && \text{(Phillips curve)} \\
 i_t &= \rho i_{t-1} + [\gamma_\pi \mathbb{E}_t \pi_{t+1} + \gamma_g (g_t - \bar{g})] + \underbrace{\sigma_{mp} \eta_{mp,t}}_{\text{Mon. Pol. shock}} && \text{(Taylor rule),}
 \end{aligned} \tag{2.16}$$

where:

$$\eta_t = \begin{bmatrix} \eta_{\pi,t} \\ \eta_{d,t} \\ \eta_{mp,t} \end{bmatrix} \sim i.i.d. \mathcal{N}(0, I). \tag{2.17}$$

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 (2.16) could be rewritten in the form:

$$\begin{bmatrix} d_t \\ \pi_t \\ i_t \end{bmatrix} = \Phi(L) \begin{bmatrix} d_{t-1} \\ \pi_{t-1} \\ i_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & \bullet & 0 \\ \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet \end{bmatrix}}_{\substack{=B \\ =\varepsilon_t}} \eta_t \tag{2.18}$$

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 (up to the signs of its columns) as soon as $\Omega = BB'$ is known.

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.

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) the first two endogenous variables (say, $y_{n_1,t}$ 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 , that is a lower triangular matrix B that is such that:

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

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: the industrial production, the consumer price index, a commodity price index, the short-term rate, monetary aggregate and the effective exchange rate (except for the US). This ordering implies that monetary policy reacts to the shocks affecting the first three variables but that the latter react to monetary policy shocks with a one-period lag only.

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

$$\begin{bmatrix} \varepsilon_{S,t} \\ \varepsilon_{r,t} \\ \varepsilon_{F,t} \end{bmatrix} = \begin{bmatrix} B_{SS} & 0 & 0 \\ B_{rS} & B_{rr} & 0 \\ B_{FS} & B_{Fr} & B_{FF} \end{bmatrix} \begin{bmatrix} \eta_{S,t} \\ \eta_{r,t} \\ \eta_{F,t} \end{bmatrix},$$

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 $\text{Var}(\eta_{r,t}) = 1$, 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 $\eta_{r,t}$.

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(AEC)
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, # This is used in the parameter
                                confidence.interval = 0.90, # expressed in pp.
                                indic.plot = 1 # Plots are displayed if = 1.
)
```

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 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 can be easily formulated as a function of B and of the matrices Φ_i when y_t (including ΔGDP_t) follows a VAR process.

Without loss of generality, we will only consider the VAR(1) case. Indeed, one can always write a VAR(p) as a VAR(1). To see that, stack the last p values of

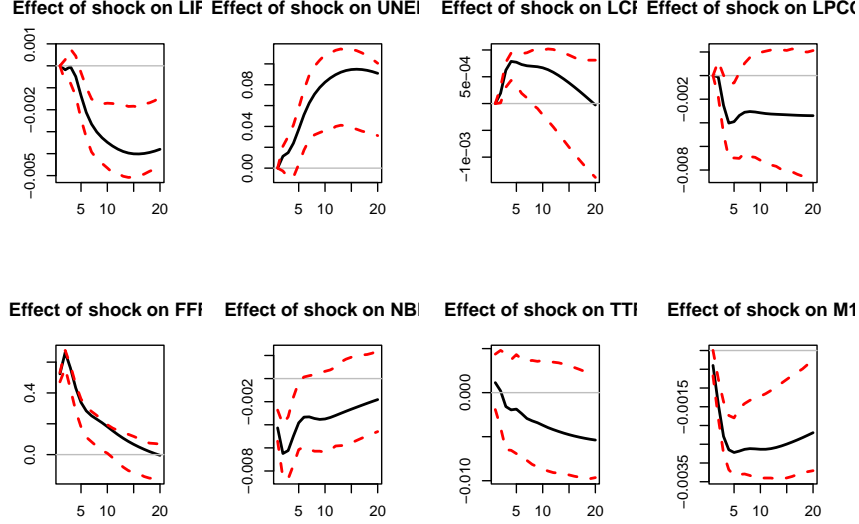


Figure 2.2: Response to a monetary-policy shock. Identification approach of Christiano, Eichenbaum and Evans (1996).

vector y_t in vector $y_t^* = [y_t', \dots, y_{t-p+1}']'$; Eq. (2.1) can then be rewritten in its companion form:

$$y_t^* = \underbrace{\begin{bmatrix} c \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{=c^*} + \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix}}_{=\Phi} y_{t-1}^* + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{=\varepsilon_t^*}, \quad (2.19)$$

where matrices Φ and $\Omega^* = \text{Var}(\varepsilon_t^*)$ are of dimension $np \times np$; Ω^* is filled with zeros, except the $n \times n$ upper-left block that is equal to $\Omega = \text{Var}(\varepsilon_t)$. (Matrix Φ had been introduced in Eq. (2.7).)

Focusing on the VAR(1) case:

$$y_t = c + \Phi y_{t-1} + \varepsilon_t \quad (2.20)$$

$$\begin{aligned} &= c + \varepsilon_t + \Phi(c + \varepsilon_{t-1}) + \dots + \Phi^k(c + \varepsilon_{t-k}) + \dots \\ &= \mu + \varepsilon_t + \Phi \varepsilon_{t-1} + \dots + \Phi^k \varepsilon_{t-k} + \dots \end{aligned} \quad (2.21)$$

$$= \mu + B \eta_t + \Phi B \eta_{t-1} + \dots + \Phi^k B \eta_{t-k} + \dots, \quad (2.22)$$

The sequence of shocks $\{\eta_t\}$ determines the sequence $\{y_t\}$. What if $\{\eta_t\}$ is replaced by $\{\tilde{\eta}_t\}$, 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 $t \geq s$):

$$\begin{aligned} (\tilde{y}_t - y_t) + (\tilde{y}_{t-1} - y_{t-1}) + \dots + (\tilde{y}_s - y_s) &= \\ (Id + \Phi + \Phi^2 + \dots + \Phi^{t-s}) B \gamma. \end{aligned} \quad (2.23)$$

Consider a shock on $\eta_{1,t}$, with a magnitude of 1. This shock corresponds to $\gamma = [1, 0, \dots, 0]'$. Given Eq. (2.23), the long-run cumulative effect of this shock on the endogenous variables is given by:

$$\underbrace{(Id + \Phi + \dots + \Phi^k + \dots)}_{=(Id - \Phi)^{-1}} B \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

that is the first column of $\Theta \equiv (Id - \Phi)^{-1} B$.

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

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

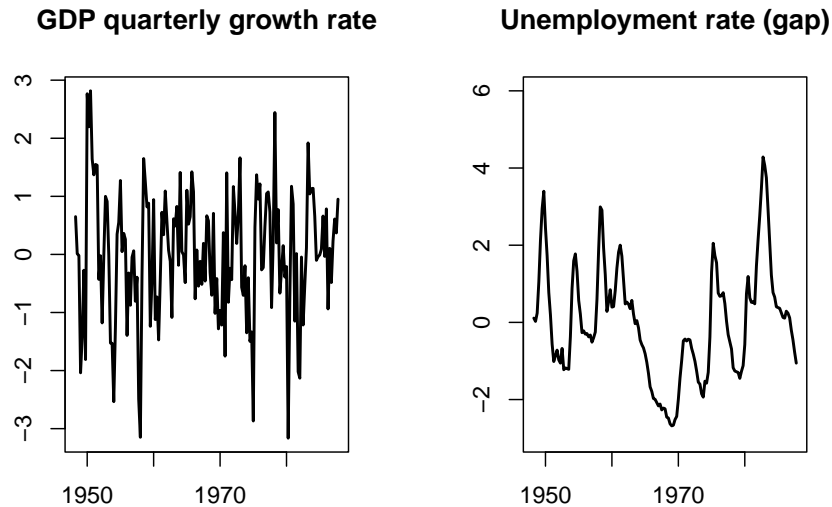
where Θ_{ij} is the element (i, j) of Θ .

Blanchard and Quah (1989) implement 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. The 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(AEC)
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)
```

¹The 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).



Estimate a reduced-form VAR(8) model:

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

Now, let us define a loss function (`loss`) that is equal to zero if (a) $BB' = \Omega$ and (b) the element (1,1) of ΘB 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 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))
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")
```

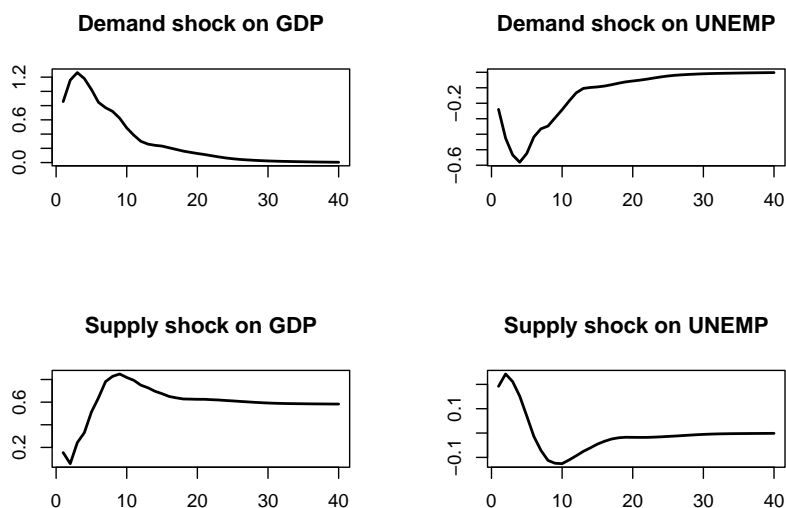


Figure 2.3: IRF of GDP and unemployment to demand and supply shocks.

2.1.7 Sign restrictions

To identify the structural shocks, we need to find a matrix B that satisfies $\Omega_\varepsilon = BB'$ (with $\Omega = \text{Var}(\varepsilon_t)$) and other restrictions. Indeed, $\Omega_\varepsilon = BB'$ is not sufficient to identify B since, if we take any orthogonal matrix Q (see Def. 2.3), then $\mathcal{P} = BQ$ also satisfies $\Omega = \mathcal{P}\mathcal{P}'$.

Definition 2.3 (Orthogonal matrix). An orthogonal matrix Q is a matrix such that $QQ' = I$, i.e., all columns (rows) of Q are orthogonal and unit vectors:

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

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

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:

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

(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.

This 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.

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:²

- 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.

²Standard 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.

```

library(AEC);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[[1]][1,3] <- 1
sign.restrictions[[1]][2,5] <- -1
sign.restrictions[[1]][3,6] <- 1
horizon[[1]] <- 1:5
res.svar.signs <- svar.signs(y,p=3,
                             nb.shocks = 1, #number of identified shocks
                             nb.periods.IRF = 20,
                             bootstrap.replications = 1, # = 0 if no bootstrap
                             confidence.interval = 0.80, # 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 recur.
)

```

```
nb.rotations <- res.svar.signs$xx
```

A remarkable characteristic of sign restrictions is that it does not lead to a unique IRF. In that sense, we say that it is set-identified, not point-identified. By drawing the rotation matrices (Q) from a distribution and taking the median, we adopt an agnostic approach: “pure sign restriction approach”.

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*:

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

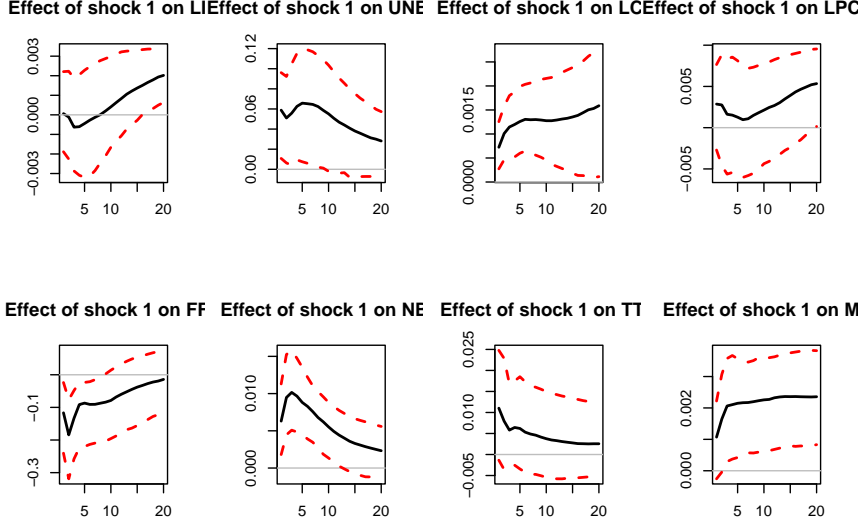


Figure 2.4: IRF associated with a monetary policy shock; sign-restriction approach.

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

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

The problem is solved numerically.

2.1.8 Forecast error variance maximization

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

Consider a process $\{y_t\}$ that admits the infinite MA representation of Eq. (2.3). Let Q be an orthogonal matrix, an alternative decomposition is:

$$y_t = \sum_{h=0}^{+\infty} \Psi_h \underbrace{\eta_{t-h}}_{Q\tilde{\eta}_{t-h}} = \sum_{h=0}^{+\infty} \underbrace{\Psi_h Q}_{\tilde{\Psi}_h} \tilde{\eta}_{t-h} = \sum_{h=0}^{+\infty} \tilde{\Psi}_h \tilde{\eta}_{t-h}, \quad (2.24)$$

where $\tilde{\eta}_{t-h} = Q'\eta_{t-h}$ are the white-noise shocks associated with the new MA representation. (They also satisfy $\text{Var}(\tilde{\eta}_t) = Id$.)

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

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

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 $h \in [\underline{h}, \bar{h}]$.

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

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

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:

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

where

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

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 $S q_1 = \lambda q_1$ (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 $\sigma^2(\underline{h}, \bar{h}, q_1) = \lambda$. 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 XXXXXXXX). 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: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$, then $\sigma^2(\underline{h}, \bar{h}, q_1)$ is maximized for $q_1 = p_1$, where p_1 is the first column of \mathcal{P} .

2.1.9 Identification based on non-normality of the shocks

In this section, we show that the non-identification of the structural shocks (η_t) is very specific to the Gaussian case. We propose consistent estimation approaches in the context of non-Gaussian shocks. In a first part, we focus on non-Gaussian SVAR models; in a second part, we discuss the case of non-Gaussian SVARMA models.

Non-Gaussian SVAR models

The estimation of (S)VARs is more common than that of (S)VARMA. (Simpler estimation of Φ_k 's). We have seen in what precedes that, in the Gaussian case, we cannot identify B in the Gaussian case. That is, even if we observe an infinite number of i.i.d. $B\eta_t$, we cannot recover B if 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 $BQ\eta_t \sim \mathcal{N}(0, \Omega)$.

Example: Bivariate Gaussian case (Eq. (2.4), with $\Theta_1 = 0$)

$\begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix} \sim \mathcal{N}(0, Id)$, $B = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$ and $Q = \begin{bmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{bmatrix}$ (rotation).

\Rightarrow Distribution of $B\eta_t$ versus that of $BQ\eta_t$?

```
theta.angle <- pi/3
Q <- matrix(c(cos(theta.angle),sin(theta.angle),-sin(theta.angle),cos(theta.angle)),2,2)
#nb.sim <- 10^4
nb.sim <- 10^2
distri.1 <- list(type=c("gaussian"),name="Panel (a) Gaussian",name.4.table="Gaussian")
distri.2 <- list(type=c("mixt.gaussian"),mu=0,sigma=5,p=.03,name="Panel (b) Mixture of")
distri.3 <- list(type=c("student"),df=c(5),name="Panel (c) Student (df: 5)",name.4.table="Student")
distri.4 <- list(type=c("student"),df=c(10),name="Panel (d) Student (df: 10)",name.4.table="Student")
x.lim <- c(-7,7)
y.lim <- c(-5,5)
nb.points <- 100
x.points <- seq(x.lim[1],x.lim[2],length.out=nb.points)
y.points <- x.points
all.x <- c(matrix(x.points,nb.points,nb.points))
all.y <- c(t(matrix(x.points,nb.points,nb.points)))
eps <- cbind(all.x,all.y)
par(plt=c(.25,.9,.25,.8))
eta.1 <- simul.distri(distri.1,nb.sim)
eta.2 <- simul.distri(distri.1,nb.sim)
epsilon.C <- cbind(eta.1,eta.2) %*% t(C)
epsilon.CQ <- cbind(eta.1,eta.2) %*% t(C %*% Q)
Model$distri <- list(type=c("gaussian","gaussian"),df=c(NaN,NaN))
par(mfrow=c(1,2))
plot(epsilon.C[,1],epsilon.C[,2],pch=19,
     xlim=x.lim,ylim=y.lim,col="#00000044",
     xlab=expression(epsilon[1]),
     ylab=expression(epsilon[2]),cex.lab=1.6,cex.main=1.6,
     main=expression(paste("Distribution of ",epsilon[t]," = ",B,eta[t],sep="")))
z <- matrix(exp(g(eps,Model)),nb.points,nb.points)
```



```

par(new=TRUE)
max.z <- max(z)
levels <- c(.01,.1,.3,.6,.9)*max.z
contour(x.points,y.points,z,levels=levels,xlim=x.lim,ylim=y.lim,col="red",lwd=2)
plot(epsilon.CQ[,1],epsilon.CQ[,2],pch=19,
      xlim=x.lim,ylim=y.lim,col="#00000044",
      xlab=expression(epsilon[1]),
      ylab=expression(epsilon[2]),cex.lab=1.6,cex.main=1.6,
      main=expression(paste("Distribution of ",epsilon[t]," = ",BQ,eta[t],sep="")))
z <- matrix(exp(g(eps,Model)),nb.points,nb.points)
par(new=TRUE)
max.z <- max(z)
levels <- c(.01,.1,.3,.6,.9)*max.z
contour(x.points,y.points,z,levels=levels,xlim=x.lim,ylim=y.lim,col="red",lwd=2)

```

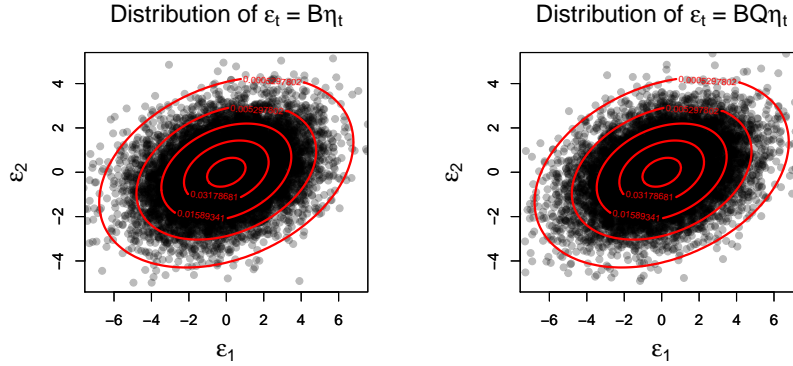


Figure 2.5: XXXX.

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.

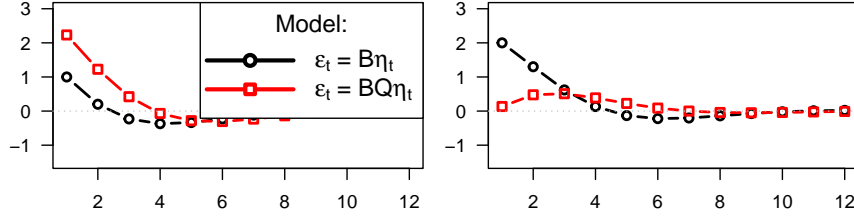
Rigobon (2003) Normandin and Phaneuf (2004) Lanne and Lutkepohl (2008)

Example: Bivariate Gaussian + Student (5) case:

$$\eta_{1,t} \sim \mathcal{N}(0,1), \eta_{2,t} \sim t(5), B = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \text{ and } Q = \begin{bmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{bmatrix}.$$

Distribution of $B\eta_t$ versus that of $BQ\eta_t$?

Response of y_1 to a one-unit increase in η Response of y_1 to a one-unit increase in η :



Response of y_2 to a one-unit increase in η Response of y_2 to a one-unit increase in η :

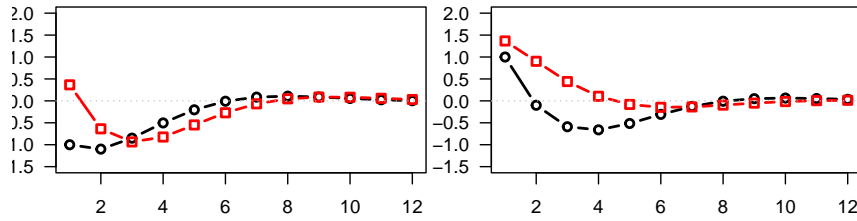


Figure 2.6: XXXX.

```
theta.angle <- pi/3
Q <- matrix(c(cos(theta.angle),sin(theta.angle),-sin(theta.angle),cos(theta.angle)),2,2)
#nb.sim <- 10^4
nb.sim <- 10^2
distri.1 <- list(type=c("gaussian"),name="Panel (a) Gaussian",name.4.table="Gaussian")
distri.2 <- list(type=c("mixt.gaussian"),mu=0,sigma=5,p=.03,name="Panel (b) Mixture of
distri.3 <- list(type=c("student"),df=c(5),name="Panel (c) Student (df: 5)",name.4.tab
distri.4 <- list(type=c("student"),df=c(10),name="Panel (d) Student (df: 10)",name.4.t
x.lim <- c(-7,7)
y.lim <- c(-5,5)
nb.points <- 100
x.points <- seq(x.lim[1],x.lim[2],length.out=nb.points)
y.points <- x.points
all.x <- c(matrix(x.points,nb.points,nb.points))
all.y <- c(t(matrix(x.points,nb.points,nb.points)))
eps <- cbind(all.x,all.y)
par(plt=c(.25,.9,.25,.8))
eta.1 <- simul.distri(distri.1,nb.sim)
eta.2 <- simul.distri(distri.3,nb.sim)
epsilon.C <- cbind(eta.1,eta.2) %*% t(C)
epsilon.CQ <- cbind(eta.1,eta.2) %*% t(C %*% Q)
Model$distri <- list(type=c("gaussian","student"),df=c(NaN,5))
```

```

par(mfrow=c(1,2))
plot(epsilon.C[,1],epsilon.C[,2],pch=19,
      xlim=x.lim,ylim=y.lim,col="#00000044",
      xlab=expression(epsilon[1]),
      ylab=expression(epsilon[2]),cex.lab=1.6,cex.main=1.6,
      main=expression(paste("Distribution of ",epsilon[t]," = ",B,eta[t],sep="")))
z <- matrix(exp(g(eps,Model)),nb.points,nb.points)
par(new=TRUE)
max.z <- max(z)
levels <- c(.01,.1,.3,.6,.9)*max.z
contour(x.points,y.points,z,levels=levels,xlim=x.lim,ylim=y.lim,col="red",lwd=2)
plot(epsilon.CQ[,1],epsilon.CQ[,2],pch=19,
      xlim=x.lim,ylim=y.lim,col="#00000044",
      xlab=expression(epsilon[1]),
      ylab=expression(epsilon[2]),cex.lab=1.6,cex.main=1.6,
      main=expression(paste("Distribution of ",epsilon[t]," = ",BQ,eta[t],sep="")))
z <- matrix(exp(g(eps,Model)),nb.points,nb.points)
par(new=TRUE)
max.z <- max(z)
levels <- c(.01,.1,.3,.6,.9)*max.z
contour(x.points,y.points,z,levels=levels,xlim=x.lim,ylim=y.lim,col="red",lwd=2)

```

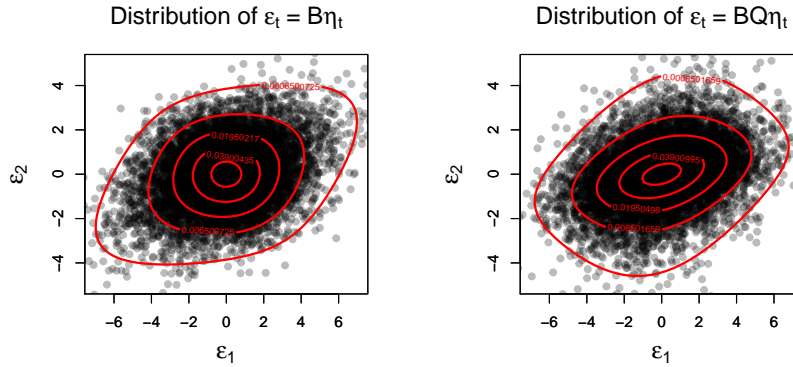


Figure 2.7: XXXX.

NB: In both cases, we have $\text{Var}(\varepsilon_t) = BB'$.

Example: Bivariate Student (5) case

$$\eta_{1,t} \sim t(5), \eta_{2,t} \sim t(5), B = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \text{ and } Q = \begin{bmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{bmatrix}.$$

\Rightarrow Distribution of $B\eta_t$ versus that of $BQ\eta_t$?

```

theta.angle <- pi/3
Q <- matrix(c(cos(theta.angle),sin(theta.angle),-sin(theta.angle),cos(theta.angle)),2,2)
#nb.sim <- 10^4
nb.sim <- 10^2
distri.1 <- list(type=c("gaussian"),name="Panel (a) Gaussian",name.4.table="Gaussian")
distri.2 <- list(type=c("mixt.gaussian"),mu=0,sigma=5,p=.03,name="Panel (b) Mixture of
distri.3 <- list(type=c("student"),df=c(5),name="Panel (c) Student (df: 5)",name.4.tab
distri.4 <- list(type=c("student"),df=c(10),name="Panel (d) Student (df: 10)",name.4.t
x.lim <- c(-7,7)
y.lim <- c(-5,5)
nb.points <- 100
x.points <- seq(x.lim[1],x.lim[2],length.out=nb.points)
y.points <- x.points
all.x <- c(matrix(x.points,nb.points,nb.points))
all.y <- c(t(matrix(x.points,nb.points,nb.points)))
eps <- cbind(all.x,all.y)
par(plt=c(.25,.9,.25,.8))
eta.1 <- simul.distri(distri.3,nb.sim)
eta.2 <- simul.distri(distri.3,nb.sim)
epsilon.C <- cbind(eta.1,eta.2) %*% t(C)
epsilon.CQ <- cbind(eta.1,eta.2) %*% t(C %*% Q)
Model$distri <- list(type=c("student","student"),df=c(5,5))
par(mfrow=c(1,2))
plot(epsilon.C[,1],epsilon.C[,2],pch=19,
      xlim=x.lim,ylim=y.lim,col="#00000044",
      xlab=expression(epsilon[1]),
      ylab=expression(epsilon[2]),cex.lab=1.6,cex.main=1.6,
      main=expression(paste("Distribution of ",epsilon[t]," = ",B,eta[t],sep="")))
z <- matrix(exp(g(eps,Model)),nb.points,nb.points)
par(new=TRUE)
max.z <- max(z)
levels <- c(.01,.1,.3,.6,.9)*max.z
contour(x.points,y.points,z,levels=levels,xlim=x.lim,ylim=y.lim,col="red",lwd=2)
plot(epsilon.CQ[,1],epsilon.CQ[,2],pch=19,
      xlim=x.lim,ylim=y.lim,col="#00000044",
      xlab=expression(epsilon[1]),
      ylab=expression(epsilon[2]),cex.lab=1.6,cex.main=1.6,
      main=expression(paste("Distribution of ",epsilon[t]," = ",BQ,eta[t],sep="")))
z <- matrix(exp(g(eps,Model)),nb.points,nb.points)
par(new=TRUE)
max.z <- max(z)
levels <- c(.01,.1,.3,.6,.9)*max.z
contour(x.points,y.points,z,levels=levels,xlim=x.lim,ylim=y.lim,col="red",lwd=2)

```

NB: In both cases, we have $\mathbb{V}ar(\varepsilon_t) = BB'$.

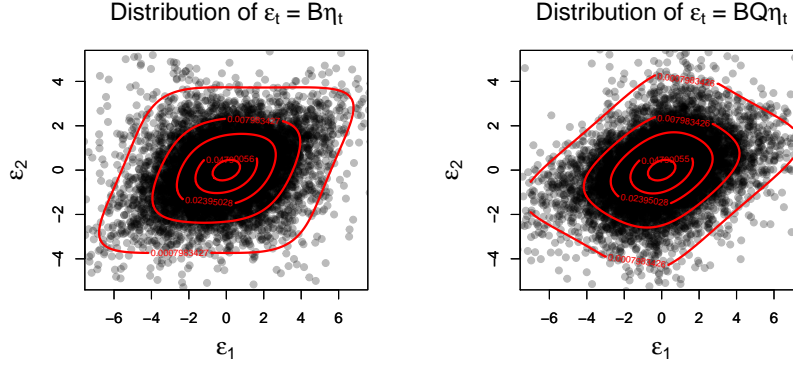


Figure 2.8: XXXX.

Relationship with Signal Processing Literature

Task to be performed:

you observe n linear combinations of n **independent signals**; you want to estimate these linear combinations (\Leftrightarrow recover independent signals).

Without loss of generality, we can assume that $BB' = Id$ (i.e. B is orthogonal)
(If not the case, i.e. if $\text{Var}(\varepsilon_t) = \Omega \neq Id$, pre-multiply the data by $\Omega^{-1/2}$).

Classical signal processing problem: Independent Component Analysis (ICA)

Find B such that $\varepsilon_t = B\eta_t$ (or $\eta_t = B'\varepsilon_t$) given that i. you observe the ε_t 's, ii. The components of η_t are independent, iii. $BB' = Id$ (B is orthogonal).

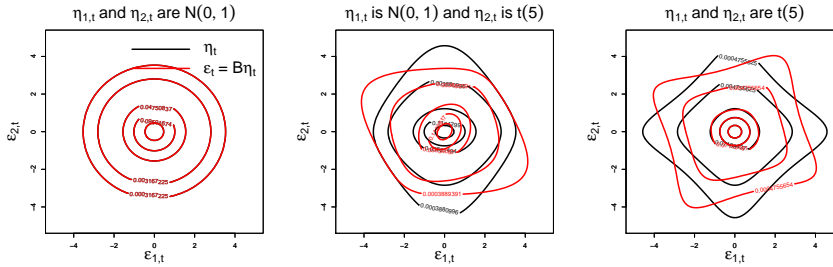


Figure 2.9: XXXX.

In all cases, we have $\text{Var}(\varepsilon_t) = \text{Var}(\eta_t) = Id$. Assume you observe the distribution of ε_t (in red). How to rotate ε_t to get obtain *independent signals* ($\eta_t = B'\varepsilon_t$)?

NB: $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are not independent.

For instance: We have $\mathbb{E}(\varepsilon_{2,t}|\varepsilon_{1,t} > 4) < 0$ (whereas $\mathbb{E}(\eta_{2,t}|\eta_{1,t} > 4) = 0$).

Hypothesis 2.1. We have:

- i. The shocks η_t are i.i.d. with $\mathbb{E}(\eta_t) = 0$ and $\text{Var}(\eta_t) = Id$.
- ii. The components $\eta_{1,t}, \dots, \eta_{n,t}$ are mutually independent.
- iii We have

$$Y_t = B_0 \eta_t,$$

with $\mathbb{E}(Y_t) = Id$ (i.e. B_0 is orthogonal).

Theorem 2.1 (Eriksson, Koivunen (2004)). *If Hypothesis 2.1 is satisfied and if at most one of the components of η is Gaussian, then matrix B_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 .*

The PML approach

We introduce a set of p.d.f. $g_i(\eta_i), i = 1, \dots, n$, and consider the **pseudo log-likelihood function**:

$$\log \mathcal{L}_T(B) = \sum_{t=1}^T \sum_{i=1}^n \log g_i(b'_i Y_t), \quad (2.25)$$

where b_i is the i^{th} column of matrix B (or b'_i is the i^{th} row of B^{-1} since $B^{-1} = B'$).

The log-likelihood function (2.25) is computed as if the errors $\eta_{i,t}$ had the p.d.f. $g_i(\eta_i)$.

A **pseudo maximum likelihood (PML) estimator** of matrix B maximizes the pseudo log-likelihood function:

$$\widehat{B}_T = \arg \max_B \sum_{t=1}^T \sum_{i=1}^n \log g_i(b'_i Y_t), \quad (2.26)$$

s.t. $B' B = Id$.

The restrictions $B' B = Id$ can be eliminated by parameterizing B .

Cayley's representation:

Any orthogonal matrix with no eigenvalue equal to -1 can be written as

$$B(A) = (Id + A)(Id - A)^{-1}, \quad (2.27)$$

where A is a skew symmetric (or antisymmetric) matrix, such that $A' = -A$.

There is a one-to-one relationship with A , since:

$$A = (B(A) + Id)^{-1}(B(A) - Id). \quad (2.28)$$

PML estimator of matrix B : $\widehat{B}_T = B(\hat{A}_T)$, where:

$$\hat{A}_T = \arg \max_{a_{i,j}, i > j} \sum_{t=1}^T \sum_{i=1}^n \log g_i[b_i(A)'Y_t]. \quad (2.29)$$

Asymptotic properties of the PML approach

Hypothesis 2.2. We have:

- i. The functions $\log g_i$, $i = 1, \dots, n$, are twice continuously differentiable.
- ii. $\sup_{B: B'B = Id} |\sum_{i=1}^n \log g_i(b'_i y)| \leq h(y)$, where $\mathbb{E}_0[h(Y)] < \infty$.

Hypothesis 2.3 (Identification from the asymptotic FOC). The only solutions of the system of equations:

$$\begin{cases} \mathbb{E}_0 \left[b'_j Y_t \frac{d \log g_i}{d \eta}(b'_i Y_t) \right] = 0, & i \neq j, \\ B'B = Id, \end{cases}$$

are the elements of $\mathcal{P}_0 \equiv \mathcal{P}(B_0)$, which is the set of matrices obtained by permutation and sign change of the columns of B_0 .

Hypothesis 2.4 (Local concavity). The asymptotic objective function is locally concave in a neighbourhood of a matrix B of $\mathcal{P}(B_0)$, which is the case if and only if

$$\mathbb{E}_0 \left[\frac{d^2 \log g_i(\eta_{i,t})}{d\eta^2} + \frac{d^2 \log g_j(\eta_{j,t})}{d\eta^2} - \eta_{j,t} \frac{d \log g_j(\eta_{j,t})}{d\eta} - \eta_{i,t} \frac{d \log g_i(\eta_{i,t})}{d\eta} \right] < 0, \forall i < j,$$

where $\eta_{i,t}$ is the i^{th} component of the η_t associated with this particular element B of $\mathcal{P}(B_0)$.

This condition is in particular satisfied under the following set of conditions: derived in Hyvarinen (1997) XXX

$$\mathbb{E}_0 \left[\frac{d^2 \log g_i(\eta_{i,t})}{d\eta^2} - \eta_{i,t} \frac{d \log g_i(\eta_{i,t})}{d\eta} \right] < 0, \quad i = 1, \dots, n. \quad (2.30)$$

Hyperbolic secant and the subgaussian distributions (see table on next slide): either one, or the other satisfy the inequality (2.30) [Hyvarinen, Karhunen, Oja, 2001 XXXX].

XXXXXXXXXXXXXXXXXX

Note: Except for the Gaussian distribution, we have $E[d^2 \log g(X)/d\varepsilon^2 - X d \log g(X)/d\varepsilon] < 0$ (i.e. Assumption 4 is satisfied) when these pseudo distributions coincide to the distribution of X . The subGaussian distribution is a mixture of Gaussian distributions: X is drawn from this distribution if it is equal to $BY - (1 - B)Y$, where B is drawn from a Bernoulli distribution of parameter $1/2$ and $Y \sim \mathcal{N}(\sqrt{(\pi - 2)/\pi}, 2/\pi)$.

Under Hypotheses ??-??, the PML estimator \widehat{B}_T of B_0 is consistent (in \mathcal{P}_0) and asymptotically normal, with speed of convergence $1/\sqrt{T}$.

The asymptotic variance-covariance matrix of $\text{vec}\sqrt{T}(\widehat{B}_T - B_0)$ is $A^{-1} \begin{bmatrix} \Gamma & 0 \\ 0 & 0 \end{bmatrix} (A')^{-1}$, where matrices A and Γ are detailed in Gouri'eroux, Monfort and Renne (2017) XXX.

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

The previous proposition can be exploited to build a test whose null hypothesis is:

H_0 : B belongs to \mathcal{P}_0 , where \mathcal{P}_0 is the set of orthogonal matrices obtained by permuting and changing the signs of the columns of a given orthogonal matrix B_0 .

Application: Structural VAR model

Three dependent variables: inflation (π_t), economic activity (z_t) and the nominal short-term interest rate (r_t).

Structural shocks are posited monetary-policy, demand and supply shocks.

W_t : 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:

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

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

U.S. data. 1959:IV to 2015:I at the quarterly frequency ($T = 224$). Source: Federal Reserve Economic Database (FRED).

Two different measures of economic activity are considered: the output gap and the unemployment gap. Change in the log of oil prices added as an exogenous variable (x_t).

μ , Φ_i , Θ and Σ are consistently estimated by OLS.

Jarque-Bera tests support the hypothesis of non-normality for all residuals.

We want to estimate the orthogonal matrix B such that $u_t = SB\eta_t$, where

- S results from the Cholesky decomposition of Σ and
- the components of η_t are independent, zero-mean with unit variance.

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

$$\hat{S}_T^{-1} \underbrace{[y_t - a(W_t; \hat{\theta}_T)]}_{\text{VAR residuals}}.$$

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

Pseudo density functions: Distinct and asymmetric mixtures of Gaussian distributions.

Once B has been estimated, it remains to associate the structural shocks (monetary-policy, supply or demand) with the different components of η_t .

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

Important: This method does not rely on the assumption that specific impacts (e.g. contemporaneous or long-run) are null.

Comparison of the previous IRFs with those stemming from “recursive” identification approaches based on specific short-run restrictions (SRRs).

SRRs approach (Section,??) are based on the assumptions that

- a. $Cov(\eta_t) = Id$,
- b. the k^{th} structural shock does not contemporaneously affects the first $k-1$ endogenous variables and
- c. the contemporaneous effect of the k^{th} structural shock on the k^{th} dependent variable is positive.

Under these assumptions, the structural shocks are given by $S^{-1}u_t$.

SRR approaches assume –potentially wrongly– that the contemporaneous impacts of some structural shocks on given variables are null.

The null hypothesis of these tests is $H_0 = (P \in \mathcal{P}(Id))$. The null hypothesis H_0 stating that the true value of B belongs to \mathcal{P}_0 is not standard since it is a finite union of simple hypotheses $H_{0,j} = (B = B_{j,0})$.

First testing procedure:

- Define the Wald statistics $\hat{\xi}_{j,T}$, $j \in J$:

$$\hat{\xi}_{j,T} = T[vec\hat{B}_T - vecB_{j,0}]' \hat{A}_T' \begin{bmatrix} \hat{\Omega}_T^{-1} & 0 \\ 0 & 0 \end{bmatrix} \hat{A}_T [vec\hat{B}_T - vecB_{j,0}], \quad (2.31)$$

\hat{A}_T and $\hat{\Omega}_T$ being consistent estimators of the matrices A and Ω .

Since the dimension of the asymptotic distribution of $\sqrt{T}[vec\hat{B}_T - vecB_{j,0}]$ is $\frac{1}{2}n(n-1)$, the asymptotic distribution of $\hat{\xi}_{j,T}$ under $H_{0,j}$ is $\chi^2(\frac{1}{2}n(n-1))$.

- Define $\hat{\xi}_T = \min_{j \in J} \hat{\xi}_{j,T}$ as the test statistic for H_0 .
- Under H_0 , \hat{B}_T converges to $B_{j_0,0}$ (say).
- By the asymptotic properties of the Wald statistics for simple hypotheses:

$$\hat{\xi}_{j_0,T} \xrightarrow{D} \chi^2\left(\frac{n(n-1)}{2}\right) \quad \text{and} \quad \hat{\xi}_{j,T} \rightarrow \infty, \text{ if } j \neq j_0. \quad (2.32)$$

Under the null hypothesis, $\hat{\xi}_T = \min_j \hat{\xi}_{j,T}$ is asymptotically equal to $\hat{\xi}_{j_0,T}$ and its asymptotic distribution, $\chi^2(\frac{1}{2}n(n-1))$, does not depend on j_0 . Therefore $\hat{\xi}_T$ is asymptotically a pivotal statistic for the null hypothesis H_0 and the test of critical region $\hat{\xi}_T \geq \chi_{1-\alpha}^2(\frac{1}{2}n(n-1))$ is of asymptotic level α and is consistent.

Second testing procedure

Define $B_{0,T} = \underset{B \in \mathcal{P}_0}{\text{Argmin}} d(\hat{B}_T, B)$ where d is any distance, for instance the Euclidean one.

Under the null hypothesis H_0 : ($B \in \mathcal{P}_0$), \hat{B}_T converges almost surely to an element of \mathcal{P}_0 denoted by $B_{j_0,0}$ and it is also the case for $B_{0,T}$ since, asymptotically, we have $B_{0,T} = B_{j_0,0}$.

Moreover:

$$\sqrt{T}(\hat{B}_T - B_{0,T}) = \sqrt{T}(\hat{B}_T - B_{j_0,0}) + \sqrt{T}(B_{j_0,0} - B_{0,T}),$$

and, since $B_{0,T}$ is almost surely asymptotically equal to $B_{j_0,0}$, the asymptotic distribution of $\sqrt{T}(\hat{B}_T - B_{0,T})$ under H_0 is the same as that of $\sqrt{T}(\hat{B}_T - B_{j_0,0})$.

This implies that

$$\tilde{\xi}_T = T[vec\hat{B}_T - vecB_{0,T}]' \hat{A}_T' \begin{bmatrix} \hat{\Omega}_T^{-1} & 0 \\ 0 & 0 \end{bmatrix} \hat{A}_T[vec\hat{B}_T - vecB_{0,T}]$$

is asymptotically distributed as $\chi^2(\frac{1}{2}n(n-1))$ under H_0 .

An advantage of this second method is that it necessitates the computation of only one Wald test statistic.

We consider two specific SRR schemes: * In both of them, it is assumed that the monetary-policy shock has no contemporaneous impact on π_t and y_t . * In SRR Scheme 1: Inflation is contemporaneously impacted by one structural shock only. * In SRR Scheme 2: Economic activity is contemporaneously impacted by one structural shock only. \end{itemize} * When economic activity is measured by means of the output gap, both SRRs are rejected at the 5% level.

Relation with the Heteroskedasticity Identification

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

Lanne, Lutkepohl and Maciejowska (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:

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

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, Lutkepohl and Maciejowska (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” [Lutkepohl and Netsunajev (2017)(<https://www.sciencedirect.com/science/article/pii/S2452306216300223>)].

If the regimes s_t are exogenous and serially independent, then this situation is consistent with the “non-Gaussian” situation described in this section.

2.1.10 Factor-Augmented VAR (FAVAR)

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** can be employed to estimate the relevant factors (Bernanke et al. (2005)).

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 $[F_t', x_t']'$:

$$\begin{bmatrix} F_t \\ x_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ x_{t-1} \end{bmatrix} + v_t. \quad (2.33)$$

Standard identification techniques of structural shocks can be employed in Eq. (2.33): 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 macro-variables (in x_t).

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 contemporaneously 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.

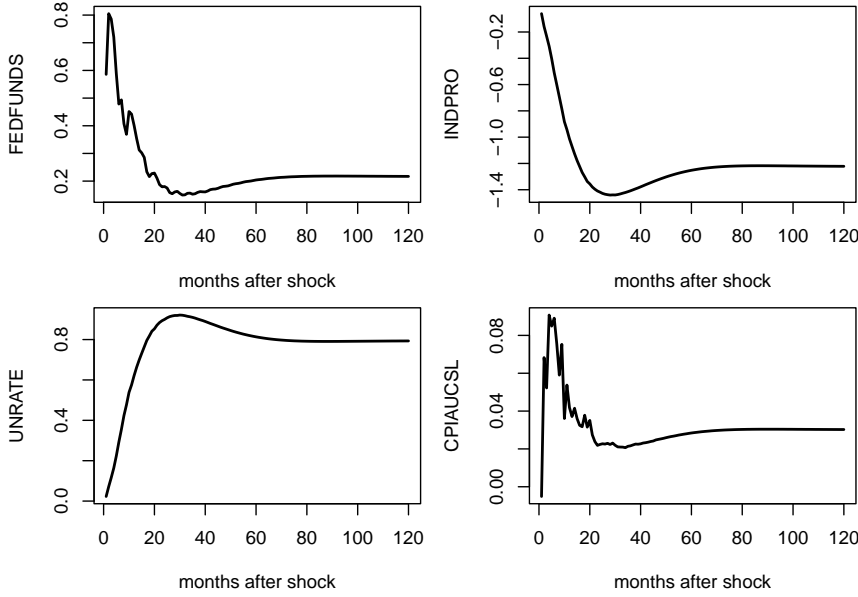


Figure 2.10: Responses of a monetary-policy shock. FAVAR approach of Bernanke, Boivin, and Elias (2005). FRED-MD dataset.

2.1.11 Projections Methods

Consider the infinite MA representation of y_t (Eq. (2.3)):

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

As seen in Section 2.1.2, 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 $\eta_{j,t}$:

$$y_{i,t+h} = \mu_i + \Psi_{i,j,h} \eta_{j,t} + u_{i,j,t+h}. \quad (2.34)$$

Because the residuals $u_{i,j,t+h}$ are autocorrelated (for $h > 0$), estimates of the covariance of the OLS estimators of the $\Psi_{i,j,h}$ then have to be based on robust estimators (e.g. Newey-West, see Eq. (??)). This is the core idea of the **local projection approach** proposed by Jordà (2005).

Now, how to proceed in the (usual) case where $\eta_{j,t}$ is not observed? We consider two situations.

Situation A: Without IV

This corresponds to the original Jordà (2005)'s approach.

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 $\eta_{1,t}$:

$$\eta_{1,t} = x_t - \mathbb{E}(x_t|w_t), \quad (2.35)$$

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. (2.35) 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. (2.34)) 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

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

This is the *conditional mean independence* case.

Let's rewrite Eq. (2.34) as follows:

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

What precedes implies that the expectation of the blue term, conditional on x_t and w_t , is linear in w_t . Standard results in the conditional mean independence case imply that the regression of $y_{i,t+h}$ on x_t , controlling for w_t , provides a consistent estimate of $\Psi_{i,1,h}$:

$$y_{i,t+h} = \alpha_i + \Psi_{i,1,h}x_t + \beta'w_t + v_{i,t+h}. \quad (2.36)$$

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 do not contemporaneously affect ΔGDP_t and π_t .

The IRFs can 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 $k \leq n$, 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 .³

Situation B: IV approach

³This is reminiscent of the distinction between 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).

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

$$\begin{cases} (IV.i) & \mathbb{E}(z_t \eta_{1,t}) \neq 0 & \text{(relevance condition)} \\ (IV.ii) & \mathbb{E}(z_t \eta_{j,t}) = 0 \text{ for } j > 1 & \text{(exogeneity condition)} \end{cases} \quad (2.37)$$

The instrument z_t can be used to identify the structural shock. Eq. (2.37) 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 $\rho = \frac{\mathbb{E}(\eta_{1,t} z_t)}{\text{var}(z_t)}$ and $\xi_t = \eta_{1,t} - \rho z_t$. It is easily seen that ξ_t satisfies the moment restrictions given above. \square

Ramey (2016) reviews the different approaches employed to construct monetary policy-shocks (the two main approaches are presented in 2.1 and ?? below). She has also collected time series of such shocks, see her website.

Example 2.1 (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:

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

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 2.11, 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).

Example 2.2 (Identification of Monetary-Policy Shocks Based on the Narrative Approach). Romer and Romer (2004) propose a two-step approach:

Figure 1. Intraday Trading in Federal Funds Futures Contracts

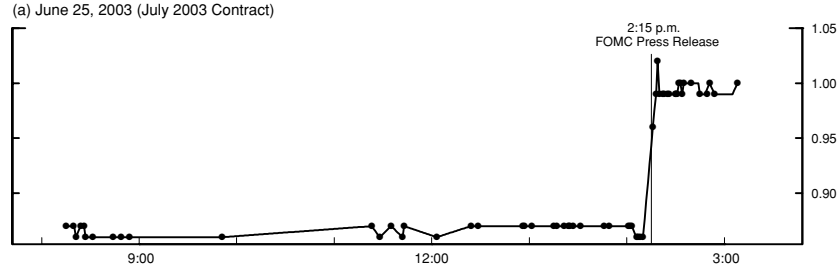


Figure 2.11: Source: Gurkaynak, Sack and Swanson (2005). Transaction rates of Federal funds futures on June 25, 2003, day on which a regularly scheduled FOMC meeting was scheduled. At 2:15 p.m., the FOMC announced that it was lowering its target for the federal funds rate from 1.25% to 1%, while many 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.

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

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

- a. The LP-IV approach, where y_t ’s DGP is left unspecified,
- b. The SVAR-IV approach.

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.

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. You have, for $i \in \{1, \dots, n\}$:

$$\begin{aligned}\varepsilon_{i,t} &= b_{i,1}\eta_{1,t} + u_{i,t}(\#eq : eps_rho) \\ &= b_{i,1}\rho z_t + \underbrace{b_{i,1}\xi_t + u_{i,t}}_{\perp z_t}.\end{aligned}\tag{2.38}$$

($u_{i,t}$ 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 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 one-unit contemporaneous effect on $y_{1,t}$. Denoting by $\tilde{B}_{i,1}$ the contemporaneous impact of $\tilde{\eta}_{1,t}$, we get:

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

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

Eq. @ref(eq:eps_rho) 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 $\varepsilon_{1,t}$, using z_t as an instrument.

What about inference? One 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 $\{\hat{\Phi}_1, \dots, \hat{\Phi}_p, \hat{B}_1\}$ using the SVAR-IV approach based on a size- T sample. Generate N (where N is large) size- T samples from the following VAR:

$$\begin{bmatrix} \hat{\Phi}(L) & 0 \\ 0 & \hat{\rho}(L) \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ e_t \end{bmatrix},$$

$$\text{where } \begin{bmatrix} \varepsilon_t \\ e_t \end{bmatrix} \sim i.i.d. \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Omega & S'_{\varepsilon,e} \\ S_{\varepsilon,e} & \sigma_e^2 \end{bmatrix} \right),$$

where $\hat{\rho}(L)$ and σ_e^2 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 $\{\tilde{\Phi}_1, \dots, \tilde{\Phi}_p, \tilde{B}_1\}$ and associated $\tilde{\Psi}_{i,1,h}$. This provides e.g. a sequence of N estimates of $\Psi_{i,1,h}$, from which quantiles and conf. intervals can be deduced.

```
# Load vars package:
library(vars)
library(AEC)
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 <- "FF4_TC" # "FF4_TC", "ED2_TC", "ff1_vr", "rrshock83b"
indic.shock.name <- which(names(USmonthly)==shock.name)
Z <- matrix(USmonthly[,indic.shock.name],ncol=1)
plot(USmonthly$DATES,Z,type="l",xlab="",ylab="")
```

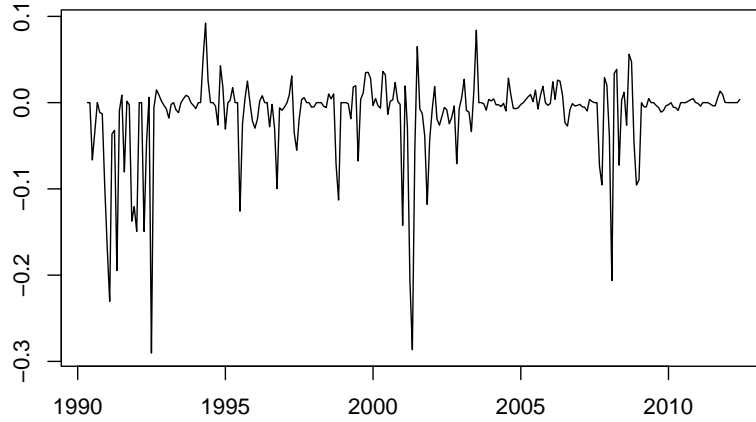


Figure 2.12: Gertler-Karadi monthly shocks, fed funds futures 3 months.

```

considered.variables <- c("GS1","LIP","LCPI","EBP")
y <- as.matrix(USmonthly[,considered.variables])
n <- length(considered.variables)
colnames(y) <- considered.variables
res.svar.iv <- svar.iv(y,Z,p = 4,
                      names.of.variables=considered.variables,
                      nb.periods.IRF = 20,
                      z.AR.order=1, # This is used in the parametric bootstrap only
                      nb.bootstrap.replications = 100, # This is used in the parametric bootstrap only
                      confidence.interval = 0.90, # This is used in the parametric bootstrap only
                      indic.plot=1)

```

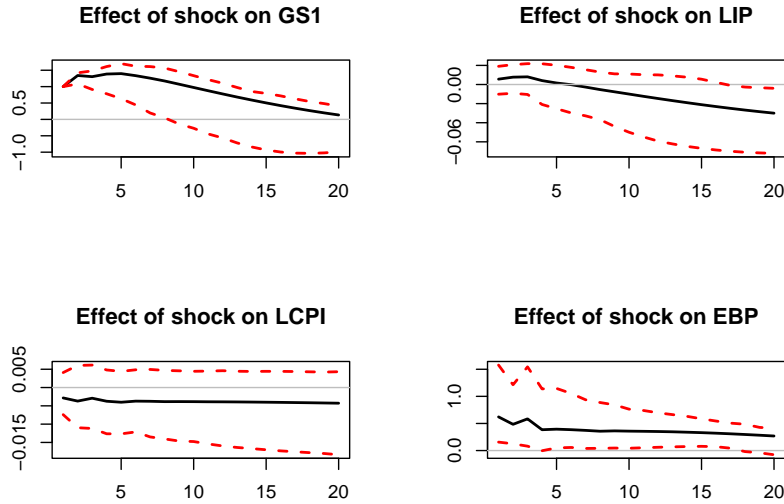


Figure 2.13: Responses to a monetary-policy shock, SVAR-IV approach.

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. (2.37)) have to be complemented with (IV.iii):

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

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:

$$y_{i,t+h} = \alpha_i + \tilde{\Psi}_{i,1,h} y_{1,t} + \nu_{i,t+h}, \quad (2.39)$$

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} are supposed to be correlated to the residuals ε .

Moreover, the ε are supposed to be possibly heteroskedastic and auto-correlated.

We consider the instruments \mathbf{Z} , 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} .

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

where \mathbf{S} is the long-run variance of $\mathbf{z}_t \varepsilon_t$ (see next slide).

The asymptotic covariance matrix of $\sqrt{T}\mathbf{b}_{iv}$ is $\mathbf{Q}_{xz}\mathbf{S}\mathbf{Q}_{xz}'$.

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} (see Eq. (??))

(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 (2.39). 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.

As noted by Stock and Watson (2018), the relevant variance is the long-run variance of the instrument-times-error term. They also recommend (p.926) using leads and lags of z_t to improve efficiency.

```
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)
```

```
## [1] "LP-IV approach, Currently working on horizon h=0"
## [1] "LP-IV approach, Currently working on horizon h=1"
## [1] "LP-IV approach, Currently working on horizon h=2"
## [1] "LP-IV approach, Currently working on horizon h=3"
## [1] "LP-IV approach, Currently working on horizon h=4"
## [1] "LP-IV approach, Currently working on horizon h=5"
## [1] "LP-IV approach, Currently working on horizon h=6"
## [1] "LP-IV approach, Currently working on horizon h=7"
## [1] "LP-IV approach, Currently working on horizon h=8"
## [1] "LP-IV approach, Currently working on horizon h=9"
## [1] "LP-IV approach, Currently working on horizon h=10"
## [1] "LP-IV approach, Currently working on horizon h=11"
## [1] "LP-IV approach, Currently working on horizon h=12"
## [1] "LP-IV approach, Currently working on horizon h=13"
## [1] "LP-IV approach, Currently working on horizon h=14"
## [1] "LP-IV approach, Currently working on horizon h=15"
## [1] "LP-IV approach, Currently working on horizon h=16"
## [1] "LP-IV approach, Currently working on horizon h=17"
## [1] "LP-IV approach, Currently working on horizon h=18"
## [1] "LP-IV approach, Currently working on horizon h=19"
## [1] "LP-IV approach, Currently working on horizon h=20"
```

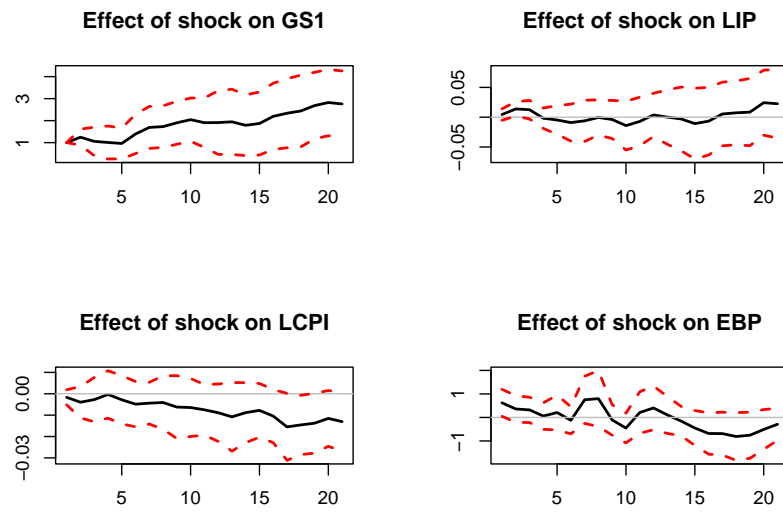


Figure 2.14: Responses to a monetary-policy shock, LP-IV approach.

Chapter 3

Appendix

3.1 Statistical Tables

3.2 Statistics: definitions and results

Definition 3.1 (Partial correlation). The **partial correlation** between y and z , controlling for some variables \mathbf{X} is the sample correlation between y^* and z^* , where the latter two variables are the residuals in regressions of y on \mathbf{X} and of z on \mathbf{X} , respectively.

This correlation is denoted by $r_{yz}^{\mathbf{X}}$. By definition, we have:

$$r_{yz}^{\mathbf{X}} = \frac{\mathbf{z}^{*'} \mathbf{y}^*}{\sqrt{(\mathbf{z}^{*'} \mathbf{z}^*)(\mathbf{y}^{*'} \mathbf{y}^*)}}. \quad (3.1)$$

Definition 3.2 (Skewness and kurtosis). Let Y be a random variable whose fourth moment exists. The expectation of Y is denoted by μ .

- The skewness of Y is given by:

$$\frac{\mathbb{E}[(Y - \mu)^3]}{\{\mathbb{E}[(Y - \mu)^2]\}^{3/2}}.$$

- The kurtosis of Y is given by:

$$\frac{\mathbb{E}[(Y - \mu)^4]}{\{\mathbb{E}[(Y - \mu)^2]\}^2}.$$

Table 3.1: 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 3.2: 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 3.3: 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 3.4: 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 |

Definition 3.3 (Eigenvalues). The eigenvalues of a matrix M are the numbers λ for which:

$$|M - \lambda I| = 0,$$

where $|\bullet|$ is the determinant operator.

Proposition 3.1 (Properties of the determinant). *We have:*

- $|MN| = |M| \times |N|$.
- $|M^{-1}| = |M|^{-1}$.
- If M admits the diagonal representation $M = TDT^{-1}$, where D is a diagonal matrix whose diagonal entries are $\{\lambda_i\}_{i=1,\dots,n}$, then:

$$|M - \lambda I| = \prod_{i=1}^n (\lambda_i - \lambda).$$

Definition 3.4 (Moore-Penrose inverse). If $M \in \mathbb{R}^{m \times n}$, then its Moore-Penrose pseudo inverse (exists and) is the unique matrix $M^* \in \mathbb{R}^{n \times m}$ that satisfies:

- i. $MM^*M = M$
- ii. $M^*MM^* = M^*$
- iii. $(MM^*)' = MM^*$.iv $(M^*M)' = M^*M$.

Proposition 3.2 (Properties of the Moore-Penrose inverse). • If M is invertible then $M^* = M^{-1}$.

- The pseudo-inverse of a zero matrix is its transpose. *
- *

The pseudo-inverse of the pseudo-inverse is the original matrix.

Definition 3.5 (F distribution). Consider $n = n_1 + n_2$ i.i.d. $\mathcal{N}(0, 1)$ r.v. X_i . If the r.v. F is defined by:

$$F = \frac{\sum_{i=1}^{n_1} X_i^2}{\sum_{j=n_1+1}^{n_1+n_2} X_j^2} \frac{n_2}{n_1}$$

then $F \sim \mathcal{F}(n_1, n_2)$. (See Table 3.4 for quantiles.)

Definition 3.6 (Student-t distribution). Z follows a Student-t (or t) distribution with ν degrees of freedom (d.f.) if:

$$Z = X_0 / \sqrt{\frac{\sum_{i=1}^{\nu} X_i^2}{\nu}}, \quad X_i \sim i.i.d. \mathcal{N}(0, 1).$$

We have $\mathbb{E}(Z) = 0$, and $\mathbb{V}ar(Z) = \frac{\nu}{\nu-2}$ if $\nu > 2$. (See Table 3.2 for quantiles.)

Definition 3.7 (Chi-square distribution). Z follows a χ^2 distribution with ν d.f. if $Z = \sum_{i=1}^{\nu} X_i^2$ where $X_i \sim i.i.d. \mathcal{N}(0, 1)$. We have $\mathbb{E}(Z) = \nu$. (See Table 3.3 for quantiles.)

Definition 3.8 (Idempotent matrix). Matrix M is idempotent if $M^2 = M$.

If M is a symmetric idempotent matrix, then $M'M = M$.

Proposition 3.3 (Roots of an idempotent matrix). *The eigenvalues of an idempotent matrix are either 1 or 0.*

Proof. If λ is an eigenvalue of an idempotent matrix M then $\exists x \neq 0$ s.t. $Mx = \lambda x$. Hence $M^2x = \lambda Mx \Rightarrow (1 - \lambda)Mx = 0$. Either all element of Mx are zero, in which case $\lambda = 0$ or at least one element of Mx is nonzero, in which case $\lambda = 1$. \square

Proposition 3.4 (Idempotent matrix and chi-square distribution). *The rank of a symmetric idempotent matrix is equal to its trace.*

Proof. The result follows from Prop. 3.3, combined with the fact that the rank of a symmetric matrix is equal to the number of its nonzero eigenvalues. \square

Proposition 3.5 (Constrained least squares). *The solution of the following optimisation problem:*

$$\begin{aligned} \min_{\beta} \quad & ||\mathbf{y} - \mathbf{X}\beta||^2 \\ \text{subject to } & \mathbf{R}\beta = \mathbf{q} \end{aligned}$$

is given by:

$$\beta^r = \beta_0 - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\{\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\}^{-1}(\mathbf{R}\beta_0 - \mathbf{q}),$$

where $\beta_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

Proof. See for instance Jackman, 2007. \square

Proposition 3.6 (Chebychev's inequality). *If $\mathbb{E}(|X|^r)$ is finite for some $r > 0$ then:*

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X - c| > \varepsilon) \leq \frac{\mathbb{E}[|X - c|^r]}{\varepsilon^r}.$$

In particular, for $r = 2$:

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X - c| > \varepsilon) \leq \frac{\mathbb{E}[(X - c)^2]}{\varepsilon^2}.$$

Proof. Remark that $\varepsilon^r \mathbb{I}_{\{|X| \geq \varepsilon\}} \leq |X|^r$ and take the expectation of both sides. \square

Definition 3.9 (Convergence in probability). The random variable sequence x_n converges in probability to a constant c if $\forall \varepsilon, \lim_{n \rightarrow \infty} \mathbb{P}(|x_n - c| > \varepsilon) = 0$.

It is denoted as: $\text{plim } x_n = c$.

Definition 3.10 (Convergence in the L^r norm). x_n converges in the r -th mean (or in the L^r -norm) towards x , if $\mathbb{E}(|x_n|^r)$ and $\mathbb{E}(|x|^r)$ exist and if

$$\lim_{n \rightarrow \infty} \mathbb{E}(|x_n - x|^r) = 0.$$

It is denoted as: $x_n \xrightarrow{L^r} c$.

For $r = 2$, this convergence is called **mean square convergence**.

Definition 3.11 (Almost sure convergence). The random variable sequence x_n converges almost surely to c if $\mathbb{P}(\lim_{n \rightarrow \infty} x_n = c) = 1$.

It is denoted as: $x_n \xrightarrow{a.s.} c$.

Definition 3.12 (Convergence in distribution). x_n is said to converge in distribution (or in law) to x if

$$\lim_{n \rightarrow \infty} F_{x_n}(s) = F_x(s)$$

for all s at which F_X —the cumulative distribution of X —is continuous.

It is denoted as: $x_n \xrightarrow{d} x$.

Proposition 3.7 (Rules for limiting distributions (Slutsky)). *We have:*

i. **Slutsky's theorem:** If $x_n \xrightarrow{d} x$ and $y_n \xrightarrow{p} c$ then

$$\begin{aligned} x_n y_n &\xrightarrow{d} xc \\ x_n + y_n &\xrightarrow{d} x + c \\ x_n / y_n &\xrightarrow{d} x/c \quad (\text{if } c \neq 0) \end{aligned}$$

ii. **Continuous mapping theorem:** If $x_n \xrightarrow{d} x$ and g is a continuous function then $g(x_n) \xrightarrow{d} g(x)$.

Proposition 3.8 (Implications of stochastic convergences). *We have:*

$$\begin{array}{ccccc} \boxed{L^s} & & \boxed{L^r} & & \\ \rightarrow & \Rightarrow_{1 \leq r \leq s} & \rightarrow & & \\ & & \Downarrow & & \\ \boxed{a.s.} & \Rightarrow & \boxed{p} & \Rightarrow & \boxed{d} \\ \rightarrow & & \rightarrow & & \rightarrow \end{array}$$

Proof. (of the fact that $\left(\xrightarrow{p}\right) \Rightarrow \left(\xrightarrow{d}\right)$). Assume that $X_n \xrightarrow{p} X$. Denoting by F and F_n the c.d.f. of X and X_n , respectively:

$$F_n(x) = \mathbb{P}(X_n \leq x, X \leq x + \varepsilon) + \mathbb{P}(X_n \leq x, X > x + \varepsilon) \leq F(x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon). \quad (3.2)$$

Besides,

$$F(x-\varepsilon) = \mathbb{P}(X \leq x-\varepsilon, X_n \leq x) + \mathbb{P}(X \leq x-\varepsilon, X_n > x) \leq F_n(x) + \mathbb{P}(|X_n - X| > \varepsilon),$$

which implies:

$$F(x-\varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \leq F_n(x). \quad (3.3)$$

Eqs. (3.2) and (3.3) imply:

$$F(x-\varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \leq F_n(x) \leq F(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon).$$

Taking limits as $n \rightarrow \infty$ yields

$$F(x-\varepsilon) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x+\varepsilon).$$

The result is then obtained by taking limits as $\varepsilon \rightarrow 0$ (if F is continuous at x). \square

Proposition 3.9 (Convergence in distribution to a constant). *If X_n converges in distribution to a constant c , then X_n converges in probability to c .*

Proof. If $\varepsilon > 0$, we have $\mathbb{P}(X_n < c - \varepsilon) \xrightarrow{n \rightarrow \infty} 0$ i.e. $\mathbb{P}(X_n \geq c - \varepsilon) \xrightarrow{n \rightarrow \infty} 1$ and $\mathbb{P}(X_n < c + \varepsilon) \xrightarrow{n \rightarrow \infty} 1$. Therefore $\mathbb{P}(c - \varepsilon \leq X_n < c + \varepsilon) \xrightarrow{n \rightarrow \infty} 1$, which gives the result. \square

Example of $plim$ but not L^r convergence: Let $\{x_n\}_{n \in \mathbb{N}}$ be a series of random variables defined by:

$$x_n = nu_n,$$

where u_n are independent random variables s.t. $u_n \sim \mathcal{B}(1/n)$.

We have $x_n \xrightarrow{p} 0$ but $x_n \not\xrightarrow{L^r} 0$ because $\mathbb{E}(|X_n - 0|) = \mathbb{E}(X_n) = 1$.

Theorem 3.1 (Cauchy criterion (non-stochastic case)). *We have that $\sum_{i=0}^T a_i$ converges ($T \rightarrow \infty$) iff, for any $\eta > 0$, there exists an integer N such that, for all $M \geq N$,*

$$\left| \sum_{i=N+1}^M a_i \right| < \eta.$$

Theorem 3.2 (Cauchy criterion (stochastic case)). *We have that $\sum_{i=0}^T \theta_i \varepsilon_{t-i}$ converges in mean square ($T \rightarrow \infty$) to a random variable iff, for any $\eta > 0$, there exists an integer N such that, for all $M \geq N$,*

$$\mathbb{E} \left[\left(\sum_{i=N+1}^M \theta_i \varepsilon_{t-i} \right)^2 \right] < \eta.$$

Definition 3.13 (Characteristic function). For any real-valued random variable X , the characteristic function is defined by:

$$\phi_X : u \rightarrow \mathbb{E}[\exp(iuX)].$$

Theorem 3.3 (Law of large numbers). *The sample mean is a consistent estimator of the population mean.*

Proof. Let's denote by ϕ_{X_i} the characteristic function of a r.v. X_i . If the mean of X_i is μ then the Talyor expansion of the characteristic function is:

$$\phi_{X_i}(u) = \mathbb{E}(\exp(iuX)) = 1 + iu\mu + o(u).$$

The properties of the characteristic function (see Def. 3.13) imply that:

$$\phi_{\frac{1}{n}(X_1 + \dots + X_n)}(u) = \prod_{i=1}^n \left(1 + i\frac{u}{n}\mu + o\left(\frac{u}{n}\right)\right) \rightarrow e^{iu\mu}.$$

The facts that (a) $e^{iu\mu}$ is the characteristic function of the constant μ and (b) that a characteristic function uniquely characterises a distribution imply that the sample mean converges in distribution to the constant μ , which further implies that it converges in probability to μ . \square

Theorem 3.4 (Lindberg-Levy Central limit theorem, CLT). *If x_n is an i.i.d. sequence of random variables with mean μ and variance σ^2 ($\in]0, +\infty[$), then:*

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad \text{where } \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i.$$

Proof. Let us introduce the r.v. $Y_n := \sqrt{n}(\bar{X}_n - \mu)$. We have $\phi_{Y_n}(u) = \left[\mathbb{E} \left(\exp(i \frac{1}{\sqrt{n}} u (X_1 - \mu)) \right) \right]^n$. We have:

$$\begin{aligned} \left[\mathbb{E} \left(\exp \left(i \frac{1}{\sqrt{n}} u (X_1 - \mu) \right) \right) \right]^n &= \left[\mathbb{E} \left(1 + i \frac{1}{\sqrt{n}} u (X_1 - \mu) - \frac{1}{2n} u^2 (X_1 - \mu)^2 + o(u^2) \right) \right]^n \\ &= \left(1 - \frac{1}{2n} u^2 \sigma^2 + o(u^2) \right)^n. \end{aligned}$$

Therefore $\phi_{Y_n}(u) \xrightarrow{n \rightarrow \infty} \exp(-\frac{1}{2} u^2 \sigma^2)$, which is the characteristic function of $\mathcal{N}(0, \sigma^2)$. \square

Proposition 3.10 (Inverse of a partitioned matrix). *We have:*

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})^{-1} & -\mathbf{A}_{11}^{-1} \mathbf{A}_{12} (\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12})^{-1} \\ -(\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12})^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & (\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12})^{-1} \end{bmatrix}.$$

Proposition 3.11. *If \mathbf{A} is idempotent and if \mathbf{x} is Gaussian, \mathbf{Lx} and $\mathbf{x}'\mathbf{Ax}$ are independent if $\mathbf{LA} = \mathbf{0}$.*

Proof. If $\mathbf{LA} = \mathbf{0}$, then the two Gaussian vectors \mathbf{Lx} and \mathbf{Ax} are independent. This implies the independence of any function of \mathbf{Lx} and any function of \mathbf{Ax} . The results then follow from the observation that $\mathbf{x}'\mathbf{Ax} = (\mathbf{Ax})'(\mathbf{Ax})$, which is a function of \mathbf{Ax} . \square

Proposition 3.12 (Inner product of a multivariate Gaussian variable). *Let X be a n -dimensional multivariate Gaussian variable: $X \sim \mathcal{N}(0, \Sigma)$. We have:*

$$X'\Sigma^{-1}X \sim \chi^2(n).$$

Proof. Because Σ is a symmetrical definite positive matrix, it admits the spectral decomposition PDP' where P is an orthogonal matrix (i.e. $PP' = Id$) and D is a diagonal matrix with non-negative entries. Denoting by $\sqrt{D^{-1}}$ the diagonal matrix whose diagonal entries are the inverse of those of D , it is easily checked that the covariance matrix of $Y := \sqrt{D^{-1}}P'X$ is Id . Therefore Y is a vector of uncorrelated Gaussian variables. The properties of Gaussian variables imply that the components of Y are then also independent. Hence $Y'Y = \sum_i Y_i^2 \sim \chi^2(n)$.

It remains to note that $Y'Y = X'PD^{-1}P'X = X'\mathbb{V}ar(X)^{-1}X$ to conclude. \square

Theorem 3.5 (Cauchy-Schwarz inequality). *We have:*

$$|\text{Cov}(X, Y)| \leq \sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}$$

and, if $X \neq 0$ and $Y \neq 0$, the equality holds iff X and Y are the same up to an affine transformation.

Proof. If $\mathbb{V}ar(X) = 0$, this is trivial. If this is not the case, then let's define Z as $Z = Y - \frac{\text{Cov}(X, Y)}{\mathbb{V}ar(X)}X$. It is easily seen that $\text{Cov}(X, Z) = 0$. Then, the variance of $Y = Z + \frac{\text{Cov}(X, Y)}{\mathbb{V}ar(X)}X$ is equal to the sum of the variance of Z and of the variance of $\frac{\text{Cov}(X, Y)}{\mathbb{V}ar(X)}X$, that is:

$$\mathbb{V}ar(Y) = \mathbb{V}ar(Z) + \left(\frac{\text{Cov}(X, Y)}{\mathbb{V}ar(X)} \right)^2 \mathbb{V}ar(X) \geq \left(\frac{\text{Cov}(X, Y)}{\mathbb{V}ar(X)} \right)^2 \mathbb{V}ar(X).$$

The equality holds iff $\mathbb{V}ar(Z) = 0$, i.e. iff $Y = \frac{\text{Cov}(X, Y)}{\mathbb{V}ar(X)}X + cst$. \square

Definition 3.14 (Matrix derivatives). Consider a fonction $f : \mathbb{R}^K \rightarrow \mathbb{R}$. Its first-order derivative is:

$$\frac{\partial f}{\partial \mathbf{b}}(\mathbf{b}) = \begin{bmatrix} \frac{\partial f}{\partial b_1}(\mathbf{b}) \\ \vdots \\ \frac{\partial f}{\partial b_K}(\mathbf{b}) \end{bmatrix}.$$

We use the notation:

$$\frac{\partial f}{\partial \mathbf{b}'}(\mathbf{b}) = \left(\frac{\partial f}{\partial \mathbf{b}}(\mathbf{b}) \right)'.$$

Proposition 3.13. *We have:*

- If $f(\mathbf{b}) = A'\mathbf{b}$ where A is a $K \times 1$ vector then $\frac{\partial f}{\partial \mathbf{b}}(\mathbf{b}) = A$.
- If $f(\mathbf{b}) = \mathbf{b}'A\mathbf{b}$ where A is a $K \times K$ matrix, then $\frac{\partial f}{\partial \mathbf{b}}(\mathbf{b}) = 2A\mathbf{b}$.

Definition 3.15 (Asymptotic level). An asymptotic test with critical region Ω_n has an asymptotic level equal to α if:

$$\sup_{\theta \in \Theta} \lim_{n \rightarrow \infty} \mathbb{P}_\theta(S_n \in \Omega_n) = \alpha,$$

where S_n is the test statistic and Θ is such that the null hypothesis H_0 is equivalent to $\theta \in \Theta$.

Definition 3.16 (Asymptotically consistent test). An asymptotic test with critical region Ω_n is consistent if:

$$\forall \theta \in \Theta^c, \quad \mathbb{P}_\theta(S_n \in \Omega_n) \rightarrow 1,$$

where S_n is the test statistic and Θ^c is such that the null hypothesis H_0 is equivalent to $\theta \notin \Theta^c$.

Definition 3.17 (Kullback discrepancy). Given two p.d.f. f and f^* , the Kullback discrepancy is defined by:

$$I(f, f^*) = \mathbb{E}^* \left(\log \frac{f^*(Y)}{f(Y)} \right) = \int \log \frac{f^*(y)}{f(y)} f^*(y) dy.$$

Proposition 3.14 (Properties of the Kullback discrepancy). *We have:*

- i. $I(f, f^*) \geq 0$
- ii. $I(f, f^*) = 0$ iff $f \equiv f^*$.

Proof. $x \rightarrow -\log(x)$ is a convex function. Therefore $\mathbb{E}^*(-\log f(Y)/f^*(Y)) \geq -\log \mathbb{E}^*(f(Y)/f^*(Y)) = 0$ (proves (i)). Since $x \rightarrow -\log(x)$ is strictly convex, equality in (i) holds if and only if $f(Y)/f^*(Y)$ is constant (proves (ii)). \square

Proposition 3.15 (Square and absolute summability). *We have:*

$$\underbrace{\sum_{i=0}^{\infty} |\theta_i| < +\infty}_{\text{Absolute summability}} \quad \Rightarrow \quad \underbrace{\sum_{i=0}^{\infty} \theta_i^2 < +\infty}_{\text{Square summability}}.$$

Proof. See Appendix 3.A in Hamilton. Idea: Absolute summability implies that there exist N such that, for $j > N$, $|\theta_j| < 1$ (deduced from Cauchy criterion, Theorem 3.1 and therefore $\theta_j^2 < |\theta_j|$). \square

3.3 Some properties of Gaussian variables

Proposition 3.16 (Bayesian update in a vector of Gaussian variables). *If*

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}\right),$$

then

$$Y_2|Y_1 \sim \mathcal{N}(\Omega_{21}\Omega_{11}^{-1}Y_1, \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}).$$

$$Y_1|Y_2 \sim \mathcal{N}(\Omega_{12}\Omega_{22}^{-1}Y_2, \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}).$$

Proposition 3.17 (Truncated distributions). *If X is a random variable distributed according to some p.d.f. f , with c.d.f. F , with infinite support. Then the p.d.f. of $X|a \leq X < b$ is*

$$g(x) = \frac{f(x)}{F(b) - F(a)} \mathbb{I}_{\{a \leq x < b\}},$$

for any $a < b$.

In particular, for a Gaussian variable $X \sim \mathcal{N}(\mu, \sigma^2)$, we have

$$f(X = x|a \leq X < b) = \frac{\frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right)}{Z}.$$

with $Z = \Phi(\beta) - \Phi(\alpha)$, where $\alpha = \frac{a - \mu}{\sigma}$ and $\beta = \frac{b - \mu}{\sigma}$.

Moreover:

$$\mathbb{E}(X|a \leq X < b) = \mu - \frac{\phi(\beta) - \phi(\alpha)}{Z} \sigma. \quad (3.4)$$

We also have:

$$\begin{aligned} & \text{Var}(X|a \leq X < b) \\ = & \sigma^2 \left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{Z} - \left(\frac{\phi(\beta) - \phi(\alpha)}{Z} \right)^2 \right] \end{aligned} \quad (3.5)$$

In particular, for $b \rightarrow \infty$, we get:

$$\text{Var}(X|a < X) = \sigma^2 [1 + \alpha\lambda(-\alpha) - \lambda(-\alpha)^2], \quad (3.6)$$

with $\lambda(x) = \frac{\phi(x)}{\Phi(x)}$ is called the **inverse Mills ratio**.

Consider the case where $a \rightarrow -\infty$ (i.e. the conditioning set is $X < b$) and $\mu = 0$, $\sigma = 1$. Then Eq. (3.4) gives $\mathbb{E}(X|X < b) = -\lambda(b) = -\frac{\phi(b)}{\Phi(b)}$, where λ is the function computing the inverse Mills ratio.

Proposition 3.18 (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:*

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

3.4 Proofs

3.4.1 Proof of Proposition ??

Proof. Assumptions (i) and (ii) (in the set of Assumptions ??) imply that θ_{MLE} exists ($= \text{argmax}_{\theta} (1/n) \log \mathcal{L}(\theta; \mathbf{y})$).

$(1/n) \log \mathcal{L}(\theta; \mathbf{y})$ can be interpreted as the sample mean of the r.v. $\log f(Y_i; \theta)$ that are i.i.d. Therefore $(1/n) \log \mathcal{L}(\theta; \mathbf{y})$ converges to $\mathbb{E}_{\theta_0}(\log f(Y; \theta))$ – which exists (Assumption iv).

Because the latter convergence is uniform (Assumption v), the solution θ_{MLE} almost surely converges to the solution to the limit problem:

$$\text{argmax}_{\theta} \mathbb{E}_{\theta_0}(\log f(Y; \theta)) = \text{argmax}_{\theta} \int_{\mathcal{Y}} \log f(y; \theta) f(y; \theta_0) dy.$$

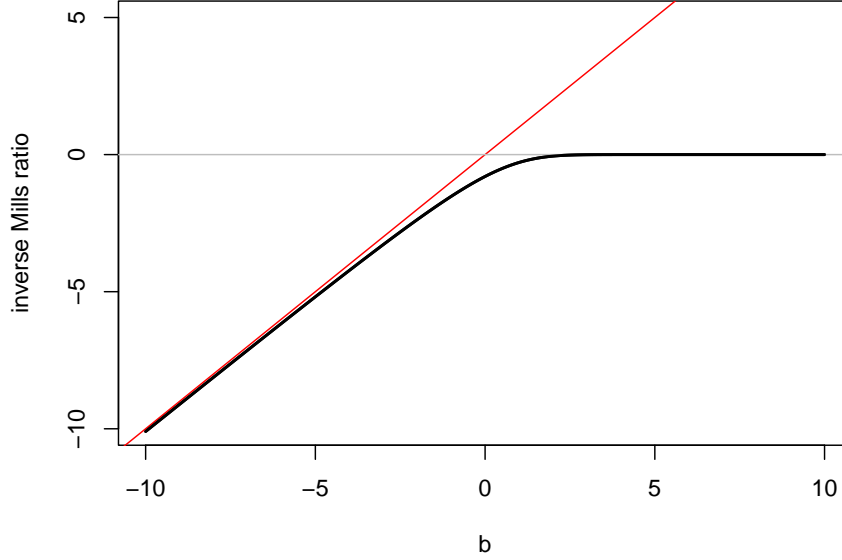


Figure 3.1: $\mathbb{E}(X|X < b)$ as a function of b when $X \sim \mathcal{N}(0,1)$ (in black).

Properties of the Kullback information measure (see Prop. 3.14), together with the identifiability assumption (ii) implies that the solution to the limit problem is unique and equal to θ_0 .

Consider a r.v. sequence θ that converges to θ_0 . The Taylor expansion of the score in a neighborhood of θ_0 yields to:

$$\frac{\partial \log \mathcal{L}(\theta; \mathbf{y})}{\partial \theta} = \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} + \frac{\partial^2 \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta \partial \theta'} (\theta - \theta_0) + o_p(\theta - \theta_0)$$

θ_{MLE} converges to θ_0 and satisfies the likelihood equation $\frac{\partial \log \mathcal{L}(\theta; \mathbf{y})}{\partial \theta} = \mathbf{0}$. Therefore:

$$\frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} \approx - \frac{\partial^2 \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta \partial \theta'} (\theta_{MLE} - \theta_0),$$

or equivalently:

$$\frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} \approx \left(-\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(y_i; \theta_0)}{\partial \theta \partial \theta'} \right) \sqrt{n} (\theta_{MLE} - \theta_0),$$

By the law of large numbers, we have: $\left(-\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(y_i; \theta_0)}{\partial \theta \partial \theta'} \right) \rightarrow \frac{1}{n} \mathbf{I}(\theta_0) = \mathcal{I}_Y(\theta_0)$.

Besides, we have:

$$\begin{aligned} \frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} &= \sqrt{n} \left(\frac{1}{n} \sum_i \frac{\partial \log f(y_i; \theta_0)}{\partial \theta} \right) \\ &= \sqrt{n} \left(\frac{1}{n} \sum_i \left\{ \frac{\partial \log f(y_i; \theta_0)}{\partial \theta} - \mathbb{E}_{\theta_0} \frac{\partial \log f(Y_i; \theta_0)}{\partial \theta} \right\} \right) \end{aligned}$$

which converges to $\mathcal{N}(0, \mathcal{I}_Y(\theta_0))$ by the CLT.

Collecting the preceding results leads to (b). The fact that θ_{MLE} achieves the FDCR bound proves (c). \square

3.4.2 Proof of Proposition ??

Proof. We have $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1})$ (Eq. ??eq:normMLE). A Taylor expansion around θ_0 yields to:

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta_0)) \xrightarrow{d} \mathcal{N} \left(0, \frac{\partial h(\theta_0)}{\partial \theta'} \mathcal{I}(\theta_0)^{-1} \frac{\partial h(\theta_0)'}{\partial \theta} \right). \quad (3.7)$$

Under H_0 , $h(\theta_0) = 0$ therefore:

$$\sqrt{n}h(\hat{\theta}_n) \xrightarrow{d} \mathcal{N} \left(0, \frac{\partial h(\theta_0)}{\partial \theta'} \mathcal{I}(\theta_0)^{-1} \frac{\partial h(\theta_0)'}{\partial \theta} \right). \quad (3.8)$$

Hence

$$\sqrt{n} \left(\frac{\partial h(\theta_0)}{\partial \theta'} \mathcal{I}(\theta_0)^{-1} \frac{\partial h(\theta_0)'}{\partial \theta} \right)^{-1/2} h(\hat{\theta}_n) \xrightarrow{d} \mathcal{N}(0, Id).$$

Taking the quadratic form, we obtain:

$$nh(\hat{\theta}_n)' \left(\frac{\partial h(\theta_0)}{\partial \theta'} \mathcal{I}(\theta_0)^{-1} \frac{\partial h(\theta_0)'}{\partial \theta} \right)^{-1} h(\hat{\theta}_n) \xrightarrow{d} \chi^2(r).$$

The fact that the test has asymptotic level α directly stems from what precedes.

Consistency of the test: Consider $\theta_0 \in \Theta$. Because the MLE is consistent, $h(\hat{\theta}_n)$ converges to $h(\theta_0) \neq 0$. Eq. (3.7) is still valid. It implies that ξ_n^W converges to $+\infty$ and therefore that $\mathbb{P}_\theta(\xi_n^W \geq \chi_{1-\alpha}^2(r)) \rightarrow 1$. \square

3.4.3 Proof of Proposition ??

Proof. Notations: “ \approx ” means “equal up to a term that converges to 0 in probability”. We are under H_0 . $\hat{\theta}^0$ is the constrained ML estimator; $\hat{\theta}$ denotes the unconstrained one.

We combine the two Taylor expansion: $h(\hat{\theta}_n) \approx \frac{\partial h(\theta_0)}{\partial \theta'}(\hat{\theta}_n - \theta_0)$ and $h(\hat{\theta}_n^0) \approx \frac{\partial h(\theta_0)}{\partial \theta'}(\hat{\theta}_n^0 - \theta_0)$ and we use $h(\hat{\theta}_n^0) = 0$ (by definition) to get:

$$\sqrt{n}h(\hat{\theta}_n) \approx \frac{\partial h(\theta_0)}{\partial \theta'}\sqrt{n}(\hat{\theta}_n - \hat{\theta}_n^0). \quad (3.9)$$

Besides, we have (using the definition of the information matrix):

$$\frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} \approx \frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} - \mathcal{I}(\theta_0)\sqrt{n}(\hat{\theta}_n^0 - \theta_0) \quad (3.10)$$

and:

$$0 = \frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\hat{\theta}_n; \mathbf{y})}{\partial \theta} \approx \frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} - \mathcal{I}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0). \quad (3.11)$$

Taking the difference and multiplying by $\mathcal{I}(\theta_0)^{-1}$:

$$\sqrt{n}(\hat{\theta}_n - \hat{\theta}_n^0) \approx \mathcal{I}(\theta_0)^{-1} \frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} \mathcal{I}(\theta_0). \quad (3.12)$$

Eqs. (3.9) and (3.12) yield to:

$$\sqrt{n}h(\hat{\theta}_n) \approx \frac{\partial h(\theta_0)}{\partial \theta'} \mathcal{I}(\theta_0)^{-1} \frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta}. \quad (3.13)$$

Recall that $\hat{\theta}_n^0$ is the MLE of θ_0 under the constraint $h(\theta) = 0$. The vector of Lagrange multipliers $\hat{\lambda}_n$ associated to this program satisfies:

$$\frac{\partial \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} + \frac{\partial h'(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} \hat{\lambda}_n = 0. \quad (3.14)$$

Substituting the latter equation in Eq. (3.13) gives:

$$\sqrt{n}h(\hat{\theta}_n) \approx -\frac{\partial h(\theta_0)}{\partial \theta'} \mathcal{I}(\theta_0)^{-1} \frac{\partial h'(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} \frac{\hat{\lambda}_n}{\sqrt{n}} \approx -\frac{\partial h(\theta_0)}{\partial \theta'} \mathcal{I}(\theta_0)^{-1} \frac{\partial h'(\theta_0; \mathbf{y})}{\partial \theta} \frac{\hat{\lambda}_n}{\sqrt{n}}$$

which yields:

$$\frac{\hat{\lambda}_n}{\sqrt{n}} \approx -\left(\frac{\partial h(\theta_0)}{\partial \theta'} \mathcal{I}(\theta_0)^{-1} \frac{\partial h'(\theta_0; \mathbf{y})}{\partial \theta} \right)^{-1} \sqrt{n}h(\hat{\theta}_n). \quad (3.15)$$

It follows, from Eq. (3.8), that:

$$\frac{\hat{\lambda}_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, \left(\frac{\partial h(\theta_0)}{\partial \theta'} \mathcal{I}(\theta_0)^{-1} \frac{\partial h'(\theta_0; \mathbf{y})}{\partial \theta} \right)^{-1}\right).$$

Taking the quadratic form of the last equation gives:

$$\frac{1}{n} \hat{\lambda}'_n \frac{\partial h(\hat{\theta}_n^0)}{\partial \theta'} \mathcal{J}(\hat{\theta}_n^0)^{-1} \frac{\partial h'(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} \hat{\lambda}_n \xrightarrow{d} \chi^2(r).$$

Using Eq. (3.14), it appears that the left-hand side term of the last equation is ξ^{LM} as defined in Eq. (??). Consistency: see Remark 17.3 in Gouriéroux and Monfort (1995). \square

3.4.4 Proof of Proposition ??

Proof. Let us first demonstrate the asymptotic equivalence of ξ^{LM} and ξ^{LR} .

The second-order Taylor expansions of $\log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})$ and $\log \mathcal{L}(\hat{\theta}_n; \mathbf{y})$ are:

$$\begin{aligned} \log \mathcal{L}(\hat{\theta}_n; \mathbf{y}) &\approx \log \mathcal{L}(\theta_0; \mathbf{y}) + \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta'} (\hat{\theta}_n - \theta_0) - \frac{n}{2} (\hat{\theta}_n - \theta_0)' \mathcal{J}(\theta_0) (\hat{\theta}_n - \theta_0) \\ \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y}) &\approx \log \mathcal{L}(\theta_0; \mathbf{y}) + \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta'} (\hat{\theta}_n^0 - \theta_0) - \frac{n}{2} (\hat{\theta}_n^0 - \theta_0)' \mathcal{J}(\theta_0) (\hat{\theta}_n^0 - \theta_0). \end{aligned}$$

Taking the difference, we obtain:

$$\xi_n^{LR} \approx 2 \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta'} (\hat{\theta}_n - \hat{\theta}_n^0) + n (\hat{\theta}_n^0 - \theta_0)' \mathcal{J}(\theta_0) (\hat{\theta}_n^0 - \theta_0) - n (\hat{\theta}_n - \theta_0)' \mathcal{J}(\theta_0) (\hat{\theta}_n - \theta_0).$$

Using $\frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} \approx \mathcal{J}(\theta_0) \sqrt{n} (\hat{\theta}_n - \theta_0)$ (Eq. (3.11)), we have:

$$\xi_n^{LR} \approx 2n (\hat{\theta}_n - \theta_0)' \mathcal{J}(\theta_0) (\hat{\theta}_n - \hat{\theta}_n^0) + n (\hat{\theta}_n^0 - \theta_0)' \mathcal{J}(\theta_0) (\hat{\theta}_n^0 - \theta_0) - n (\hat{\theta}_n - \theta_0)' \mathcal{J}(\theta_0) (\hat{\theta}_n - \theta_0).$$

In the second of the three terms in the sum, we replace $(\hat{\theta}_n^0 - \theta_0)$ by $(\hat{\theta}_n^0 - \hat{\theta}_n + \hat{\theta}_n - \theta_0)$ and we develop the associated product. This leads to:

$$\xi_n^{LR} \approx n (\hat{\theta}_n^0 - \hat{\theta}_n)' \mathcal{J}(\theta_0)^{-1} (\hat{\theta}_n^0 - \hat{\theta}_n). \quad (3.16)$$

The difference between Eqs. (3.10) and (3.11) implies:

$$\frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} \approx \mathcal{J}(\theta_0) \sqrt{n} (\hat{\theta}_n - \hat{\theta}_n^0),$$

which, associated to Eq. (3.10), gives:

$$\xi_n^{LR} \approx \frac{1}{n} \frac{\partial \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta'} \mathcal{J}(\theta_0)^{-1} \frac{\partial \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} \approx \xi_n^{LM}.$$

Hence ξ_n^{LR} has the same asymptotic distribution as ξ_n^{LM} .

Let's show that the LR test is consistent. For this, note that:

$$\frac{\log \mathcal{L}(\hat{\theta}, \mathbf{y}) - \log \mathcal{L}(\hat{\theta}^0, \mathbf{y})}{n} = \frac{1}{n} \sum_{i=1}^n [\log f(y_i; \hat{\theta}_n) - \log f(y_i; \hat{\theta}_n^0)] \rightarrow \mathbb{E}_0[\log f(Y; \theta_0) - \log f(Y; \theta_\infty)],$$

where θ_∞ , the pseudo true value, is such that $h(\theta_\infty) \neq 0$ (by definition of H_1). From the Kullback inequality and the asymptotic identifiability of θ_0 , it follows that $\mathbb{E}_0[\log f(Y; \theta_0) - \log f(Y; \theta_\infty)] > 0$. Therefore $\xi_n^{LR} \rightarrow +\infty$ under H_1 .

Let us now demonstrate the equivalence of ξ^{LM} and ξ^W .

We have (using Eq. ??eq:multiplier):

$$\xi_n^{LM} = \frac{1}{n} \hat{\lambda}'_n \frac{\partial h(\hat{\theta}_n^0)}{\partial \theta'} \mathcal{J}(\hat{\theta}_n^0)^{-1} \frac{\partial h'(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} \hat{\lambda}_n.$$

Since, under H_0 , $\hat{\theta}_n^0 \approx \hat{\theta}_n \approx \theta_0$, Eq. (3.15) therefore implies that:

$$\xi^{LM} \approx nh(\hat{\theta}_n)' \left(\frac{\partial h(\hat{\theta}_n)}{\partial \theta'} \mathcal{J}(\hat{\theta}_n)^{-1} \frac{\partial h'(\hat{\theta}_n; \mathbf{y})}{\partial \theta} \right)^{-1} h(\hat{\theta}_n) = \xi^W,$$

which gives the result. \square

3.4.5 Proof of Eq. (??)

Proof. We have:

$$\begin{aligned} & T\mathbb{E}[(\bar{y}_T - \mu)^2] \\ &= T\mathbb{E}\left[\left(\frac{1}{T} \sum_{t=1}^T (y_t - \mu)\right)^2\right] = \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^T (y_t - \mu)^2 + 2 \sum_{s < t \leq T} (y_t - \mu)(y_s - \mu)\right] \\ &= \gamma_0 + \frac{2}{T} \left(\sum_{t=2}^T \mathbb{E}[(y_t - \mu)(y_{t-1} - \mu)] \right) + \frac{2}{T} \left(\sum_{t=3}^T \mathbb{E}[(y_t - \mu)(y_{t-2} - \mu)] \right) + \dots \\ &\quad + \frac{2}{T} \left(\sum_{t=T-1}^T \mathbb{E}[(y_t - \mu)(y_{t-(T-2)} - \mu)] \right) + \frac{2}{T} \mathbb{E}[(y_T - \mu)(y_{T-(T-1)} - \mu)] \\ &= \gamma_0 + 2 \frac{T-1}{T} \gamma_1 + \dots + 2 \frac{1}{T} \gamma_{T-1}. \end{aligned}$$

Therefore:

$$T\mathbb{E}[(\bar{y}_T - \mu)^2] - \sum_{j=-\infty}^{+\infty} \gamma_j = -2 \frac{1}{T} \gamma_1 - 2 \frac{2}{T} \gamma_2 - \dots - 2 \frac{T-1}{T} \gamma_{T-1} - 2\gamma_T - 2\gamma_{T+1} + \dots$$

And then:

$$\left| T\mathbb{E}[(\bar{y}_T - \mu)^2] - \sum_{j=-\infty}^{+\infty} \gamma_j \right| \leq 2\frac{1}{T}|\gamma_1| + 2\frac{2}{T}|\gamma_2| + \dots + 2\frac{T-1}{T}|\gamma_{T-1}| + 2|\gamma_T| + 2|\gamma_{T+1}| + \dots$$

For any $q \leq T$, we have:

$$\begin{aligned} \left| T\mathbb{E}[(\bar{y}_T - \mu)^2] - \sum_{j=-\infty}^{+\infty} \gamma_j \right| &\leq 2\frac{1}{T}|\gamma_1| + 2\frac{2}{T}|\gamma_2| + \dots + 2\frac{q-1}{T}|\gamma_{q-1}| + 2\frac{q}{T}|\gamma_q| + \\ &\quad 2\frac{q+1}{T}|\gamma_{q+1}| + \dots + 2\frac{T-1}{T}|\gamma_{T-1}| + 2|\gamma_T| + 2|\gamma_{T+1}| + \dots \\ &\leq \frac{2}{T}(|\gamma_1| + 2|\gamma_2| + \dots + (q-1)|\gamma_{q-1}| + q|\gamma_q|) + \\ &\quad 2|\gamma_{q+1}| + \dots + 2|\gamma_{T-1}| + 2|\gamma_T| + 2|\gamma_{T+1}| + \dots \end{aligned}$$

Consider $\varepsilon > 0$. The fact that the autocovariances are absolutely summable implies that there exists q_0 such that (Cauchy criterion, Theorem 3.1):

$$2|\gamma_{q_0+1}| + 2|\gamma_{q_0+2}| + 2|\gamma_{q_0+3}| + \dots < \varepsilon/2.$$

Then, if $T > q_0$, it comes that:

$$\left| T\mathbb{E}[(\bar{y}_T - \mu)^2] - \sum_{j=-\infty}^{+\infty} \gamma_j \right| \leq \frac{2}{T}(|\gamma_1| + 2|\gamma_2| + \dots + (q_0-1)|\gamma_{q_0-1}| + q_0|\gamma_{q_0}|) + \varepsilon/2.$$

If $T \geq 2(|\gamma_1| + 2|\gamma_2| + \dots + (q_0-1)|\gamma_{q_0-1}| + q_0|\gamma_{q_0}|) / (\varepsilon/2)$ ($= f(q_0)$, say) then

$$\frac{2}{T}(|\gamma_1| + 2|\gamma_2| + \dots + (q_0-1)|\gamma_{q_0-1}| + q_0|\gamma_{q_0}|) \leq \varepsilon/2.$$

Then, if $T > f(q_0)$ and $T > q_0$, i.e. if $T > \max(f(q_0), q_0)$, we have:

$$\left| T\mathbb{E}[(\bar{y}_T - \mu)^2] - \sum_{j=-\infty}^{+\infty} \gamma_j \right| \leq \varepsilon.$$

□

3.4.6 Proof of Proposition ??

Proof. We have:

$$\begin{aligned} \mathbb{E}([y_{t+1} - y_{t+1}^*]^2) &= \mathbb{E}([\{y_{t+1} - \mathbb{E}(y_{t+1}|x_t)\} + \{\mathbb{E}(y_{t+1}|x_t) - y_{t+1}^*\}]^2) \\ &= \mathbb{E}([y_{t+1} - \mathbb{E}(y_{t+1}|x_t)]^2) + \mathbb{E}([\mathbb{E}(y_{t+1}|x_t) - y_{t+1}^*]^2) \\ &\quad + 2\mathbb{E}([y_{t+1} - \mathbb{E}(y_{t+1}|x_t)][\mathbb{E}(y_{t+1}|x_t) - y_{t+1}^*]). \quad (3.17) \end{aligned}$$

Let us focus on the last term. We have:

$$\begin{aligned}
& \mathbb{E}([y_{t+1} - \mathbb{E}(y_{t+1}|x_t)][\mathbb{E}(y_{t+1}|x_t) - y_{t+1}^*]) \\
&= \mathbb{E}(\underbrace{(\mathbb{E}([y_{t+1} - \mathbb{E}(y_{t+1}|x_t)][\mathbb{E}(y_{t+1}|x_t) - y_{t+1}^*]|x_t))}_{\text{function of } x_t}) \\
&= \mathbb{E}((\mathbb{E}(y_{t+1}|x_t) - y_{t+1}^*)\mathbb{E}([y_{t+1} - \mathbb{E}(y_{t+1}|x_t)]|x_t)) \\
&= \mathbb{E}(\underbrace{(\mathbb{E}(y_{t+1}|x_t) - y_{t+1}^*)[\mathbb{E}(y_{t+1}|x_t) - \mathbb{E}(y_{t+1}|x_t)]}_{=0}) = 0.
\end{aligned}$$

Therefore, Eq. (3.17) becomes:

$$\begin{aligned}
& \mathbb{E}([y_{t+1} - y_{t+1}^*]^2) \\
&= \underbrace{\mathbb{E}([y_{t+1} - \mathbb{E}(y_{t+1}|x_t)]^2)}_{\geq 0 \text{ and does not depend on } y_{t+1}^*} + \underbrace{\mathbb{E}([\mathbb{E}(y_{t+1}|x_t) - y_{t+1}^*]^2)}_{\geq 0 \text{ and depends on } y_{t+1}^*}.
\end{aligned}$$

This implies that $\mathbb{E}([y_{t+1} - y_{t+1}^*]^2)$ is always larger than $\mathbb{E}([y_{t+1} - \mathbb{E}(y_{t+1}|x_t)]^2)$, and is therefore minimized if the second term is equal to zero, that is if $\mathbb{E}(y_{t+1}|x_t) = y_{t+1}^*$. \square

3.4.7 Proof of Proposition 2.1

Proof. Using Proposition ?? (in Appendix ??), we obtain that, conditionally on x_1 , the log-likelihood is given by

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

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

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

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:

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

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

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

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

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

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 $\Omega \xrightarrow{\ell} L(Y_T; \hat{\Pi}, \Omega)$. We have:

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

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. \square

3.4.8 Proof of Proposition 2.2

Proof. Let us drop the i subscript. Rearranging Eq. (2.12), 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_j^v . Using the fact that x_t is a linear combination of past ε_t s and that ε_t is a white noise, we get that $\mathbb{E}(\varepsilon_t x_t) = 0$. Therefore

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

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:

$$\gamma_0^v = \mathbb{E}(\varepsilon_t^2 x_t x'_t) = \mathbb{E}(\varepsilon_t^2) \mathbb{E}(x_t x'_t) = \sigma^2 \mathbf{Q}.$$

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_j^v computed above. \square

3.5 Additional codes

3.5.1 Simulating GEV distributions

The following lines of code have been used to generate Figure ??.

```
n.sim <- 4000
par(mfrow=c(1,3),
    plt=c(.2,.95,.2,.85))
all.rhos <- c(.3,.6,.95)
for(j in 1:length(all.rhos)){
  theta <- 1/all.rhos[j]
  v1 <- runif(n.sim)
  v2 <- runif(n.sim)
  w <- rep(.000001,n.sim)
  # solve for f(w) = w*(1 - log(w)/theta) - v2 = 0
  for(i in 1:20){
    f.i <- w * (1 - log(w)/theta) - v2
    f.prime <- 1 - log(w)/theta - 1/theta
    w <- w - f.i/f.prime
  }
  u1 <- exp(v1^(1/theta) * log(w))
  u2 <- exp((1-v1)^(1/theta) * log(w))

  # Get eps1 and eps2 using the inverse of the Gumbel distribution's cdf:
  eps1 <- -log(-log(u1))
  eps2 <- -log(-log(u2))
  cbind(cor(eps1,eps2),1-all.rhos[j]^2)
  plot(eps1,eps2,pch=19,col="#FF000044",
```

```

    main=paste("rho = ",toString(all.rhos[j]),sep=""),
    xlab=expression(epsilon[1]),
    ylab=expression(epsilon[2]),
    cex.lab=2,cex.main=1.5)
}

```

3.5.2 Computing the covariance matrix of IRF using the delta method

```

irf.function <- function(THETA){
  c <- THETA[1]
  phi <- THETA[2:(p+1)]
  if(q>0){
    theta <- c(1,THETA[(1+p+1):(1+p+q)])
  }else{
    theta <- 1
  }
  sigma <- THETA[1+p+q+1]
  r <- dim(Matrix.of.Exog)[2] - 1
  beta <- THETA[(1+p+q+1+1):(1+p+q+1+(r+1))]

  irf <- sim.arma(0,phi,beta,sigma=sd(Ramey$ED3_TC,na.rm=TRUE),T=60,y.0=rep(0,length(x),
    X=NaN,beta=NaN)
  return(irf)
}

IRF.0 <- 100*irf.function(x$THETA)
eps <- .00000001
d.IRF <- NULL
for(i in 1:length(x$THETA)){
  THETA.i <- x$THETA
  THETA.i[i] <- THETA.i[i] + eps
  IRF.i <- 100*irf.function(THETA.i)
  d.IRF <- cbind(d.IRF,
    (IRF.i - IRF.0)/eps
  )
}
mat.var.cov.IRF <- d.IRF %*% x$I %*% t(d.IRF)

```

Bibliography

- 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 Elias, 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.
- 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.
- 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. and Monfort, A. (1995). *Statistics and Econometric Models*, volume 1 of *Themes in Modern Econometrics*. Cambridge University Press.
- 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.
- 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.
- 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.
- Korobilis, D. (2013). Var forecasting using bayesian variable selection. *Journal of Applied Econometrics*, 28(2):204–230.
- 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.

- 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.
- Pfaff, B. (2008). Var, svar and svec models: Implementation within R package vars. *Journal of Statistical Software*, 27(4).
- 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. (2016). Macroeconomic Shocks and Their Propagation. NBER Working Papers 21978, National Bureau of Economic Research, Inc.
- Romer, C. D. and Romer, D. H. (2004). A New Measure of Monetary Shocks: Derivation and Implications. *American Economic Review*, 94(4):1055–1084.
- Ruibo-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. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.
- Sims, C. A. (1986). Are forecasting models usable for policy analysis? *Quarterly Review*, 10(Win):2–16.
- Stock, J. and Watson, M. (2016). Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2 of *Handbook of Macroeconomics*, chapter 0, pages 415–525. Elsevier.
- 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.