Advanced Econometrics

Jean-Paul Renne

2022-09-05

Contents

Chapter 1

Prerequisites

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")
# or the development version
# devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading `#`.

To compile this example to PDF, you need XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): https://yihui.name/tinytex/.

Chapter 2

Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter **??**.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2022) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

Below is an example borrowed from Petersen.

```
library(sandwich)
## Petersen's data
data("PetersenCL", package = "sandwich")
m <- lm(y ~ x, data = PetersenCL)
```

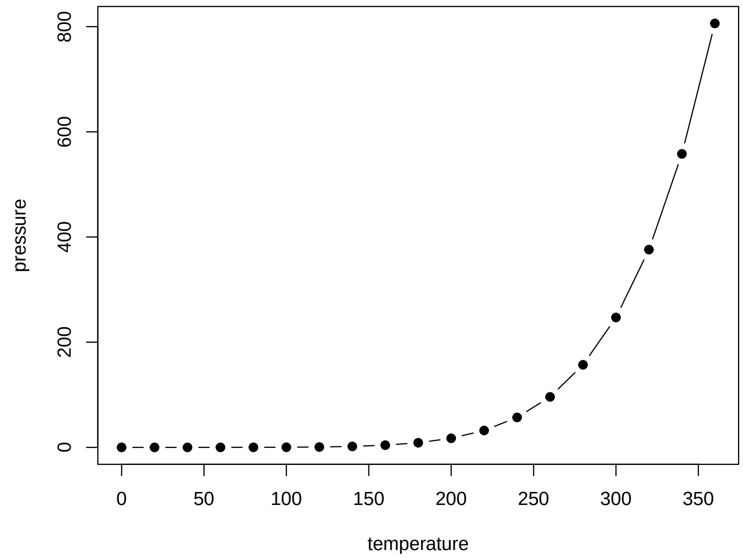

Figure 2.1: Here is a nice figure!

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

```r
## clustered covariances
## one-way
vcovCL(m, cluster = ~ firm)
```

```
##               (Intercept)             x
## (Intercept)  4.490702e-03 -6.473517e-05
## x           -6.473517e-05  2.559927e-03
```

```r
vcovCL(m, cluster = PetersenCL$firm) ## same
```

```
##               (Intercept)             x
## (Intercept)  4.490702e-03 -6.473517e-05
## x           -6.473517e-05  2.559927e-03
```

```r
## one-way with HC2
vcovCL(m, cluster = ~ firm, type = "HC2")
```

```
##               (Intercept)             x
## (Intercept)  4.494487e-03 -6.592912e-05
## x           -6.592912e-05  2.568236e-03
```

```r
## two-way
vcovCL(m, cluster = ~ firm + year)
```

```
##               (Intercept)             x
## (Intercept)  4.233313e-03 -2.845344e-05
## x           -2.845344e-05  2.868462e-03
```

```r
vcovCL(m, cluster = PetersenCL[, c("firm", "year")]) ## same
```

```
##               (Intercept)             x
## (Intercept)  4.233313e-03 -2.845344e-05
## x           -2.845344e-05  2.868462e-03
```

XXXX

Sargan-Hansen () test. Sargan (1958) and Hansen (1982)

Durbin-Wu-Hausman test: Durbin (1954) / Wu (1973) / Hausman (1978)

Use R! is an excellent tutorial. (notably for `plm` and the Arellano-Bond example, 140 UK firms)

Program evaluation (very good survey): Abadie and Cattaneo (2018) Mostly harmless: Angrist and Pischke (2008)

Diff-in-Diff: Card and Krueger (1994)

Diff-in-Diff: Meyer et al. (1995) with data in the `wooldridge` package. See this page

Chapter 3

Vector Auto-Regressive (VAR) models

Kilian (1998) See this page

Sign restrictions: package, Danne (2015).

Pfaff (2008)

Hlavac (2022)

```
library(VAR.etp)
library(vars) #standard VAR models
data(dat) # part of VAR.etp 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,] -0.3177971 0.1318571 1.016190 -0.1310161 0.1495442 0.9587964 -0.01935124
## [2,] -0.3196310 0.1459888 0.961219 -0.1605511 0.1146050 0.9343938 -0.01672199
## [3,] -0.3196310 0.1459888 0.961219 -0.1605511 0.1146050 0.9343938 -0.01672199
```

VARs 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. In economics, VAR models are often employed in order to identify *structural* shocks. First, we will present VAR models. Second, we will study its *structural* extension (SVAR models).

Definition 3.1 ((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

$$\begin{array}{rrcll} VAR: & y_t & = & c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \varepsilon_t, & \varepsilon_t: \text{ (correlated) innovation} \\ SVAR: & y_t & = & c + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + B\eta_t, & \eta_t: \text{ ($\perp$) structural shock,} \end{array} \tag{3.1}$$

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

The first line of Eq. (3.1) corresponds to the **reduced-form** of the VAR model (**structural form** for the second line).

Eq. (3.1) can also be written:

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

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

Consequently:

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

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

$$y_t = \Pi' x_t + \varepsilon_t. \tag{3.2}$$

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

3.0.1 VAR estimation

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

Proposition 3.1 (MLE of a Gaussian VAR)**.** *If y_t follows a VAR(p) (see Definition 3.1), and if $\varepsilon_t \sim i.i.d.\,\mathcal{N}(0,\Omega)$, then the ML estimate of Π, denoted by $\hat{\Pi}$ (see Eq. (3.2)), is given by*

$$\hat{\Pi} = \left[\sum_{t=1}^{T} x_t x'_t\right]^{-1} \left[\sum_{t=1}^{T} y'_t x_t\right] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{3.3}$$

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:

$$y_{i,t} = \beta'_i x_t + \varepsilon_{i,t}, \tag{3.4}$$

(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. (3.4)*, i.e.:*

$$\hat{\Omega} = \frac{1}{T}\sum_{i=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t', \quad \textit{with } \hat{\varepsilon}_t = y_t - \hat{\Pi}' x_t. \tag{3.5}$$

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

Proof. See Appendix 4.4.8. □

As stated by Proposition 4.4.9, 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 3.1, we have:

$$\mathbf{b}_i = \beta_i + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon_i. \tag{3.6}$$

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

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

Proposition 3.2 (Asymptotic distribution of the OLS estimate of β_i)**.** *If y_t follows a VAR model, as defined in Definition 3.1, we have:*

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

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

$$\mathbf{Q} = \left[\begin{array}{ccccc} 1 & \mu' & \mu' & \dots & \mu' \\ \mu & \gamma_0 + \mu\mu' & \gamma_1 + \mu\mu' & \dots & \gamma_{p-1} + \mu\mu' \\ \mu & \gamma_1 + \mu\mu' & \gamma_0 + \mu\mu' & \dots & \gamma_{p-2} + \mu\mu' \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mu & \gamma_{p-1} + \mu\mu' & \gamma_{p-2} + \mu\mu' & \dots & \gamma_0 + \mu\mu' \end{array}\right]. \tag{3.7}$$

Proof. See Appendix 4.4.9. □

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 3.3 (Asymptotic distribution of the OLS estimates)**.** *If y_t follows a VAR model, as defined in Definition 3.1, we have:*

$$\sqrt{T}\left[\begin{array}{c} vec(\hat{\Pi}-\Pi) \\ vec(\hat{\Omega}-\Omega) \end{array}\right] \sim \mathcal{N}\left(0, \left[\begin{array}{cc} \Omega\otimes\mathbf{Q}^{-1} & 0 \\ 0 & \Sigma_{22} \end{array}\right]\right), \tag{3.8}$$

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 $\hat{\mathbf{Q}} = \frac{1}{T}\sum_{t=p}^{T} x_t x_t'$ and Σ with the matrix whose components are of the form $\hat{\sigma}_{i,l}\hat{\sigma}_{j,m}+\hat{\sigma}_{i,m}\hat{\sigma}_{j,l}$, where the $\hat{\sigma}_{i,l}$'s are the components of $\hat{\Omega}$.

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

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

The last term is:

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

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

$$\log\mathcal{L}(Y_T;\hat{\Pi},\hat{\Omega}) = -(Tn/2)\log(2\pi)+(T/2)\log\left|\hat{\Omega}^{-1}\right| - Tn/2. \tag{3.9}$$

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:

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

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:

$$\left[\begin{array}{c} y_t^{(1)} \\ y_t^{(2)} \end{array}\right] = \left[\begin{array}{cc} \Phi^{(1,1)} & \Phi^{(1,2)} \\ \Phi^{(2,1)} & \Phi^{(2,2)} \end{array}\right] \left[\begin{array}{c} y_{t-1}^{(1)} \\ y_{t-1}^{(2)} \end{array}\right] + \varepsilon_t.$$

One can easily test for block exogeneity of $y_t^{(2)}$ (say). The null assumption ca n be expressed as $\Phi^{(2,1)} = 0$ and $\Sigma^{(2,1)} = 0$.

Lag selection

In a VAR, adding 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, so-called **selection criteria** can be used (see Definition **??**). These criteria have to be minimized. That is, the best specification is the one giving the lowest criteria.

In the context of VAR models, using Eq. (3.9), we have:

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

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

Companion Form and Stability of a VAR process

Let us introduce vector y_t^*, whihc stacks the last p values of y_t:

$$y_t^* = \left[\begin{array}{cccc} y_t' & y_{t-1}' & \dots & y_{t-p+1}' \end{array}\right]',$$

Eq. (3.1) can then be rewritten in its companion form:

$$y_t^* = \underbrace{\left[\begin{array}{c} c \\ 0 \\ \vdots \\ 0 \end{array}\right]}_{=c^*} + \underbrace{\left[\begin{array}{cccc} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ I & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I & 0 \end{array}\right]}_{=\Phi} y_{t-1}^* + \underbrace{\left[\begin{array}{c} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{array}\right]}_{\varepsilon_t^*} \qquad (3.10)$$

Matrices Φ and $\Sigma^* = \mathbb{V}ar(\varepsilon_t^*)$ are of dimension $np \times np$. Σ^* is filled with zeros, except the $n \times n$ upper-left block that is equal to $\Sigma = \mathbb{V}ar(\varepsilon_t)$.

We then have:

$$
\begin{aligned}
y_t^* &= c^* + \Phi\left(c^* + \Phi y_{t-2}^* + \varepsilon_{t-1}^*\right) + \varepsilon_t^* \\
&= c^* + \varepsilon_t^* + \Phi(c^* + \varepsilon_{t-1}^*) + \dots + \Phi^k(c^* + \varepsilon_{t-k}^*) + \Phi^k y_{t-k}^*.
\end{aligned}
$$

If the eigenvalues of Φ are strictly within the unit circle, then Φ^k geometrically decays to the zero matrix and we get the following Wold decomposition for y_t:

$$
\begin{aligned}
y_t^* &= c^* + \varepsilon_t^* + \Phi(c^* + \varepsilon_{t-1}^*) + \dots + \Phi^k(c^* + \varepsilon_{t-k}^*) + \dots \\
&= \mu^* + \varepsilon_t^* + \Phi\varepsilon_{t-1}^* + \dots + \Phi^k\varepsilon_{t-k}^* + \dots, \qquad (3.11)
\end{aligned}
$$

where $\mu^* = (I - \Phi)^{-1}c^*$.

(It can also be seen that $\mu^* = [\mu', \dots, \mu']'$, where $\mu = (I - \Phi_1 - \dots - \Phi_p)^{-1}c$).

The unconditional variance of y_t can be derived from Eq. (3.11), exploiting the fact that the ε_t^* are serially uncorrelated:

$$
\mathbb{V}ar(y_t^*) = \Omega^* + \Phi\Omega^*\Phi' + \dots + \Phi^k\Omega^*\Phi'^k + \dots,
$$

with $\mathbb{V}ar(\varepsilon_t^*) = \Omega^*$.

The unconditional variance of y_t is the upper-left $n \times n$ block of matrix $\mathbb{V}ar(y_t^*)$.

Eq. (3.11) also implies that the Ψ_k matrices defining the IRFs (see Eq. (**??**)) are given by: $\Psi_k = \widetilde{\Phi^k}B$, where $\widetilde{\Phi^k}$ is the upper-left matrix block of Φ^k.

Granger Causality

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

Formally, let us denote three information sets:

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

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

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

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

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

where $\Phi_k^{(ij)}$ denotes the element (i, j) of Φ_k.

Then, $y_{1,t}$ is said not to Granger-cause $y_{2,t}$ if

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

Therefore the hypothesis testing is

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

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

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

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 (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 q-dimensional subset of y_t (with $q \ll n$). 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.

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

$$\left[\begin{array}{c} f_t \\ x_t \end{array}\right] = \Phi(L) \left[\begin{array}{c} f_{t-1} \\ x_{t-1} \end{array}\right] + v_t. \tag{3.12}$$

f_t e.g. $\equiv$ the first k principal components of y_t.

Standard identification techniques of structural shocks can be employed in Eq. (3.12): 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 y_t).

3.1 Identification Problem and Standard Identification Techniques

The Identification Issue*

In the previous section, we have seen how to estimate Ω 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. 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. Example for $n = 2$:

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

We 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)).

A Simple Example

Consider the following stylized economic dynamics:

$$\begin{array}{rcll}
g_t & = & \bar{g} - \lambda(i_{t-1} - \mathbb{E}_{t-1}\pi_t) + \underbrace{\sigma_d \eta_{d,t}}_{\text{demand shock}} & \text{(IS curve)} \\
\Delta\pi_t & = & \beta(g_t - \bar{g}) + \underbrace{\sigma_\pi \eta_{\pi,t}}_{\text{cost push shock}} & \text{(Phillips curve)} \\
i_t & = & \rho i_{t-1} + \left[\gamma_\pi \mathbb{E}_t \pi_{t+1} + \gamma_g (g_t - \bar{g})\right] & \\
 & & + \underbrace{\sigma_{mp} \eta_{mp,t}}_{\text{Mon. Pol. shock}} & \text{(Taylor rule),}
\end{array} \tag{3.13}$$

where:

$$\eta_t = \begin{bmatrix} \eta_{\pi,t} \\ \eta_{d,t} \\ \eta_{mp,t} \end{bmatrix} \sim i.i.d.\, \mathcal{N}(0, I). \tag{3.14}$$

Vector η_t is a vector of Gaussian {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 (3.13) could be rewritten in the form:

$$\begin{bmatrix} d_t \\ \pi_t \\ i_t \end{bmatrix} = \Phi(L) \begin{bmatrix} d_{t-1} \\ \pi_{t-1} \\ i_{t-1} + \end{bmatrix} + \underbrace{\underbrace{\begin{bmatrix} 0 & \bullet & 0 \\ \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet \end{bmatrix}}_{=B} \eta_t}_{=\varepsilon_t} \tag{3.15}$$

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 decompositions of $\mathbb{V}ar(\varepsilon_t) = \Omega_\varepsilon$ 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), ...

Without loss of generality, we can set $n_1 = n$, $n_2 = n-1$, ... 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.

Importantly, the Cholesky approach can be useful if you are essentially interested in one 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:

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

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

An equivalent approach consists in computing the Cholesky decomposition of BB' and keep only the column corresponding to the policy variable.

Long-run restrictions

A second type of restriction relates to 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 when y_t (including ΔGDP_t) follows a VAR(MA) process.

As was shown previously (Eq. (3.10)), one can always write a VAR(p) as a VAR(1). Consequently, let us focus on the VAR(1) case:

$$\begin{aligned}
y_t &= c + \Phi y_{t-1} + \varepsilon_t && (3.16)\\
&= c + \varepsilon_t + \Phi(c + \varepsilon_{t-1}) + \ldots + \Phi^k(c + \varepsilon_{t-k}) + \ldots \\
&= \mu + \varepsilon_t + \Phi\varepsilon_{t-1} + \ldots + \Phi^k\varepsilon_{t-k} + \ldots && (3.17)\\
&= \mu + B\eta_t + \Phi B\eta_{t-1} + \ldots + \Phi^k B\eta_{t-k} + \ldots, && (3.18)
\end{aligned}$$

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 $t \geq s$):

$$\begin{aligned}
(\tilde{y}_t - y_t) + (\tilde{y}_{t-1} - y_{t-1}) + \dots + (\tilde{y}_s - y_s) &= \\
(Id + \Phi + \Phi^2 + \dots + \Phi^{t-s})B\gamma. && (3.19)
\end{aligned}$$

Consider a shock on $\eta_{1,t}$, with a magnitude of 1. This shock is $\gamma = [1, 0, \ldots, 0]'$. Given Eq. @ref{eq:cumul), the long-run cumulative effect of this shock on the endogenous variables is given by:

$$(Id + \Phi + \ldots + \Phi^k + \ldots)B\left[\begin{array}{c}1\\0\\\vdots\\0\end{array}\right],$$

that is the first column of $\Theta \equiv (I + \Phi + \ldots + \Phi^k + \ldots)B$.

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

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

$$
\begin{aligned}
Y_t &= M_t - P_t + a.\theta_t && (3.20)\\
Y_t &= N_t + \theta_t && (3.21)\\
P_t &= W_t - \theta_t && (3.22)\\
W_t &= W \mid \left\{\mathbb{E}_{t-1} N_t = \overline{N}\right\}. && (3.23)
\end{aligned}
$$

To close the model, the authors assume the following dynamics for the money supply and the productivity:

$$
\begin{aligned}
M_t &= M_{t-1} + \varepsilon_t^d\\
\theta_t &= \theta_{t-1} + \varepsilon_t^s.
\end{aligned}
$$

In this context, it can be shown that

$$
\begin{aligned}
\Delta Y_t &= (\varepsilon_t^d - \varepsilon_{t-1}^d) + a.(\varepsilon_t^s - \varepsilon_{t-1}^s) + \varepsilon_t^s\\
u_t &= -\varepsilon_t^d - a\varepsilon_t^s
\end{aligned}
$$

Then, it appears that the demand shocks have no long-run cumulative impact on ΔY_t, the GDP growth, i.e. no long-term impact on output Y_t. The vector of endogenous variables is $y_t = [\Delta Y_t \quad u_t]'$ where ΔY_t denotes the GDP growth. Estimation data are quarterly, and span the period from 1950:2 to 1987:4; 8 lags are used in the VAR model.

Chapter 4

Appendix

4.1 Statistical Tables

4.2 Statistics: definitions and results

Definition 4.1 (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:

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

Definition 4.2 (Eigenvalues)**.** The eigenvalues of of a matrix M are the numbers λ for which:

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

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

Proposition 4.1 (Properties of the determinant)**.** *We have:*

- $|MN| = |M| \times |N|$.
- $|M^{-1}| = |M|^{-1}$.

Table 4.1: 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 4.2: 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 4.3: 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 4.4: 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

- *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 4.3 (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 4.2 (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 4.4 (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 4.4 for quantiles.)

Definition 4.5 (Student-t distribution)**.** Z follows a Student-t (or t) distribution with ν degrees of freedom (d.f.) if:

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

We have $\mathbb{E}(Z) = 0$, and $\mathbb{V}ar(Z) = \frac{\nu}{\nu-2}$ if $\nu > 2$. (See Table 4.2 for quantiles.)

Definition 4.6 (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 4.3 for quantiles.)

Definition 4.7 (Idempotent matrix)**.** Matrix M is idempotent if $M^2 = M$.

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

Proposition 4.3 (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 4.4 (Idempotent matrix and chi-square distribution)**.** *The rank of a symmetric idempotent matrix is equal to its trace.*

Proof. The result follows from Prop. 4.3, combined with the fact that the rank of a symmetric matrix is equal to the number of its nonzero eigenvalues. □

Proposition 4.5 (Constrained least squares)**.** *The solution of the following optimisation problem:*

$$\min_{\beta} \quad ||\mathbf{y} - \mathbf{X}\beta||^2$$
$$\textit{subject to } \mathbf{R}\beta = \mathbf{q}$$

is given by:

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

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

Proof. See for instance Jackman, 2007. □

Proposition 4.6 (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 $\varepsilon^r \mathbb{1}_{\{|X| \geq \varepsilon\}} \leq |X|^r$ and take the expectation of both sides. □

Definition 4.8 (Convergence in probability)**.** The random variable sequence x_n converges in probability to a constant c if $\forall \varepsilon$, $\lim_{n\to\infty} \mathbb{P}(|x_n - c| > \varepsilon) = 0$.

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

Definition 4.9 (Convergence in the Lr 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

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

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

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

Definition 4.10 (Almost sure convergence)**.** The random variable sequence x_n converges almost surely to c if $\mathbb{P}(\lim_{n\to\infty} x_n = c) = 1$.

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

Definition 4.11 (Convergence in distribution)**.** x_n is said to converge in distribution (or in law) to x if

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

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

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

Proposition 4.7 (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 4.8 (Implications of stochastic convergences)**.** *We have:*

$$\begin{array}{ccccc} \boxed{\xrightarrow{L^s}} & \underset{1\le r\le s}{\Rightarrow} & \boxed{\xrightarrow{L^r}} & & \\ & & \Downarrow & & \\ \boxed{\xrightarrow{a.s.}} & \Rightarrow & \boxed{\xrightarrow{p}} & \Rightarrow & \boxed{\xrightarrow{d}}. \end{array}$$

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:

$$F_n(x) = \mathbb{P}(X_n \le x, X \le x+\varepsilon) + \mathbb{P}(X_n \le x, X > x+\varepsilon) \le F(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon). \tag{4.1}$$

Besides,

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

which implies:

$$F(x-\varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \le F_n(x). \tag{4.2}$$

Eqs. (4.1) and (4.2) imply:

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

Taking limits as $n \to \infty$ yields

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

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

Proposition 4.9 (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 $\mathbb{P}(X_n < c-\varepsilon) \underset{n\to\infty}{\to} 0$ i.e. $\mathbb{P}(X_n \geq c-\varepsilon) \underset{n\to\infty}{\to} 1$ and $\mathbb{P}(X_n < c+\varepsilon) \underset{n\to\infty}{\to} 1$. Therefore $\mathbb{P}(c-\varepsilon \leq X_n < c+\varepsilon) \underset{n\to\infty}{\to} 1$, 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 $x_n \overset{p}{\to} 0$ but $x_n \overset{L^r}{\nrightarrow} 0$ because $\mathbb{E}(|X_n - 0|) = \mathbb{E}(X_n) = 1$.

Theorem 4.1 (Cauchy criterion (non-stochastic case)). *We have that $\sum_{i=0}^{T} a_i$ converges ($T \to \infty$) 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 4.2 (Cauchy criterion (stochastic case)). *We have that $\sum_{i=0}^{T} \theta_i \varepsilon_{t-i}$ converges in mean square ($T \to \infty$) 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 4.12 (Characteristic function). For any real-valued random variable X, the characteristic function is defined by:

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

Theorem 4.3 (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. 4.12) imply that:

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

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. □

Theorem 4.4 (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:*

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

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 $\phi_{Y_n}(u) \underset{n\to\infty}{\to} \exp\left(-\frac{1}{2}u^2\sigma^2\right)$, which is the characteristic function of $\mathcal{N}(0,\sigma^2)$. □

Proposition 4.10 (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 4.11. *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 follows from the observation that $\mathbf{x}'\mathbf{Ax} = (\mathbf{Ax})'(\mathbf{Ax})$, which is a function of $\mathbf{Ax}$. □

Proposition 4.12 (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. □

Theorem 4.5 (Cauchy-Schwarz inequality)**.** *We have:*

$$|\mathbb{C}ov(X, Y)| \leq \sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}$$

and, if $X \neq=$ 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 $Z = Y - \frac{\mathbb{C}ov(X,Y)}{\mathbb{V}ar(X)}X$. It is easily seen that $\mathbb{C}ov(X, Z) = 0$. Then, the variance of $Y = Z + \frac{\mathbb{C}ov(X,Y)}{\mathbb{V}ar(X)}X$ is equal to the sum of the variance of Z and of the variance of $\frac{\mathbb{C}ov(X,Y)}{\mathbb{V}ar(X)}X$, that is:

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

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

Definition 4.13 (Matrix derivatives)**.** Consider a fonction $f : \mathbb{R}^K \to \mathbb{R}$. Its first-order derivative is:

$$\frac{\partial f}{\partial \mathbf{b}}(\mathbf{b}) = \left[\begin{array}{c} \frac{\partial f}{\partial b_1}(\mathbf{b}) \\ \vdots \\ \frac{\partial f}{\partial b_K}(\mathbf{b}) \end{array}\right].$$

We use the notation:

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

Proposition 4.13. *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 4.14 (Asymptotic level)**.** An asymptotic test with critical region Ω_n has an asymptotic level equal to α if:

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

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

Definition 4.15 (Asymptotically consistent test)**.** An asymptotic test with critical region Ω_n is consistent if:

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

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 4.16 (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 4.14 (Properties of the Kullback discrepancy)**.** *We have:*

i. $I(f, f^*) \geq 0$

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

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

Proposition 4.15 (Square and absolute summability)**.** *We have:*

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

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 4.1 and therefore $\theta_j^2 < |\theta_j|$. □

4.3 Some properties of Gaussian variables

Proposition 4.16 (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

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

Proposition 4.17 (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 partiucular, 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 $\alpha = \dfrac{a-\mu}{\sigma}$ and $\beta = \dfrac{b-\mu}{\sigma}$.

Moreover:

$$\mathbb{E}(X|a \leq X < b) \quad = \quad \mu - \frac{\phi(\beta) - \phi(\alpha)}{Z}\sigma. \tag{4.3}$$

We also have:

$$\begin{aligned} & \mathbb{V}ar(X|a \leq X < b) \\ = \quad & \sigma^2\left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{Z} - \left(\frac{\phi(\beta) - \phi(\alpha)}{Z}\right)^2\right] \end{aligned} \tag{4.4}$$

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

$$\mathbb{V}ar(X|a < X) = \sigma^2\left[1 + \alpha\lambda(-\alpha) - \lambda(-\alpha)^2\right], \tag{4.5}$$

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

Consider the case where $a \to -\infty$ (i.e. the conditioning set is $X < b$) and $\mu = 0$, $\sigma = 1$. Then Eq. (4.3) gives $\mathbb{E}(X|X < b) = -\lambda(b) = -\dfrac{\phi(b)}{\Phi(b)}$, where λ is the function computing the inverse Mills ratio.

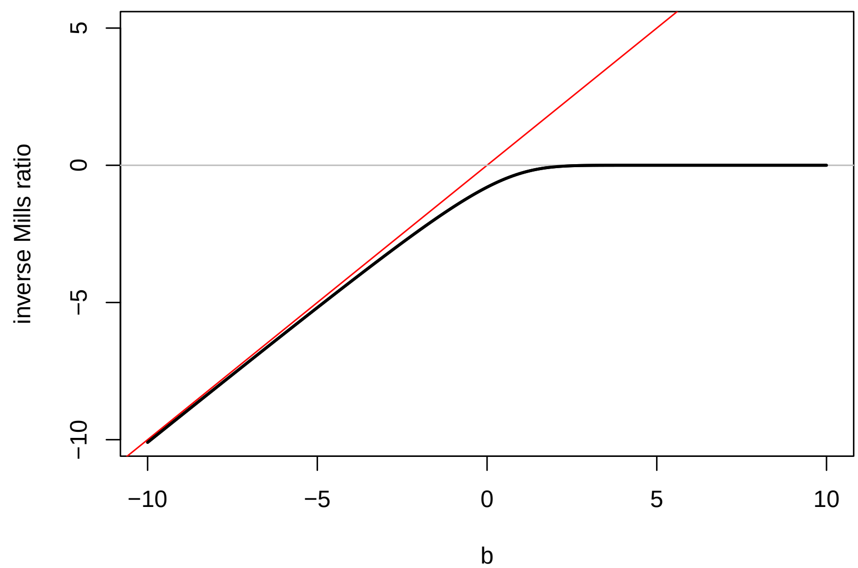

Figure 4.1: $\mathbb{E}(X|X < b)$ as a function of b when $X \sim \mathcal{N}(0,1)$ (in black).

Proposition 4.18 (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:*

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

4.4 Proofs

4.4.1 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:

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

Properties of the Kullback information measure (see Prop. 4.14), 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 neighborood 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). □

4.4.2 Proof of Proposition ??

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

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta_0)) \overset{d}{\rightarrow} \mathcal{N}\left(0, \frac{\partial h(\theta_0)}{\partial \theta'}\mathcal{I}(\theta_0)^{-1}\frac{\partial h(\theta_0)'}{\partial \theta}\right). \tag{4.6}$$

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

$$\sqrt{n}h(\hat{\theta}_n) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial h(\theta_0)}{\partial \theta'}\mathcal{I}(\theta_0)^{-1}\frac{\partial h(\theta_0)'}{\partial \theta}\right). \tag{4.7}$$

Hence

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

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. (4.6) is still valid. It implies that ξ_n^W converges to $+\infty$ and therefore that $\mathbb{P}_\theta(\xi_n^W \geq \chi^2_{1-\alpha}(r)) \to 1$. □

4.4.3 Proof of Proposition ??

Proof. Notations: "≈" 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:

$$\sqrt{n}h(\hat{\theta}_n) \approx \frac{\partial h(\theta_0)}{\partial \theta'}\sqrt{n}(\hat{\theta}_n - \hat{\theta}_n^0). \tag{4.8}$$

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

$$\frac{1}{\sqrt{n}}\frac{\partial \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta} \approx \frac{1}{\sqrt{n}}\frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} - \mathcal{I}(\theta_0)\sqrt{n}(\hat{\theta}_n^0 - \theta_0) \tag{4.9}$$

and:

$$0 = \frac{1}{\sqrt{n}}\frac{\partial \log \mathcal{L}(\hat{\theta}_n; \mathbf{y})}{\partial \theta} \approx \frac{1}{\sqrt{n}}\frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} - \mathcal{I}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0). \tag{4.10}$$

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

$$\sqrt{n}(\hat{\theta}_n - \hat{\theta}_n^0) \approx \mathcal{I}(\theta_0)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \log \mathcal{L}(\hat{\theta}_n^0; \mathbf{y})}{\partial \theta}\mathcal{I}(\theta_0). \tag{4.11}$$

Eqs. (4.8) and (4.11) yield to:

$$\sqrt{n}h(\hat{\theta}_n) \approx \frac{\partial h(\theta_0)}{\partial \theta'}\mathcal{I}(\theta_0)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \log \mathcal{L}(\hat{\theta}^0_n;\mathbf{y})}{\partial \theta}. \tag{4.12}$$

Recall that $\hat{\theta}^0_n$ 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:

$$\frac{\partial \log \mathcal{L}(\hat{\theta}^0_n;\mathbf{y})}{\partial \theta} + \frac{\partial h'(\hat{\theta}^0_n;\mathbf{y})}{\partial \theta}\hat{\lambda}_n = 0. \tag{4.13}$$

Substituting the latter equation in Eq. (4.12) gives:

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

which yields:

$$\frac{\hat{\lambda}_n}{\sqrt{n}} \approx -\left(\frac{\partial h(\theta_0)}{\partial \theta'}\mathcal{I}(\theta_0)^{-1}\frac{\partial h'(\theta_0;\mathbf{y})}{\partial \theta}\right)^{-1}\sqrt{n}h(\hat{\theta}_n). \tag{4.14}$$

It follows, from Eq. (4.7), that:

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

Taking the quadratic form of the last equation gives:

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

Using Eq. (4.13), 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). □

4.4.4 Proof of Proposition ??

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

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

Since, under H_0, $\hat{\theta}^0_n \approx \hat{\theta}_n \approx \theta_0$, Eq. (4.14) therefore implies that:

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

which gives the result. □

4.4.5 Proof of Proposition ??

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

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

Taking the difference, we obtain:

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

Using $\frac{1}{\sqrt{n}} \frac{\partial \log \mathcal{L}(\theta_0; \mathbf{y})}{\partial \theta} \approx \mathcal{I}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0)$ (Eq. (4.10)), we have:

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

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

$$\xi^{LR}_n \approx n(\hat{\theta}^0_n - \hat{\theta}_n)' \mathcal{I}(\theta_0)^{-1}(\hat{\theta}^0_n - \hat{\theta}_n). \tag{4.15}$$

The difference between Eqs. (4.9) and (4.10) implies:

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

which, associated to Eq. @**??**eq:lr10), gives:

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

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

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

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

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 $\xi^{LR}_n \to +\infty$ under H_1. □

4.4.6 Proof of Eq. (??)

Proof. We have:

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

Therefore:

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

And then:

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

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

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

Consider $\varepsilon > 0$. The fact that the autocovariances are absolutely summable implies that there exists q_0 such that (Cauchy criterion, Theorem 4.1):

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

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

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

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

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

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

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

$\square$

4.4.7 Proof of Proposition ??

Proof. We have:

$$\begin{aligned} \mathbb{E}([y_{t+1} - y^*_{t+1}]^2) &= \mathbb{E}\left([\{y_{t+1} - \mathbb{E}(y_{t+1}|x_t)\} + \{\mathbb{E}(y_{t+1}|x_t) - y^*_{t+1}\}]^2\right) \\ &= \mathbb{E}\left([y_{t+1} - \mathbb{E}(y_{t+1}|x_t)]^2\right) + \mathbb{E}\left([\mathbb{E}(y_{t+1}|x_t) - y^*_{t+1}]^2\right) \\ & \quad +2\mathbb{E}\left([y_{t+1} - \mathbb{E}(y_{t+1}|x_t)][\mathbb{E}(y_{t+1}|x_t) - y^*_{t+1}]\right). \end{aligned} \quad (4.16)$$

Let us focus on the last term. We have:

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

Therefore, Eq. (4.16) becomes:

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

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 $\mathbb{E}(y_{t+1}|x_t) = y^*_{t+1}$. $\square$

4.4.8 Proof of Proposition 3.1

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

which leads to the result. □

4.4.9 Proof of Proposition 3.2

Proof. Let us drop the i subscript. Rearranging Eq. (3.6), 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_ts 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. □

4.5 Additional codes

4.5.1 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)
}
```

4.5.2 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

Abadie, A. and Cattaneo, M. D. (2018). Econometric Methods for Program Evaluation. *Annual Review of Economics*, 10(1):465–503.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

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

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

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

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.

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.

Durbin, J. (1954). Errors in variables. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 22(1/3):23–32.

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.

Gouriéroux, C. and Monfort, A. (1995). *Statistics and Econometric Models*, volume 1 of *Themes in Modern Econometrics*. Cambridge University Press.

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

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

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Hlavac, M. (2022). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Social Policy Institute, Bratislava, Slovakia. R package version 5.2.3.

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

Meyer, B. D., Viscusi, W. K., and Durbin, D. L. (1995). Workers' Compensation and Injury Duration: Evidence from a Natural Experiment. *American Economic Review*, 85(3):322–340.

Pfaff, B. (2008). Var, svar and svec models: Implementation within R package vars. *Journal of Statistical Software*, 27(4).

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

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, 26(3):393–415.

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.

Wu, D.-M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica*, 41(4):733–750.

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2022). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.27.