Thesis

Novel Bayesian methods on multivariate cointegrated time series

University of Sheffield

Basile Marquier

October 4, 2017

Acknowledgements

Firstly I would like to thank everyone who supported me during the development of this thesis. I am very thankful to my parents, brothers and sisters in law, to my girlfriend Belle and to my little nieces and nephews. Of course, a lot of thoughts also go to my friends and colleagues I have met during these four years.

I would like to thank the University of Sheffield and the School of Mathematics and Statistics for having welcome me and funded me in order to do this project. In particular, I am very thankful to Dr. Keith Harris for having supported me and helped me in finalising this thesis. I would also like to thank Pr. Eleanor Stillman, Pr. Paul Blackwell and Pr. Caitlin Buck for their support throughout my studies in the department. I would like to thank my supervisor Dr. Kostas Triantafyllopoulos and my co-supervisor Dr. Miguel Juarez. Finally I would like to thank MASH, in particular Dr. Chetna Patel, Dr. Jennifer Freeman and Ellen Marshall, for having offered me the opportunity to teach statistics during this Ph.D.

Je remercie chaleureusement toutes les personnes qui m'ont aidé pendant l'élaboration de ma thèse. Je suis fortement redevable à mes parents, mes frères et mes belle-soeurs, ma compagne Belle ainsi que mes petites nièces et neveux. Je n'oublie pas bien sûr les amis et collègues que j'ai rencontrés au cours de ma vie.

Aussi, ce projet n'aurait pu voir le jour sans l'accueil chaleureux de l'Université de Sheffield, qui m'a permis, grâce à une allocation de recherches et diverses aides financières, de me consacrer à ce projet. Je suis tout particulièrement reconnaissant envers Dr. Keith Harris pour son soutien et son aide afin de réaliser ce travail. Je voudrais également remercier Pr. Eleanor Stillman, Pr. Paul Blackwell et Pr. Caitlin Buck pour leur soutien tout au long de ces études. Je voudrais ensuite remercier mon directeur de thèse Dr. Kostas Triantafyllopoulos et mon co-directeur de thèse Dr. Miguel Juarez. Enfin, je remercie MASH, en particulier Dr. Chetna Patel, Dr. Jennifer Freeman et Ellen Marshall, pour m'avoir offert la possibilité d'enseigner et de développer encore plus de compétences en statistiques durant ce doctorat.

Abstract

Many economic time series exhibit random walk or trend dynamics and other persistent non-stationary behaviour (*e.g.* stock prices, exchange rates, unemployment rate and net trading). If a time series is not stationary, then any shock can be permanent and there is no tendency for its level to return to a constant mean over time; moreover, in the long run, the volatility of the process is expected to grow without bound, and the time series cannot be predicted based on historical observations, see Diebold and Kilian (2001). Cointegration allows the identification of economic integrated time series that exhibit similar dynamics in the long run and the estimation of their relationships, by exploiting the stationary linear combinations of these time series, see Granger (1981).

This thesis proposes three Bayesian estimation methods of the well-known Vector Error Correction Model (VECM) about difference stationary time series in order to extract the long-run equilibrium relationships. Each method used in this thesis is implemented using Markov Chain Monte Carlo (MCMC) and illustrated on synthetic data, and then on real economic data sets. The first method consists of a static model, where we compare comovements between Eurozone economic time series comprising net trading, long-term interest rates and the harmonised unemployment rate. Primiceri (2005) established a time-varying model for the vector autoregressive model. Following Primiceri and the idea of the static model seen in the first method, we are constructing a time-varying model for our VECM, from which we extract information about the time-varying cointegration matrix, and more interestingly about its time-varying rank (i.e. the cointegration rank) and independent cointegration relationships. These two first methods are based on the singular value decomposition of the cointegration matrix from the error correction model and the so-called irrelevance criterion, a flexible thresholding approach to determine its rank. In these two methods, the joint estimation of the cointegration rank and the cointegration relationships is deducted from synthetic data sets before applying them to real data sets (European economies and major stock market exchange indices). The last main chapter of this thesis

covers the use of a prior singular distribution on the long-run relationship matrix of the VECM given the cointegration rank. Based on the definition of the singular matrix normal distribution proposed by Gupta and Nagar (2000), we also learn about the space definition and the density of such a distribution based on the work of Uhlig (1994) and Díaz-García *et al.* (2006). Gupta and Nagar (2000), Díaz-García *et al.* (1997) and Díaz-García and Gutiérrez-Jáimez (1997) also define the singular Inverse-Wishart distribution and in our discussion, we eventually open the issues arising in implementing a dynamic model, by developing the idea of a singular Inverse-Wishart distribution on the variance covariance matrix of the transition equation (see Chapter 6).

Contents

Notation

$p \in \mathbb{N}^\star$, $p \geq 2$, $1 \leq r < p$. In this thesis, p denotes the number of time series used in a data set. $T > p$ denotes the total length of the time series.

$\mathcal{M}_{p,n}(\mathbb{R})$ is the vector space of the real matrices of dimension $p \times n$ with $n \in \mathbb{N}^\star$.

$\mathcal{M}_{p,p}(\mathbb{R})$ is the vector space of the real square matrices of dimension $p \times p$.

$\mathcal{D}_p(\mathbb{R})$ is the vector space of the real diagonal matrices of dimension $p \times p$.

$\mathcal{GL}_p(\mathbb{R})$ is the vector space of the real invertible matrices of dimension $p \times p$.

$\mathcal{S}_p^+(r)$ is the set of $p \times p$ semidefinite positive matrices of rank r.

0_p is the null element of $\mathcal{M}_{p,p}(\mathbb{R})$. I_p represents the identity matrix of $\mathcal{M}_{p,p}(\mathbb{R})$.

$\forall A \in \mathcal{M}_{p,n}(\mathbb{R})$, A^+ will define the Moore-Penrose inverse of A. $\mathcal{O}_p$ is the group of orthogonal $p \times p$ matrices H, i.e. respecting $HH^T = H^T H = I_p$.

$\mathcal{V}_{r,p}$ is the set of matrices $H \in \mathcal{M}_{p,r}$ of full rank r and such that $H^T H = I_r$. $\mathcal{V}_{r,p}$ is called the Stiefel manifold.

$N_{p\times n}(M, Q, P)$ with $M \in \mathcal{M}_{p,n}(\mathbb{R})$, $P \in \mathcal{M}_{p,p}(\mathbb{R})$ and $Q \in \mathcal{M}_{n,n}(\mathbb{R})$ represents the matrix variate normal distribution seen in Chapter 2 of Gupta and Nagar (2000) with mean M and covariance matrix $P \otimes Q$.

$t_{p\times n}(M, P, Q, n)$ represents the matrix variate t-distribution seen in Chapter 4 of Gupta and Nagar (2000) with location matrix $M \in \mathcal{M}_{p,n}(\mathbb{R})$, scale matrices $P \in \mathcal{M}_{p,p}(\mathbb{R})$ and $Q \in \mathcal{M}_{n,n}(\mathbb{R})$ and degrees of freedom n.

$TN(\eta_1, \eta_2, \mu, s)$ with $\eta_1 < \eta_2$ denotes the truncated normal distribution with mean μ and standard deviation s and where η_1 and η_2 are respectively, the lower bound and the upper bound.

Chapter 1

Introduction

1.1 Background

Economic time series are in general considered as trend dynamics or having a non-stationary behaviour over time. Such economic time series include stock market prices, foreign exchange rates, or macroeconomic variables, such as the unemployment rate, net trading and others. If a time series evolves as a random walk, then in the long run, the process will not be stable and the time series will grow or decrease without any limit. In this case it will become hard to predict the behaviour of these time series based on historical data, see Diebold and Kilian (2001). The principle of cointegration established by Granger (1981) allows the identification of economic time series that exhibit similar dynamics in the long run and the estimation of their relationships. This similarity is studied via the mean-reversion or stationarity of linear combinations of several time series.

Cointegration occurs for a set of integrated time series of order m if we can find a linear combination between them, that is integrated of lower order $d < m$, see Engle and Granger (1987), Johansen (1996) and Johansen (1997). But in this thesis we will only consider difference stationary time series (integrated of order 1), as it is often the case in econometrics, and we will propose methodologies to establish cointegration relationships between them that are stationary

(integrated of order 0). The cointegration rank, denoted generally as r in this thesis, represents the number of independent linear combinations satisfying the property of stationarity. Our set of economic time series will then be said to be cointegrated of rank r, see Engle and Granger (1987). After the groundbreaking work of Granger (1981) and Engle and Granger (1987), there has been a large literature on cointegration, see Johansen (1988), Johansen (1991), Johansen (1997), Phillips and Perron (1988) and Phillips (1991). Among Bayesian analysts, one can mention the works of Villani (2000), Strachan (2003), Kleibergen and van Dijk (1994) and Bauwens and Lubrano (1996). As for non-Bayesian analysts, Johansen (1997) developed two tests in order to evaluate the cointegration rank in a set of time series. These two tests are widely used for any study about cointegration.

A method that can come to mind in order to study the comovement between difference stationary time series is to conduct an Ordinary Least Squares (OLS) regression. For instance let us consider an OLS regression between 2 non-stationary time series x_t and y_t:

$$y_t = \beta x_t + u_t$$

with u_t representing the error terms and β the slope of the regression. If our time series are difference stationary, then the error terms may be non-stationary and the regression would therefore incorrectly reject the null hypothesis $H_0 : \ \beta = 0$. This comes from the fact that the estimator $\hat{\beta}$ of the slope would not actually follow a Student distribution, leading to a spurious regression, see Banerjee *et al.* (1993), Damghani *et al.* (2012) and Granger and Newbold (1974). The regression results are then wrong, leading to a misinterpretation of the value of the coefficient β. However, if we regress instead Δy_t with Δx_t by OLS: $\Delta y_t = \gamma \Delta x_t + v_t$ with v_t as a white-noise process, then because Δy_t and Δx_t are stationary, the estimator $\hat{\gamma}$ will be consistent.

By estimating the VECM, we actually take the lag difference Δx_t of our VAR model x_t and determine the cointegrating matrix with a Bayesian approach (see Section 2.3). We will then avoid the case of spurious regression. One of the key points in the Bayesian approach is to choose a suitable prior distribution for the parameters of the VECM we want to estimate. We will see that the prior distribution implies the use of other parameters called hyperparameters, of which

one has to choose suitable estimates by appealing to certain methods (see for instance Section 3.3.2). Luetkepohl (2006) gives a method to estimate the parameters of the model that we will use to initialize the parameters in this thesis (see Section 2.5). The main parameter of interest in this thesis is the long-run relationships matrix obtained from the VECM. In this thesis, we focus on Bayesian inference of the cointegration matrix, by choosing a suitable prior and then derive a posterior distribution given the data and other parameters of the model. On the other hand, the likelihood can easily be derived from the distribution of the error terms in the VECM (see Section 3.3.1). Once both the likelihood and suitable priors are obtained, we can determine the full conditional posterior distributions of our parameters and run a Gibbs Sampler so that at the end we can get adequate estimates of the distribution of the parameters of the VECM.

1.2 Aim of the thesis and layout

This thesis introduces three methods to estimate the cointegration matrix: two methods are used to estimate static parameters (static methods) and the other method is about estimating time-varying parameters of the VECM (dynamic method). For the static methods as well as for the dynamic method, we will consider non-singular and singular Bayesian inferences around the cointegration matrix. In the non-singular Bayesian methods, we estimate the cointegration rank based on the singular values of the cointegration matrix (see Chapters 3 and 4). In the singular method, we define a singular prior for the long-run impact matrix conditional on the rank (see Chapter 5).

First of all, a static model is used to compare the economies of the 4 biggest countries of the eurozone before and after the introduction of the single currency (see Chapter 3). This method starts from the VECM, from which priors are initialised on the parameters. In that model, we consider the long-run relationships matrix and the matrix of lag parameters as each having a non-singular multivariate normal prior distribution each, given the covariance matrix of the errors. This latter parameter will have an Inverse-Wishart prior distribution. Based on the likelihood of our model, we find the three full conditional distributions wanted. But we decide to integrate

out the covariance matrix, in order to obtain matrix t-distributions for both the cointegrating matrix and the lag parameters matrix. This allows the Markov Chain Monte Carlo algorithm to run faster. For each cointegrating matrix simulated, an estimation of the rank is given by assessing the number of its most irrelevant singular values.

The second method consists of estimating a time-varying VECM (see Chapter 4), that we call the dynamic model, in order to differentiate it from the static model, seen in the previous chapter. For that, a Forward Filtering Backward Recursion algorithm is employed. In that method we still define a non-singular distribution for the cointegrating matrix. Since the cointegration matrix is time-varying, then the cointegration rank is also evolving over time. Then by using the same estimation of the rank on the time-varying cointegrating matrix as in the static method of Chapter 3, we are able to estimate a dynamic rank. Then from this dynamic rank and the dynamic cointegration matrix, we can easily obtain dynamic independent cointegrating relations. We can then obtain, for each time, an estimation of the rank and of the independent cointegrating relations. We test this novel method on a set of simulated data where we change on purpose the number of cointegrating relations over time. An application is then carried out on real data sets such as the European panel data seen in Chapter 3. We also decide to study the evolution of the cointegration rank in 3 sectors of the Dow Jones data set (see Chapter 4) from the year 2001 until 2009.

In Chapter 5, we go back to a static model but we establish a singular prior distribution on the long-run relationships matrix. Such a distribution may be called a reduced rank distribution. However, the prior singular distribution used on the cointegration matrix is based on knowledge of the rank. By fixing the rank of the prior, we achieve the property of conjugacy and obtain a full conditional singular posterior distribution for the cointegrating matrix, with reduced rank the same as defined in the prior. The prior distribution of the cointegrating matrix is a singular normal matrix distribution, see Gupta and Nagar (2000) and Díaz-García *et al.* (2006). The lag parameter matrix still has a non-singular normal prior and the covariance matrix of the errors will have an Inverse-Wishart distribution. In this chapter, we will not integrate out the

variance of the errors and we will therefore have three full conditional distributions: the singular posterior distribution of the cointegrating matrix, the non-singular posterior distribution of the lag parameter matrix and the non-singular posterior of the covariance matrix of the errors. Unlike the 2 previous methods, we cannot estimate the rank in the MCMC procedure. We use Johansen tests, see Johansen (1988), to assess the rank of the data, before running the algorithm. The property of conjugacy is verified and the MCMC algorithm uses a full conditional posterior singular distribution on the long-run impact matrix. The methods are applied on simulated data sets with a comparison with Chapter 3 and some real economic and financial data sets.

Chapter 2

Literature Review

2.1 Introduction to cointegration

In general, most financial time series are difference stationary. Over many decades econometricians have been interested in developing models to study economic time series behaviours, see Keynes (1936). If we think about mathematical models such as autoregressive models, a white-noise process is used to represent the error terms in such models. This section introduces the advantages in choosing a Gaussian distribution for the error terms. In general, analysis of market data has shown that we can model economic time series as random walks, evolving as difference stationary or unit root processes. The future of such time series cannot be predicted. This theory goes with the efficient-market hypothesis studied by Fama (1970):

$$x_t = x_{t-1} + u_t \ , \ u_t \sim N(0, \sigma^2)$$

where u_t is a sequence of i.i.d. random variables following a normal distribution with mean 0 and variance σ^2. With this hypothesis in mind, we can state that financial markets are informationally efficient and therefore, we cannot predict returns given the information available at the time of investment. But there also exist many other types of difference stationary time series. For example, the GDP of a country or its interest rate is not stationary, as it is typically exhibiting a trend.

However, in this thesis we believe in mathematical models in order to retrieve the facts that have occured in reality whether it is about the comovements in the Eurozone before and after the Euro (see Chapters 3 and 4) or in stock market indices (see Chapters 4 and 5). Financial data cannot be accurately predicted or evaluated only from geopolitical situations and intuition. Mathematical models are one of the most rational ways of predicting markets and often involves the knowledge of the trend in the past, or in a pre-sample before predicting more or less the behaviour of stock trends afterwards (e.g. comparing a moving-average time series and the real time series). For that reason and throughout many decades, investors have chosen to employ statisticians in order to make a decision on which stocks they should invest. Knowing events in the world and reading newspapers about finance are more than necessary in order to invest well, but it is not enough to understand comovements between such or such stocks. We can also bring that same argument to governments and their economies in the world.

The models proposed in this thesis are used to give an idea of which stocks or macroeconomic variables are potentially coevolving. Cointegration occurs when there exists a linear combination of these non-stationary time series, that is stationary. Cointegrating relations and their cointegrating coefficients can inform us about the coevolution between these trend time series, see Johansen (2005), and inform a lot on market investment strategies. One of the recurring methods in which cointegration occurs is the pair trading strategy, see Schmidt (2008) and Rad *et al.* (2016). For instance, let us consider a pair of 2 stocks (Coca Cola and Pepsi) that are cointegrated over a training sample. Suppose now that at this present date, the Pepsi index goes down compared to the Coca-Cola index. We therefore think that since both of these stocks are cointegrated, there is a high chance that the Pepsi index will go back up again, in order to catch up with the Coca Cola index and to satisfy the long-run cointegrating relationship. In that case, the investor might want to short sell a certain amount of stocks from Coca-Cola (when the value of Pepsi suddenly goes downwards), in order to buy at the same time for the same price a certain amount of stocks from Pepsi. At the time when Pepsi goes back up again, the investor can then sell the stocks from Pepsi and pocket a profit.

Cointegration in other areas than economics and finance

Cointegration is often used in econometrics or finance due to the frequency of integrated time series we encounter in these two fields. However, it is interesting to know that numerous works about cointegration have been applied to other fields such as, for instance, biology, socio-economic and environmental science panel data. Many fields can contain integrated time series and therefore cointegration techniques come naturally to assess relations between variables.

For instance, Chintrakarn and Herzer (2012) used panel cointegration techniques to investigate the effect of income inequality on crime in the United States. Kaufmann and Stern (2002) used cointegration to study co-movements between hemispheric temperature and the radiative forcing: solar irradiance, greenhouse gases, and tropospheric sulfates. Ostergaard *et al.* (2017) applied cointegration to a system of linearly phase coupled oscillating processes.

2.2 Definitions of cointegration

In this section, we recall general definitions on weakly stationary time series and the order of integration for a given time series. Any time series $(x_t)_{1\leq t\leq T}$ of length T will be denoted as x_t in this section. We also assume that any time series has a known initial value x_0.

Definition 1. ***Stationary time series***

Let x_t be a real time series. x_t is said to be weakly stationary, if: $\forall\ t \in [\![1, T]\!]$,

$$\exists\, \mu \in \mathbb{R}\ ,\ \textit{such that } E[x_t] = \mu$$

$$\exists\, \sigma \in \mathbb{R}\ ,\ \textit{such that } Var[x_t] = \sigma^2$$

$$\forall\ h \in \mathbb{Z},\ \exists\, \gamma_h \in \mathbb{R}\ ,\ \textit{such that } Cov[x_{t+h}, x_t] = \gamma_h$$

Definition 2. ***Order of integration*** 0***: Stationary***

Let x_t be a real time series. x_t is said to be integrated of order 0 , denoted $x_t \sim I(0)$, or trend stationary if x_t is stationary.

Definition 3. ***Order of integration*** 1***: Difference stationary***

Let x_t be a real time series. x_t is said to be integrated of order 1 , denoted $x_t \sim I(1)$, or difference stationary if $\Delta x_t = x_t - x_{t-1}$ is stationary, i.e., $x_t \sim I(1)$ iff $\Delta x_t \sim I(0)$.

Remark 1. *$x_t \sim I(0) \Rightarrow x_t \sim I(1)$ is true **but** $x_t \sim I(1) \Rightarrow x_t \sim I(0)$ is wrong!*

A vector u_t of time series integrated of order 1 is a vector for which each component is integrated of order 1. It is denoted as $u_t \sim I(1)$. We will use mostly the vector x_t to describe our time series $x_{i,t}$.

2.2.1 Theory of cointegration

In this section, we consider a set of p difference time series represented as a vector $x_t = (x_{it})_{1 \leq i \leq p}$, in which each component x_{it} represents the i^{th} time series of our group at time t.

Definition 4. ***Cointegration***

*A p-vector of **difference stationary time series** x_t ($x_t \sim I(1)$) is said to be **cointegrated** if there exists at least one non-zero p-vector β such that $\beta' x_t$ is trend stationary (i.e. $\beta' x_t \sim I(0)$). β is called a **cointegrating vector**.*

Definition 5. ***Cointegration rank***

*If there exists $r \in [\![1, p-1]\!]$ linearly independent vectors β_i, $i \in [\![1, r]\!]$, such that ${\beta_i}' x_t$ is stationary, then x_t is said to have **cointegration rank** r.*

*The r linearly independent p-vectors β_i are called **independent cointegrating relations**. They are stacked in a **cointegration matrix** denoted as $\beta = (\beta_1, \beta_2, ..., \beta_r)$ in this thesis.*

Granger (1981) and Engle and Granger (1987) were the first to develop the notion of cointegration. Later on, Johansen (1988, 1991, 1996, 1997, 2005, 2006) will introduce several statistical tests to determine the cointegration rank. In addition, the cointegration rank can be thought as an index of how well time series are co-evolving: the bigger the cointegration rank is in a set of time series, the more comovements will be present in that set of time series.

2.2.2 The Dickey-Fuller test: Testing stationarity of time series

Establishing cointegrating relationships requires the need to test if the linear combination of our time series, i.e. the cointegrating relation, is stationary. For that matter, several statistical tests are employed in order to evaluate if a time series is stationary or not: Phillips and Perron (1988), Kwiatkowski *et al.* (1992) (KPSS) and Dickey and Fuller (1979). We focus in this chapter on the Dickey Fuller test.

Let us first consider a simple autoregressive model AR(1) represented by:

$$x_t = \phi x_{t-1} + e_t \ , \ \forall t \in [\![1, T]\!]$$

where ϕ and x_0 are real numbers and where $e_t \overset{iid}{\sim} N(0, \sigma^2)$ with $\sigma^2 > 0$.

Let us now study the following three different cases leading to the conclusion if x_t is stationary or not:

1. $|\phi| < 1 \Rightarrow x_t$ is stationary:

   Since $|\phi| < 1$, then one has:

   $$x_t = \phi x_{t-1} + e_t \iff x_t - \phi x_{t-1} = e_t \iff (1 - \phi B) x_t = e_t$$

   with B being the backward shift operator. If $|\phi| < 1$ then $(1 - \phi B)$ is invertible and:

   $$(1 - \phi B)^{-1} = \sum_{i=0}^{\infty} \phi^i B^i$$

   Then we can write:

   $$x_t = (1 - \phi B)^{-1} e_t = \sum_{i=0}^{\infty} \phi^i B^i e_t = \sum_{i=0}^{\infty} \phi^i e_{t-i}$$

   Since all the e_t are independent we have $\forall t \in \mathbb{N}^*$,

   (a) $E[x_t] = \sum_{i=0}^{\infty} \phi^i E[e_{t-i}] = 0$,

   (b) $Var[x_t] = \sum_{i=0}^{\infty} \phi^{2i} Var[e_{t-i}] = \sum_{i=0}^{\infty} \phi^{2i} \sigma^2 = \sigma^2 \sum_{i=0}^{\infty} \phi^{2i}$ and then $Var[x_t] = \sigma^2 \frac{1}{1-\phi^2} = \frac{\sigma^2}{1-\phi^2}$.

(c) $\forall h \in \{1, ..., T-t\}$,

$$Cov[x_{t+h}, x_t] = Cov[\sum_{j=0}^{\infty} \phi^j e_{t+h-j}, \sum_{i=0}^{\infty} \phi^i e_{t-i}]$$
$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \phi^{i+j} Cov[e_{t+h-j}, e_{t-i}]$$

Then, since $Cov[e_{t-i}, e_{t+h-j}] \neq 0 \iff t-i = t+h-j \iff j = h+i$, then we have $\forall\ t$,

$$Cov[x_{t+h}, x_t] = \sum_{i=0}^{\infty} \phi^{2i+h} \sigma^2 = \phi^h \sigma^2 \sum_{i=0}^{\infty} \phi^{2i} = \frac{\phi^h \sigma^2}{1-\phi^2}$$

proving thereby stationarity.

2. $|\phi| = 1 \Rightarrow x_t$ is not stationary.

If $|\phi| = 1$, then $\phi = \pm 1$ and then we can write:

$$x_t = \sum_{i=0}^{t-1} \phi^i e_{t-i} + \phi^t x_0$$

Then we have:

$$\forall t > 0, \quad E[x_t] = \sum_{i=0}^{t-1} \phi^i E[e_{t-i}] + \phi^t x_0 = \phi^t x_0$$

and we can see that the expectation of x_t is not constant if and only if $x_0 \neq 0$, and then the vector of time series x_t is not stationary.

If $x_0 = 0$, then the expectation is constant and equal to 0. But in that case if we write the variance of x_t, we get:

$$Var[x_t] = Var[\sum_{i=0}^{t-1} \phi^i e_{t-i} + \phi^t x_0] = \sum_{i=0}^{t-1} \phi^{2i} Var[e_{t-i}] = \sigma^2 \sum_{i=0}^{t-1} 1 = \sigma^2 t$$

and we can clearly see that the variance of x_t is not constant and thus the time series is not stationary.

3. $|\phi| > 1 \Rightarrow x_t$ is not stationary.

In that case we can use similar computations to write the variance of x_t as:

$$Var[x_t] = \sum_{i=0}^{t-1} \phi^{2i} Var[e_{t-i}] = \sigma^2 \sum_{i=0}^{t-1} \phi^{2i}$$

Since $|\phi| > 1$, then $\phi^2 > 1$ and therefore the above time series diverges as t tends to infinity. Therefore we conclude that x_t is non-stationary.

The Dickey-Fuller distribution

Dickey and Fuller (1979) developed a test for detecting the presence of a unit root in an autoregressive model. If a unit root is present then it means that the process is not stationary. This test uses critical values corresponding to a distribution called the Dickey-Fuller distribution which depends also on the sample size.

Let us take the example of a process without deterministic terms based on T observations x_1, x_2,..., x_T:

$$x_t = \phi x_{t-1} + e_t \ , \ \forall t \in [\![1, T]\!]$$

where $e_t \overset{iid}{\sim} N(0, \sigma^2)$.

Given these T observations $x_1, x_2, ..., x_T$, the maximum likelihood estimator of ϕ is obtained from the moment condition, i.e. that the error terms e_t and the time series x_t are uncorrelated:

$$E[e_t x_{t-1}] = 0 \Longrightarrow \sum_{t=1}^{T} e_t x_{t-1} = \sum_{t=1}^{T} (x_t - \hat{\phi} x_{t-1}) x_{t-1} = 0$$

Hence:

$$\hat{\phi} = \frac{\sum_{t=1}^{T} x_t x_{t-1}}{\sum_{t=1}^{T} {x^2}_{t-1}}$$

The Dicky-Fuller test is based on the following statistic:

$$\hat{f} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})}$$

where $SE(\hat{\phi})$ is the usual standard error estimate. Then we can write $\hat{f}$ as:

$$\hat{f} = \frac{\hat{\phi} - 1}{S/(\sqrt{\sum_{t=2}^{T} {x^2}_{t-1}})} = \frac{\sum_{t=2}^{T} (\Delta x_t) x_{t-1}}{S\sqrt{\sum_{t=2}^{T} {x^2}_{t-1}}}$$

where S is the unbiased estimator of σ^2: $S^2 = \frac{1}{T-2}\sum_{t=2}^{T}(x_t - \hat{\phi}x_{t-1})^2$. Using the fact that S^2 is consistent, that is, S^2 converges in probability to σ^2 (as T takes large values), we obtain:

$$\hat{f} = \frac{\frac{1}{T}\sum_{t=2}^{T}(\Delta x_t)x_{t-1}}{S\sqrt{\frac{1}{T^2}\sum_{t=2}^{T}x^2{}_{t-1}}} \tag{2.1}$$

Phillips and Perron (1988) proved that the sample moments of $\{x_t\}$ converge to functions of Wiener processes:

$$A = T^{-1}\sum_{t=1}^{T} x_{t-1}\epsilon_t \xrightarrow{d} \sigma^2 \int_0^1 W(r)dW(r)$$

$$B = T^{-2}\sum_{t=1}^{T} x^2{}_{t-1} \xrightarrow{d} \sigma^2 \int_0^1 W(r)^2 dr$$

where $\xrightarrow{d}$ means "converges in distribution to". Now, as T tends to infinity we have:

$$\hat{f} = \frac{A}{S\sqrt{B}} \xrightarrow{d} X = \frac{\int_0^1 W(r)dW(r)}{\sqrt{\int_0^1 W(t)^2 dt}} \tag{2.2}$$

where $\{W(t), t \geq 0\}$ is the standard Wiener process, see Phillips and Perron (1988).

The asymptotic distribution of the Dickey-Fuller statistic $\hat{f}$ (2.1) is in fact a functional of the Wiener process. This asymptotic distribution of $\hat{f}$ is called the Dickey-Fuller (DF) distribution and does not have any closed form representation. The quantiles and critical values are therefore derived from numerical approximations or simulations (see Figure 2.1).

In order to approximate the Dickey-Fuller distribution, we will need to simulate several times a random walk of a given time length T:

$$x_t = x_{t-1} + e_t \ , \ \forall t \in [\![1, T]\!]$$

with $e_t \overset{iid}{\sim} N(0, \sigma^2)$.

If we simulate N random walks, then we can extract N values $\hat{f}$ (2.1), that will build our Dickey-Fuller distribution (see Figure 2.1). The approximation is more accurate when the number N gets bigger, that is the number of random walks simulated. From the N different values of $\hat{f}$ obtained, it is straightforward to derive some quantiles for the distribution created.

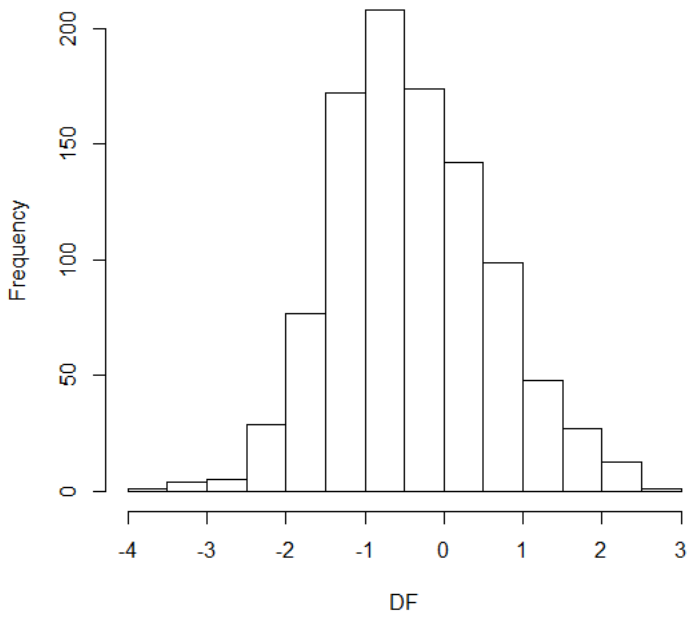

Figure 2.1: Simulated Dickey-Fuller distribution

The histogram in Figure 2.1 is constructed based on $1,000$ simulations of X from the expression derived in (2.2). Based on this histogram and with a lot of simulations of the random variable X, we are able to determine an approximation of the 10%, 5% and 1% quantiles.

Instruction and explanation of the test

From a data set measured from $t = 1$ to $t = T$ (with $T > 1$), we can calculate the statistic $\hat{f}$ from Equation (2.1). The null hypothesis of the Dickey-Fuller test is:

$$H_0 : \hat{f} \geq 0 \quad \text{i.e. } \hat{\phi} = 1, \text{ or } x_t \text{ is not stationary}$$

$$H_1 : \hat{f} < 0 \quad \text{i.e. } \hat{\phi} < 1, \text{ or } x_t \text{ is stationary}$$

From the simulated Dickey-Fuller distribution (see Figure 2.1), we obtain the quantiles -2.451, -1.992, -1.603, that is:

$$P(X < -2.451) = 0.01$$

$$P(X < -1.992) = 0.05$$

$$P(X < -1.603) = 0.10$$

If for example $\hat{f} = -1.7$ we will reject the null hypothesis of non-stationarity (i.e. consider the process x_t to be stationary) at risk 10%. On the other hand we will not reject the fact that it is not stationary at risk 5%. We can also directly look at the p-value of the test which corresponds to the value of $P(X \leq \hat{f})$. If that p-value is lower than 10% but greater than 5% we will draw the same conclusion.

The Augmented Dickey-Fuller test

Said and Dickey (1984) augmented the basic autoregressive unit root test to time series of unknown lag order $k > 1$. This test is called the Augmented Dickey-Fuller (ADF) test. The ADF test tests the null hypothesis that a process x_t is difference stationary against the alternative that x_t is trend stationary. This ADF test is based on estimating the test regression:

$$x_t = \phi x_{t-1} + \sum_{j=1}^{k} \psi_j \Delta x_{t-j} + \gamma' D_t + \epsilon_t$$

where D_t is a vector of deterministic terms (constant, trend, etc.). But again, as mentioned above the deterministic part of the equation is not taken into account throughout this thesis. The k terms Δx_{t-j} are called the lagged difference terms. Under the null hypothesis, x_t is difference stationary, i.e. $\Delta x_t = x_t - x_{t-1}$ is stationary, which implies $\phi = 1$.

In fact, an alternative formulation of the ADF test regression is:

$$\Delta x_t = \pi x_{t-1} + \sum_{j=1}^{k} \psi_j \Delta x_{t-j} + \gamma' D_t + \epsilon_t$$

where $\pi = \phi - 1$. Under the null hypothesis, Δx_t is stationary which entails that $\pi = 0$. The ADF test statistic is then the statistic for testing $\pi = 0$:

$$ADF_t = \frac{\hat{\pi}}{SE(\hat{\pi})} = \frac{\sum_{t=2}^{T} (\Delta x_t) x_{t-1}}{S \sqrt{\sum_{t=2}^{T} {x_{t-1}}^2}}$$

with $S^2 = \frac{1}{T-2}\sum_{t=2}^{T}(x_t - \sum_{j=1}^{k}\Delta x_{t-j})^2$ which is an unbiased estimator of σ^2 because:

$$\begin{aligned} E[S^2] &= E\Big[\frac{1}{T-2}\sum_{t=2}^{T}(x_t - \sum_{j=1}^{k}\Delta x_{t-j})^2\Big] \\ &= \frac{1}{T-2}\sum_{t=2}^{T}E[(x_t - \sum_{j=1}^{k}\Delta x_{t-j})^2] \\ &= \frac{1}{T-2}\sum_{t=2}^{T}E[\epsilon^2] \\ &= \frac{1}{T-2}(T-2)\sigma^2 = \sigma^2 \end{aligned}$$

2.3 Cointegration and Vector Error Correction Model

2.3.1 Introduction to the Vector Error Correction Model

Most of the research papers about cointegration use the Vector Error Correction Model to retrieve the cointegrating relations (see Engle and Granger (1987) and Villani (2005)). This model indeed provides a cointegrating matrix, from which independent relationships are derived as well as the cointegration rank. If we consider $(x_t)_{t=1}^{t=T}$ as a realization of the p-dimensional Vector Autoregressive (VAR) process of lag length $k \in \mathbb{N}^\star$, then:

$$x_t = \sum_{i=1}^{k}\Gamma_i x_{t-i} + \epsilon_t \tag{2.3}$$

with $\epsilon_t \sim N_p(0_p, \Sigma)$ and Σ a positive definite matrix ($\Sigma > 0$).

From there, we can obtain the Vector Error Correction Model (VECM) by taking the lag difference of order 1 $\Delta x_t = x_t - x_{t-1}$:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1}\Psi_i \Delta x_{t-i} + \epsilon_t \tag{2.4}$$

with $\epsilon_t \sim N_p(0_p, \Sigma)$.

From the parameters of the VAR model (2.3), we can obtain the parameters of the VECM

(2.4):

$$\Psi_j = -(\Gamma_{j+1} - ... - \Gamma_k) = -\sum_{i=j+1}^{k} \Gamma_i, \quad \forall j \in [\![1, k-1]\!] \tag{2.5}$$

$$\Pi = -(I_p - \Gamma_1 - \Gamma_2 - ... - \Gamma_k) = -(I_p - \sum_{j=1}^{k} \Gamma_j) \tag{2.6}$$

From Equation (2.4), we can isolate the term Πx_{t-1} on one side of the equation:

$$\Pi x_{t-1} = \Delta x_t - \sum_{i=1}^{k-1} \Psi_i \Delta x_{t-i} - \epsilon_t \tag{2.7}$$

Our vector of time series $x_t = (x_{it})_{1 \leq i \leq p}$ is composed of p integrated processes x_{it} of order 1: $x_{it} \sim I(1)$. Thus, each first difference lag vector of time series $\{\Delta x_{t-j}\}_{0 \leq j \leq k-1}$ will be a vector of stationary processes. Since the error processes vector $\epsilon_t = (\epsilon_{it})_{1 \leq i \leq p}$ is also composed of stationary signals $\epsilon_{it} \sim I(0)$, then by operation, the right hand side of (2.7) is also stationary. Therefore Πx_{t-1} is equal to a vector of stationary processes:

$$\Pi x_{t-1} = v_t = (v_{it})_{1 \leq i \leq p} \text{ with each } v_{it} \sim I(0) \tag{2.8}$$

The matrix Π, called the long-run impact matrix, is a cointegrating matrix from which each row constitutes a cointegrating vector in $\mathbb{R}^p$. Then, depending on the rank of Π, we have 3 cases.

Case 1: The cointegrating matrix Π is of full rank

If the cointegrating matrix Π is of full rank, i.e. the rank of Π is p, then Π is invertible and, from equation (2.8) above, we will have for each t: $x_{t-1} = \Pi^{-1} v_t$. From this point we can deduce by operation that each component x_{it} of x_t is a stationary time series.

In this particular case, Π is invertible and the rows of Π are p independent cointegrating relations.

Case 2: The cointegrating matrix Π is of lower rank $r \in [\![1, p-1]\!]$

In this case, the rank of Π is supposed to be not full and equal to $1 \leq r < p$. This rank is defined to be the cointegration rank. From the properties of the rank of a matrix, and the full

rank decomposition theorem, see Banerjee and Roy (2014), we know that we can derive at most r independent cointegrating relationships from Π. We can indeed decompose Π into a product $\alpha\beta'$ where α and β are 2 matrices of $\mathcal{M}_{p,r}(\mathbb{R})$ of full rank r, see Puntanen *et al.* (2011).

We can obtain from (2.8) that $\Pi x_{t-1} = \alpha\beta' x_{t-1} = v_t = (v_{it})_{1\leq i\leq p}$ with each $v_{it} \sim I(0)$. Then, by operation, we can obtain:

$$\beta' x_{t-1} = u_t = (u_{it})_{1\leq i\leq r} \text{ with each } u_{it} \sim I(0) \tag{2.9}$$

Therefore, the r independent cointegrating relations are obtained from the r rows of the matrix β', or the r columns of matrix β. Matrix β will therefore give the independent cointegrating relations.

Case 3: The cointegrating matrix Π is of rank 0

If the rank of Π is 0, then Π is the null element of $\mathcal{M}_{p,p}(\mathbb{R})$, and there is no cointegration.

2.3.2 Method to stack the data of the VECM in this thesis

This section describes how we stack the data so that we can define a general likelihood of the VECM, taking into account all the data we have from time 1 to time T. We assume a lag order $k \geq 2$ to be known for the VAR model from which the VECM is defined. We have:

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{k-1} \Psi_j \Delta x_{t-j} + \epsilon_t \tag{2.10}$$

Firstly, we define the $p \times p\ (k-1)$ matrix Ψ gathering the lag parameter matrices of the VECM given by Equation (2.10):

$$\Psi = [\Psi_1, \Psi_2, \ldots, \Psi_{k-1}]$$

Then we define for each time $t \in [\![1, T]\!]$, the vector z_t, of size $p\ (k-1)$, containing respectively $[\Delta x_{t-1},\ \Delta x_{t-2},\ \ldots\ ,\ \Delta x_{t-k+1}]$, i.e.

$$z_t = \begin{bmatrix} \Delta x'_{t-1} & \Delta x'_{t-2} & \cdots & \Delta x'_{t-k+1} \end{bmatrix}'$$

We can now write (2.10) as :

$$\Delta x_t = \Pi x_{t-1} + \Psi z_t + \epsilon_t \tag{2.11}$$

Then, by denoting each Δx_t as y_t, we can transpose both sides of the expression (2.11) and obtain:

$$y_t{}' = x_{t-1}{}'\Pi' + z_t{}'\Psi' + \epsilon_t{}' \tag{2.12}$$

Let us assume now that x_{-k}, x_{-k+1}, ..., x_0 exist and are provided. Then we can create the matrix Y of size $T \times p$, that gathers all the y_ts of Equation (2.12) from $t = 1$ to $t = T$:

$$Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_T \end{bmatrix}' = \begin{bmatrix} \Delta x_1 & \Delta x_2 & \cdots & \Delta x_T \end{bmatrix}'$$

We also create the matrix X of size $T \times p$, that gathers all x_{t-1} from $t = 1$ to $t = T$:

$$X = \begin{bmatrix} x_0 & x_1 & \cdots & x_{T-1} \end{bmatrix}'$$

Then, we create the matrix Z of size $T \times p\ (k-1)$, that gathers all z_t from $t = 1$ to $t = T$:

$$Z = \begin{bmatrix} z_1 & z_2 & \cdots & z_T \end{bmatrix}'$$

Finally the matrix of the errors E of size $T \times p$, that gathers all the errors ϵ_t from $t = 1$ to $t = T$, is:

$$E = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_T \end{bmatrix}' \tag{2.13}$$

Therefore, the t-th row of X, Y, Z and E are respectively x'_{t-1}, $\Delta x'_t$, $[\Delta x'_{t-1}, \ldots, \Delta x'_{t-k+1}]$ and ϵ'_t. In this thesis, we will often define "the information brought by the data" by the set $\mathcal{D} = \{X, Y, Z\}$. We can write the total VECM system, based on the data from $\mathcal{D}$, as:

$$Y' = \Pi X' + \Psi Z' + E' \tag{2.14}$$

Then, by transposing the above expression (2.14), we can also obtain:

$$Y = X\Pi' + Z\Psi' + E \tag{2.15}$$

2.3.3 Johansen tests: Frequentist estimation of the cointegration rank

Johansen (1991) elaborated two types of tests for cointegration. These tests in fact study, for each cointegration rank r assumed, if the r independent linear combinations for a set of p time series give stationary processes.

The two tests that Johansen elaborated are the Maximum Eigenvalue test and the Trace test. The Maximum Eigenvalue test examines the null hypothesis if the cointegration rank r is equal to a certain value r_0 against the alternate hypothesis that it is $r_0 + 1$. The Trace test examines the null hypothesis that the number of linear combinations r is equal to a given value r_0 against the alternative hypothesis that the cointegration rank r is greater than r_0.

2.3.4 Cointegrating relations and common trends

If the matrix Π has rank r, then we can decompose Π into the product of two $p \times r$ full rank matrices α and β as $\Pi = \alpha\beta'$. Then, since matrix β is of full rank, the columns of β will represent r independent cointegrating relations.

According to Johansen (1988) and Johansen (1991), we can find a $p \times (p-r)$ matrix $\beta_\perp$, that is orthogonal to β, i.e. $\beta'\beta_\perp = 0$, and such that the $p \times p$ matrix $G = [\beta, \beta_\perp]$ is invertible. The matrix $\beta_\perp$ is called the common trends loading matrix, see Johansen (1988), Johansen (1991) and Stock and Watson (1996).

If we have for example a vector of p time series x_t of which we find r independent cointegrating vectors stacked as the columns of β, then the space spanned by $\beta' x_t$ is called the cointegrated space of the set of time series x_t and the space spanned by ${\beta_\perp}' x_t$ is called the unit root space of x_t. The number of common trends is in fact equivalent to the number of time series subtracted to the number of independent cointegrating relations. Stock and Watson (1996) proposed tests to evaluate the number of common trends rather than the number of cointegrating relations. Their tests are applied to U.S. postwar interest rates.

2.4 Bayesian work on cointegration

Since the work of Sims (1988), who advocated the Bayesian paradigm for unit root testing, there has been a growing interest in Bayesian cointegration as evidenced by Schotman and van Dijk (1991), Kleibergen and van Dijk (1994), Strachan (2003), Bauwens and Lubrano (1996), Villani (2005), Conigliani and Tancredi (2009) and Meligkotsidou *et al.* (2014). A good review of the Bayesian approach to cointegration is given in Koop and Tobias (2006). Considering multivariate unit root testing, there appears to be two main points of interest (a) estimating the number of cointegrating relationships (i.e. the cointegration rank) and (b) estimating the coefficients which take part in these cointegrating relationships, usually adopting the vector error correction model (VECM) introduced by Engle and Granger (1987). For example, Villani (2005) estimates the parameters of the error correction model conditional on the cointegration rank, by splitting the cointegrating matrix Π into two full rank matrices α (matrix of adjustment coefficients) and β (matrix containing the independent cointegrating vectors). Given the cointegration rank, Villani then derives a full conditional posterior distribution for the parameters α, β, Ψ and Σ. Besides, Villani (2005) derives a posterior distribution conditional on the data for the cointegration rank.

A Bayesian analysis of cointegration is very useful because it produces a distribution rather than a point estimate of the parameters used to establish cointegration. We can obtain more information (credible intervals, mode, median, mean) than a simple estimate. Multivariate cointegration methods offer an important framework of identifying relationships between financial time series, hence are often exploited in developing long term decision making, trading and portfolio management. We propose in this thesis a Bayesian analysis of the Error Correction Model, using Markov Chain Monte Carlo methods (MCMC) in a static or dynamic context.

As part of the literature on Bayesian cointegration, Koop *et al.* (2006) provides a detailed summary of Bayesian methods developed in the last thirty years. We will recapitulate them in some of the next sections in this chapter (see Sections 2.4.2, 2.4.3, 2.4.4).

2.4.1 The Gibbs Sampler

The Gibbs sampler is a Markov Chain Monte Carlo algorithm (or MCMC), that allows us to obtain a sequence of observations that are approximately sampled from a specified multivariate probability distribution. This sequence is then used to approximate the joint posterior distribution of the parameters of a model. The idea of sampling comes from the physicist and researcher J. W. Gibbs who wanted to make an analogy between the sampling algorithm and statistical physics. The work of Casella and George (1992) gives more details about the idea of Gibbs sampling. Although the principles of the Gibbs algorithm were to be used in physics, they are also widely used in econometrics, and are also applicable to the mathematical models of many fields (biology, weather forecasting, etc.). Let us assume we have a model where we want to determine parameters θ_j, $j \in [\![1, h]\!]$. We decide to represent the set of parameters in a vector of parameters θ:

$$\theta = \begin{pmatrix} \theta_1 & \theta_2 & \cdots & \theta_h \end{pmatrix}'$$

The Gibbs sampler begins with an initial vector of parameters, called $\theta^{(0)}$. Those initial values will be assumed to be known at the beginning in order to explain the Gibbs sampling algorithm. Initial values can usually be determined from a pre-sample, but also given subjectively. We will recall later some methods developed in the thesis in order to have suitable initial conditions for the Vector Error Correction Model. The i^{th} draw of a parameter in the Gibbs sampler will be denoted by $\theta^{(i)}$. The i^{th} draw of θ is obtained by collecting sequentially and in the right order the h draws from the full conditional posteriors for θ_j , with $j = 1, ..., h$, where:

$$\theta_j^{(i)} \sim p(\theta_j | \theta_1^{(i)}, ..., \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, ..., \theta_h^{(i-1)}, \mathcal{D})$$

where $j = 1, ..., h$ and $i = 1, ..., g$. In the end we collect the i^{th} drawn vector

$$\theta^{(i)} = \begin{pmatrix} \theta_1^{(i)} & \theta_2^{(i)} & \cdots & \theta_h^{(i)} \end{pmatrix}'$$

$\theta^{(i)}$ is a draw from the joint posterior distribution $p(\theta|\mathcal{D})$, where $\mathcal{D}$ represents the data or the information brought by the likelihood in Bayes' theorem. It follows that each element $\theta^{(i)}$ is

simulated at each step i from:

$$\theta^{(i)} = (\theta_1^{(i)}, \cdots, \theta_{j-1}^{(i)}, \theta_j^{(i)}, \theta_{j+1}^{(i-1)}, \cdots, \theta_h^{(i-1)}) \sim p(\theta|\mathcal{D})$$

Hence, $\theta^{(i)}$ is also a draw from the joint posterior distribution of θ. Bauwens and Giot (1998) use a Gibbs sampling approach to cointegration by applying it to a cointegrated Vector Autoregressive system and therefore derive the cointegrating relations from a Bayesian perspective. Similarly, Villani (2005) makes use of the Gibbs sampling method to estimate the parameters of the Error Correction Model.

2.4.2 A prior on the cointegrating space

It may also be of interest to mention from the literature some Bayesian works around the possibility of setting a prior on the cointegrating space, i.e. the space spanned by the independent cointegrating vectors: $sp(\beta)$. Villani (2005) and Strachan and Inder (2004) adopted this novel approach where $sp(\beta)$ becomes the centre of interest rather than the values of the cointegrating coefficients β.

The cointegrating space $sp(\beta)$ is generally denoted as $\mathfrak{b}$ in the literature. For Strachan and Inder (2004) $\mathfrak{b} = sp(\beta)$ is actually a random parameter taking values in the Grassmann manifold $G_{r,p-r}$. We recall that $\mathcal{V}_{r,p}$ is the set of matrices $H \in \mathcal{M}_{p,r}$ of full rank r and such that $H^T H = I_r$ (orthogonal matrices). If the $p \times r$ matrix $\beta \in \mathcal{V}_{r,p}$, then the space spanned by the matrix β is in the Grassmann manifold $G_{r,p-r}$: $\mathfrak{b} = sp(\beta) \in G_{r,p-r}$.

Villani (2005) and Strachan and Inder (2004) use a uniform prior on $\mathfrak{b}$, which can be obtained from a prior distribution of β on $\mathcal{V}_{r,p}$. A draw from a uniform prior over $\mathcal{V}_{r,p}$ can be obtained by the operation $\beta = Z(Z'Z)^{-1/2}$ where $Vec(Z) \sim N(0, I_{pr})$. Then the space spanned by β will be uniformly distributed over $G_{r,p-r}$. Villani (2005) in Lemma 3.4 states that if we have $\beta = (I_r \; B')'$ with $B \sim t_{(p-r)\times r}(0, I_{p-r}, I_r, 1)$, then β will be uniformly distributed over the Grassman manifold $G_{r,p-r}$. Villani then derives a matrix variate normal full conditional posterior distribution for B, see Theorem 4.5, Villani (2005).

In addition, Strachan and Inder (2004) mentioned the possibility of having an informative

prior on the cointegrating space. We will assume that we have $p = 3$ economic time series x_{1t}, x_{2t} and x_{3t} with cointegration rank $r = 2$. If we have an idea of what the 2 independent cointegrating coefficients are (for instance, $x_{1t} - \beta_1 x_{2t} \sim I(0)$ and $x_{2t} - \beta_2 x_{3t} \sim I(0)$), then we can define the matrix H as:

$$H = \begin{bmatrix} 1 & 0 \\ -\beta_1 & 1 \\ 0 & -\beta_2 \end{bmatrix}$$

We have that $\mathfrak{p}^h = sp(H)$ is a value in $G_{r,p-r}$. Strachan and Inder (2004) propose a prior for the random parameter $\mathfrak{p}$ where its mass is centered around $\mathfrak{p}^h$. For that, they define the $p \times p$ random matrix P_τ as:

$$P_\tau = HH' + H_\perp H_\perp{}' \tau$$

where τ is chosen as being a scalar random variable normally distributed as $\tau \sim N(0, {\sigma_\tau}^2)$.

Then, by defining a random $p \times r$ matrix Z distributed as $Vec(Z) \sim N(0, I_{pr})$, we construct the matrix $X = P_\tau Z$ that can be decomposed later on as $X = \beta\kappa$, where κ is an $r \times r$ lower triangular matrix. The dispersion around $\mathfrak{p}^h$ is controlled by the chosen value of the variance of τ, i.e. ${\sigma_\tau}^2$.

2.4.3 A Bayesian estimation of the Error Correction Model including the cointegration rank: Villani (2005)

A very interesting paper on Bayesian estimation of the Vector Error Correction Model was developed by Villani (2005). Conditional on the cointegration rank r, Villani (2005) splits the cointegrating matrix Π into 2 full rank $p \times r$ matrices α and β and infers these two latter.

In Villani (2005), the VECM is then constructed as the following, conditional on the cointegration rank r:

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{k-1} \Psi_i \Delta x_{t-i} + \epsilon_t \tag{2.16}$$

with $\epsilon_t \sim N_p(0_p, \Sigma)$.

Villani (2005) then infers α, β, Ψ and Σ conditional on r by building first the joint prior distribution of these four parameters. The prior of Σ is an Inverse-Wishart: $\Sigma \sim IW(A, q)$ where $A > 0$ and q are two hyperparameters. A uniform prior is defined on $\Psi = [\Psi_1, \cdots, \Psi_{k-1}]$ and the prior of α is Gaussian and conditional on β and Σ: $\alpha \sim N_{p\times r}(0, (\beta'\beta)^{-1}, v^{-1}\Sigma)$ with v as a hyperparameter. As for β, Villani (2005) sets a uniform prior on the cointegration space by using the form $\beta = (I_r \ B')'$ with $B \sim t_{(p-r)\times r}(0, I_{p-r}, I_r, 1)$. Finally the joint distribution of α , β, Ψ and Σ conditional on r is given by:

$$f(\alpha, \beta, \Psi, \Sigma | r) \propto c_r |\Sigma|^{-\frac{p+r+q+1}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}(A + v\alpha\beta'\beta\alpha'))\right) \tag{2.17}$$

where c_r is a scalar depending on r.

The likelihood is constructed on the error terms of the VECM (2.16), each having a normal distribution with covariance matrix Σ: $\epsilon_t \sim N(0, \Sigma)$. Hence,

$$f(\mathcal{D}|\alpha, \beta, \Psi, \Sigma, r) \propto f(Vec(E')|\alpha, \beta, \Psi, \Sigma, r) \propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}E'E)\right) \tag{2.18}$$

where $E' = Y' - \Pi X' - \Psi Z'$ from equation (2.14) will contain α, β (by $\Pi = \alpha\beta'$) and Ψ.

As for the cointegration rank, the prior distribution $f(r)$ is a discrete uniform distribution over the space $[\![0, p]\!]$, i.e. $f(r = \nu) = \frac{1}{p+1}$ for each $\nu \in [\![0, p]\!]$. Villani (2005) then derives the posterior distribution of the cointegration rank that is conditional on the data $\mathcal{D}$ only:

$$f(r|\mathcal{D}) = \frac{f(\mathcal{D}|r)f(r)}{\sum_{r=0}^{p} f(\mathcal{D}|r)f(r)} \tag{2.19}$$

Villani (2005) obtains this posterior distribution of the cointegration rank by integrating out Σ, Ψ, α and β, in order to obtain the marginal likelihood of the data given the cointegration rank $f(\mathcal{D}|r)$:

$$f(\mathcal{D}|r) = \iiiint f(\mathcal{D}|\alpha, \beta, \Psi, \Sigma, r) f(\alpha, \beta, \Psi, \Sigma|r) \ d\Sigma \ d\Psi \ d\alpha \ d\beta \tag{2.20}$$

The cointegration rank r will therefore have a posterior distribution conditional on the data only (2.19): $f(r|\mathcal{D})$. According to a sensitivity analysis on the hyperparameters, the posterior

distribution of r will have more mass around certain values. Villani (2005) uses a bivariate process in order to assess the probability of three different values for the rank: $f(r = 0|\mathcal{D})$, $f(r = 1|\mathcal{D})$ and $f(r = 2|\mathcal{D})$. The other parameters of the VECM such as α and β are also estimated. Following the methods of Villani (2005), an application to the demand for the Euro Area was performed by Warne (2006).

2.4.4 The embedding approach on the Error Correction Model

Let us consider an Error Correction Model as in (2.16) in which the rank is established to be $r < p$ and the matrix Π is split into 2 full rank $p \times r$ matrices α and β. For this section, we will call that model the Error Correction Cointegration model (ECC model).

The embedding approach involves the existence of a parameter matrix that will evaluate the degree of the rank of Π. Let us now construct a model called the Unrestricted Error Correction model (UEC model) by adding a $(p-r) \times (p-r)$ parameter matrix λ making the long-run relations matrix Π to be of full rank. We can write the long-run relations matrix Π as:

$$\Pi = \alpha\beta' + \begin{bmatrix} 0 \\ I_{p-r} \end{bmatrix} \lambda \begin{bmatrix} 0 & I_{p-r} \end{bmatrix} \tag{2.21}$$

where $\alpha = \begin{bmatrix} {\alpha_1}' & {\alpha_2}' \end{bmatrix}'$ and $\beta' = \begin{bmatrix} I_r & B' \end{bmatrix}$ We have α_1 of size $r \times r$, α_2 of size $(p-r) \times r$ and B of size $(p-r) \times r$.

If we now write Π as:

$$\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} \tag{2.22}$$

where Π_{11} is of size $r \times r$, Π_{12} is of size $r \times (p-r)$, Π_{21} is of size $(p-r) \times r$ and Π_{22} is of size $(p-r) \times (p-r)$, then thanks to (2.22), we can identify the expressions of α, B and λ from Equation (2.21).

We have:

$$\begin{aligned} \alpha_1 &= \Pi_{11} \\ \alpha_2 &= \Pi_{21} \\ B' &= {\Pi_{11}}^{-1}\Pi_{12} \\ \lambda &= \Pi_{22} - \Pi_{21}{\Pi_{11}}^{-1}\Pi_{12} \end{aligned} \tag{2.23}$$

However, a problem of local non-identification can occur when Π_{11} is not invertible, in which case λ and B are diverging. An embedding model will be constructed in order to nest various Error Correction Cointegration models according to different values of r. The embedding model approach was first investigated by Kleibergen and van Dijk (1994) and then by Kleibergen and Paap (2002).

Kleibergen and Paap (2002) proposed a singular value decomposition of Π in the UEC model as:

$$\Pi = USV' = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} {V_{11}}' & {V_{21}}' \\ {V_{12}}' & {V_{22}}' \end{bmatrix} \tag{2.24}$$

where by definition U and V are orthonormal matrices and S is a diagonal matrix containing the singular values of Π (of the UEC model) in descending order. This implies that S_2 is a diagonal matrix containing the $p - r$ smallest singular values of Π.

By using (2.24) and (2.21), we can retrieve the parameters:

$$\begin{aligned} \alpha' &= V_{11}S_1[{U_{11}}', {U_{21}}'] \\ B &= {V_{11}}^{-1}V_{12} \\ \lambda &= ({V_{22}}'V_{22})^{-\frac{1}{2}}V_{22}S_2{U_{22}}'({U_{22}}'U_{22})^{-\frac{1}{2}} \end{aligned}$$

In that way, there is no local non-identification issue since ${V_{22}}'V_{22}$ and ${U_{22}}'U_{22}$ are both invertible (unitary matrices). Kleibergen and Paap (2002) then use the help of Bayes factors in order to assess a posterior probability distribution for the cointegration rank r. At first, they set a prior distribution for the rank, $P[\text{rank} = r]$, e.g. a uniform distribution over $[\![0, p]\!]$, and then

construct a prior odds ratio function consisting of the ratio between the prior probability of rank r and the prior probability of full rank $r = p$:

$$PROR[r|p] = \frac{P[\text{rank} = r]}{P[\text{rank} = p]} \tag{2.25}$$

A Bayes Factor $BF[r|p]$ for each rank r is then constructed as follows:

$$BF[r|p] = \frac{P[\mathcal{D}|\text{rank} = r]}{P[\mathcal{D}|\text{rank} = p]} \tag{2.26}$$

$P[\mathcal{D}|\text{rank} = r]$ is computed by integrating out Σ, Ψ, α and β from the joint posterior of the parameters of the ECC model $f_{ECC}(\alpha, \beta, \Psi, \Sigma|\mathcal{D})$ (conditional on the rank):

$$P[\mathcal{D}|\text{rank} = r] = \iiiint f_{ECC}(\alpha, \beta, \Psi, \Sigma|\mathcal{D})\ d\Sigma\ d\Psi\ d\alpha\ d\beta$$

$P[\mathcal{D}|\text{rank} = p]$ is computed by integrating out Σ, Ψ, α, β and λ from the posterior of the parameters of the UEC model $f_{UEC}(\lambda, \alpha, \beta, \Psi\Sigma|\mathcal{D})$:

$$P[\mathcal{D}|\text{rank} = p] = \int \cdots \int f_{UEC}(\lambda, \alpha, \beta, \Psi, \Sigma|\mathcal{D})\ d\Sigma\ d\Psi\ d\alpha\ d\beta\ d\lambda$$

If the Bayes Factor (2.26) is larger than 1, then the model of rank r is preferred to the full rank model. Thanks to Equations (2.25) and (2.26), we can obtain the posterior odds ratio (2.27) below:

$$POR[r|p] = PROR[r|p] \times BF[r|p] \tag{2.27}$$

Then, from all the posterior odds ratios $POR[r|p]$ derived for each $r \in [\![0, p]\!]$, we can obtain a posterior distribution for the rank r:

$$P[\text{rank} = r|\mathcal{D}] = \frac{POR[r|p]}{\sum_{r=0}^{p} POR[r|p]} \tag{2.28}$$

Kleibergen and Paap (2002) estimate the cointegration rank between real money supply M_2, real income, price level and costs of holding money in Denmark by using the posterior distribution of the rank given the data, i.e. Equation (2.28). They find a cointegration rank in favor of 1 for an Error Correction model considering a restricted constant by using the fact that the posterior probability $P[\text{rank} = 1|\mathcal{D}]$ is the highest. Besides, all restricted models are more likely than the full rank model because each of their Bayes factors (2.26) are higher than 1.

2.4.5 Bayesian estimation of the lag order of the model

As for the lag order k, it can enter into the model as a parameter on which Bayesian inference can be performed. Phillips (1996), Corander and Villani (2004) and Chao and Phillips (1999) estimate the posterior distribution for the lag order jointly with the cointegration rank.

2.4.6 Time-varying Bayesian estimation of the VECM

In the literature about time-varying Bayesian cointegration, it is worth mentioning a few works on the time-varying Error Correction Model (ECM). To start with, Granger and Lee (1991) introduced a time-varying cointegrated process that they applied to US data prices and wages. Bierens and Martins (2010) propose a time-varying ECM where the cointegrating relations change smoothly over a certain time period. These works assume a constant cointegration rank and focus more on the values of the cointegrating coefficients evolving over time.

Koop *et al.* (2011) also developed a dynamic ECM in which he considers the cointegrating space (see Section 2.4.2) evolving over time. He introduces a Markov Chain Monte Carlo procedure and an algorithm for state space models in order to infer the time-varying ECM. Koop *et al.* (2011) showed the behaviour over time of one cointegrating relation between US economic variables.

2.4.7 Bayesian cointegration on other models than the VECM

A significant amount of Bayesian works around cointegration do not involve the use of the Error Correction model , such as DeJong (1992), Dorfman (1994), Koop (1991) and Koop (1994). Let us consider a VAR model (2.3) for the p-dimensional process x_t composed of difference-stationary elements such as $x_t = \sum_{j=1}^{k} \Gamma_j x_{t-j} + \epsilon_t$ with $t \in [\![1, T]\!]$ and we can write the VAR representation as:

$$(I_p - \sum_{j=1}^{k} \Gamma_j L^j) x_t = \epsilon_t \tag{2.29}$$

where L^j is the lag operator raised to the integer power j.

The number of independent cointegrating relations is estimated from the number of non-stationary roots of the VAR model: the elements of x_t will be cointegrated if $(I_p - \sum_{j=1}^{k} \Gamma_j z^j)$ has $0 < p - r < p$ non-stationary unit roots. In such a case, there exist r cointegrating vectors β_i such that $\beta_i' x_t$ is stationary.

DeJong (1992) uses non-informative uniform priors on the Γ_j coefficients in order to derive posterior distributions from which draws are made thanks to Monte Carlo integration methods. We can obtain the roots of the VAR representation by constructing their posteriors based on the simulated Γ_j coefficients. By still using the approach of the number of non-stationary roots in a VAR model, Dorfman (1994) develops a Bayesian cointegration test focusing on the posterior odds of the number of unit roots in the set of integrated processes. DeJong (1992) and Koop (1991) use this approach to verify common behaviours between stock prices and dividends. Koop (1994) also detects common trends between spot and forward exchange rates from different countries (USA, Canada, Germany and the UK). Apart from between stocks and dividends, DeJong (1992) investigates these methods of finding cointegration between other pairs of bivariate processes: consumption and income, short and long term interest rates, GNP and money supply M2.

2.5 Estimation from a pre-sample

Bayesian analyses covered in this thesis use a pre-sample or historical data in order to initialize the parameters or give an objective estimate of some hyperparameters. In this section, we present the method on how to estimate the essential parameters of the VECM based on a pre-sample: the long-run relations matrix Π, the lag parameters matrix Ψ and the covariance matrix Σ of the errors.

Let us assume we have a collection of data for p difference stationary time series over a period of time length T. From that data set, we can extract a small period from time 1 until a certain time $\tau < T$. This time-period $[\![1, \tau]\!]$ of size τ corresponds to the time period of the pre-sample (see Figure 2.2). Then based on this pre-sample, we will obtain the parameter estimates $\widehat{\Pi}$, $\widehat{\Psi}$

and $\widehat{\Sigma}$ of the VECM model that will be used to initialize our algorithms in Chapter 3, Chapter 4 and Chapter 5 (see Algorithms 1, 4 and 8). The initial parameters will then be used sometimes to evaluate the values of certain hyperparameters (scale matrix A of the covariance matrix of the errors in the VECM in Chapters 3, 4 and 5).

The sample corresponds to the time-period $[\![\tau+1, T]\!]$ (see Figure 2.2). The size of the sample is $T-\tau$ and it will be the time period on which Bayesian inference on the parameters of the VECM will be made in Chapters 3, 4 and 5. The sample and the pre-sample do not overlap and the size of the pre-sample is usually much smaller than the sample.

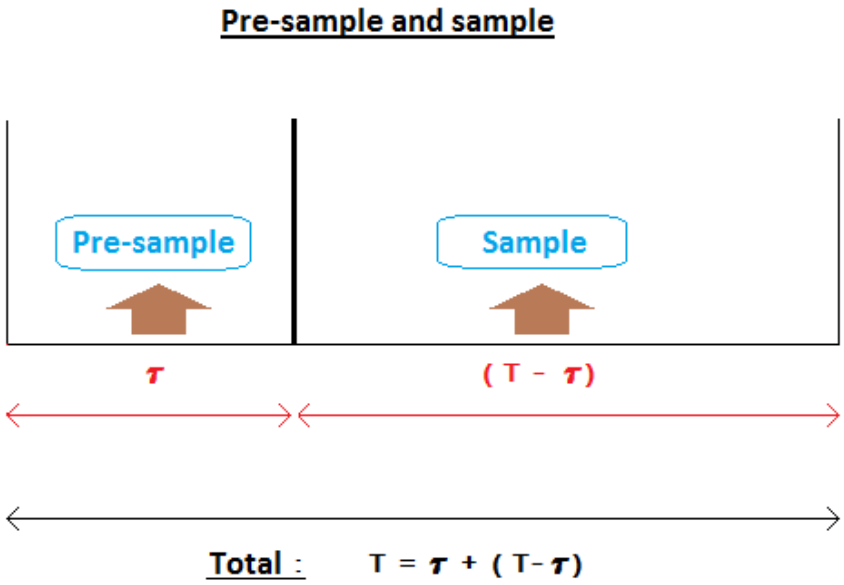

Figure 2.2: Arrangement of the data: Pre-sample and sample.

Based on a small pre-sample of size $\tau < T$, we then construct and estimate the parameters of the VECM. The detailed methodology is given in Section 7.2 of Luetkepohl (2006). We obtain estimates of the parameters of the VECM (2.4): $\widehat{\Pi}$, $\widehat{\Psi}$ and the variance of the errors $\widehat{\Sigma}$. We recall how to obtain their least squares estimates below.

We have:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Psi_i \Delta x_{t-i} + \epsilon_t$$

where $\epsilon_t \sim N(0, \Sigma)$. Furthermore, we assume that for each vector x_t, $x_{-k+1}, \cdots, x_0$ are available ($t = 1$ is the first time of the pre-sample). We can then build the following matrices:

$$\Delta X = [\Delta x_1, \cdots, \Delta x_\tau]$$

$$X_{-1} = [x_0, \cdots, x_{\tau-1}]$$

$$\Delta Z = [\Delta Z_0, \cdots, \Delta Z_{\tau-1}] \text{ with } \Delta Z_{t-1} = \begin{bmatrix} \Delta x_{t-1} \\ \vdots \\ \Delta x_{t-k+1} \end{bmatrix}$$

$$U = [\epsilon_1, \cdots, \epsilon_\tau]$$

Then, we have the following VECM for $t \in [\![1, \tau]\!]$:

$$\Delta X = \Pi X_{-1} + \Psi \Delta Z + U$$

and we obtain the least squares estimators of Π, Ψ and Σ by:

$$[\widehat{\Pi}, \widehat{\Psi}] = [\Delta X {X_{-1}}', \Delta X \Delta Z'] \begin{bmatrix} X_{-1}{X_{-1}}' & X_{-1}\Delta Z' \\ \Delta Z {X_{-1}}' & \Delta Z \Delta Z' \end{bmatrix}^{-1} \tag{2.30}$$

$$\widehat{\Sigma} = (\tau - pk)^{-1} (\Delta X - \widehat{\Pi} X_{-1} - \widehat{\Psi} \Delta Z)(\Delta X - \widehat{\Pi} X_{-1} - \widehat{\Psi} \Delta Z)'$$

Chapter 3

Estimation of the cointegration rank and the coefficients in a static model

3.1 Introduction

This chapter develops Bayesian cointegration methods for a set of time series where cointegration is assumed. This chapter has two aims: to estimate first the cointegration rank by avoiding reliance upon Johansen tests and to find the cointegrating relationships by operations based on the long-run impact matrix of the Error Correction Model (see Section 2.3). We decide to determine the cointegration rank within an MCMC procedure based on the singular values of the cointegration matrix from the Error Correction Model. Based on that rank r, we then derive r independent cointegrating relations from the cointegration matrix. The proposed methodology is tested on simulated data sets and then illustrated with a panel data set of Eurozone economic time series consisting of net trading, long-term interest rates and the harmonized unemployment rate. Our cointegration methods will try to establish their performance, co-evolution and long-run relationships.

In this chapter, we propose to set-up a weakly informative Gaussian prior on the cointegra-

tion matrix, asserting that *a priori* we expect that the cointegration matrix has zero mean (the case of no cointegration), but there is wide uncertainty around this. We propose a new determination of the cointegration rank by the number of irrelevant singular values of the estimated cointegration matrix. The estimation commences by Markov chain Monte Carlo, which provides posterior samples of the cointegration rank. Thus, we can have access to an approximation of the posterior distribution of the cointegration rank, which provides also the associated uncertainty around this estimation. Comparisons with Johansen's test indicate that this approach works reasonably well. The resulting cointegration relationships are derived by determining first the cointegration rank based on the singular values of the long-run relations matrix (during the MCMC), and then decomposing the latter into two full rank matrices (after the MCMC). The proposed methodology is illustrated by considering two simulated data sets and panel data on several macroeconomic variables (net trading, long-term interest rates and unemployment rate) across four Euro zone countries (Germany, France, Italy and Spain).

The approach adopted in this chapter is somewhat associated with the embedding approach of the Error Correction Model, see Kleibergen and van Dijk (1994) and Kleibergen and Paap (2002), recalled briefly in the literature review (see Section 2.4.4). In this approach, they build an Unrestricted Error Correction model by adding a $(p-r)\times(p-r)$ parameter matrix λ to the lower rank product of matrices $\alpha\beta'$ of rank r. The long-run relations matrix Π of this unrestricted ECM is then of full rank: while matrices α and β are $p\times r$ full rank matrices of rank r, it is the matrix parameter λ which controls the evaluation of the rank (see Section 2.4.4). Posterior probability distributions for the cointegration rank r can be derived by the use of Bayes factors, see Kleibergen and Paap (2002). In this chapter, the assumption of having a non-singular prior for the cointegrating matrix Π implies a prior for its singular values. We discuss briefly the prior of these singular values and how it can imply a prior for the rank (see Section 3.3.3).

The chapter is organised as follows: Section 3.2 introduces the approximation method for the determination of the cointegration rank. The Bayesian inference is in Section 3.3, which includes the operations on the long-run relationships matrix Π for the determination of the cointegrating

relationships. Section 3.4 gives the application of the methodology to the simulated data sets and to the real panel data sets of European economies. This chapter concludes with closing comments in Section 3.5.

Following standard notation seen in Chapter 2 we use the notation $x_t = (x_{it})_{1 \leq i \leq p}$ to be a vector of economic $I(1)$ time series represented by a p-dimensional vector autoregressive (VAR) process of lag length $k = 2$: an argument for this value is given in Section 3.4.3 and in Appendix B.

3.2 Approximation of the rank of Π

This section describes a method to obtain the rank of the long-run relationships matrix by numerical approximations based on its singular values. Finding the rank of any matrix is in fact equivalent to finding the number of non-zero singular values of that matrix. Suppose that Π has been specified or estimated and is of size $p \times p$. We can always obtain p real positive or null singular values for Π by its singular value decomposition (SVD):

$$\Pi = UDV^{\star}$$

where U and V are $p \times p$ unitary matrices. U (resp. V) represents the left (resp. right) singular vectors. D contains the p singular values, which are called μ_1, μ_2, ... , μ_p.

Let $\Pi^{\star}$ be the conjugate transpose of Π. The singular values of Π are actually the square roots of the eigenvalues of the matrix $\Pi^{\star}\Pi$, which is semidefinite positive if Π has lower rank, so all the singular values are either strictly positive or equal to 0. The rank of the matrix Π is the same as the rank of $\Pi^{\star}\Pi$ and thus equal to the number of non-zero singular values of Π.

We are using the programming language R Core Team (2013) to obtain those p singular values of Π: $(\mu_1, \mu_2, \ldots, \mu_p)$ such that $\{\mu_1 \geq \mu_2 \geq \cdots \geq \mu_p \geq 0\}$ thanks to the command svd. When computations are involved in the evaluation of those values (e.g. in the programming language R Core Team (2013)) of Π having rank $r < p$, it is not uncommon that the $p - r$ singular values that are supposed to be zero are close to zero, but not exactly zero. If in addition Π is estimated,

there might be some uncertainty around the zero (or near zero) singular values.

Based on this observation, we propose that the rank of Π will be $r < p$ if the sum of the singular values from μ_{r+1} to μ_p is "close" to 0. In other words, the rank of Π is considered to be r once $\sum_{i=r+1}^{p} \mu_i$ represents a small percentage of the total sum of the singular values $\sum_{i=1}^{p} \mu_i$. We then propose to use a threshold, an insignificance criterion ε, corresponding to the percentage below which the singular values are considered too small. Let us define the function K that represents the percentage of this contribution, i.e. $\forall j \in \{1, ..., p-1\}$, $K(j)$ represents the contribution brought by the smallest $p - j$ singular values:

$$K(j) = \frac{\sum_{i=j+1}^{p} \mu_i}{\sum_{i=1}^{p} \mu_i}. \tag{3.1}$$

When $K(j) \leq \varepsilon$, we suggest that the rank r is equal to j. The insignificance criterion ε is taken preferably to be below 10%. Let us assume for instance that we are provided with a 15×15 matrix Π with 15 eigenvalues. If we rely on ε to be 5% and if for any $i \leq 9$, $K(i) \geq 0.05$ but $K(9) < 0.05$, then an estimation of the rank will be 9. In this chapter we propose to estimate Π in a Gibbs sampler and then apply the above insignificance criterion to approximate the rank r. From each simulated cointegrating matrix Π, a rank will be determined by this method. At the end of the Gibbs sampler, we can either make an approximation of the rank by taking the median of all the rank estimations or make a histogram representing the frequency of each value of r (see Section 3.4.3). The value of the irrelevance criterion ε can be determined by conducting some sensitivity analysis and by comparison with Johansen's tests for consistency (see Section 3.4.3).

3.3 Bayesian inference

3.3.1 The likelihood of the VECM

This section describes the steps to find the likelihood of the Vector Error Correction Models used in this thesis. For the VECM used in this chapter, the error terms are normally distributed

with mean 0 and covariance matrix $\Sigma > 0$ and we can write the total VECM system (2.14) as:

$$Y' = \Pi X' + \Psi Z' + E'$$

where $E = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_T \end{bmatrix}'$ contains all the error terms ϵ_t.

For any $t \in [\![1, T]\!]$, we have $\epsilon_t \sim N(0, \Sigma)$. It is then straightforward to see that the matrix of the errors E' grouping all the ϵ_t from Equation (2.13) (see Section 2.3.2), will be distributed as:

$$E' = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_T \end{bmatrix} \sim N(0, \Sigma, I_T)$$

which gives:

$$Vec(E') \sim N(0, I_T \otimes \Sigma) \tag{3.2}$$

Recall that $E' = Y' - \Pi X' - \Psi Z'$ depends on the data $\mathcal{D} = \{X, Y, Z\}$ and the parameters of the Error Correction Model $\{\Pi, \Psi, \Sigma\}$. Therefore from Equation (3.2), we have:

$$f(Vec(E')|\mathcal{D}, \Pi, \Psi, \Sigma) = (2\pi)^{-\frac{Tp}{2}} |\Sigma|^{-\frac{T}{2}} |I_T|^{-\frac{p}{2}} \exp\left(-\frac{1}{2} Vec(E')'(I_T \otimes \Sigma)^{-1} Vec(E')\right)$$

Then we have:

$$(I_T \otimes \Sigma)^{-1} Vec(E') = (I_T \otimes \Sigma^{-1}) Vec(E') = Vec(\Sigma^{-1} E')$$

Now since $Vec(E')'Vec(\Sigma^{-1}E') = \mathrm{Tr}(\Sigma^{-1}E'E)$, we obtain:

$$f(Vec(E')|\mathcal{D}, \Pi, \Psi, \Sigma) = (2\pi)^{-\frac{Tp}{2}} |\Sigma|^{-\frac{T}{2}} |I_T|^{-\frac{p}{2}} \exp\left(-\frac{1}{2} \mathrm{Tr}(\Sigma^{-1} E'E)\right)$$

We can therefore define the likelihood function of our model by:

$$L(\mathcal{D}; \Pi, \Psi, \Sigma) \propto f(Vec(E')|\mathcal{D}, \Pi, \Psi, \Sigma) \tag{3.3}$$

which gives:

$$L(\mathcal{D}; \Pi, \Psi, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \mathrm{Tr}(\Sigma^{-1} E'E)\right) \tag{3.4}$$

3.3.2 The prior distributions

Unlike the derivation of the likelihood in Section 3.3.1, finding a joint prior distribution for our parameters requires us to make a good and suitable choice of distributions. In Bayesian analysis, this choice of prior distributions constitutes perhaps a dangerous but compulsory step in order to provide a good posterior distribution after that. We can clearly see that all our analysis will be affected by this choice.

The general prior

Let $(x_t)_{t=1}^{t=T}$ be a realisation of a p-dimensional Vector Autoregressive process of lag length k from which a Vector Error Correction Model in (2.10) is derived. We have then 3 unknown parameters to determine: Π, Ψ and the covariance matrix of the errors Σ.

Works of Bauwens and Lubrano (1996), Geweke (1996) and Kleibergen and van Dijk (1994) directly impose a reduced rank r for the cointegrating matrix Π and decompose this later into two full rank $p \times r$ matrices α and β. They infer α and β to analyse the cointegration relations. For each model, the Bayesian analysis is conditioned on the knowledge of the rank r beforehand. They actually try different models according to different values of the cointegration rank. They generally use posterior or predictive Bayesian odds ratios to assess the value of the cointegration rank.

On the other hand, the method used by Villani (2005) consists of inferring the parameters α, β, Ψ and Σ of the Error Correction Model conditional on the rank and to develop a posterior distribution for the rank, that is conditional on the data only (see Section 2.4.3).

In this chapter, we want to determine the cointegration rank by including it in the MCMC procedure with all the other parameters. For that, we decided not to do any decomposition of Π and to use a non-singular prior distribution on Π, that is a distribution from which we can only simulate a non-singular (or invertible) matrix. The idea of this chapter is to simulate at each step of the MCMC procedure the matrix Π and to estimate the rank of that matrix based on the irrelevance of some of its singular values. We consider in this thesis that the number of

independent cointegration relations, i.e. the cointegration rank, is a one-to-one function of the cointegrating matrix Π, and thus does not need to have a prior distribution.

A non-singular prior distribution for Π given Σ

This section is about giving a non-singular prior distribution for Π and explaining the choice of non-singularity for Π despite the fact that it is theoretically singular.

Since our time series are considered as $I(1)$ and that at least one of them is not stationary, Π has lower rank r. Therefore in principle, we should not choose a non-singular distribution for Π. However, the general assumption of this chapter is to consider the cointegrating matrix Π as being a full rank matrix. We can firstly assume that Π has a non-singular prior distribution. A non-singular posterior distribution will then be derived and under the programming language R Core Team (2013), we will see that for each simulated cointegrating matrix Π, some singular values will be close to 0. These singular values will therefore be considered as irrelevant and we can then have an estimation of the cointegration rank by the number of singular values that are not considered as irrelevant.

A reasonable non-singular prior for Π given Σ is to consider a matrix normal prior distribution:

$$\Pi|\Sigma \sim N_{p\times p}(0, v^{-1}\Sigma, I_p) \tag{3.5}$$

The aim of this section is to motivate now the choice of this normal distribution. First of all, the property of conjugacy is witnessed, i.e. the posterior distribution will also be a matrix normal distribution. Now let us analyze the form of the prior distribution:

$$\Pi|\Sigma \sim N_{p\times p}(0, v^{-1}\Sigma, I_p) \iff Vec(\Pi)|\Sigma \sim N_{p^2}(0, I_p \otimes v^{-1}\Sigma) \tag{3.6}$$

We have chosen the prior mean to be equal to 0 in order to have no influence on the values of the coefficients at the beginning. The scalar v is a regulatory fixed hyperparameter reflecting how much the probability distribution of Π is concentrated around 0 (i.e. the mean of Π). But on the other hand, since we have no information and do not want to emphasize the initial non-cointegration assumption too much, we will increase the value of the variance covariance matrix

of Π. In that way, we will use a weakly informative prior on Π. For that, we will use the scalar v in order to get a bigger covariance matrix. We can input a small value for $v = 0.001$ so that the variance is increased by $v^{-1} = 1000$. The regular Bayesian updates will anyway shift the mean of Π towards a more actual value thanks to the information we add from the data and the other parameters.

Let us now have a look at the variance of $Vec(\Pi)$ in order to explain in more detail our choice:

$$I_p \otimes v^{-1}\Sigma = \begin{pmatrix} v^{-1}\Sigma & 0 & \cdots & 0 \\ 0 & v^{-1}\Sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v^{-1}\Sigma \end{pmatrix} = v^{-1} \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma \end{pmatrix}$$

This variance must depend on Σ since it is reasonable to think that Π depends on the error terms of our model. In order to stay the most objective as possible, we decided to create a block diagonal matrix containing the same amount (v^{-1}) of Σ for each block diagonal element. We set all matrix elements not in the diagonal to be 0 because we assume that the columns of Π are uncorrelated, that is, if we let π_i denote the i^{th} column of Π, then we have that $Cov[\pi_i, \pi_k] = 0_{p\times p}$ with $i \neq k$.

A prior for Ψ given Σ

Ψ is a random matrix of size $d \times p$ (with $d = (k-1)p$), that will depend on Σ. This prior is chosen to be a matrix normal distribution, a change from Villani (2005), where he sets a uniform prior on Ψ for more simplicity. In addition, like for the distribution of Π, we can introduce a scale w in order to control the weakness of prior information about Ψ. However, in all our results in this thesis, a scale of $w = 1$ was used without any problem. This scale hyperparameter was created by convention, in the potential case we needed it.

The prior distribution of $\Psi|\Sigma$ is given in that case by:

$$\Psi|\Sigma \sim N_{d\times p}(0, w^{-1}\Sigma, I_d) \iff Vec(\Psi)|\Sigma \sim N_{p\times d}(0, I_d \otimes w^{-1}\Sigma) \tag{3.7}$$

A prior for Σ: Inverse-Wishart distribution

The definition below recapitulates the probability density function of an Inverse-Wishart distribution:

Definition 6. ***Probability density function of the Inverse-Wishart distribution***

Let $V \sim IW_p(B, m)$ where B is a positive definite scale matrix, and m and p are non-zero integers. Then V is ***positive definite*** *and has the probability density function:*

$$f(V) = \frac{|B|^{m/2}}{|V|^{\frac{m+p+1}{2}} 2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \exp\left(-\frac{1}{2}\operatorname{Tr}(BV^{-1})\right)$$

where $\Gamma_p(.)$ is the multivariate gamma function.

The Inverse-Wishart distribution is commonly used in Bayesian statistics to infer covariance matrices of normally distributed data. For instance, we can consider ϵ_1, ϵ_2, ..., ϵ_N to be a sequence of N random variables where each random p-vector ϵ_i has a multivariate normal distribution with mean 0 and covariance matrix Σ. Then if an Inverse-Wishart $IW_p(B, m)$ prior is defined for Σ, we shall achieve the property of conjugacy and obtain an Inverse-Wishart posterior distribution $IW_p(B + S, m + N)$ where S represents the sample sums of squares $\sum_{i=1}^{N} \epsilon_i \epsilon_i'$. The equivalent univariate distribution is the Inverse-Gamma, which is also used to infer the variance of a univariate random variable in Bayesian statistics.

In this chapter, the prior of Σ is chosen to be an Inverse-Wishart distribution with parameters A and q:

$$\Sigma \sim IW_p(A, q) \tag{3.8}$$

The parameter A is called the scale matrix and q represents the degrees of freedom of the Inverse-Wishart distribution. A and q will then consist of hyperparameters which we must set to suitable values. The hyperparameter A is estimated from a pre-sample (a training data set) of the data (see Section 3.3.6). As for q, it must be strictly higher than $p + 3$ in order for the variance of Σ (see equation (3.33) in Section 3.3.6) to stay positive definite, see also the definition of the density of an Inverse-Wishart distribution in Gupta and Nagar (2000). We subjectively took a value of $q = p + 4$.

The prior for all the parameters

In this chapter, we will assume that $\Pi|\Sigma$, $\Psi|\Sigma$ and Σ are independent. Thus, we can easily obtain the joint prior distribution of these parameters, which in other words constitutes the prior of our model:

$$f(\Pi, \Psi, \Sigma) = f(\Pi|\Sigma)\ f(\Psi|\Sigma)\ f(\Sigma) \tag{3.9}$$

According to the chosen prior distributions of $\Pi|\Sigma$, $\Psi|\Sigma$ and Σ, we can write the relation of proportionality that their respective densities verify:

$$f(\Pi|\Sigma) \propto |\Sigma|^{-p/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1} v \Pi\Pi')\right) \tag{3.10}$$

$$f(\Psi|\Sigma) \propto |\Sigma|^{-d/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1} w \Psi\Psi')\right) \tag{3.11}$$

$$f(\Sigma) \propto |\Sigma|^{-(p+q+1)/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1} A)\right) \tag{3.12}$$

Therefore by using (3.9) and by multiplying $f(\Pi|\Sigma)\ f(\Psi|\Sigma)$ and $f(\Sigma)$, we immediately obtain the relation of proportionality, that the full prior of the VECM verifies:

$$f(\Pi, \Psi, \Sigma) \propto |\Sigma|^{-(2p+d+q+1)/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1} v \Pi\Pi' + \Sigma^{-1} A + \Sigma^{-1} w \Psi\Psi')\right) \tag{3.13}$$

where $|\cdot|$ denotes determinant and $\operatorname{Tr}(\cdot)$ denotes trace.

3.3.3 A prior of the rank implied by the prior of the singular values of the cointegrating matrix

In this section, we focus on the prior for the singular values implied by the prior of $\Pi|\Sigma$ (3.5). We can first rewrite the density of the prior for $\Pi|\Sigma$ given in (3.10) by:

$$f(\Pi|\Sigma) \propto \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1} v \Pi\Pi')\right) \tag{3.14}$$

where $\Pi\Pi'$ is symmetric and thus can be decomposed as the following:

$$\Pi\Pi' = PDP' \tag{3.15}$$

where P is an orthogonal matrix and D contains the p singular values of Π.

Since a non-singular prior is set on Π, then the product matrix $\Pi\Pi'$ is positive definite and the p singular values are strictly positive. If we denote by μ_i each singular value of Π, we can derive the prior for μ_i from (3.14) by:

$$\begin{aligned} f(\Pi|\Sigma) &\propto \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}v\Pi\Pi')\right) \\ &\propto \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}vPDP^{-1})\right) \\ &\propto \exp\left(-\frac{1}{2}\operatorname{Tr}(vP^{-1}\Sigma^{-1}PD)\right) \end{aligned}$$

Now by calling $M = vP^{-1}\Sigma^{-1}P = (M_{ij})_{1\leq i,j\leq p}$, we can derive:

$$f(\mu_i|P,\Sigma) \propto \exp\left(-\frac{1}{2}M_{ii}\mu_i\right) \tag{3.16}$$

Expression (3.16) is then proportional to an exponential distribution of parameter $\frac{M_{ii}}{2}$. Thus the prior of μ_i given Σ and P is:

$$\mu_i \mid P,\Sigma \sim Exp\left(\frac{M_{ii}}{2}\right) \tag{3.17}$$

The prior of the singular values of Π (3.17) would indeed imply a prior for the cointegration rank but the goal of this chapter is not to infer the rank. The approach adopted in this chapter is quite similar to the embedding approach detailed in Section 2.4.4 where we define the unrestricted ECM by adding a parameter λ that will make the ECM of full rank. In this approach, it is the parameter λ that is in fact estimating the cointegration rank, see Kleibergen and van Dijk (1994) and Kleibergen and Paap (2002).

3.3.4 The posterior distributions

The full conditional posterior distribution is the posterior distribution for one parameter (or possibly a reduced group of parameters) conditional on the data and the rest of the parameters.

Let us define a set of h ($h > 2$) parameters by the vector θ:

$$\theta = \begin{pmatrix}\theta_1 & \theta_2 & \cdots & \theta_h\end{pmatrix}'$$

If we denote the information from the data by $\mathcal{D}$, then the full conditional posterior of θ_j ($\forall\, j \in [\![1, h]\!]$) is denoted by $f(\theta_j|\theta_1, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_h, \mathcal{D})$.

Villani (2005) develops an approach in which the impact matrix Π is split into two full rank matrices α and β of size $p \times r$, where r is the cointegration rank. Inference about α and β are then conditional on r. Villani then sets priors on the parameters of the equation and finds the full conditional posterior distributions for α, β, Ψ and Σ, conditional on r, see Theorem 4.5 of Villani (2005). The posterior distribution of the rank r is evaluated as a function of $\sigma = v^{-1/2}$ where v is the hyperparameter of the covariance matrix of the prior of α in this case (see Section 2.4.3). According to a certain value of v, Villani (2005) then assesses the rank to be $r = 1$ in a bivariate process.

In this chapter, we will establish the posterior distributions of the long-run impact matrix Π and Ψ, unconditional on the covariance matrix Σ.

3.3.5 The posterior distributions of Π and Ψ unconditional on the variance matrix Σ

In this section we describe the unconditional posterior distributions of Π and Ψ on the variance covariance matrix Σ. The covariance matrix Σ can indeed slow down the Gibbs Sampler if we simulate it as an additional parameter. The next result gives the conditional posteriors of Π and Ψ given the data $\mathcal{D}$ and the cointegration rank r. The Gibbs Sampler is in fact generating samples directly from $f(\Pi, \Psi|\mathcal{D}, r)$.

Based on the priors of Section 3.3.2, the full conditional posterior distributions of Π and Ψ are first derived by an application of Bayes' theorem.

Result 1. *We have the following posterior distributions for* Π *and* Ψ *unconditional on* Σ*:*

1. $\Psi|\Pi, r, \mathcal{D} \sim t_{p\times d}(M, S, \Omega, n)$.

2. $\Pi|\Psi, r, \mathcal{D} \sim t_{p\times p}(N, R, P, n')$

with

$$Q = Y - X\Pi' \tag{3.18}$$
$$W = Y - Z\Psi'$$

$$\Omega = (Z'Z + wI_d)^{-1} \tag{3.19}$$
$$M = Q'Z\Omega$$
$$S = A + Q'Q + v\Pi\Pi' - M'\Omega^{-1}M$$
$$n = T + q + 1$$

$$P = (vI_p + X'X)^{-1} \tag{3.20}$$
$$R = A + w\Psi\Psi' + W'W - NP^{-1}N'$$
$$N = W'XP$$
$$n' = T + d + q - p + 1$$

In the application of the above Gibbs sampler, the posterior samples of Π will be non-singular, but with singular values close to zero. In principle, we apply the rank approximation after obtaining the posterior sample of Π and this effectively means that we can obtain a sample for the cointegration rank r, given the data. This allows us to provide summary statistics about the rank.

Proofs

Proof. Unconditional posterior distribution of Ψ on Σ (Result 1):

We will prove now the first distribution of Result 1:

$$\Psi|\Pi, r, \mathcal{D} \sim t_{p\times d}(M, S, \Omega, n)$$

with:

$$\begin{cases} \Omega = (wI_d + Z'Z)^{-1} \\ M = Q'Z\Omega \\ S = A + Q'Q + v\Pi\Pi' - M'\Omega^{-1}M \\ n = T + q + 1 \end{cases}$$

First of all, let us write down Bayes' theorem for the joint posterior distribution of the three parameters Π, Ψ and Σ:

$$f(\Pi, \Psi, \Sigma|\mathcal{D}) \propto L(\Pi, \Psi, \Sigma; \mathcal{D}) f(\Pi, \Psi, \Sigma) \tag{3.21}$$

Now, the unconditional distribution of Π and Ψ is written as:

$$f(\Pi, \Psi|\mathcal{D}) \propto \int_{\Sigma>0} L(\Pi, \Psi, \Sigma; \mathcal{D}) f(\Pi, \Psi, \Sigma) d\Sigma \tag{3.22}$$

From Equation (3.4) the likelihood $L(\Pi, \Psi, \Sigma; \mathcal{D})$ in (3.21) and (3.22) can be written as:

$$L(\Pi, \Psi, \Sigma; \mathcal{D}) \propto |\Sigma|^{-T/2} \exp\left(-\frac{1}{2}\mathrm{Tr}(\Sigma^{-1}E'E)\right)$$

The joint prior (3.13) can also be written as the product of successively $f(\Pi|\Sigma)$ (3.10), $f(\Psi|\Sigma)$ (3.11) and $f(\Sigma)$ (3.12):

$$f(\Pi, \Psi, \Sigma) \propto |\Sigma|^{-\frac{2p+d+q+1}{2}} \exp\left(-\frac{1}{2}\,\mathrm{Tr}(\Sigma^{-1}v\Pi\Pi' + \Sigma^{-1}A + \Sigma^{-1}\Psi\Psi')\right)$$

Now if we write $g(\Sigma) = |\Sigma|^{-(T+2p+d+q+1)/2} \exp\left(-\frac{1}{2}\,\mathrm{Tr}(\Sigma^{-1}(A + \Psi\Psi' + v\Pi\Pi' + E'E))\right)$, we then have:

$$f(\Pi, \Psi, \Sigma|\mathcal{D}) \propto g(\Sigma)$$

In order to obtain the joint posterior distribution of Π and Ψ unconditional on Σ, written as $f(\Pi, \Psi|\mathcal{D})$, we have to integrate out Σ:

$$f(\Pi, \Psi|\mathcal{D}) \propto \int_{\Sigma>0} g(\Sigma) d\Sigma$$

Furthermore, we have:

$$f(\Psi|\Pi, \mathcal{D}) \propto \frac{f(\Psi, \Pi|\mathcal{D})}{f(\Pi)} \propto f(\Psi, \Pi|\mathcal{D})$$

Finally, we can also derive the unconditional posterior distribution of Ψ by integrating:

$$f(\Psi|\Pi, \mathcal{D}) \propto \int_{\Sigma>0} g(\Sigma) d\Sigma$$

In order to integrate out Σ, let us first recall the expression of the probability density function of an Inverse-Wishart distribution:

If $\Sigma \sim IW_p(B, m)$, then the p.d.f. of Σ is proportional to:

$$f(\Sigma) \propto |\Sigma|^{-\frac{m+p+1}{2}} |B|^{\frac{m}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}B)\right) \tag{3.23}$$

If we look at the expression of $g(\Sigma)$, we have:

$$g(\Sigma) = |\Sigma|^{-(T+2p+d+q+1)/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}(A + \Psi\Psi' + v\Pi\Pi' + E'E))\right)$$

so we can integrate out Σ by using $m = T + p + d + q$ and $B = A + w\Psi\Psi' + v\Pi\Pi' + E'E$ and write $g(\Sigma)$ as:

$$g(\Sigma) \propto f(\Sigma)|B|^{-\frac{m}{2}}$$

Now we can start by integrating out Σ:

$$\begin{aligned} \int_{\Sigma>0} g(\Sigma) d\Sigma &\propto \int_{\Sigma>0} f(\Sigma) d\Sigma \times |B|^{-\frac{m}{2}} \\ &\propto |B|^{-\frac{m}{2}} = |A + w\Psi\Psi' + v\Pi\Pi' + E'E|^{-\frac{m}{2}} \end{aligned} \tag{3.24}$$

Finally,

$$\int_{\Sigma>0} g(\Sigma) d\Sigma \propto |A + \Psi\Psi' + v\Pi\Pi' + E'E|^{-\frac{T+p+d+q}{2}} \tag{3.25}$$

Let us now consider the term post-multiplied by Σ^{-1} in the trace of the exponential:

$$\begin{aligned} A + w\Psi\Psi' + v\Pi\Pi' + E'E &= A + w\Psi\Psi' + v\Pi\Pi' + (Q - Z\Psi')'(Q - Z\Psi') \\ &= A + w\Psi\Psi' + v\Pi\Pi' + (Q' - \Psi Z')(Q - Z\Psi') \\ &= A + w\Psi\Psi' + v\Pi\Pi' + Q'Q - Q'Z\Psi' - \Psi Z'Q + \Psi Z'Z\Psi' \\ &= A + Q'Q + v\Pi\Pi' + w\Psi\Psi' - Q'Z\Psi' - \Psi Z'Q + \Psi Z'Z\Psi' \\ &= A + Q'Q + v\Pi\Pi' + \Psi(wI_d + Z'Z)\Psi' - Q'Z\Psi' - \Psi Z'Q \end{aligned}$$

Then we have:

$$(\Psi - M)\Omega^{-1}(\Psi - M)' = \Psi\Omega^{-1}\Psi' - \Psi\Omega^{-1}M' - M\Omega^{-1}\Psi' + M\Omega^{-1}M'$$

Then we can identify the parameters:

$$\Omega^{-1} = wI_d + Z'Z \Rightarrow \Omega = (wI_d + Z'Z)^{-1}$$

$$M\Omega^{-1} = Q'Z \Rightarrow M = Q'Z\Omega = Q'Z(wI_d + Z'Z)^{-1}$$

$$\Psi(wI_d + Z'Z)\Psi' - Q'Z\Psi' - \Psi Z'Q = (\Psi - M)\Omega^{-1}(\Psi - M)' - M\Omega^{-1}M'$$

Now, by calling $S = A + Q'Q + v\Pi\Pi' - M\Omega^{-1}M'$, we can obtain:

$$A+Q'Q+v\Pi\Pi'-M\Omega^{-1}M' = S+(\Psi-M)\Omega^{-1}(\Psi-M)' = S(I_p+S^{-1}(\Psi-M)\Omega^{-1}(\Psi-M)') \quad (3.26)$$

Finally, we can start distinguishing the form of the probability density function of a matrix t-distribution:

$$f(\Psi|\Pi,\mathcal{D}) \propto |A+w\Psi\Psi'+v\Pi\Pi'+E'E|^{-\frac{T+p+d+q}{2}} \propto |I_p+S^{-1}(\Psi-M)\Omega^{-1}(\Psi-M)'|^{-\frac{T+p+d+q}{2}} \quad (3.27)$$

Now if we deal with the degrees of freedom, n must satisfy the equation $n+p+d-1 = T+p+d+q$, so $n = T + q + 1$. Finally, we can conclude that the distribution of Ψ is:

$$\Psi|\Pi,\mathcal{D} \sim t_{p\times d}(M, S, \Omega, n) \quad (3.28)$$

with:

$$\begin{aligned}
&\Omega = (wI_d + Z'Z)^{-1} \\
&M = Q'Z\Omega \\
&S = A + Q'Q + v\Pi\Pi' - M'\Omega^{-1}M \\
&n = T + q + 1
\end{aligned}$$

□

Proof. Unconditional posterior distribution of Π on Σ (Result 1):

We will now prove:

$$\Pi|\Psi, r, \mathcal{D} \sim t_{p\times p}(N, R, P, n')$$

with:

$$N = W'XP$$
$$R = A + w\Psi\Psi' + W'W - NP^{-1}N'$$
$$P = (vI_p + X'X)^{-1}$$
$$n' = T + d + q - p + 1$$

In order to obtain the unconditional distribution of Π on Σ, we start again from:

$$f(\Pi, \Psi|\mathcal{D}) \propto \int_{\Sigma>0} g(\Sigma)d\Sigma$$

We have of course:

$$f(\Pi|\Psi, \mathcal{D}) \propto f(\Pi, \Psi|\mathcal{D})f(\Psi) \propto f(\Pi, \Psi|\mathcal{D})$$

so in fact:

$$f(\Pi|\Psi, \mathcal{D}) \propto \int_{\Sigma>0} g(\Sigma)d\Sigma \tag{3.29}$$

However, we will now deal differently with the term post-multiplied by Σ^{-1} in the trace of the exponential: $A + w\Psi\Psi' + v\Pi\Pi' + E'E$, that is we will try to write $A + w\Psi\Psi' + v\Pi\Pi' + E'E$ under the form $R + (\Pi - N)P^{-1}(\Pi - N)'$ where R, N and P are matrices to be determined. For that:

$$\begin{aligned} A + w\Psi\Psi' + v\Pi\Pi' + E'E &= A + w\Psi\Psi' + v\Pi\Pi' + (W - X\Pi')'(W - X\Pi') \\ &= A + w\Psi\Psi' + v\Pi\Pi' + (W' - \Pi X')(W - X\Pi') \\ &= A + w\Psi\Psi' + v\Pi\Pi' + W'W - W'X\Pi' - \Pi X'W + \Pi X'X\Pi' \\ &= A + w\Psi\Psi' + W'W + \Pi(vI_p + X'X)\Pi' - W'X\Pi' - \Pi X'W \end{aligned}$$

If we expand $R + (\Pi - N)P^{-1}(\Pi - N)'$ we get:

$$R + (\Pi - N)P^{-1}(\Pi - N)' = R + \Pi P^{-1}\Pi' - \Pi P^{-1}N' - NP^{-1}\Pi' + NP^{-1}N' \tag{3.30}$$

Now, we recognize from expression (3.30):

$$P^{-1} = vI_p + X'X \Rightarrow P = (vI_p + X'X)^{-1}$$

$$NP^{-1} = W'X \Rightarrow N = W'XP = W'X(vI_p + X'X)^{-1}$$

Now:

$$\begin{aligned} A + w\Psi\Psi' + v\Pi\Pi' + E'E &= A + w\Psi\Psi' + W'W + \Pi(vI_p + X'X)\Pi' - W'X\Pi' - \Pi X'W \\ &= A + w\Psi\Psi' + W'W - NP^{-1}N' + (\Pi - N)P^{-1}(\Pi - N)' \\ &= \underbrace{(A + w\Psi\Psi' + W'W - NP^{-1}N')}_{R} + (\Pi - N)P^{-1}(\Pi - N)' \end{aligned}$$

and we find:

$$R = A + w\Psi\Psi' + W'W - NP^{-1}N'$$

Now we can write our proportionality relation (3.29) as:

$$\begin{aligned} f(\Pi|\Psi, \mathcal{D}) &\propto \int_{\Sigma>0} g(\Sigma)d\Sigma \\ &\propto |A + w\Psi\Psi' + v\Pi\Pi' + E'E|^{-\frac{1}{2}(T+p+d+q)} \\ &\propto |R + (\Pi - N)P^{-1}(\Pi - N)'|^{-\frac{1}{2}(T+p+d+q)} \\ &\propto |R(I_p + R^{-1}(\Pi - N)P^{-1}(\Pi - N)')|^{-\frac{1}{2}(T+p+d+q)} \\ &\propto |I_p + R^{-1}(\Pi - N)P^{-1}(\Pi - N)'|^{-\frac{1}{2}(T+p+d+q)} \end{aligned} \tag{3.31}$$

Now, we can recognize a matrix t-distribution for $\Pi|\Psi, r, \mathcal{D}$:

$$\Pi|\Psi, r, \mathcal{D} \sim t_{p\times p}(N, R, P, n')$$

where the only parameter now to determine is the degrees of freedom n'. From the expression of the probability density function of a multivariate t-distribution, we can derive the relation: $n' + p + p - 1 = T + p + d + q$, which leads to the following expression for n':

$$n' = T + d + q - p + 1$$

□

3.3.6 Pre-sample and hyperparameters choice

For each of the periods studied we split the data into two parts. The first part of the data (of length $\tau < T$) is used as a pre-sample in order to determine the initial parameters (Σ_0, Π_0

and Ψ_0) by estimating the VECM model based on the methods described by Luetkepohl (2006) and recapitulated in Section 2.5.

We also decided to use the first third of the data to estimate the hyperparameter matrix A from the Inverse-Wishart distribution. The value of this hyperparameter is not obvious to decide because it is a $p \times p$ matrix and we do not want to put a subjective choice such as the identity matrix simply because it would imply no covariance between the errors of the time series ($\Sigma = Var(\epsilon_t)$). Since the mean of the random variable Σ is $A/(q-p-1)$, see Gupta and Nagar (2000), then we can estimate A by $(q-p-1) \times \Sigma_0$.

After estimating from the pre-sample we use the rest of the data (the other two thirds) to run the Gibbs Sampler and finally obtain the parameter estimates (see Figure 2.2). In the VAR and VECM model, we have:

$$\epsilon_t \overset{i.i.d.}{\sim} N_p(0_p, \Sigma)$$

with Σ as one of the unknown parameters of the Bayesian analysis. The prior chosen for the parameter Σ is an Inverse-Wishart distribution of parameters A (matrix of size $p \times p$) and q (a real number). We will first start by finding an unbiased estimator of Σ so that we have good hyperparameters A and q.

First of all, since $\Sigma \sim IW(A, q)$, we get:

$$E[\Sigma] = \frac{A}{q-p-1} \tag{3.32}$$

and:

$$\begin{aligned} Var[\Sigma_{ij}] &= \frac{(q-p+1)A_{ij}^2 + (q-p-1)A_{ii}A_{jj}}{(q-p)(q-p-1)^2(q-p-3)} \\ Cov[\Sigma_{ij}, \Sigma_{kl}] &= \frac{2A_{ij}A_{kl} + (q-p-1)(A_{ik}A_{jl} + A_{il}A_{kj})}{(q-p)(q-p-1)^2(q-p-3)} \end{aligned} \tag{3.33}$$

Then by observing the denominator of the variance in (3.33), we can see the terms "$q-p-3$" and "$q-p$". We have to make sure that the variance in (3.33) is positive. Since $A > 0$, if we set the condition that "$q > p+3$", then the positivity of the variance is ensured. Therefore we will fix a value of q to be strictly larger than $p+3$. We will first pick the value of $p+4$. From now

on, we call by $\hat{q}$ any arbitrary number taken for q. For example to start the program, we will set $\hat{q} = p + 4$.

Then we have $\hat{A} = (\hat{q} - p - 1) \times E[\Sigma] = (\hat{q} - p - 1) \times \hat{\Sigma}$ since we will say that the unbiased estimator $\hat{\Sigma}$ previously calculated estimates $E[\Sigma]$.

Finally, v is a hyperparameter fixed to the value 0.001 in order to establish a weakly informative prior for Π as said in Section 3.3.2.

3.3.7 Obtaining the linearly independent cointegrating relationships from the matrix Π

Let us now assume that Π is a $p \times p$ cointegrating matrix of the multivariate time series x_t with $p \geq 2$: $\Pi x_t = y_t = (y_{it})_{1 \leq i \leq p}$ where $\forall i \in [\![1, p]\!], x_{it} \sim I(1)$ and $\forall i \in [\![1, r]\!], y_{it} \sim I(0)$. If Π is of rank $r \leq p$, then by using the full rank decomposition theorem, Π can be written as $\Pi = \alpha\beta'$ where α and β are full rank matrices of size $p \times r$.

Let us now write the matrix Π in terms of blocks where the first block $\Pi_{r,r}$ on the top left-hand side represents an $r \times r$ square matrix corresponding to the first r rows and the first r columns of Π.

$$\Pi = \left[\begin{array}{c|c} \Pi_{r,r} & \Pi_{r,p-r} \\ \hline \Pi_{p-r,r} & \Pi_{p-r,p-r} \end{array}\right] = \alpha\beta'$$

Then the matrix β is in general written as: $\beta = \begin{bmatrix} I_r & B' \end{bmatrix}'$ in which each column of β will in fact represent an independent cointegrating relation since β is of full rank. The crucial matrix to find in order to obtain the independent cointegrating relations is actually the matrix B, which contains all the coefficients we need to perform a study of cointegration.

We can let matrix α be equal to the first r columns of Π, which actually constitute a rectangular submatrix of Π of full rank. From this value of α, we can then easily obtain the matrix β from the matrix Π by the following operation:

$$\alpha = \begin{bmatrix} \Pi_{r,r} \\ \Pi_{p-r,r} \end{bmatrix} \quad \Longrightarrow \quad \beta' = (\alpha'\alpha)^{-1}\alpha'\Pi = \begin{bmatrix} I_r & B' \end{bmatrix} \tag{3.34}$$

Therefore we can estimate the independent cointegrating relations (given by matrix β) just from the long-run impact matrix Π of size $p \times p$ and from its rank estimated by the Gibbs Sampler.

3.3.8 General Gibbs for a static Error Correction Model using a non-singular posterior distribution for Π and Ψ conditional on Σ

This section outlines the initialization, Gibbs sampling and post-processing algorithms used in this chapter. A Directed Acyclic Graph is constructed in order to have a better view of the algorithms and of the Bayesian network built in this chapter (see Figure 3.1):

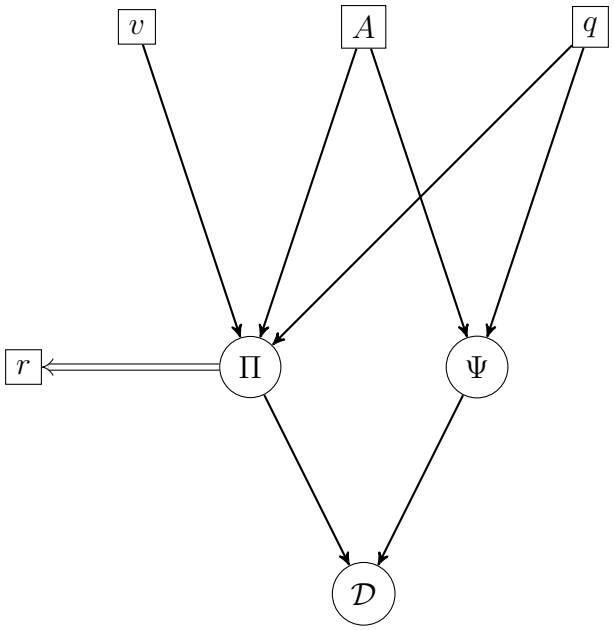

Figure 3.1: Directed Acyclic Graph: Square boxes contain the fixed parameters, circles contain the random parameters. The cointegration rank r is estimated from the simulated cointegrating matrix Π (Double arrow).

Algorithm 1 Initialization based on the pre-sample.

Set the size of the pre-sample which contains observations: $\tau = [T/3]$, where the square brackets represent the nearest integer function.

For the pre-sample data set $[\![1, \tau]\!]$:

- Create matrices ΔX, X_{-1}, ΔZ following the instructions in Section 2.5.

- Initialize $\Pi^{(0)} \leftarrow \widehat{\Pi}$, $\Psi^{(0)} \leftarrow \widehat{\Psi}$, $\Sigma^{(0)} \leftarrow \widehat{\Sigma}$ from the LS estimates (2.30) seen in Section 2.5.

Set the values of the hyperparameters:

- $v = 1$, $q = p + 4$, $A = (q - p - 1) \times \Sigma^{(0)}$.
- $r^{(0)}$ is estimated from the singular values of $\Pi^{(0)}$ and by choosing an irrelevance criterion ε: see Section 3.2.

Algorithm 2 Gibbs sampler on the sample.

For the sample data set which contains observations $[\![\tau + 1, T]\!]$:

- Create matrices Y, X, Z and E following the instructions in Section 2.3.2 for $t \in [\![\tau+1, T]\!]$.

We have: $\mathcal{D} = \{X, Y, Z\}$.

Set the number of iterations $m = 50{,}000$.

for $i \in [\![1, m]\!]$ **do**

- Sample $\Pi^{(i)}$ from the posterior distribution of Result 1.2. using $\Psi = \Psi^{(i-1)}$ and $\mathcal{D}$.
- Sample $\Psi^{(i)}$ from the posterior distribution of Result 1.1. using $\Pi = \Pi^{(i)}$ and $\mathcal{D}$.
- $r^{(i)}$ is estimated from the singular values of $\Pi^{(i)}$ and by choosing an irrelevance criterion ε: see Section 3.2.

end for

Algorithm 3 Final results: Obtaining the independent cointegrating relations.

Final results:

- $\Pi_{mean} \leftarrow \underset{\{m-10{,}000 \leq i \leq m\}}{\text{mean}} [\Pi^{(i)}]$
- $r_{median} \leftarrow \underset{\{m-10{,}000 \leq i \leq m\}}{\text{median}} [r^{(i)}]$
- The independent cointegrating relations (i.e. β) are then obtained from Π_{mean} and r_{median}, by using the operation of Section 3.3.7.

3.4 Application

We analyse two kinds of data sets, the first are synthetic and we use then to assess accuracy in the estimation of the cointegration rank and to aid model evaluation. The second is a panel of long term economic variables comprising net trading, long-term interest rates and unemployment from four European countries, before and after the introduction of the Euro.

3.4.1 Synthetic data sets and implementation

Description of the data sets

We have created two data sets with $T = 350$ data points each. The first data set (P_1) consists of seven time series $x_{1t}, x_{2t}, ..., x_{7t}$ with 4 independent cointegrating relations $y_{1t}, y_{2t}, y_{3t}, y_{4t}$. The process u_{it} is defined as a white noise process for any $i = 1, \ldots, 7$ below and we have:

$$x_{1t} = \sum_{k=1}^{t} u_{1k} \sim I(1) \ , \qquad x_{2t} = \sum_{k=1}^{t} u_{2k} \sim I(1)$$

$$x_{3t} = x_{2t} + x_{1t} + u_{3t} \sim I(1) \Longrightarrow y_{1t} = x_{3t} - x_{2t} - x_{1t} \sim I(0)$$

$$x_{4t} = x_{2t} + u_{4t} \sim I(1) \Longrightarrow y_{2t} = x_{4t} - x_{2t} \sim I(0)$$

$$x_{5t} = x_{1t} + u_{5t} \sim I(1) \Longrightarrow y_{3t} = x_{5t} - x_{1t} \sim I(0)$$

$$x_{6t} = \sum_{k=1}^{t} u_{6k} \sim I(1)$$

$$x_{7t} = x_{6t} - x_{2t} + u_{7t} \sim I(1) \Longrightarrow y_{4t} = x_{7t} - x_{6t} + x_{2t} \sim I(0)$$

The second data set (P_2) consists of five time series $x_{1t}, x_{2t}, ..., x_{5t}$ with three independent cointegrating relations y_{1t}, y_{2t}, y_{3t}. Letting v_{it} be a white noise process for any $i = 1, 2, 3$, we have:

$$x_{1t} = \sum_{k=1}^{t} v_{1k} \sim I(1) \ , \qquad x_{2t} = \sum_{k=1}^{t} v_{2k} \sim I(1)$$

$$x_{3t} = x_{2t} + x_{1t} + v_{3t} \sim I(1) \Longrightarrow y_{1t} = x_{3t} - x_{2t} - x_{1t} \sim I(0)$$

$$x_{4t} = x_{2t} + v_{4t} \sim I(1) \Longrightarrow y_{2t} = x_{4t} - x_{2t} \sim I(0)$$

$$x_{5t} = x_{1t} + v_{5t} \sim I(1) \Longrightarrow y_{3t} = x_{5t} - x_{1t} \sim I(0)$$

Implementation of the code on the synthetic data sets

We create these two data sets in R Core Team (2013) and we take the size of the pre-sample to be $\tau = [T/3]$ (see Algorithm 1). On the pre-sample data set containing observations $[\![1, \tau]\!]$, we estimate $\widehat{\Pi}$, $\widehat{\Psi}$ and $\widehat{\Sigma}$ from the methods of Luetkepohl recapitulated in Section 2.5 which will be our initial parameters: $\Pi^{(0)} \leftarrow \widehat{\Pi}$, $\Psi^{(0)} \leftarrow \widehat{\Psi}$, $\Sigma^{(0)} \leftarrow \widehat{\Sigma}$. The initial cointegration rank $r^{(0)}$ is estimated from $\Pi^{(0)}$ on the method based on the irrelevance criterion seen in Section 3.2. In order to estimate $r^{(0)}$, we take an irrelevance criterion $\varepsilon = 5\%$. We then set the hyperparameters as $v = 0.001$, $q = p + 4$ and $A = (q - p - 1) \times \Sigma^{(0)}$ (see Section 3.3.6).

For each data set, we will run four Gibbs samplers each using a different level for the insignificance criterion: $\varepsilon = 3, 4, 5, 8\%$ on the sample data set corresponding to the period $[\![\tau + 1, T]\!]$. The MCMC procedure where the parameters of the VECM are simulated at each iteration is presented in Algorithm 2. For both of our data sets, we consider a burn-in set of 40,000 iterations, leaving the 10,000 last simulations (of $m = 50{,}000$ iterations) to estimate the parameters of the VECM. At the end of the MCMC, we will compare the estimated rank with the rank determined by the Johansen tests on the data. The determined rank is equal to the median rank of the last 10,000 estimated ranks.

Finally, we will determine the independent cointegrating relations by using the operation seen in Section 3.3.7 on the mean of the last 10,000 simulated cointegrating matrices Π (see Algorithm 3).

Study of the cointegration rank and the cointegration relations for the simulated data

We apply Johansen trace tests to the simulated data sets P_1 and P_2, see Johansen (1991). Results are presented in Table 3.1. For the first simulated data set P_1, the test does not reject a rank less than or equal to 4 (Test: $20.03 < 31.52$) but rejects a rank less than or equal to 3 (Test: $131.25 > 48.28$). We then retrieve a rank equal to 4 for data set P_1. For the simulated data set P_2, the test does not reject a rank less than or equal to 3 (Test: $10.62 < 17.95$) but rejects a rank less than or equal to 2 (Test: $143.53 > 31.52$). The cointegration rank for the data set P_2 is then 3.

Table 3.1: Johansen trace tests for the cointegration rank of the simulated data of cointegration rank 4 (P_1) and 3 (P_2): For each data set, the first column corresponds to the test statistics and the second column corresponds to the 95% critical values from Johansen trace tests.

Test: $r \leq r_0$	P_1	Critical values (95%)	P_2	Critical values (95%)
0	580.30	124.25	441.66	70.60
1	413.14	90.39	284.96	48.28
2	262.46	70.60	143.53	31.52
3	131.25	48.28	10.62	17.95
4	20.03	31.52	4.18	8.18
5	11.38	17.95		
6	4.66	8.18		

Based on an insignificance criterion $\varepsilon = 5\%$ or 4% (see Table 3.2 below) we find a credible interval of $(3, 4, 4)$ and $(4, 4, 4)$ respectively for the first simulated data set P_1. As for the second

simulated data set, we find a credible interval of $(2, 3, 3)$ for $\varepsilon = 5\%$ and $(3, 3, 3)$ for $\varepsilon = 4\%$. We can conclude that the cointegration rank determined by Johansen and the rank determined by our methods based on the insignificance criterion are in agreement. Big insignificance criteria ($\varepsilon = 5\%$ and $\varepsilon = 8\%$) tend to neglect some singular values that are relevant, giving then a smaller rank (see Table 3.2).

Table 3.2: Cointegration rank posterior credible intervals of size 0.95 and median from the synthetic data sets P_1 and P_2 from different levels of the insignificance criterion.

ε	P_1	P_2
3%	(4,4,5)	(3,3,4)
4%	(4,4,4)	(3,3,3)
5%	(3,4,4)	(2,3,3)
8%	(3,4,4)	(2,3,3)

Figure 3.2 below shows the contribution of the singular values of a simulated cointegrating matrix Π from the data set P_2 at iteration 10,000. We can see on Figure 3.2 the drop off in the contribution (3.1) of the singular values. In particular, after the third singular value, we can see that the contributions of the 2 last singular values are irrelevant, suggesting a rank of 3 for that simulated matrix at the $10{,}000^{th}$ iteration.

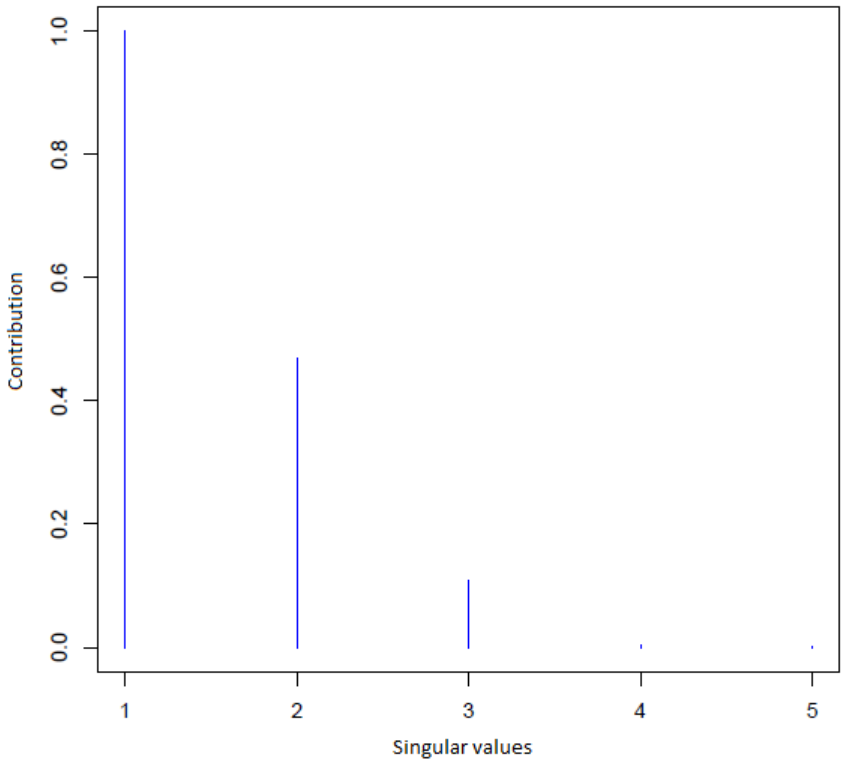

Figure 3.2: Contribution of the singular values of the simulated cointegrating matrix $\Pi^{(10,000)}$ for the data set P_2.

The cointegrating relations of the first simulated data set are presented in Table 3.3 below. We have, by descending row by row in the table, the cointegrating relations $y_{1t}, y_{2t}, y_{3t}, y_{4t}$:

Table 3.3: Cointegrating relations for the first simulated data set P_1

	x_{7t}	x_{5t}	x_{4t}	x_{3t}	x_{6t}	x_{2t}	x_{1t}
y_{1t}	1	0	0	0	-0.941	1.009	0.043
y_{2t}	0	1	0	0	-0.058	0.043	-0.971
y_{3t}	0	0	1	0	-0.014	-0.965	0.047
y_{4t}	0	0	0	1	-0.056	-0.961	-0.961

As for the second simulated data set, the results are presented in Table 3.4 below, and likewise, we obtain the three cointegrating relations y_{1t}, y_{2t}, y_{3t} by descending row by row:

Table 3.4: Cointegrating relations for the second simulated data set P_2

	x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
y_{1t}	1	0	0	-0.012	-1.009
y_{2t}	0	1	0	-0.983	0.014
y_{3t}	0	0	1	-1.019	-1.009

We can retrieve quite accurately the theoretical cointegrating relations of the two synthetic data sets in Tables 3.3 and 3.4.

3.4.2 The economic data sets

Net trading (NX) (also known as national current account) is the difference between the value of exports and imports for an economy over a certain period of time (in general, a month or a year). It is measured in the currency of that economy (in our case, the four countries studied share the same currency). A positive balance consists of exporting more than importing and thus leads to economic growth. A negative net trading brings a deficit to the economy.

Long-term interest rates (IR) can be considered as an index of business investment. Low long-term interest rates encourage investment whereas high interest rates discourage it. Consequently, since investment is a major source of economic growth then we can consider long-term interest rates to be related to it. Long-term interest rates refer to government bonds with a maturity of T = 10 years. They are not the interest rates at which the loans were issued, but the interest rates implied by the prices at which these government bonds are traded on financial markets.

The **unemployment rate (UR)** is a measure of the prevalence of unemployment and is the ratio between unemployed individuals and all individuals in the labour force. The unemployment

rate in a country is considered to be linked to economic growth and therefore could be related to the long-term interest rate and net trading.

Implementation of the real data sets

For both of our data sets about the European economies we will use the same implementation of the MCMC algorithm as seen in Section 3.4.1 for the two simulated data sets. In particular the size of the pre-sample will be equal to $\tau = [T/3]$ and with that pre-sample, we will estimate the initial parameters of the VECM: $\Pi^{(0)}$, $\Psi^{(0)}$ and $\Sigma^{(0)}$. The hyperparameters will be estimated using the same methods: $v = 0.001$, $q = p+4$ and $A = (q-p-1) \times \Sigma^{(0)}$. The first cointegration rank $r^{(0)}$ is estimated with an irrelevance criterion $\varepsilon = 5\%$ from matrix $\Pi^{(0)}$.

The number of iterations is chosen to be $m = 50{,}000$ again and likewise, we choose a burn-in set of 40,000 iterations, leaving the last 10,000 iterations to estimate the mean of the parameters of the VECM and the median of the cointegration ranks. Based on the median rank of the last 10,000 iterations, we determine the independent cointegrating relations from the mean of the last 10,000 cointegrating matrices Π.

3.4.3 Comparison with Johansen tests for the European panel data sets

This section describes a comparison between the cointegration rank estimated by our methods and the cointegration rank found by using Johansen tests. We recall that one of the aims of this chapter is to propose another way to determine the cointegration rank other than by using Johansen tests. The estimation of the cointegration rank is based on the insignificance criterion ε (see Section 3.2). From the function K seen in Section 3.2, we can see that if ε is small, then the smallest singular values will be less likely to be rejected, therefore it will increase the estimate of the rank.

By using the MCMC algorithm, we run several Gibbs samplers and test four insignificance criteria: $\varepsilon = 3\%$, $\varepsilon = 4\%$, $\varepsilon = 5\%$ and $\varepsilon = 8\%$. In this section, we will compare our method based

on the insignificance criterion with Johansen cointegration tests by studying the two European panel data sets corresponding to the time periods $(1991 - 1998)$ and $(1999 - 2008)$.

We used a lag order of 2 for both of the European data sets by using Luetkepohl's methods in comparing the AIC values, see Luetkepohl (2006). In any case, Appendix B gives reasons for the choice of the lag order. A lag order of 2 permits the MCMC algorithm to run faster and does not really affect the resulting cointegrating relations of our methods, should the data set actually have a lag order of 3 or 4 (see Appendix B). In our VECM, we restricted the constant term to be 0, so we will use critical values of the Johansen trace tests (see Table 3.6) defined for constant term 0. For the first time period $(1991 - 1998)$, Johansen tests reject the null hypotheses $r = 0$ to $r \leq 5$ but do not reject $r \leq 6$. For the second time period $(1999 - 2008)$, the tests suggest that we do not reject the hypothesis $r \leq 5$. We then have a smaller cointegration rank than in the first data set.

Histograms 3.3 and 3.4 both represent the distribution of the rank for the time period before and the time period after the Euro. Each histogram shows an estimation of the cointegration rank r throughout the Gibbs sampler. The median rank from all the simulations is eventually taken as the decision value for the cointegration rank.

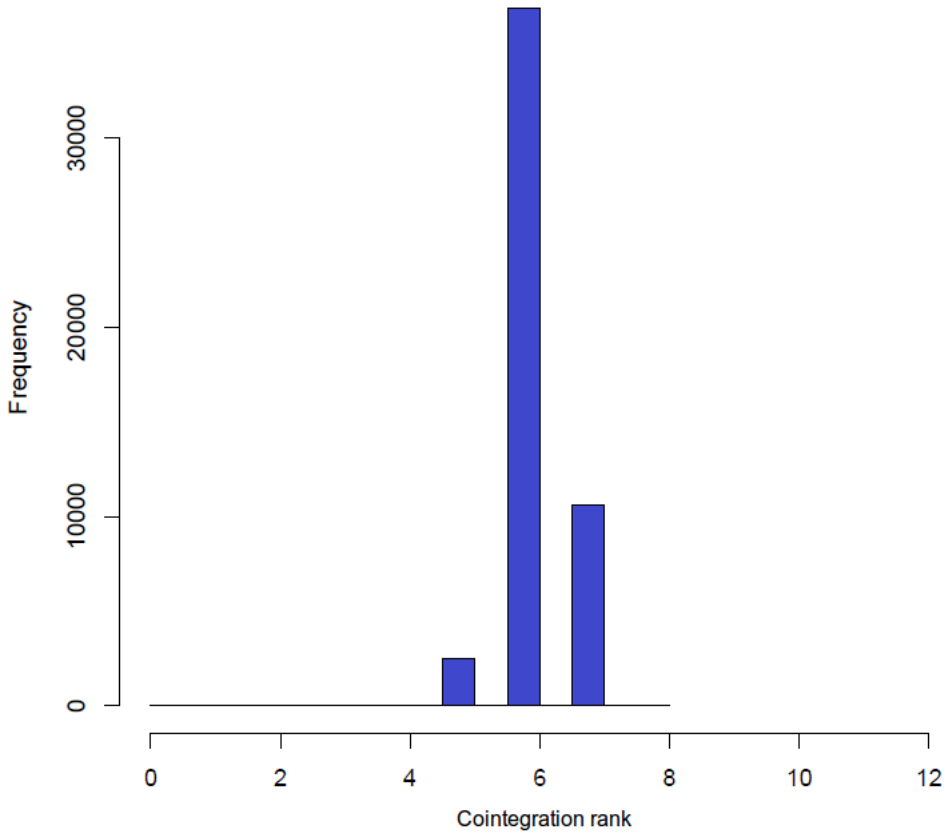

Figure 3.3: Histogram representing the distribution of the rank for the period pre-Euro (1991-1998).

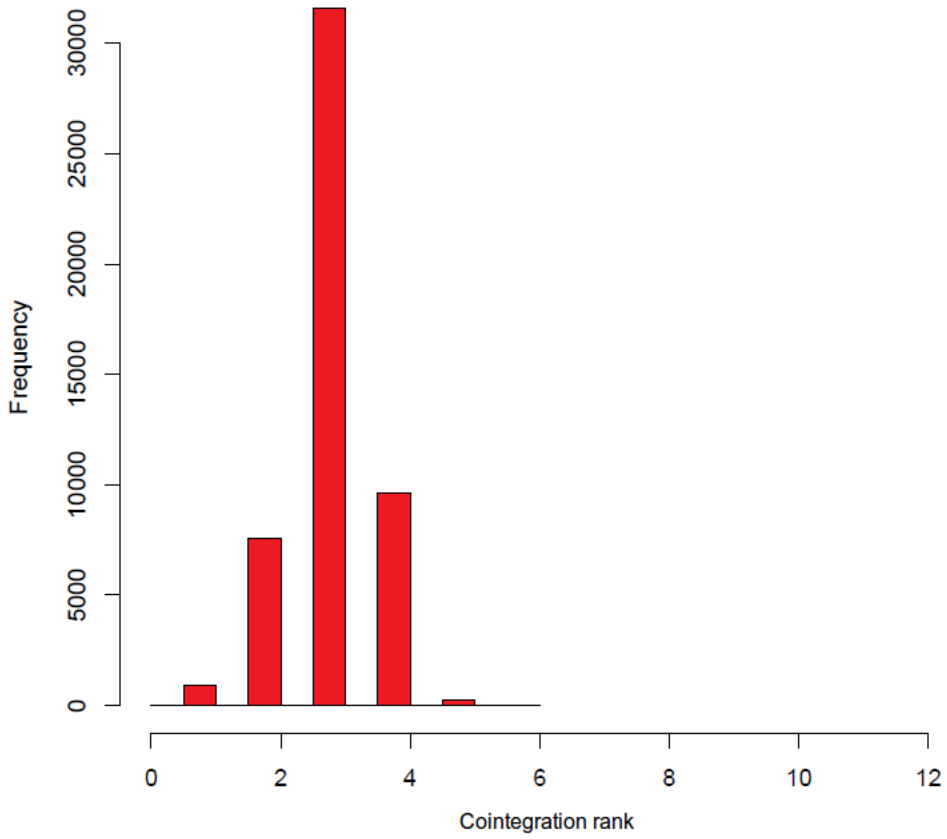

Figure 3.4: Histogram representing the distribution of the rank for the period post-Euro (1999-2008).

From Table 3.5, we can see that for any irrelevance criterion taken, the median rank is 3 for the time period after the Euro. As for the data before the Euro, we find a median rank of 6 by using an insignificance criterion of $\varepsilon = 4\%$ or 5%. Like for the Johansen tests, Table 3.5 suggests a smaller cointegration rank during the period after the Euro, hence less independent cointegrating relations. However, a little difference is pointed out compared with Johansen tests: indeed whatever the value of the insignificance criterion, the cointegration rank should be no more than 3 for the post-Euro data set. Thus, on the one hand Johansen tests reveal a cointegration

rank of 5 whilst on the other hand we find a cointegration rank of 3 by using the insignificance criteria. An insignificance criterion $\varepsilon = 8\%$ is definitely too large and tends to reject singular values that should be taken as "big". Indeed, for the real pre-Euro data set with cointegration rank of 6, we find 5 for the median rank. An insignificance criterion of $\varepsilon = 4\%$ or 5% seems to be closer to reality than for $\varepsilon = 8\%$.

Table 3.5: Cointegration rank posterior credible intervals of size 0.95 and median for the European panel data set (PrE = 1991–1998, PoE= 1999–2008) from different levels of the insignificance criterion.

ε	PrE	PoE
3%	(6,7,8)	(2,3,4)
4%	(5,6,7)	(2,3,4)
5%	(5,6,7)	(2,3,4)
8%	(4,5,6)	(2,3,3)

Table 3.6: Johansen tests for the two periods of the European panel data set (PrE = 1991–1998, PoE= 1999–2008).

Test: $r \leq r_0$	Period 1991-1998	Period 1999-2008	Critical values (95%)
0	536.64	441.09	301.95
1	394.50	353.29	277.39
2	308.73	278.59	232.49
3	245.90	218.50	192.84
4	184.84	164.43	157.11
5	135.27	117.57	124.25
6	90.18	81.49	90.39
7	61.47	53.54	70.60
8	39.20	34.81	48.28
9	22.08	18.64	31.52
10	8.61	9.09	17.95
11	0.24	0.41	8.18

According to Table 3.2 and Table 3.5, the insignificance criterion of 4% is the one that is in the most agreement with our simulated data and Johansen. The only problem is that it shows a cointegration rank of 3 for the period after the euro (instead of 4 as Johansen tests show in Table 3.6). We prefer to trust our method using the insignificance criterion and take $\varepsilon = 4\%$. In addition, it is less risky to reduce the cointegration rank to 3. Therefore we will use a cointegration rank of 6 for the data before the Euro $(1991 - 1998)$ and 3 for the period after the Euro $(1999 - 2008)$. According to the results we obtained in our new method, an insignificance criterion of $\varepsilon = 4\%$ may be more in agreement with Johansen and the simulated data sets. However, for more details, we decide to present in the next section a sensitivity analysis that explores more irrelevance criteria.

A more detailed sensitivity analysis on the irrelevance criterion

Table 3.7 below gives a more detailed sensitivity analysis on the irrelevance criterion by looking at values covering a range from 1% to 10%. This table presents the credible intervals of the rank for the two real data sets of the decade before the Euro (PrE= 1991–1998) and the decade after the Euro (PoE= 1999–2008) as well as the two synthetic data sets P_1 and P_2 previously studied in Section 3.4.1.

Table 3.7: Cointegration rank posterior credible intervals of size 0.95 and median for the synthetic VAR and the European panel data sets (PrE = 1991–1998, PoE= 1999–2008) from different levels of the irrelevance criterion.

ε	P_1	P_2	PrE	PoE
1%	(5,5,6)	(3,4,4)	(8,8,9)	(3,5,6)
1.5%	(4,5,6)	(3,4,4)	(7,8,8)	(3,4,5)
2%	(4,5,5)	(3,3,4)	(7,7,8)	(3,4,5)
2.5%	(4,4,5)	(3,3,4)	(6,7,8)	(3,4,5)
3%	(4,4,5)	(3,3,4)	(6,7,8)	(2,3,4)
3.5%	(4,4,5)	(3,3,3)	(6,6,7)	(2,3,4)
4%	(4,4,5)	(3,3,3)	(5,6,7)	(2,3,4)
4.5%	(4,4,5)	(2,3,3)	(5,6,7)	(2,3,4)
5%	(4,4,5)	(2,3,3)	(5,6,7)	(2,3,4)
5.5%	(3,4,5)	(2,3,3)	(5,6,7)	(2,3,4)
6%	(3,4,4)	(2,3,3)	(5,6,6)	(2,3,4)
7%	(3,4,4)	(2,3,3)	(5,5,6)	(2,3,3)
8%	(3,4,4)	(2,3,3)	(4,5,6)	(2,3,3)
9%	(3,4,4)	(2,3,3)	(4,5,6)	(2,2,3)
10%	(3,4,4)	(2,3,3)	(4,5,5)	(2,2,3)

From Table 3.7, we can see in each column that the credible intervals show a range of

decreasing values as the value of the irrelevance criterion becomes more important. Based on the simulated data sets P_1 and P_2, a too small irrelevance criterion is more likely to affect the results than a high value: high irrelevance criteria indeed give the correct cointegration rank for both P_1 and P_2. A too small irrelevance criterion (less than 2.5%) on the other hand leads to incorrect results. For the real data set, we retrieve the fact that the period before the Euro has a higher cointegration rank than the period post-Euro for any irrelevance criterion used.

3.4.4 Comparison of the independent cointegrating relations with Villani

The idea in this section is to compare the independent cointegrating relations between the method developed in this chapter and the inference of β proposed by Villani (2005). On the one hand we use our posterior distributions on Π and Ψ unconditional on Σ and estimate the cointegration rank in the MCMC procedure. We will only use the full conditional posterior distributions of α, β, Ψ and Σ in Theorem 4.5 Villani (2005) and not infer the cointegration rank based on its posterior distribution (2.19) and recapitulated in Section 2.4.3. The rank will be a constant and fixed to the rank found by the methods of this chapter. Thus, we will only compare the independent cointegrating relations between the two methods.

We will now study the independent cointegrating relations between the four net tradings of Germany, France, Italy and Spain for the time period before the introduction of the Euro. Based on the methods of this chapter, the cointegration rank is derived from the MCMC procedure (one rank estimated per simulation of Π) by taking the median rank of the last 10,000 iterations. From the last 10,000 simulations, we determine a median rank of 3. We are then going to run a Gibbs sampler using the full conditional distributions of Theorem 4.5 from Villani (2005) conditional on rank 3 with the same number of iterations and same burn-in period. We will therefore simulate the two full rank 4×3 matrices α and β composing the matrix Π.

Table 3.8 shows the results obtained by using both methods. We can clearly see that the two methods are in close agreement. But while the method employed by Villani (on the left) uses

four full conditional distributions on Σ, α, β and Ψ, here with our new method (on the right) we only use two posterior distributions of Π and Ψ unconditional on Σ, avoiding the simulation of Σ.

Table 3.8: Estimated relations between net trading, pre Euro (1991–1998): Villani and new method

Villani				Irrelevance criterion			
FraNX	GerNX	ItaNX	SpaNX	FraNX	GerNX	ItaNX	SpaNX
1	0	0	-1.233	1	0	0	-1.234
0	1	0	-1.624	0	1	0	-1.630
0	0	1	-1.374	0	0	1	-1.383

3.4.5 Interpretation of cointegrating relations

Let us assume that there exists a cointegrating relation over a period of time between the French and the German net trading ($(Fra)_{NT}$ and $(Ger)_{NT}$ respectively) with α and β being the 2 respective coefficients of France and Germany. Then we can write:

$$\alpha(Fra)_{NT} + \beta(Ger)_{NT} = u_t \sim I(0) \Rightarrow (Fra)_{NT} = \frac{-\beta}{\alpha}(Ger)_{NT} + u_t \sim I(0) \tag{3.35}$$

The sign is very important to reveal if the time series are coevolving (positively) in the same direction or (negatively) in the opposite direction. This section will explain this by taking the two cases in which the coevolution is positive and negative. If in the example above, α and β are both of the same sign, then $\frac{-\beta}{\alpha}$ is negative. In that case, if the German net trading is increasing, then according to (3.35) we have that $\frac{-\beta}{\alpha}(Ger)_{NT}$ is decreasing and then $(Fra)_{NT}$ is decreasing and we obtain a **negative coevolution** between the two countries. If now, for instance, α and β are of opposite sign, then $\frac{-\beta}{\alpha}$ is positive. In that case if the German net trading is increasing, then $\frac{-\beta}{\alpha}(Ger)_{NT}$ is increasing and $(Fra)_{NT}$ is increasing. We would then obtain a positive coevolution between the two countries.

By studying the value of the cointegrating coefficients, we can see how fast one variable is

evolving compared to another. If in the example above we have $|\beta| < |\alpha|$, then the ratio $\frac{|\beta|}{|\alpha|}$ is lower than 1. Assuming that the German net trading coevolves at a certain rate ρ with the French net trading, then France would evolve at a slower rate $\frac{|\beta|}{|\alpha|}\rho$ than Germany.

3.4.6 Study of the cointegrating relations before and after the Euro

Like for the simulated data set and as the algorithms 1, 2 and 3 suggest, we used a burn-in period of 40,000 iterations (the number of iterations used is 50,000) for both of our real data sets. The first data set, consisting of the economies in the Euro-zone before the introduction of the Euro, has a cointegration rank of 6. Table 3.9 describes the six independent cointegrating relationships derived from the methods explained in Section 3.3.7. This choice in the order of the variables is to explain better the comovements between the countries of the Euro-zone.

Table 3.9: Cointegrating relations before the introduction of the Euro: 1991-1998

GerNX	ItaNX	SpaNX	ItaIR	GerUR	ItaUR	FraNX	GerIR	FraIR	SpaIR	FraUR	SpaUR
1	0	0	0	0	0	-2.27	0.51	0.69	-0.54	0.23	0.35
0	1	0	0	0	0	-0.69	-1.18	1.91	-0.41	-0.36	-0.15
0	0	1	0	0	0	-1.45	-0.72	0.48	0.30	1.51	-0.50
0	0	0	1	0	0	0.31	0.30	-0.68	-0.76	0.07	-0.24
0	0	0	0	1	0	-0.90	-0.14	-0.44	0.46	0.27	-0.02
0	0	0	0	0	1	-1.04	-0.02	0.75	-0.34	0.03	0.01

Let us first look at the first three rows of Table 3.9. These rows compare the French current account (seventh column) with each of the other net tradings seperately (each of coefficient 1 for column 1, 2 or 3). Each coefficient of France is of negative sign when comparing with Germany (-2.27), Italy (-0.69) and Spain (-1.45). It means that the French net trading is evolving positively with each of those countries (since its coefficient is of opposite sign to the others, see Johansen (1988)). Figure 3.5 gives the four net tradings during both periods before and after the Euro. The straight line separates the two time periods (January 1999). By looking at the period

before the introduction of the Euro, we can see the common increasing movement between the four countries.

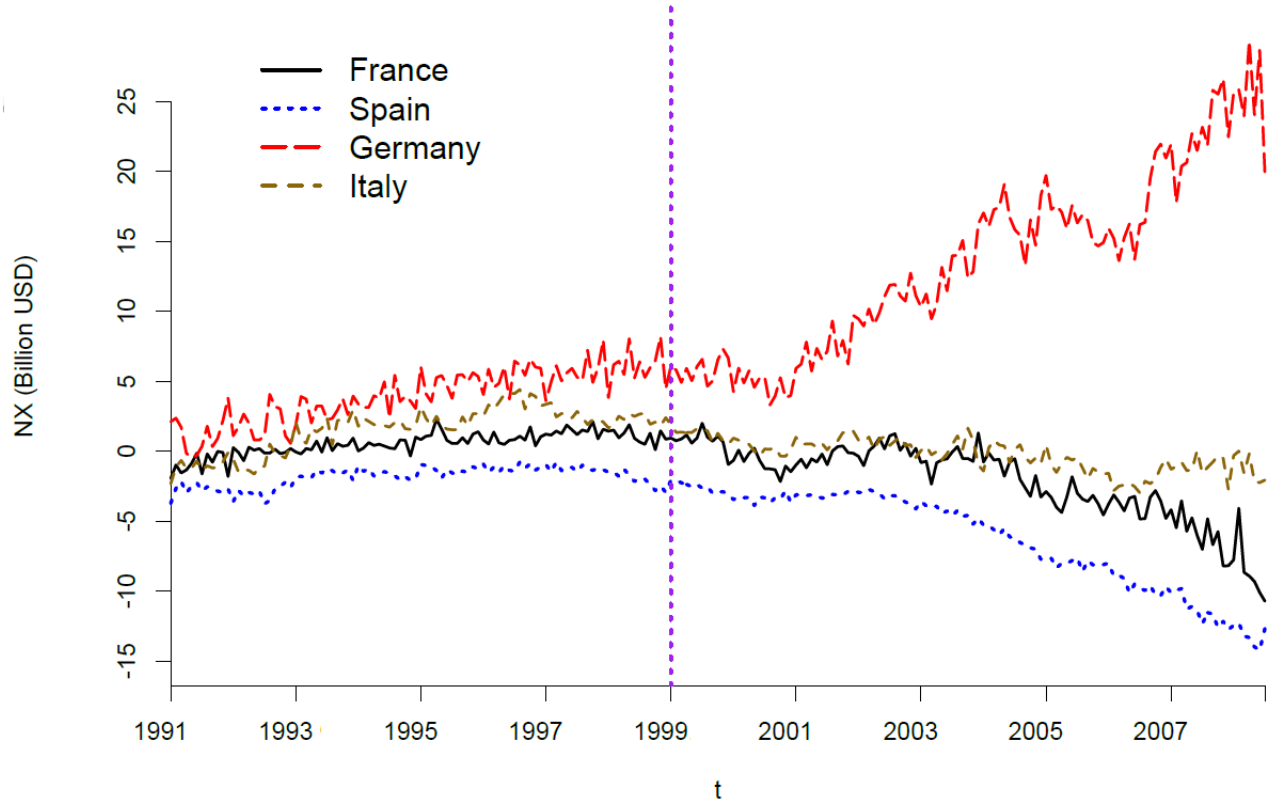

Figure 3.5: Traces depicting net trading of Germany, France, Italy and Spain for the period 1991-2008. The vertical line marks the start of the Euro. Comparing both sides, a common increasing movement between these countries can be appreciated.

Now we can have a look at the effect of the German and the Italian current accounts on the unemployment rates of France and Spain by looking at the two first rows of Table 3.9. The first row suggests that as the German net trading is increasing, the unemployment rates of France and Spain are decreasing (positive coefficient of $+0.23$ for the French unemployment rate and $+0.35$ for the Spanish unemployment rate). But the second row, on the other hand, leads to the fact that an increasing Italian current account has a bad effect on the French and the Spanish unemployment rates (due to the coefficients -0.36 and -0.15 being of opposite sign).

In order to summarize the period before the introduction of the Euro, we can retain the fact

that there is a positive comovement between the four countries current accounts, which is a good sign of convergence between the economies. The behaviour of the unemployment rates are, on the other hand, harder to describe if there is any comovement between them.

As for the data set after the introduction of the Euro, we found a cointegration rank of 3. Table 3.10 describes the three independent cointegrating relationships derived from the same methods as in Section 3.3.7. We chose the same order of variables as in Table 3.9.

Table 3.10: Cointegrating relations after the introduction of the Euro: 1999-2008

GerNX	ItaNX	SpaNX	ItaIR	GerUR	ItaUR	FraNX	GerIR	FraIR	SpaIR	FraUR	SpaUR
1	0	0	3.97	0.31	1.11	2.78	-61.60	77.53	-23.18	-3.98	-2.17
0	1	0	3.80	0.25	-0.66	0.26	-14.97	21.12	-10.55	-0.71	-0.40
0	0	1	0.94	-0.27	0.24	-1.50	23.15	-27.24	3.83	0.82	0.13

If we compare the coefficients of the net tradings this time, we have that the German and the French net tradings are coevolving negatively: the coefficient of France is +2.78, which suggests that the French net trading is evolving in the opposite direction as the German current account, with a quite fast speed rate of 2.78. The graph on Figure 3.5 indeed shows the German net trading increasing while the French current account is decreasing. However, the cointegrating relations inform more about the speed of this decrease than by just looking at the graph. On the other hand the French current account still coevolves positively with the Spanish net trading (-1.50). As for the Italian net trading, it is difficult to draw conclusions. The Italian current account is coevolving negatively with the French one. In Figure 3.5, we notice that the Italian current account is actually the hardest to define (in terms of tendency). The graph also shows vaguely a tendency for the Italian net trading to decrease after the introduction of the Euro, but less faster than the Spanish and the French net tradings. In addition the cointegrating coefficient of the French current account is below 1 and is quite small (+0.26). For this reason, it is unclear to state if Italy is actually more similar to France and Spain, rather than Germany, in terms of

evolution of the current account.

Figure 3.5 indeed shows a decreasing movement for France and Spain after the Euro (on the right handside of the straight line separating the two periods). A decrease is noticeable also for Italy but we notice a tendency to stabilize at the end and the cointegrating relations actually reveal an increasing tendency. Compared to the data before the introduction of the single currency, we can separate the set of countries into two groups: Germany on the one side that is clearly increasing, but France and Spain clearly decreasing. Italy is a country of which the net trading was indeed decreasing as well in the beginning of the decade, but "stabilized" itself more at the end of that period. A reason suggested would be that the Italian economy focused more on heavy industries, more exportable by using the Euro.

If we look at the interest rates in Figure 3.6, we can say that they are almost "equal" and conclude a quasi perfect positive comovement between the four interest rates. By looking at the three rows of Table 3.10, a complementarity can indeed be noticed between the 4 coefficients corresponding to the interest rates. If for each row we sum all the coefficients corresponding to the interest rates we will obtain a very small number close to 0. This tells us that the coefficients are complementary and go with the fact that these four variables are almost equal (see Figure 3.6 after 1999). We can almost say that the four interest rates are cointegrated independently from the other variables.

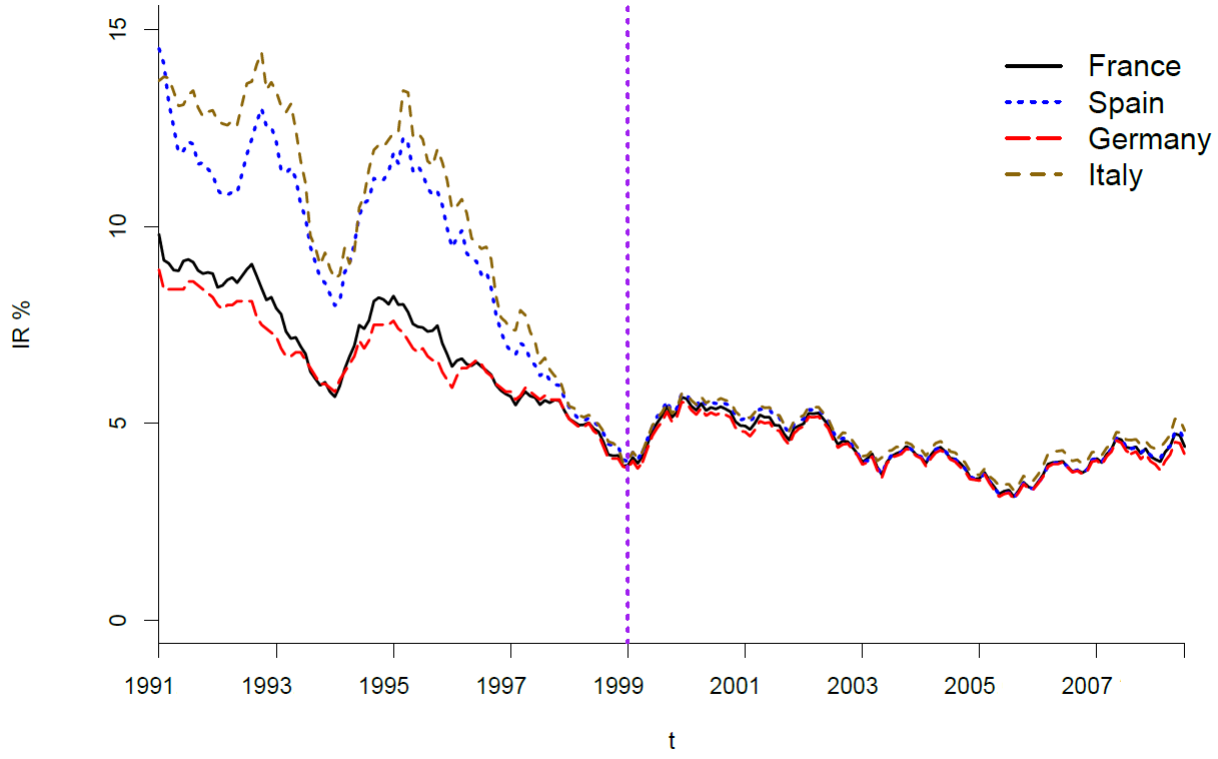

Figure 3.6: Long-term interest rates in France, Germany, Italy and Spain as a percentage. The vertical dashed line marks the periods before and after the Euro. Convergence is apparent as the inception of the Euro approaches.

After the Euro, the German unemployment rate is difficult to interpret due to a steady growth before 2005 and then a sudden decrease as the years of the financial crisis are approaching (see Figure 3.7). By looking at the first row of Table 3.10, as the German current account is increasing, the French and the Spanish unemployment rates are increasing (positive coevolution due to negative coefficients for the unemployment rates of those two countries: -3.98 and -2.17, respectively). On the other hand, the Italian and the German unemployment rates are decreasing, due to positive coefficients ($+1.11$ and $+0.31$, respectively), as the German net balance is increasing. If we look at the comparison with the Italian net trading in the second row of Table 3.10, the unemployment rates of Italy (-0.66), France (-0.71) and Spain (-0.40) increase as the Italian net trading increases, but the German unemployment rate decreases ($+0.25$). The Italian current account thus has a negative effect on its own unemployment rate but not on the German

unemployment rate. Eventually, when comparing with the Spanish current account in the third row of Table 3.10, we notice that the German unemployment rate is coevolving positively (-0.27) but the French unemployment rate's coefficient ($+0.82$) is coevolving negatively with the Spanish net trading. This suggests that the Spanish net trading has a positive impact on the French unemployment rate but not on the German one. In that same row, the Spanish unemployment rate is positive ($+0.13$) and thus it suggests that the Spanish unemployment rate increases as the Spanish net trading decreases. The Spanish net trading also has a positive impact on the Italian unemployment rate due to a positive coefficient ($+0.24$).

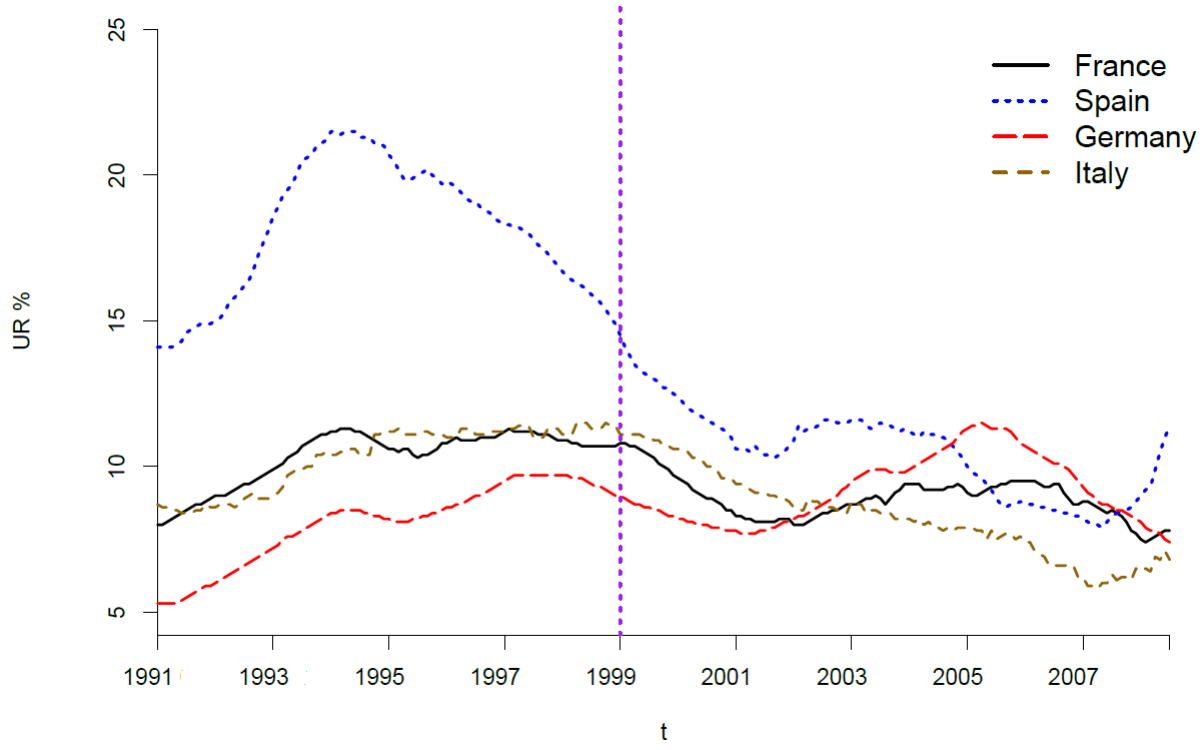

Figure 3.7: Unemployment rates in France, Germany, Italy and Spain as a percentage. The vertical dashed line marks the periods before and after the Euro. Although slowly, there are signs of convergence after the inception of the single currency.

In order to sum up the results over the period following the adoption of the Euro, the German current account is the only one doing well in increasing while the French and Spanish net tradings

decrease. The Italian net trading is decreasing at the beginning of the decade but starts to increase by the end of the time period studied (according to the analysis of both the graph and the cointegrating coefficients). From the cointegrating relations after the adoption of the Euro in Table 3.10, we can say that the French and the Spanish economies are coevolving well: indeed the net tradings are coevolving positively and the Spanish net trading has a positive impact on the French and Spanish unemployment rates. Finally it is also good to point out that the cointegration rank dropped to the value of 3 in this second period, which proves that there is in fact less comovements between the economies after the introduction of the Euro.

3.4.7 Posterior summaries on the real data sets

This section presents traceplots from the MCMC procedure of some coefficients of the VECM. Figure 3.8 shows the traceplot of the coefficients Π_{63}, Π_{81}, Ψ_{23} and Ψ_{93} after running the Gibbs sampler for the European panel data set before the introduction of the Euro. We can see immediate convergence of the Markov chain for these coefficients. Likewise, we obtain the trace plots of other coefficients for the European panel data set for the time period post-Euro (see Figure 3.9). Again, the convergence is very visible in the chain for those particular coefficients (Π_{32}, Π_{99}, Ψ_{21} and Ψ_{78}).

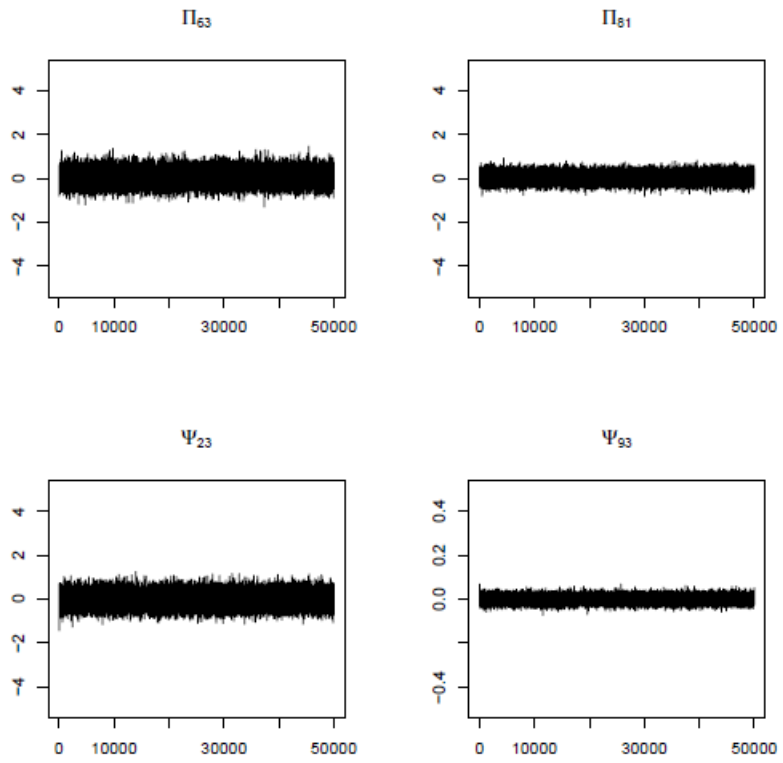

Figure 3.8: Trace plots of the coefficients Π_{63}, Π_{81}, Ψ_{23} and Ψ_{93} after running the Gibbs sampler for the data set of the period pre-Euro (1991-1998).

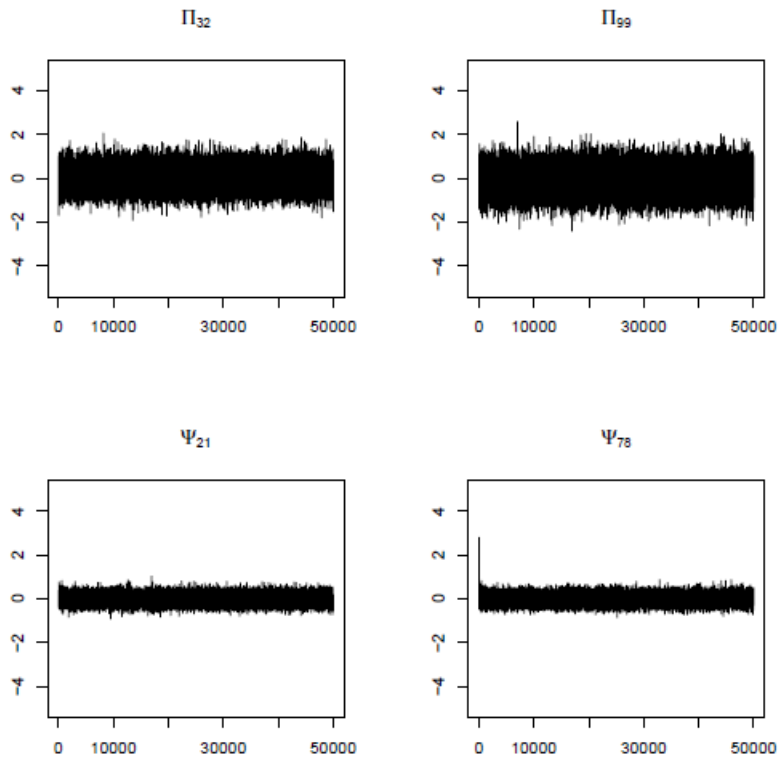

Figure 3.9: Trace plots of the coefficients Π_{32}, Π_{99}, Ψ_{21} and Ψ_{78} after running the Gibbs sampler for the data set of the period post-Euro (1999-2008).

Figures 3.10 and 3.11 present the posterior and prior distributions obtained for the same coefficients presented in the traceplots of Figures 3.8 and 3.9, after running the MCMC described by Algorithms 1 and 2. Prior distributions are represented in blue whereas posterior distributions are in red. Priors of Π and Ψ are simulated at each step of the MCMC from the priors defined in (3.5) and (3.7) where the covariance matrix Σ is simulated from the Inverse-Wishart prior defined in (3.8). The same hyperparameters described in Algorithm 1 are used. For both of our data sets, we obtain fairly symmetric distributions with a sharp shape for the posterior distributions compared to the flat shape of the priors.

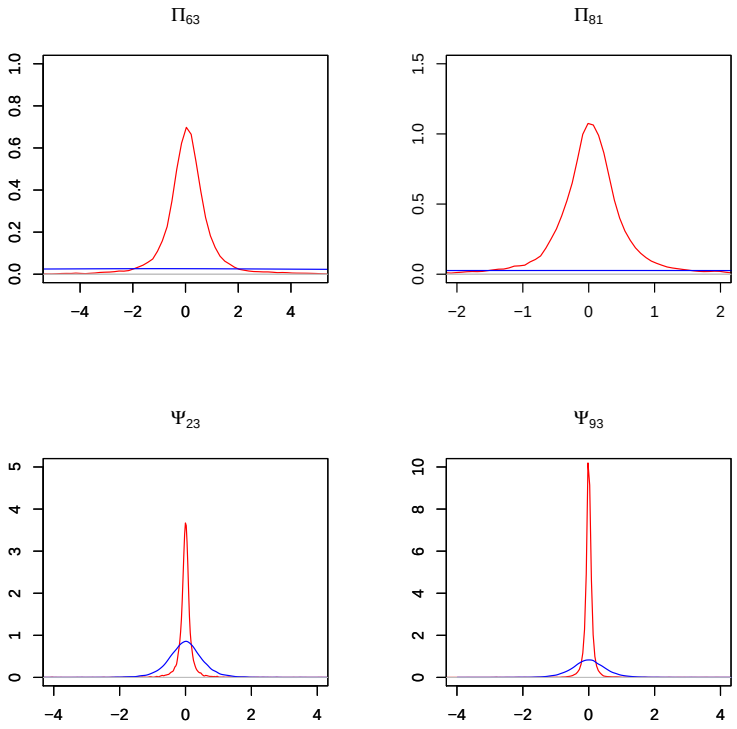

Figure 3.10: Posterior and prior densities of the coefficients Π_{63}, Π_{81}, Ψ_{23} and Ψ_{93} after running the Gibbs sampler for the data set of the period pre-Euro (1991-1998): Priors in blue and posteriors in red.

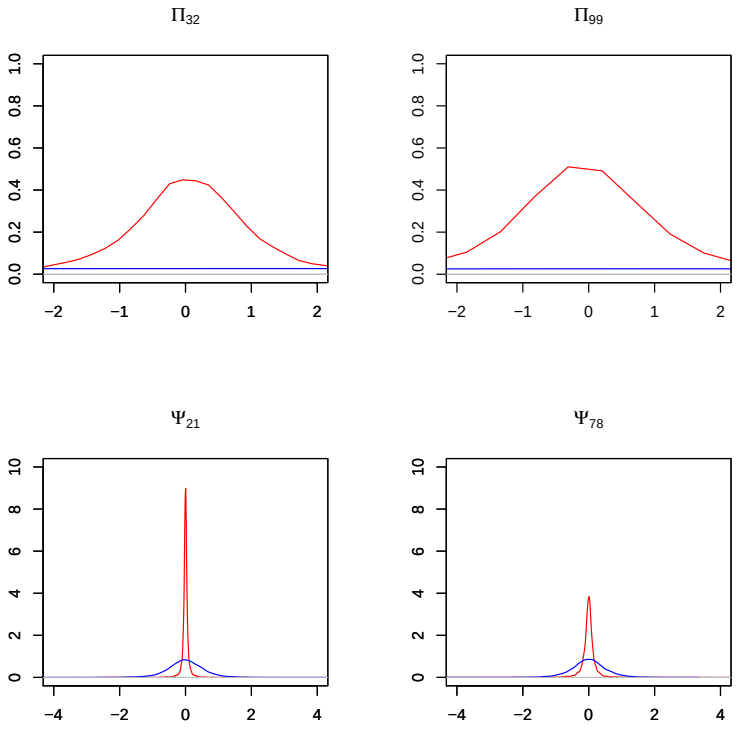

Figure 3.11: Posterior and prior densities of the coefficients Π_{32}, Π_{99}, Ψ_{21} and Ψ_{78} after running the Gibbs sampler for the data set of the period post-Euro (1999-2008): Priors in blue and posteriors in red.

3.5 Discussion

This chapter is based on the use of a non-singular Gaussian prior for the cointegrating matrix Π and on determining the cointegration rank by assessing the number of irrelevant singular values of Π throughout the Gibbs sampler. The lag-parameters Ψ of the VECM are inferred thanks to a non-singular Gaussian prior as well. While we set an Inverse-Wishart prior distribution for the covariance matrix Σ of the errors, we derive unconditional distributions for Π and Ψ from Σ. In the Gibbs sampler, we then only have to simulate Π and Ψ from these unconditional distributions.

The core of our methods develops Gibbs sampling based on Gaussian priors, while the coin-

tegration space is taken into account when the posterior samples are trimmed using a threshold, the insignificance criterion. Our methodology benefits by avoiding the requirements of applying Johansen cointegration tests and of pre-ordering the data before Gibbs sampling takes place. This allows assessment of uncertainty in the estimation of the cointegration rank and results in significant computational savings. Our proposal via the insignificance criterion provides flexibility in the stringency the analyst may apply for rank estimation. It is expected that the threshold specification will depend on the particular application and will require some amount of experimentation. It benefits from flexibility and subjectivity without being difficult to set and interpret, hence it is expected to appeal to practicing econometricians and economists alike. Furthermore, the proposed rank determination is applied post-hoc which can allow the adoption of other MCMC estimation procedures of the cointegration matrix, adding further flexibility and adaptability to this procedure.

The methodology is illustrated with simulated data sets and a panel of macro-economic variables of Euro-zone countries before and after the introduction of the Euro. We found that the cointegration rank was significantly smaller for the period after the adoption of the Euro (rank of 3 after and 6 before), suggesting less cointegration relations, and therefore less comovement among these economies after the introduction of the Euro.

We can associate this Error Correction Model to the Unrestricted Error Correction model of the embedding approach, see Kleibergen and van Dijk (1994) and Kleibergen and Paap (2002). Such a prior on the cointegrating matrix Π implies a prior for its singular values, and therefore we can have a prior for the rank (see Section 3.3.3). For future work, we could try to derive a posterior distribution for the cointegration rank in order to evaluate it in a Bayesian methodology.

The methodology seen in this chapter will be extended to time-varying cointegration systems by allowing MCMC estimation of the time-varying cointegration matrix in the next chapter (see Chapter 4). Indeed, a forward filtering backward sampling MCMC scheme may be adopted, while the cointegration rank can be determined using the insignificance criterion proposed in

this chapter, given that our methodology allows estimating the cointegration matrix as if it were non-singular and then determining its reduced rank post-hoc. This extension of the time-varying cointegration systems is proposed in the next chapter of this thesis: "Time-varying cointegration".

Chapter 4

Time-varying cointegration

4.1 Time-varying Vector Error Correction Model

In models involving macroeconomic or financial time series, some parameters may change over time (e.g. Ang and Bekaert (2002), Cogley (2005), Cogley and Sargent (2001), Cogley and Sargent (2005) and Stock and Watson (1996)). For instance, a model on financial time series can be used to take into account the data before the crisis of the subprimes, but that model will not be applicable on the period following the crisis. The theory of time varying parameters can actually make the parameters of that model change over time and thus adapt them on different time periods. The parameters of such a model are then said to be time varying or dynamic. Estimating time-varying parameters in an economic model also makes it possible to find out similarities between time periods in historical data.

Among works about time-varying cointegration, we can first mention Granger and Lee (1991) who described a time-varying cointegrated process applied to US data prices and wages. Bierens and Martins (2010) also developed a model in which a smooth evolution of the cointegrating relations over time is seen in the Purchasing Power Parity between different countries. As for Choi and Saikkonen (2004) they developed statistical tests about the linearity in cointegrating smooth transition regressions and applied them to a U.K. money demand function.

Among works about cointegration, we can mention Koop *et al.* (2011) who developed a time varying Error Correction model. In particular, they proposed a method, that allowed the cointegrating space (i.e. a vector space spanned by the independent cointegrating vectors) to evolve over time. They then use a state space model algorithm in order to infer the time-varying ECM and highlight the evolution of some cointegrating relations between US macroecomic variables.

Besides, it is also important to mention the work of Primiceri (2005) who proposed an efficient MCMC procedure applied to a Vector Autoregressive model in order to model US monetary policies and its private sector behavior. Following the work of Primiceri, we start this chapter by introducing the VAR model with time varying coefficients. In our case we assume the lag order k to be known and fixed.

$$\begin{aligned} x_t &= \Gamma_{1,t}x_{t-1} + \Gamma_{2,t}x_{t-2} + ... + \Gamma_{k,t}x_{t-k} + u_t \\ x_t &= \sum_{j=1}^{k} \Gamma_{j,t}x_{t-j} + u_t, \quad u_t \sim N(0, \Sigma) \end{aligned} \tag{4.1}$$

The Vector Error Correction Model (VECM) is derived from the time-varying VAR model (4.1). We eventually have a time-varying VECM from which we can see how cointegration evolves over time by studying the movements of the time-varying long-run impact matrix Π_t. Papers about Bayesian time-varying cointegration have assumed a constant cointegration rank so far and focus more on the values of the cointegrating coefficients evolving over time, see Granger and Lee (1991), Bierens and Martins (2010) and Koop *et al.* (2011). In this chapter, we can use the same idea of applying a non-singular distribution on the cointegrating matrix and assess the time varying cointegration rank from it. From the expression of the VAR model (4.1) we obtain the following VECM:

$$\Delta x_t = \Pi_t x_{t-1} + \sum_{j=1}^{k-1} \Psi_{j,t}\Delta x_{t-j} + u_t, \quad u_t \sim N(0, \Sigma) \tag{4.2}$$

with:

$$\Pi_t = -(I_p - \Gamma_{1,t} - \Gamma_{2,t} - ... - \Gamma_{k,t}) = -(I_p - \sum_{j=1}^{k} \Gamma_{j,t}) \tag{4.3}$$

$$\Psi_{j,t} = -(\Gamma_{j+1,t} - ... - \Gamma_{k,t}) = -\sum_{i=j+1}^{k} \Gamma_{i,t}, \quad \forall j \in [\![1, k-1]\!]$$

However, Primiceri (2005) considered in his model a time-varying error variance matrix Σ_t for the VAR (4.1) in which he defines the following triangular reduction:

$$L_t \Sigma_t L_t' = D_t D_t' \tag{4.4}$$

where L_t is a lower triangular matrix with diagonal elements all equal to 1 and D_t is a diagonal matrix. According to Smith and Kohn (2002), Pinheiro and Bates (1996) and Pourahmadi (2000), such a decomposition (4.4) allows to estimate more efficiently a variance matrix. Inferring Σ_t at each time can indeed present a huge computational task and they consequently need to use the decomposition (4.4). With this triangular decomposition, the error correction model can be written as:

$$\Delta x_t = \Pi_t x_{t-1} + \sum_{j=1}^{k-1} \Psi_{j,t} \Delta x_{t-j} + {L_t}^{-1} D_t \epsilon_t, \quad \epsilon_t \sim N(0, I_p)$$

In this chapter, we make the choice of Σ to be time-invariant. Some researchers insist on having a time-varying variance matrix of the errors, see Primiceri (2005) and Koop (1991). However in this thesis, we think that a time-varying covariance matrix will affect the cointegrating relations. Indeed by definition, a cointegrating relation is stationary, and the variance of such a relation should therefore be constant over time. For instance, assuming a random walk process for Σ_t as Uhlig (1994) suggests would break the assumption of stationarity and lead to an inaccurate estimation of the cointegrating relations. We also think that a time-invariant matrix Σ in the VECM (4.2) will allow a smooth and gradual evolution of the parameters and the data, which is what can be seen in practice. Finally, having a time-invariant covariance matrix will considerably ameliorate the speed of our algorithms, especially with high dimensional data. We will therefore assume for our model that each $u_t \sim N(0, \Sigma)$ and we will infer Σ from all the data u_1, u_2, ..., u_p.

Section 4.2 gives a reminder of the general state space model used in the literature and the state space model used for the Vector Error Correction Model (VECM). We then recapitulate in Section 4.2.3 the algorithm of the Forward Filtering and Backward Recursion in order to

estimate the dynamic parameters of the VECM (4.3). Based on the singular values of the time-varying cointegration matrix Π_t, we describe an estimate of the cointegration rank at each time in Section 4.5.1. We also give the method for estimating the time-varying cointegrating independent relations in Section 4.5.2.

In Section 4.7, this novel methodology is exploited on two synthetic data sets where we define different time periods with a different number of independent cointegrating relations. In these simulated processes, the cointegrating relations also evolve over time. Section 4.8.1 will be an application of this new approach on real data sets: a part of the European panel data set seen in Chapter 3 and three sectors of the Dow Jones, introduced in Section 4.8.2.

4.2 State space models and estimating the parameters

This section presents the general state space model and how we apply it to the Vector Error Correction Model (VECM).

4.2.1 The general state space model

The general state space model used in this chapter is composed of two main equations, the measurement equation and the transition equation (4.5). We use the same notation as for the state space model introduced by Primiceri (2005).

$$\begin{aligned} y_t &= H_t\beta_t + \epsilon_t \ , \quad \textit{Measurement equation} \\ \beta_t &= F\beta_{t-1} + u_t \ , \quad \textit{Transition equation} \end{aligned} \tag{4.5}$$

$$\text{where} \quad \begin{bmatrix} \epsilon_t \\ u_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R_t & 0 \\ 0 & G \end{bmatrix} \right) \tag{4.6}$$

Now let:

$$\beta_{t|s} = E(\beta_t | Y_s, H_s, R_s, G)$$

$$V_{t|s} = Var(\beta_t | Y_s, H_s, R_s, G)$$

Then, given $\beta_{0|0}$ and $V_{0|0}$, we use the following standard Kalman filter:

$$\begin{aligned}
\beta_{t|t-1} &= F\beta_{t-1|t-1} \\
V_{t|t-1} &= FV_{t-1|t-1}F' + G \\
K_t &= V_{t|t-1}H_t'(H_tV_{t|t-1}H_t' + R_t)^{-1} \\
\beta_{t|t} &= \beta_{t|t-1} + K_t(y_t - H_t\beta_{t|t-1}) \\
V_{t|t} &= V_{t|t-1} - K_tH_tV_{t|t-1}
\end{aligned} \tag{4.7}$$

The backward recursion is defined as follows. We first start to simulate β_T from its moment and variance given the information at time T: $\beta_{T|T}$ and $V_{T|T}$. Then recursively, for each t from $T-1$ to 1, we draw β_t from $\beta_{t|t+1}$ and $V_{t|t+1}$ where:

$$\begin{aligned}
\beta_{t|t+1} &= \beta_{t|t} + V_{t|t}F'V_{t+1|t}^{-1}(\beta_{t+1} - F\beta_{t|t}) \\
V_{t|t+1} &= V_{t|t} - V_{t|t}F'V_{t+1|t}^{-1}FV_{t|t}
\end{aligned} \tag{4.8}$$

4.2.2 State space model of the Vector Error Correction Model

We can rewrite the VECM (4.2) as:

$$y_t = \Theta_t Z_t + u_t, \quad u_t \sim N(0, \Sigma) \tag{4.9}$$

where:

$$\begin{aligned}
y_t &= \Delta x_t \text{ , (size } p \times 1) \\
\Theta_t &= (\Pi_t, \Psi_{1,t}, ..., \Psi_{k-1,t}) \text{ , (size } p \times pk) \\
Z_t &= (x_{t-1}', \Delta x_{t-1}', ..., \Delta x_{t-k+1}')' \text{ , (size } pk \times 1)
\end{aligned}$$

The VECM model (4.9) constitutes in fact the measurement equation of our state space model. Let us now define $\theta_t = Vec(\Theta_t)$ and $b = p \times pk = p^2k$ where:

$$\theta_t = Vec(\Theta_t) = Vec(\Pi_t, \Psi_t) \tag{4.10}$$

Then, from the measurement equation (4.9), we have:

$$\begin{aligned} Vec(y_t) = y_t &= Vec(\Theta_t Z_t) + Vec(u_t) \\ &= Vec(\Theta_t Z_t) + u_t \\ &= (Z_t{}' \otimes I_p) Vec(\Theta_t) + u_t \\ &\text{(because } Vec(AXB) = (B' \otimes A) Vec(X)) \\ &= (Z_t{}' \otimes I_p)\theta_t + u_t \end{aligned}$$

Hence, the measurement equation can be written as:

$$y_t = (Z_t{}' \otimes I_p)\theta_t + u_t, \quad u_t \sim N(0, \Sigma) \tag{4.11}$$

We now need to construct the dynamics of the state equation in the state space model, which implies that a $b \times b$ parameter matrix F should be introduced like in the transition equation in the general state space model shown in Equation (4.5). In our model we decided to simplify the lag parameter matrix F by a scale variable ρ to which we will apply a Bayesian inference later on in this chapter (see Section 4.3.2). This choice is made in order to avoid lengthy computations in the Forward Filtering Backward Recursion algorithm.

The time-varying models that are in the literature consider a constant variance of the errors for the transition equation, meaning that the expected evolution of the parameters θ_t is the same for all the time periods. A constant variance of the errors then allows gradual and smooth evolution of the parameters, which happens to be generally the case. We will denote by Q the variance matrix of the errors in the transition equation. Therefore the $b \times 1$ vector θ_t will have the following dynamics:

$$\theta_t = \rho\theta_{t-1} + \nu_t, \quad \nu_t \sim N(0, Q) \tag{4.12}$$

Finally, our state space model is summarised as follows:

$$\begin{aligned} y_t = (Z_t{}' \otimes I_p)\theta_t + u_t, \quad u_t \sim N(0, \Sigma) \quad &\textit{Measurement equation} \\ \theta_t = \rho\theta_{t-1} + \nu_t, \quad \nu_t \sim N(0, Q) \quad &\textit{Transition equation} \end{aligned} \tag{4.13}$$

Also, in the following sections, $\mathcal{D}_t$ will denote the information brought by the data at time t, i.e. the information brought by y_t and y_{t-1}. Furthermore, we can denote by $\mathcal{D}_{1:t}$ all the information brought by the data from time 1 to time t, that is, the information brought by $y_0, y_1, \cdots, y_t$.

4.2.3 Forward Filtering and Backward Recursion of the Vector Error Correction Model

This section describes the Kalman Filter applied to the Vector Error Correction Model. In this chapter, we use a two-filter smoothing algorithm in order to calculate the posterior distribution of the parameter vector θ_t (conditional on observations). The first algorithm goes forward in time from $t = 1$ to $t = T$ whereas the second algorithm goes backward in time from $t = T$ to $t = 1$ hence the notion of Forward Filtering and Backward Recursion algorithm (FFBS). This algorithm is also referred in the literature as the forward-backward smoothing algorithm. Gibbs sampling in state space models is achieved by this algorithm. In this methodology, we must rely on the posterior distributions of the backward recursion because the parameters θ_t are simulated given all the data: $\theta_t|\mathcal{D}_{0:T}$. If we take the posterior distributions of the forward filtering part, then these distributions would only depend on the previous data time points: $\theta_t|\mathcal{D}_{0:t}$, see Carter and Kohn (1994) and Fruewirth-Schnatter (1994).

The first part consists in creating the expectation and the variance of our parameter θ_t given the information at time t: $\theta_{t|t}$ and $P_{t|t}$. That first part is called the Forward Filtering algorithm (4.14). By applying the Kalman Filter (4.7) seen in Section 4.2.1 to our state space model (4.13) and given $\theta_{0|0}$ and $P_{0|0}$, we have:

$$\begin{aligned}
&\theta_{t|t-1} = \rho\theta_{t-1|t-1} \\
&P_{t|t-1} = \rho^2 P_{t-1|t-1} + Q \\
&K_t = P_{t|t-1}(Z_t \otimes I_p)((Z_t{}' \otimes I_p)P_{t|t-1}(Z_t \otimes I_p) + \Sigma)^{-1} \\
&\theta_{t|t} = \theta_{t|t-1} + K_t(y_t - (Z_t{}' \otimes I_p)\theta_{t|t-1})
\end{aligned} \tag{4.14}$$

$$P_{t|t} = P_{t|t-1} - K_t(Z_t' \otimes I_p)P_{t|t-1}$$

The following step in the estimation of the time-varying parameters is to introduce the backward recursion. This step constitutes the smoothing filtering part in which we collect the expectation and the variance of our parameter θ_t given the information at time $t+1$, $\forall t \in [\![1, T-1]\!]$. We use the same algorithm described by Primiceri (2005) for the VAR process, but applied to the Vector Error Correction Model instead. After the forward filtering steps, we first simulate θ_T from its moment and variance given the information at time T: $\theta_{T|T}$ and $P_{T|T}$. Then recursively, for each time t from $T-1$ to 1 (backward recursion), we draw θ_t from a multivariate normal distribution with mean $\theta_{t|t+1}$ and variance $P_{t|t+1}$ where:

$$\begin{aligned} \theta_{t|t+1} &= \theta_{t|t} + \rho P_{t|t} P_{t+1|t}^{-1}(\theta_{t+1} - \rho\theta_{t|t}) \\ P_{t|t+1} &= P_{t|t} - \rho^2 P_{t|t} P_{t+1|t}^{-1} P_{t|t} \end{aligned} \tag{4.15}$$

4.2.4 Bayesian inference on the covariance matrix Σ

For any time t, we assume Σ to have an Inverse-Wishart prior distribution implying two hyperparameters A and q like in the previous chapter (see Section 3.3.2):

$$\Sigma \sim IW(A, q) \tag{4.16}$$

Let us now write the likelihood of the measurement equation given the data. This likelihood is the one that takes into account all the information from time 1 to time T. It is therefore proportional to the product from time 1 to time T of the probability distribution functions of each error u_t:

$$L(\Sigma; \mathcal{D}_{1:T}, \theta_{0:T}) \propto \prod_{t=1}^{T} f(u_t|\mathcal{D}_{1:t}, \theta_t) \propto \prod_{t=1}^{T} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1} u_t u_t')\right)$$

giving:

$$L(\Sigma; \mathcal{D}_{1:T}, \theta_{0:T}) \propto |\Sigma|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1} \sum_{t=1}^{T} u_t u_t')\right) \tag{4.17}$$

where $\mathcal{D}_{1:T}$ represents all the information given by the data y_t from time 1 to T and $\theta_{0:T}$ represents all the information given by $\theta_t = Vec(\Pi_t, \Psi_t)$ from time 0 to T.

After that, it is straightforward to derive the expression of the posterior Inverse-Wishart distribution of Σ, conditional on all the parameters $\theta_{0:T}$ and all the data $\mathcal{D}_{1:T}$:

$$\Sigma|\mathcal{D}_{1:T}, \theta_{0:T} \sim IW(A + \sum_{t=1}^{T} u_t u_t', T + q) \tag{4.18}$$

where $u_t = y_t - (Z_t' \otimes I_p)\theta_t$.

4.3 Bayesian inference on the parameters of the transition equation: Q and ρ

In this section, we will begin by recalling our state space model (4.13):

$$\begin{aligned} y_t &= (Z_t' \otimes I_p)\theta_t + u_t, \;\; u_t \sim N(0, \Sigma) \quad \textit{Measurement equation} \\ \theta_t &= \rho\theta_{t-1} + \nu_t, \;\; \nu_t \sim N(0, Q) \quad \textit{Transition equation} \end{aligned} \tag{4.19}$$

The aim of this section is to present a Bayesian inference around the two parameters Q and ρ of the transition equation (Q and ρ are time invariant). We are going to assume ρ to be a priori uniformly distributed over the interval $[-1, 1]$ (see details in Section 4.3.2) on the one hand and the variance matrix Q to have an Inverse-Wishart prior on the other hand (see details in Section 4.3.3). The range of ρ was chosen to be between -1 and $+1$ in order for the process θ_t to be stationary (see the transition equation in (4.19)).

4.3.1 The likelihood of the transition equation

As previously seen in the section about the measurement equation, we need to derive the likelihood of the transition equation. In our case though, the parameters Q and ρ are not time-varying. It is then preferable to use all the information obtained over the period $[\![1, T]\!]$. For that, we need to remember the fact that the likelihood over the total period is proportional

to the product of all the individual likelihoods at time t. We decide to name the likelihood function of the transition equation as "LT" in order to avoid confusion with the likelihood of the measurement equation (4.17):

$$LT(Q, \rho|\theta_{0:T}) \propto \prod_{t=1}^{T} f(\nu_t|\mathcal{D}_{1:t}, \theta_t)$$

where for each $t \in [\![1, T]\!]$:

$$f(\nu_t|\mathcal{D}_{1:t}, \theta_t) \propto |Q|^{-1/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(Q^{-1}\nu_t\nu_t')\right)$$

with $\nu_t = \theta_t - \rho\theta_{t-1}$ (hence the need of the information at time t and $t-1$).

After that, we can derive the following expression of proportionality for the total likelihood of the transition equation:

$$LT(Q, \rho|\theta_{0:T}) \propto |Q|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{Tr}(Q^{-1}\sum_{t=1}^{T}\nu_t\nu_t')\right) \tag{4.20}$$

4.3.2 Bayesian inference on ρ: a uniform prior

A uniform prior for ρ

The condition $|\rho| < 1$, see Koop *et al.* (2011), must be assumed if the transition equation is a stationary AR(1) process, and we will set a uniform prior for ρ over $[-1, 1]$ ($\eta_1 = -1$ and $\eta_2 = +1$). Our uniform prior for ρ actually follows the suggestions from Huerta and West (1999) for prior structure and their related MCMC algorithm where they extend the possibility of having a non-stationary process ($|\rho| = 1$).

Conjugacy and posterior of ρ

In this chapter, the truncated normal distribution is denoted as "TN" and takes into account the two delimiting bounds required (finite or infinite) in the definition, between which the random variable can take its values. Here, the lower bound defined in the prior is η_1 and the upper bound is η_2 ($\eta_2 > \eta_1$). Thus, at iteration i of the MCMC procedure, we have:

$$\rho^{(i)}|Q^{(i-1)}, \theta_{0:T}^{(i)} \sim TN(\eta_1, \eta_2, \mu, s) \tag{4.21}$$

where μ and s depend both on $Q^{(i-1)}$ and $\theta_{0:T}^{(i)}$:

$$\mu = \mu(Q^{(i-1)}, \theta_{0:T}^{(i)}) = \frac{\sum_{t=1}^{T} \theta_t^{(i)\prime} Q^{-1(i-1)} \theta_{t-1}^{(i)}}{\sum_{t=1}^{T} \theta_{t-1}^{(i)\prime} Q^{-1(i-1)} \theta_{t-1}^{(i)}} \tag{4.22}$$

and

$$s = s(Q^{(i-1)}, \theta_{0:T}^{(i)}) = \frac{1}{\sqrt{\sum_{t=1}^{T} \theta_{t-1}^{(i)\prime} Q^{-1(i-1)} \theta_{t-1}^{(i)}}} \tag{4.23}$$

Due to the expression of μ (4.22), we can guess that the values of μ will be very close to 1 since the smoothing effect of the algorithm will bring the value of $\theta_t^{(i)}$ to be very close to $\theta_{t-1}^{(i)}$. Furthermore, the standard deviation (4.23) is expected to be quite small because of its denominator.

4.3.3 Bayesian inference on Q

Prior distribution for Q

Recall that Q is a positive definite covariance matrix of size $b \times b$, where $b = p^2 k$. We decide to represent it by an Inverse-Wishart distribution involving 2 hyperparameters, the $b \times b$ scale matrix B and the degrees of freedom w:

$$Q \sim IW(B, w) \tag{4.24}$$

Conjugacy and full conditional posterior distribution of Q

By multiplying the prior of Q (4.24) with the likelihood of the transition equation LT (4.20), we easily obtain the full conditional posterior distribution of Q, i.e. a conjugate Inverse-Wishart distribution:

$$Q^{(i)} | \rho^{(i)}, \theta_{0:T}^{(i)} \sim IW(B + \sum_{t=1}^{T} (\theta_t^{(i)} - \rho^{(i)} \theta_{t-1}^{(i)})(\theta_t^{(i)} - \rho^{(i)} \theta_{t-1}^{(i)})', w + T) \tag{4.25}$$

Also, for each iteration i, we decide to simulate ρ before Q (see Algorithm 5). The covariance matrix Q of the transition equation is sampled from the appropriate Inverse-Wishart distribution using the previously simulated ρ inside the same step i of the MCMC loop. In fact, the order

of the simulations of ρ and Q does not really matter (or very little). The results of this chapter would have not been significantly altered whether ρ was simulated before or after Q.

4.4 Initialization of the parameters and hyperparameters

In this section, we describe how we initialize the respective parameters of the measurement equation $\theta_t = Vec(\Pi_t, \Psi_t)$, Σ, ρ and Q: θ_0 , Σ_0 , ρ_0 and Q_0. We will also give arguments on the choice of the hyperparameters A and q for the measurement equation and B and w for the transition equation. Time $t = 0$ is taken as the initial time of the sample.

Initial parameters and hyperparameters for the measurement equation

For all the data sets studied in this chapter (see Sections 4.7, 4.8.1 and 4.8.2), we estimated the initial parameters of the VECM by the use of a pre-sample of a certain size $\tau < T$ with the methods developed by Luetkepohl (2006) (see Section 2.5). Therefore for any $t \in [\![1, \tau]\!]$ we will estimate:

$$\Delta x_t = \Pi_0 x_{t-1} + \sum_{j=1}^{k-1} \Psi_{0j} \Delta x_{t-j} + \epsilon_t$$

with $\epsilon_t \sim N(0, \Sigma_0)$.

Π_0, $(\Psi_{0j})_{1 \leq j \leq k-1}$ and Σ_0 are estimated from the methods stated in Section 2.5. The parameter vector θ_0 is obtained thanks to the estimation of Π_0 and $(\Psi_{0j})_{1 \leq j \leq k-1}$. Σ_0 will be used as the value of Σ at iteration 0 ($\Sigma^{(0)} \leftarrow \Sigma_0$) and for estimating the hyperparameter A.

Hyperparameters A and q are chosen like in the previous chapter (see Section 3.3.6): q is equal to $p + 4$ where p is the dimension of Σ and A is determined by the estimation of the error covariance matrix from the pre-sample by which we multiply $(q - p + 1)$ (see Section 3.3.6):

$$A = (q - p - 1) \times \Sigma_0$$

Initial parameters and hyperparameters in the transition equation

The parameter ρ of the transition equation is expected to have a value between -1 and $+1$, in order for the transition equation to be stable. We decide to input the initial value $\rho_0 = 1$, assuming a random walk for the transition equation at the beginning.

Values of Q should not be very large because we want to keep the smoothing effect in the estimation of the parameter θ_t of the transition equation. Therefore we need to define a sensible scale hyperparameter B and sensible degrees of freedom w for the Inverse-Wishart prior distribution of Q so that the order of magnitude of the elements contained in Q is small.

Since Q is distributed as an Inverse-Wishart, the expectation of Q in this prior is $B/(w-b-1)$. We will choose the hyperparameter w to be the usual addition of the dimension of Q and 4: $w = b + 4$. In addition, we would like the error covariance matrix of the transition equation not to be too big. From the full conditional distribution (4.25), we have that the sum of the estimated errors would be relatively small compared with B if B had large values. The value of ρ is generally very close to 1 due to the posterior distribution that concentrates around 1 (see Section 4.3.2). Therefore, a far too big value for B would neglect the part concerning the error terms $\sum_{t=1}^{T}(\theta_t^{(i)} - \rho^{(i)}\theta_{t-1}^{(i)})(\theta_t^{(i)} - \rho^{(i)}\theta_{t-1}^{(i)})'$ in the posterior distribution scale matrix of Equation (4.25). Thanks to a sensitivity analysis conducted on Q (see Section 4.7.3), we noticed that the dynamic of the cointegration rank was much more volatile over time when B was taken to be too large ($B = 10^{-2} \times I_b$ or $10^{-3} \times I_b$). Therefore, we will choose an order of magnitude of 10^{-4} or 10^{-5} for B: when starting by using this order of magnitude for $B = 10^{-4} \times I_b$, we saw that the dynamic of the cointegration rank was much more accurate with our simulated data sets (see Section 4.7.3). In our examples on the simulated data sets, we discuss the choice of the hyperparameters w and B and how they can have an impact on the final results (see Section 4.7).

Size of the pre-sample

The data sets treated in this chapter contain a lot of time points ($T \approx 1{,}000$) in general, and the size of the pre-sample should therefore be small. We decide to have a pre-sample of size $\tau = 50$ and estimate the parameters and hyperparameters of the measurement equation from this pre-sample.

4.5 Time-varying cointegration: the rank and the cointegrating relations

4.5.1 Evolution of the cointegration rank

One of the goals in introducing the VECM as a state space model is also to see if the cointegration rank is evolving over time. We have seen in Chapter 3 that the cointegration rank can be estimated from the long-run impact matrix Π of the VECM. This method of estimation was based on the singular values of the matrix Π and on the contribution of the smallest singular values. We compared this contribution of the last singular values with a parameter called the irrelevance criterion, ϵ.

With the state space model introduced in (4.9), we can extract a time-varying cointegrating matrix Π_t from each $\theta_t = Vec(\Theta_t)$ (4.10), and derive a time-varying cointegration rank r_t.

4.5.2 Evolution of the independent cointegrating relations

It is of course theoretically possible to see the evolution of the cointegrating relations over all the data period of length T. But that would mean analysing T cointegrating matrices Π_t for each t deriving its cointegration rank and its independent cointegrating relations contained in the $p \times r_t$ matrix β_t.

Therefore, for each time t and based on the last 50 iterations of Π_t, we can derive the mean cointegrating matrix $\Pi_{mean,t}$ and the median cointegration rank $r_{median,t}$. Then, by using the same

methods as in Section 3.3.7, we can obtain the independent cointegrating relations contained in a matrix β_t:

$$\beta_t{}' = (\alpha_t{}'\alpha_t)^{-1}\alpha_t{}'\Pi_t = \begin{bmatrix} I_{r_t} & B_t{}' \end{bmatrix}$$

If we were in three dimensions ($p = 3$) and the cointegration rank was 2, e.g., we could imagine the two cointegrating vectors in a 3-D graph changing progressively their length and directions over time.

4.6 Recapitulation of the algorithm

The algorithms described in this section are based on the assumption that our data set is defined for a large time period ($T > 500$). Financial data sets consist in general of daily data and imply large data sets (see Section 4.8.2). The pre-sample taken to calculate the initial parameters and to estimate some hyperparameters can then be a small part of the actual data set, of size $\tau = 50$, for example. The real data set applied to a European panel has, however, less data time points (due to monthly data) and we will therefore specify another size for the pre-sample (see Section 4.8.1).

Algorithm 4 Initialization of the parameters

Set the size of the pre-sample $\tau = 50$. The data set treated is of size $T > 500$.

For the pre-sample data set containing observations $[\![1, \tau]\!]$:

Elements of the measurement equation:

- Initialize $\Pi_{0|0}^{(0)} \leftarrow \widehat{\Pi}$, $\Psi_{0|0}^{(0)} \leftarrow \widehat{\Psi}$, $\Sigma^{(0)} \leftarrow \widehat{\Sigma}$ from the LS estimates (2.30) seen in Section 2.5.
- Thanks to $\Pi_{0|0}^{(0)}$ and $\Psi_{0|0}^{(0)}$ we can construct the $b \times 1$ vector $\theta_{0|0}^{(0)}$: $\theta_{0|0}^{(0)} = Vec(\Pi_{0|0}^{(0)}, \Psi_{0|0}^{(0)})$.

Set the values of the hyperparameters:

- $q = p + 4$ and $A = (q - p - 1) \times \Sigma^{(0)}$

Elements of the transition equation:

- Q_0 , ρ_0 are initialized as being $Q_0 = 10^{-5} \times I_b$ and $\rho_0 = 1$ (see the justification in Section 4.4).
- Initialize $P_{0|0}$ as being the identity matrix (of size $b \times b$): $P_{0|0} = I_b$.

Algorithm 5 General Gibbs for the dynamic VECM

Set the number of iterations to be $m = 300$.

for $i \in [\![1, m]\!]$ **do**

for $t \in [\![1, T]\!]$ **do**

Forward filtering:

- Obtain $\theta_{t|t}^{(i)}$ and $P_{t|t}^{(i)}$ thanks to the forward filtering part of the algorithm (i.e. from Equation (4.14) in Section 4.2.3).

end for

- Set $\theta_T^{(i)} \leftarrow \theta_{T|T}^{(i)}$. Set $P_T^{(i)} \leftarrow P_{T|T}^{(i)}$.

for $t \in \{T-1, T-2, \cdots, 1\}$ **do**

Backward recursion:

- Obtain $\theta_t^{(i)}$ and $P_t^{(i)}$ thanks to the backward recursion part of the algorithm (i.e. from Equation (4.15) in Section 4.2.3).

end for

for $t \in \{1, 2, \cdots, T\}$ **do**

- Extract $\Pi_t^{(i)}$ from $\theta_t^{(i)}$: The columns of $\Pi_t^{(i)}$ are the first column elements in $\theta_t^{(i)}$.
- Estimate $r_t^{(i)}$ from $\Pi_t^{(i)}$ based on the irrelevance of the last singular values of $\Pi_t^{(i)}$ (see Section 3.2).

end for

- Sample $\Sigma^{(i)}$ from the Inverse-Wishart posterior distribution stated in Equation (4.18) in Section 4.2.4.
- Sample $\rho^{(i)}$ from the truncated normal posterior distribution stated in Equation (4.21) in Section 4.3.2.
- Sample $Q^{(i)}$ from the Inverse-Wishart posterior distribution stated in Equation (4.25) in Section 4.3.3.

end for

Algorithm 6 Final decisions for the time-varying cointegration rank and independent relations

for $t \in [\![1, T]\!]$ **do**

- Calculate the final time-varying cointegration ranks as follows:

$r_{median,t} \leftarrow \underset{\{m-50 \leq i \leq m\}}{\text{median}} ({r_t}^{(i)})$.

- Calculate the final time-varying cointegrating matrices as follows:

$\Pi_{mean,t} \leftarrow \underset{\{m-50 \leq i \leq m\}}{\text{mean}} ({\Pi_t}^{(i)})$.

- Calculate the final independent time-varying cointegrating relationships using the operation of Chapter 3 and recapitulated for each t in Section 4.5.2 as:

${\beta_{mean,t}}' \leftarrow ({\alpha_t}'\alpha_t)^{-1}{\alpha_t}'\Pi_{mean,t} = \begin{bmatrix} I_{r_t} & {B_t}' \end{bmatrix}$

end for

A Directed Acyclic Graph (DAG) is helpful to understand better the Bayesian network built in this chapter and described by the previous algorithms. Such a graph is represented in Figure 4.1 below. In this Figure, we can see two plates with one plate embedded in the other. Outside the plates are the parameters that neither depend on the iterations of the MCMC nor the time points. The hyperparameters A, q, η_1, η_2, B and w are fixed before running the algorithm and lie outside the two plates. The biggest plate contains the fixed parameters of the state space models, that are the parameters that are simulated at each iteration step of the MCMC. The small plate, that is embedded in the big plate, contains the parameters of the VECM that are simulated at each time point for each iteration of the MCMC. In particular, the cointegrating matrix Π belongs to the small plate in which a cointegration rank is estimated and will depend on time.

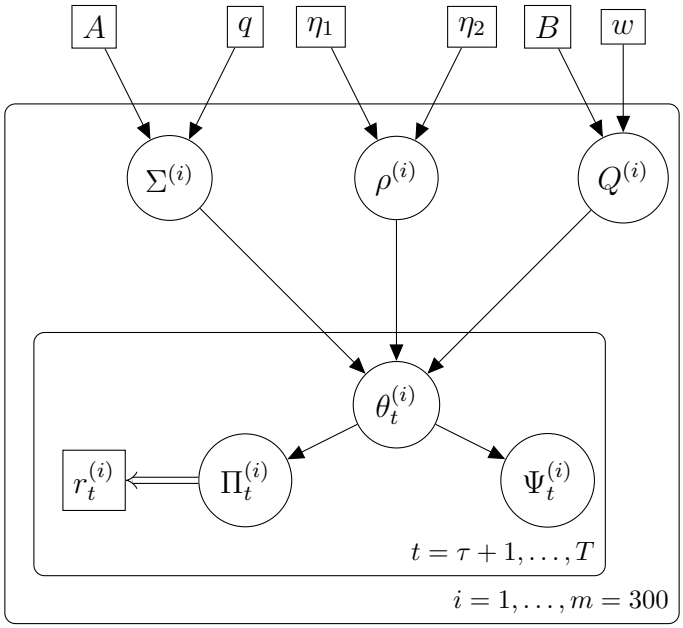

Figure 4.1: Directed Acyclic Graph of the Bayesian network established in this chapter: Square boxes contain the fixed parameters, circles contain the randomized parameters.

4.7 Simulated data

4.7.1 Description of the data

First of all, we build one simulated data set that we call D_1. This set consists of five difference stationary time series for which we create a certain number of cointegrating relations that will evolve over different time periods. Our first data set D_1 will consist of five time series of length $T = 1050$ and divided into three time periods. Each time period has its specific cointegrating relations and each time period has a different cointegration rank (i.e. number of independent cointegrating relations). In the first time-period (of length 350) we have four cointegrating relations, in the second time period (of length 300) we have three cointegrating relations and finally the last time period (of length 400) deals with four cointegrating relations.

Let $x_t = (x_{it})_{1 \leq i \leq 5}$, $\forall t \in [\![1, T = 1050]\!]$ represent our simulated data D_1. We start by simulating $v_{1t} \quad \forall t \in [\![1, \tau_1 = 350]\!]$ from the standard normal distribution, thus giving it a

stationary signal: $v_{1t} \sim I(0)$. As a consequence, $\sum_{k=1}^{t} v_{1k}$ is difference stationary or $I(1)$. We decide to give our first time series of the list, x_{1t}, a difference stationary signal, i.e. $\forall t \in [\![1, \tau_1 = 350]\!]$, $x_{1t} = \sum_{k=1}^{t} v_{1k}$. After that we can build the time series x_{2t}, x_{3t}, x_{4t} and x_{5t} as independent linear combinations of the previous time series so that we have 4 cointegrating relations:

$$\forall t \in [\![1, \tau_1]\!], \quad \begin{aligned} x_{1t} &= \textstyle\sum_{k=1}^{t} v_{1k} \sim I(1) \\ x_{2t} &= x_{1t} + v_{2t} \sim I(1) \Rightarrow y_{1t} = x_{2t} - x_{1t} \sim I(0) \\ x_{3t} &= x_{2t} + x_{1t} + v_{3t} \sim I(1) \Rightarrow y_{2t} = x_{3t} - x_{2t} - x_{1t} \sim I(0) \\ x_{4t} &= x_{2t} + v_{4t} \sim I(1) \Rightarrow y_{3t} = x_{4t} - x_{2t} \sim I(0) \\ x_{5t} &= x_{4t} + v_{5t} \sim I(1) \Rightarrow y_{4t} = x_{5t} - x_{4t} \sim I(0) \end{aligned}$$

where v_{2t}, v_{3t}, v_{4t}, and v_{5t} are simulated from the standard normal distribution.

We can see that we have four cointegrating relations giving our first period of the data set a cointegration rank of 4. These cointegrating relations are respectively:

$$\forall t \in [\![1, \tau_1]\!], \quad \begin{aligned} y_{1t} &= x_{2t} - x_{1t} = v_{2t} \sim I(0) \\ y_{2t} &= x_{3t} - x_{2t} - x_{1t} = v_{3t} \sim I(0) \\ y_{3t} &= x_{4t} - x_{2t} = v_{4t} \sim I(0) \\ y_{4t} &= x_{5t} - x_{4t} = v_{5t} \sim I(0) \end{aligned}$$

Then, our second time period has the following three cointegrating relations (therefore a cointegration rank of 3):

$$\forall t \in [\![\tau_1 + 1 = 351, \tau_2 = 650]\!], \quad \begin{aligned} x_{1t} &= \textstyle\sum_{k=1}^{t} v_{1k} \sim I(1) \\ x_{2t} &= \textstyle\sum_{k=1}^{t} v_{2k} \sim I(1) \\ x_{3t} &= x_{2t} + x_{1t} + v_{3t} \sim I(1) \Rightarrow y_{1t} = x_{3t} - x_{2t} - x_{1t} \sim I(0) \\ x_{4t} &= x_{2t} + v_{4t} \sim I(1) \Rightarrow y_{2t} = x_{4t} - x_{2t} \sim I(0) \\ x_{5t} &= x_{4t} + v_{5t} \sim I(1) \Rightarrow y_{3t} = x_{5t} - x_{4t} \sim I(0) \end{aligned}$$

Finally, the last time period of the data set D_1 has the following four independent cointegrating relations (cointegration rank 4):

$$\forall t \in [\![\tau_2 + 1 = 651, T = 1050]\!], \quad \begin{aligned} & x_{1t} = \sum_{k=1}^{t} v_{1k} \sim I(1) \\ & x_{2t} = x_{1t} + v_{2t} \sim I(1) \Rightarrow y_{1t} = x_{2t} - x_{1t} \sim I(0) \\ & x_{3t} = x_{2t} + x_{1t} + v_{3t} \sim I(1) \Rightarrow y_{2t} = x_{3t} - x_{2t} - x_{1t} \sim I(0) \\ & x_{4t} = x_{2t} + v_{4t} \sim I(1) \Rightarrow y_{3t} = x_{4t} - x_{2t} \sim I(0) \\ & x_{5t} = x_{4t} + v_{5t} \sim I(1) \Rightarrow y_{4t} = x_{5t} - x_{4t} \sim I(0) \end{aligned}$$

Let us now describe our second simulated data set called D_2. This data set is composed of $p = 4$ time series over a time period of length $T = 850$. In this data set, we distinguish two time periods. The first time period $[\![1, \tau = 450]\!]$ will have three independent cointegrating relations whereas the second time period $[\![\tau+1 = 651, T]\!]$ will have two independent cointegrating relations.

Let $x_t = (x_{it})_{1 \leq i \leq 4}$, $\forall t \in [\![1, T = 1050]\!]$, represent our simulated data. We can build the time series x_{2t}, x_{3t} and x_{4t} as independent linear combinations of the previous time series so that we obtain 4 cointegrating relations:

$$\forall t \in [\![1, \tau]\!], \quad \begin{aligned} & x_{1t} = \sum_{k=1}^{t} v_{1k} \sim I(1) \\ & x_{2t} = x_{1t} + v_{2t} \sim I(1) \Rightarrow y_{1t} = x_{2t} - x_{1t} \sim I(0) \\ & x_{3t} = x_{2t} + x_{1t} + v_{3t} \sim I(1) \Rightarrow y_{2t} = x_{3t} - x_{2t} - x_{1t} \sim I(0) \\ & x_{4t} = x_{2t} + v_{4t} \sim I(1) \Rightarrow y_{3t} = x_{4t} - x_{2t} \sim I(0) \end{aligned}$$

where v_{1t}, v_{2t}, v_{3t} and v_{4t} are simulated from the standard normal distribution.

As for the second time period, D_2 will have two independent cointegrating relations:

$$\forall t \in [\![\tau + 1 = 651, T = 1050]\!], \quad \begin{aligned} & x_{1t} = \sum_{k=1}^{t} v_{1k} \sim I(1) \\ & x_{2t} = x_{1t} + v_{2t} \sim I(1) \Rightarrow y_{1t} = x_{2t} - x_{1t} \sim I(0) \\ & x_{3t} = x_{2t} + x_{1t} + v_{3t} \sim I(1) \Rightarrow y_{2t} = x_{3t} - x_{2t} - x_{1t} \sim I(0) \\ & x_{4t} = \sum_{k=1}^{t} v_{4k} \sim I(1) \end{aligned}$$

4.7.2 Implementation of the code for the simulated data sets

For both of our simulated data sets, we take a pre-sample size of order of magnitude $\tau = 50$. In the time-period of the pre-sample $[\![1, \tau]\!]$, we then initialize the parameters of the VECM thanks to the estimation methods of Luetkepohl stated in Section 2.5. On the pre-sample data set containing observations $[\![1, \tau]\!]$, we can obtain $\widehat{\Pi}$, $\widehat{\Psi}$ and $\widehat{\Sigma}$ that will be the values given to the initial parameters of the Gibbs sampler (see Algorithm 4). These initial parameters are the parameters Π and Ψ of the VECM at iteration 0 and at time 0 given 0, where the forward filtering part of the algorithm starts: $\Pi_{0|0}^{(0)}$, $\Psi_{0|0}^{(0)}$. From these latter two parameters we can build the first element $\theta_{0|0}^{(0)}$ of the forward filtering algorithm. The covariance matrix of $\theta_{0|0}^{(0)}$ will be equal to the identity matrix, i.e. $P_{0|0} = I_b$.

In the measurement equation, the parameter Σ is time-invariant (see Section 4.1) and the value $\widehat{\Sigma}$ will be given to Σ at iteration 0: $\Sigma^{(0)} \leftarrow \widehat{\Sigma}$. As for the transition equation, the parameters Q_0 and ρ_0 are set as follows: $Q_0 = 10^{-5} \times I_b$ and $\rho_0 = 1$ (see Section 4.4).

The hyperparameters used in the measurement equation are q and A and will be given values in the same way as in the previous chapter: $q = p+4$ and $A = (q-p-1) \times \Sigma^{(0)}$ (see Section 3.3.6). The hyperparameters of the Inverse-Wishart covariance matrix Q in the transition equation are taken as being equal to $B = 10^{-5} \times I_b$ and $w = b + 4$ (i.e. the dimension of Q plus four). The bounds of the uniform prior of ρ are -1 and $+1$ (see Section 4.4). In the next section (see Section 4.7.3), we show that taking the hyperparameter scale B with too large values affects the results of finding the cointegration rank for the first simulated data set D_1. However on the same example, we show that increasing the degrees of freedom w a little will not affect the derivation of the cointegration rank (see Section 4.7.3).

At each iteration i, the Forward Filtering and Backward Recursion algorithm is then running for the sample corresponding to the time period $[\![\tau+1, T]\!]$. At the end, we obtain all the posterior distributions from which one $\theta_t^{(i)}$ is simulated (see Algorithm 5) at each time t of iteration i. We then take the cointegrating matrix $\Pi_t^{(i)}$ contained in $\theta_t^{(i)}$ and estimate the cointegration rank $r_t^{(i)}$ at time t by the methods based on the irrelevance criterion seen in Chapter 3 (see Section 3.2).

The number of iterations is equal to $m = 300$ and we have a burn-in set of 250 iterations. For each time t, we will derive the mean of the 50 last cointegrating matrices $\Pi_t^{(i)}$ and the median of the 50 last cointegration rank $r_t^{(i)}$. We will then obtain at each time point t, a cointegrating matrix $\Pi_{mean,t}$ and a cointegration rank $r_{median,t}$. Based on the cointegration rank at each time t, we will be able to obtain the $r_{median,t}$ independent cointegrating relations from $\Pi_{mean,t}$, see Section 4.5.2 and Algorithm 6.

4.7.3 Estimation of the cointegrating parameters

In Chapter 3, we found that an irrelevance criterion of 4% was actually the one that fit the most with our simulations. However it is still good to investigate several values of ϵ around 5% in this chapter as well. For the first simulated data set consisting of three different time periods with a different rank, the resulting median cointegration rank is given in Figure 4.2. The fact that the pre-sample was of size 50 implies that the cutting point from Period 1 to Period 2 is $t = 300$ and the cutting point from Period 2 to Period 3 is $t = 600$. We can see from Figure 4.2 that an irrelevance criterion of $\varepsilon = 3 - 5\%$ captures well the time points when the cointegration rank changes.

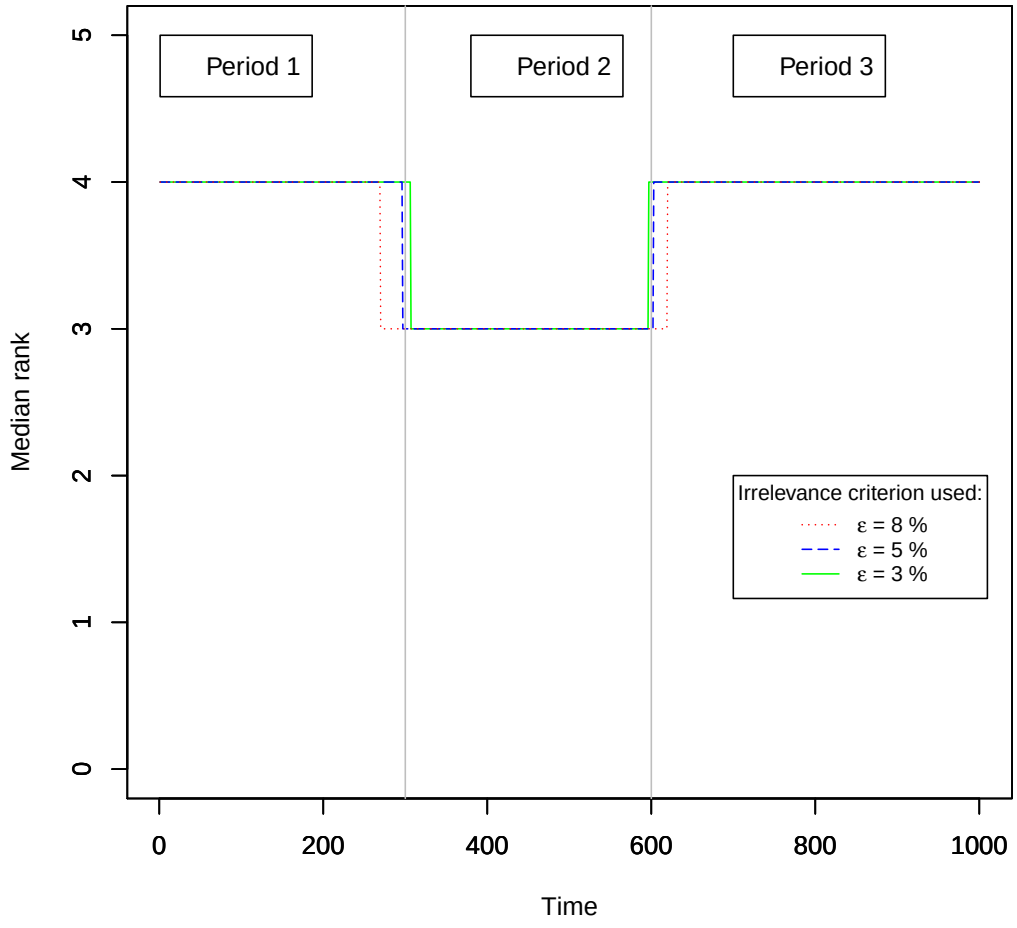

Figure 4.2: Evolution of the median cointegration rank in the simulated data set with $p = 5$ time series: cointegration rank is 4 from $t = 0$ to $t = 300$, cointegration rank 3 from $t = 300$ to $t = 600$ and cointegration rank 4 from $t = 600$ to $t = 1000$ ($m = 300$ iterations).

In Section 4.4 we discussed about asserting the value of the hyperparameters of the Inverse-Wishart prior in the transition equation. We stated that too large values for B would make the estimation of the rank volatile and therefore an order of magnitude of 10^{-5} is adapted because B would not then overwhelm the information added by the sum of squared errors in Equation (4.25). Figure 4.3 represents the time-varying estimation of the cointegration rank for the synthetic data

set D_1 when we set B to be $10^{-2} \times I_b$. We can see that the dynamics of the cointegration rank does not reflect reality. The cointegration rank keeps having the same value ($r = 4$) for irrelevance criteria of $\varepsilon = 5\%$ and 8% whereas for the small criterion $\varepsilon = 3\%$, the resulting cointegration rank seems to be very volatile and taking wrong values between 4 and 5.

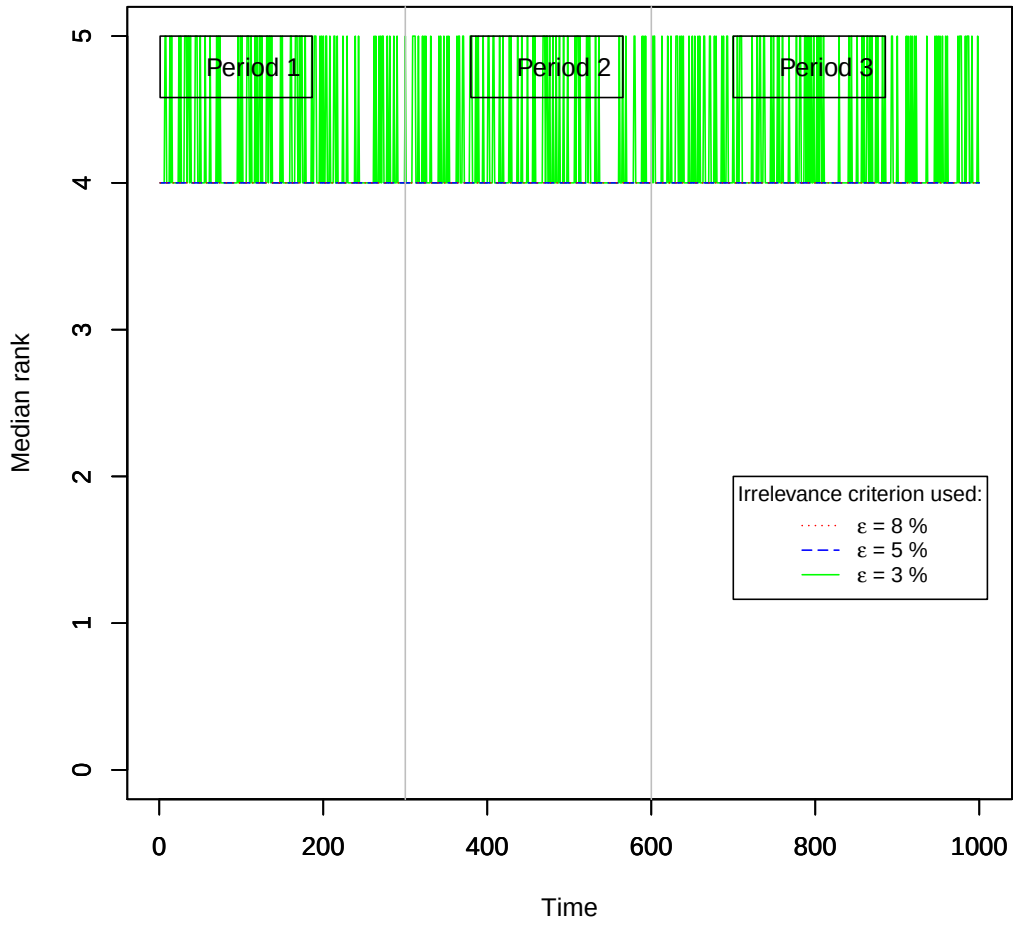

Figure 4.3: Evolution of the median cointegration rank for the simulated data set D_1 with $p = 5$ time series: cointegration rank is 4 from $t = 0$ to $t = 300$, cointegration rank 3 from $t = 300$ to $t = 600$ and cointegration rank 4 from $t = 600$ to $t = 1000$ ($m = 300$ iterations). Here we use $B = 10^{-2} \times I_b$ and $w = b + 4 = 54$.

In Figure 4.4, we have used an order of magnitude of 10^{-5} for the hyperparameter B but we have chosen a larger value of w ($w = b + 10$). We can see, on the other hand, that changing a little the value of w does not really change the dynamics of the cointegration rank over time. Therefore we conclude that a small matrix B (of magnitude 10^{-5}) will smooth the dynamics of the cointegration rank. Values of w that are not too much above the threshold $b+4$ do not have an impact on the results.

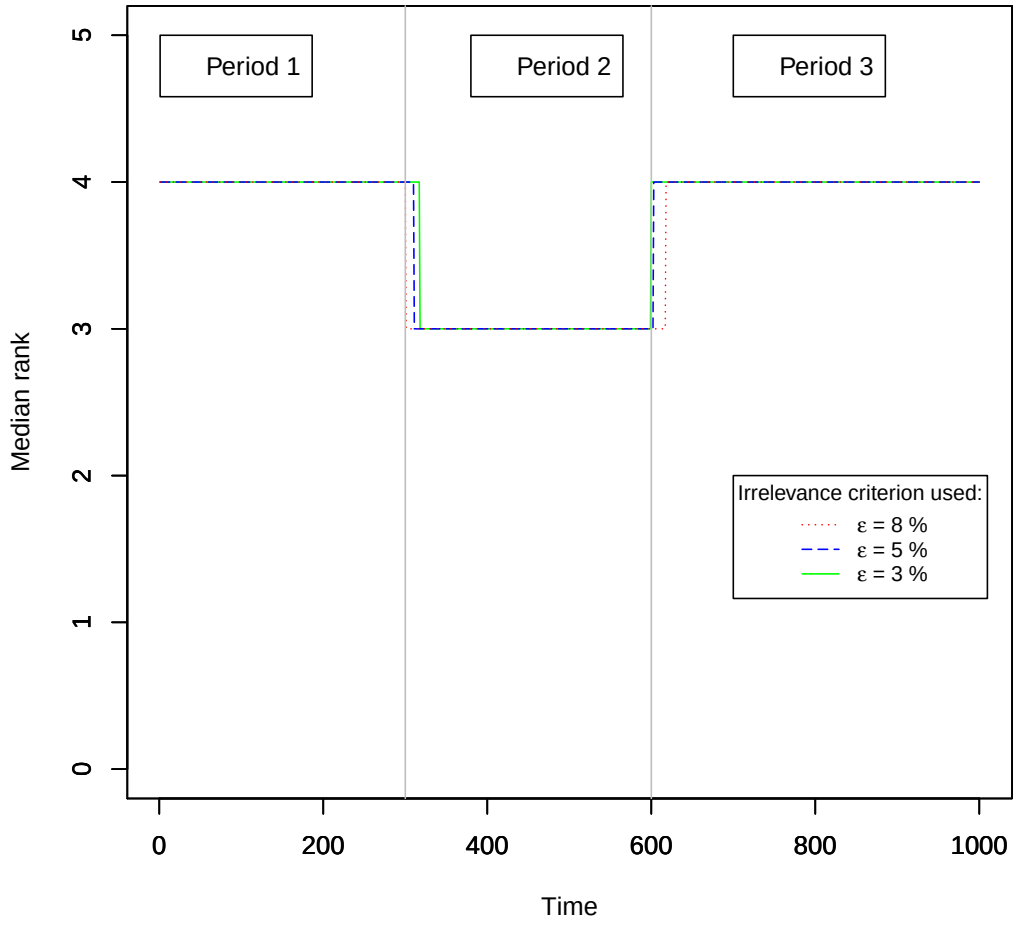

Figure 4.4: Evolution of the median cointegration rank in the simulated data set with $p = 5$ time series: cointegration rank is 4 from $t = 0$ to $t = 300$, cointegration rank 3 from $t = 300$ to $t = 600$ and cointegration rank 4 from $t = 600$ to $t = 1000$ ($m = 300$ iterations). Here we use $B = 10^{-5} \times I_b$ and $w = b + 10 = 60$.

By applying the good hyperparameters B and w on the second simulated data set, we obtain the dynamic of the cointegration rank in Figure 4.5. As we can see from that Figure, this estimation of the rank is reasonably accurate. The cutting time point between Period 1 and Period 2 is $t = 400$.

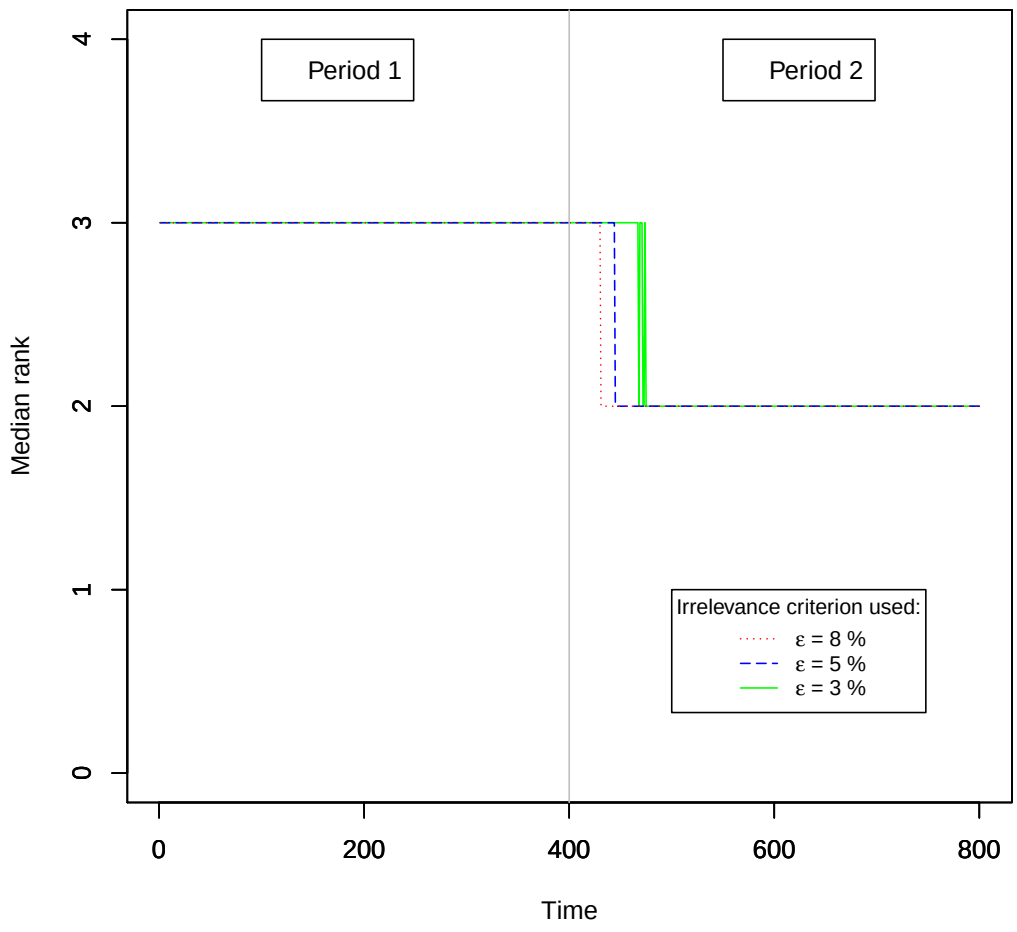

Figure 4.5: Evolution of the median cointegration rank in the simulated data set with $p = 4$ time series: cointegration rank of 3 from $t = 0$ to $t = 400$, cointegration rank of 2 from $t = 600$ to $t = 800$ ($m = 300$ iterations).

We can see the accuracy of the estimation of the cointegrating coefficients in the tables provided below. They compare the theoretical cointegrating coefficients (Tables 4.1, 4.3 and 4.5) with the estimated coefficients that are obtained from the algorithm (respectively, Tables 4.2, 4.4 and 4.6).

Table 4.1: Independent cointegrating relations for Period 1: coefficients created in the simulated data (left) transformed into the form $[I_r, B']$ (right).

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	-1	0	0	0
0	1	0	-1	0
0	0	1	-1	-1
0	0	0	1	-1

$\Longrightarrow$

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	0	0	0	-1
0	1	0	0	-1
0	0	1	0	-2
0	0	0	1	-1

Table 4.2: Independent cointegrating relations for Period 1: coefficients found for β_1 after running the algorithm.

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	0	0	0	-0.976
0	1	0	0	-0.970
0	0	1	0	-1.962
0	0	0	1	-0.982

Table 4.3: Independent cointegrating relations for Period 2: coefficients created in the simulated data (left) transformed into the form $[I_r, B']$ (right).

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	-1	0	0	0
0	1	0	-1	0
0	0	1	-1	-1

$\Longrightarrow$

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	0	0	-1	0
0	1	0	-1	0
0	0	1	-1	-1

Table 4.4: Independent cointegrating relations for Period 2: coefficients found for β_1 after running the algorithm.

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	0	0	-0.973	-0.010
0	1	0	-0.970	-0.017
0	0	1	-0.960	-0.970

Table 4.5: Independent cointegrating relations for Period 3: coefficients created in the simulated data (left) transformed into the form $[I_r, B']$ (right).

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	-1	0	0	0
0	1	0	-1	0
0	0	1	-1	-1
0	0	0	1	-1

$\Longrightarrow$

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	0	0	0	-1
0	1	0	0	-1
0	0	1	0	-2
0	0	0	1	-1

Table 4.6: Independent cointegrating relations for Period 3: coefficients found for β_1 after running the algorithm.

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	0	0	0	-0.979
0	1	0	0	-0.985
0	0	1	0	-1.946
0	0	0	1	-0.989

The dynamic Error Correction model introduced in this chapter does not pretend to be infallible but it must be seen as a first approach for estimating the cointegration rank over time, and also for deriving independent cointegration relations from the time-varying matrix Π_t. It

can happen that the rank estimated is not exactly the same as the actual rank. In our real data, where we are not supposed to know the rank before, we will see if the time-varying independent cointegration relations fit with the cointegrating rank found: We will need to check if the relations are long-run stationary over a neighbourhood about the time of interest.

4.7.4 Posterior summaries

We present in this section the posterior summaries of some time-varying cointegrating coefficients and the covariance matrix Σ of the dynamic VECM introduced in this chapter. These posterior summaries are taken from the results of the first simulated data set D_1 after applying Algorithms 4 and 5.

Figure 4.6 shows a reasonable convergence for the trace plots of some cointegrating coefficients Π_{32}, Π_{25}, Π_{15} at the specific time $t = 100$. The dynamic model indeed implies at each time t a different posterior distribution for the time-varying parameters. In this chapter, we make the assumption of having a time-invariant covariance matrix Σ, the trace plot of Σ_{43} presented on the bottom right is the same for each time t and shows a clear convergence.

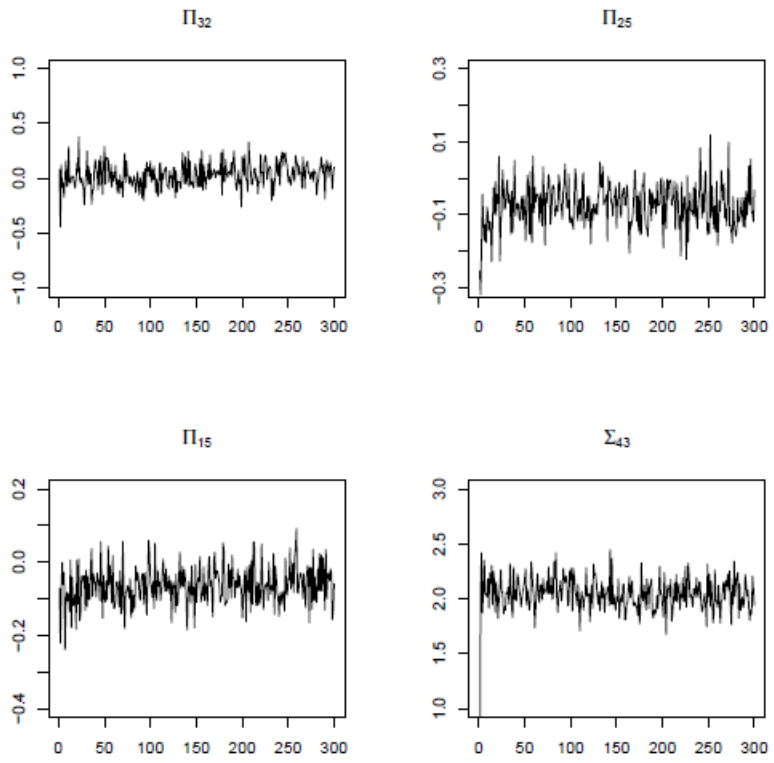

Figure 4.6: Trace plots of the coefficients Π_{32}, Π_{25}, Π_{15} at $t = 100$ and Σ_{43} for the first simulated data set D_1.

Figure 4.7 shows the posterior densities of the same coefficients for which the trace plots are presented in Figure 4.6. Figure 4.8 highlights the change of position over time of the posterior density of one of the dependent cointegrating coefficients Π_{32}. We can see the mean of the density shifting over 4 different time points: $t = 100$, $t = 300$, $t = 600$ and $t = 800$.

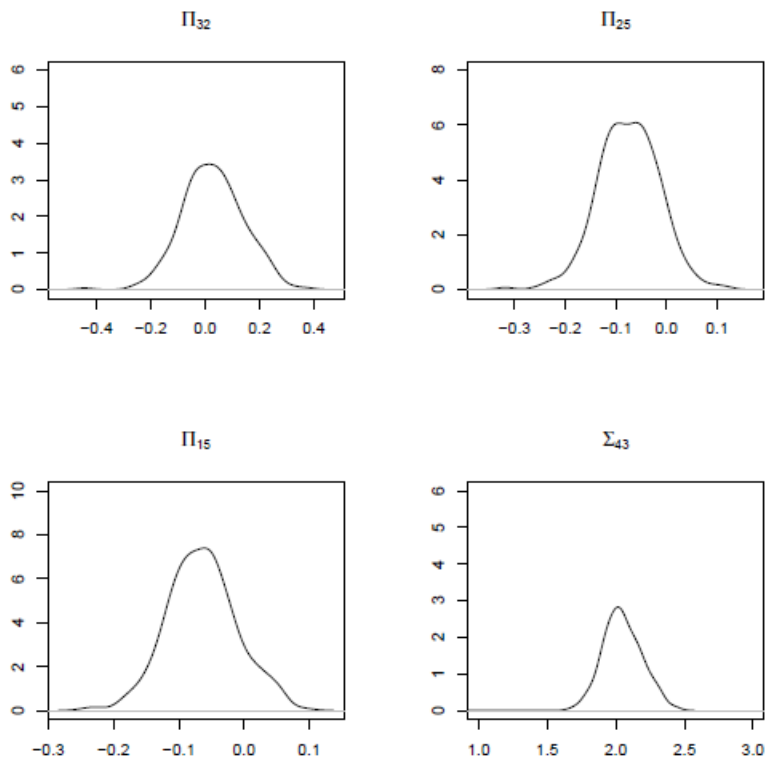

Figure 4.7: Posterior densities of the coefficients Π_{32}, Π_{25}, Π_{15} at $t = 100$ and Σ_{43} for the first simulated data set D_1.

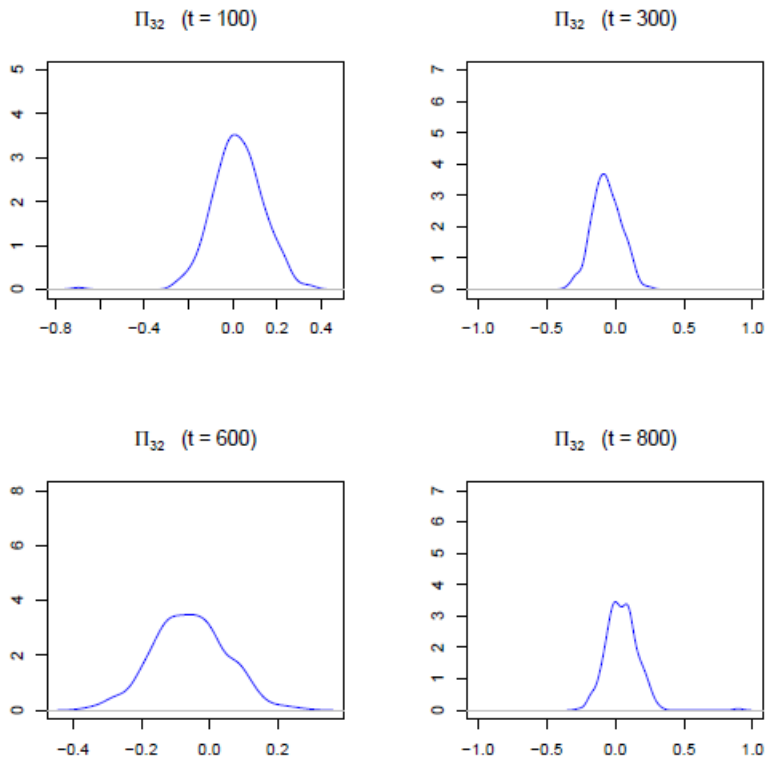

Figure 4.8: Posterior densities of the coefficient Π_{32} for the first simulated data set D_1 taken at different times: $t = 100$ (Mean $= 0.012$), $t = 300$ (Mean $= -0.041$), $t = 600$ (Mean $= -0.075$) and $t = 800$ (Mean $= 0.009$).

4.8 Application to real data sets

4.8.1 Application to the European panel data

In this section, we still study the dynamic of the cointegration rank for a part of our European panel data set seen in Chapter 3. The data set used in this section consists of three net tradings and three long-term interest rates between France, Germany and Spain over the period before and after the Euro. The data time period covers both periods of before and after introduction of the Euro: 1991-2008. The interest in this section is to see the evolution of the number of

independent cointegrating relations by grouping the two time periods of before the Euro (1991-1998) and the decade after (1999-2008). We decided to limit the number of time series to 6 rather than the entire European panel data set. If we take the 12 time series as in Chapter 3, the FFBS algorithm encounters problems of memory space. Furthermore the time for the algorithm to run becomes very long as the number of time series p increases. We decided not to take more than 6 time series for that reason.

The number of total time points in that set of time series is only $T = 211$ (covering the monthly data taken from January 1991 until July 2008). Unlike for the simulated data sets (see Section 4.7), we decide for this data set to take a pre-sample size of $\tau = 26 \approx T/8$ in order to be able to have the maximum number of data time points in our sample $[\![\tau + 1, T]\!]$. The data set starts from March 1993. For this data set, we initialize the parameters and hyperparameters in the same way as for the simulated data described in Section 4.7.2 of this chapter. The FFBS Algorithm 5 then runs and we estimate the time-moving cointegration rank as described in Algorithm 6. We can then obtain the independent cointegrating relations that we want for any time t (see Section 4.7.2).

In Figure 4.9 below, we display the evolution of the cointegration rank between the time series FraNX, GerNX, SpaNX, FraIR, GerIR and SpaIR. The whole time-period starts in April 1993 because we took a pre-sample time-period between January 1991 and March 1993 in order to estimate the initial parameters (see Algorithm 4). From June 1997, we observe that the cointegration rank is decreasing, suggesting less relations between the three economies, and thus less convergence. We can think that the cointegration rank is maybe decreasing in advance compared with what has been concluded previously in Chapter 3, that is, less convergence between European economies after the introduction of the Euro. On the other hand, we can state that the cointegration rank started to decrease in 1997 and that the adoption of the Euro may not be the principal reason as to why the European economies diverged in the post-Euro decade (1999-2008).

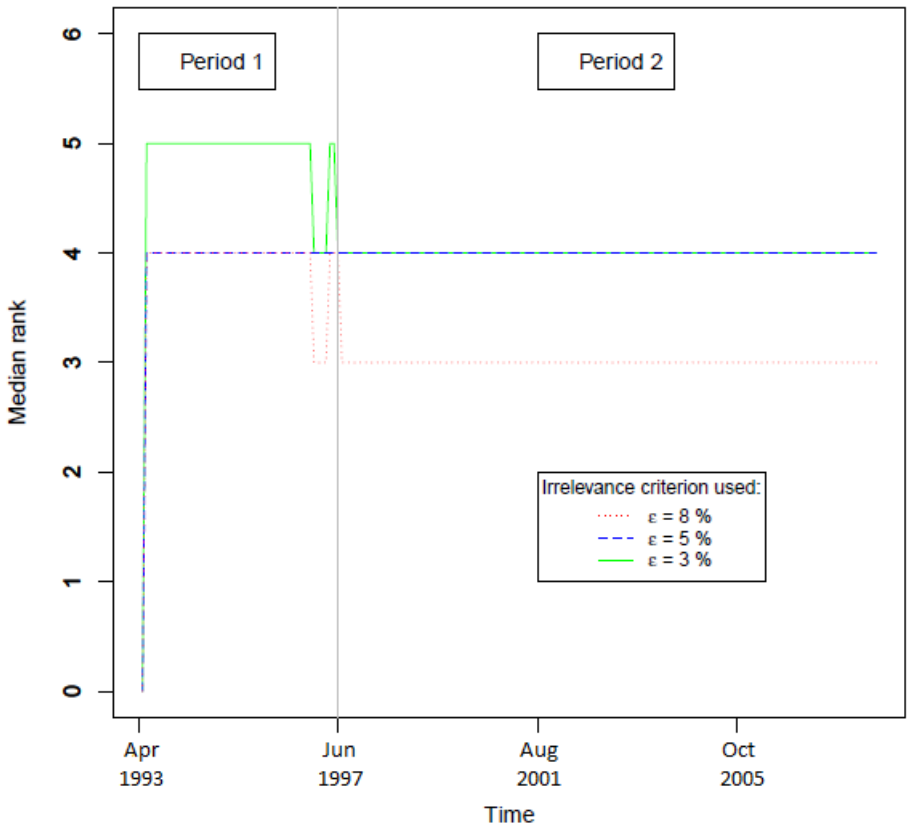

Figure 4.9: Evolution of the median cointegration rank between the net trading of France, Germany and Spain and the long term interest rate over the period before and after the introduction of the Euro (April 1993 - July 2008).

The time-varying independent cointegrating relations of the European panel data after the introduction of the Euro

In this section, we show the smoothing effect of the algorithm for the coefficients of the European panel data set during the second time-period (from June 1997). For this time-period we decided to rely on the irrelevance criterion $\varepsilon = 8\%$ and used a cointegration rank of 3 in order to establish our independent cointegrating relations.

At first, we will consider the three independent cointegrating relations taken in August 2002

(Table 4.7):

Table 4.7: Independent cointegrating relations in August 2002.

SpaNX	FraIR	GerIR	SpaIR	FraNX	GerNX
1	0	0	0.110	-0.678	0.284
0	1	0	-0.807	-0.149	-0.055
0	0	1	-0.562	-0.414	-0.158

We observe a smooth change of the cointegrating coefficients in September 2002 (Table 4.8):

Table 4.8: Independent cointegrating relations in September 2002.

SpaNX	FraIR	GerIR	SpaIR	FraNX	GerNX
1	0	0	0.114	-0.655	0.275
0	1	0	-0.823	-0.149	-0.040
0	0	1	-0.575	-0.414	-0.135

Let us now look at the relations one year later in August 2003 (Table 4.9) and the relations in January 2006 (Table 4.10).

Table 4.9: Independent cointegrating relations in August 2003.

SpaNX	FraIR	GerIR	SpaIR	FraNX	GerNX
1	0	0	0.108	-0.805	0.329
0	1	0	-0.729	-0.134	-0.141
0	0	1	-0.482	-0.402	-0.298

Table 4.10: Independent cointegrating relations in January 2006.

SpaNX	FraIR	GerIR	SpaIR	FraNX	GerNX
1	0	0	0.101	-0.715	0.285
0	1	0	-0.781	-0.114	-0.081
0	0	1	-0.557	-0.364	-0.208

We observe that the cointegrating relations have changed quite a bit but not very abruptly in the time period after the introduction of the single currency. Overall, we retrieve the results of Chapter 3, that is, that the Spanish net trading is coevolving positively with the French net trading, but negatively with the German net trading. The coefficient of the French net trading is negative while the coefficient of the German net trading is positive: August 2002, Table 4.7, (FraNX:-0.678, GerNX: 0.284), September 2002, Table 4.8, (FraNX: -0.655, GerNX: 0.275), August 2003, Table 4.9, (FraNX: -0.805, GerNX: 0.329) and January 2006, Table 4.10, (FraNX: -0.715, GerNX: 0.285).

We notice that the coefficients of the first row in Table 4.10 are closer to the coefficients in 2002: the French net trading has a coefficient close to -0.70 in 2002 while in August 2003 it is approaching -0.80 (see Tables 4.7 and 4.9). However, the second row looks more similar between August 2003 and January 2006 (see Tables 4.9 and 4.10). On the other hand, the values of the coefficients of the third row stay stable over time (August 2002, September 2002, August 2003 and January 2006). We observe overall a smooth evolution in the values of the coefficients over time.

The time-varying independent cointegrating relations of the European panel data before the introduction of the Euro

If we rely on the irrelevance criterion $\varepsilon = 8\%$, we will obtain a cointegration rank of 4 for the time period before June 1997 (see Figure 4.9). Table 4.11 represents the four independent cointegrating relations derived in March 1995. The columns are built in the same order as for

Tables 4.7, 4.8, 4.9 and 4.10

Table 4.11: Independent cointegrating relations in March 1995

SpaNX	FraIR	GerIR	SpaIR	FraNX	GerNX
1	0	0	0	-1.024	0.510
0	1	0	0	1.090	-1.763
0	0	1	0	0.783	-1.738
0	0	0	1	1.746	-2.490

From the first row of the relations in March 1995, we notice that the Spanish net trading is coevolving positively with the French net trading at a higher speed (-1.024) than in the decade following the adoption of the Euro. However, we still notice a negative comovement with the German net trading (0.510) and this negative comovement actually has a faster rate in 1995 than after the adoption of the Euro. This is a little in contradiction with what we concluded in Chapter 3, however we can think of less convergence between the French and the Spanish net trading after the adoption of the Euro: According to the relations seen in this section there is a faster speed of convergence before the introduction of the Euro between the French and the Spanish net tradings.

4.8.2 An application to the stock prices of three company sectors from the Dow Jones Industrial Indices

This section is about a study of the evolution of the cointegration rank (i.e. the number of independent cointegrating relations) for three sectors of companies in the Dow Jones Industrial Average index during the first decade of the century. The Dow Jones Industrial Average is a stock market index created by the Wall Street Journal in 1896, see Stillman (1986). It is now owned by S&P Dow Jones Indices. It is the second oldest U.S. stock market index. The average is named after Dow Jones & Company co-founder Charles Dow and one of his business

statistician associates, Edward Jones. The average index is price-weighted and shows mainly the performance of the 30 largest industries in the United States.

The data consists of three sets of daily industry indices extracted from June 18, 2001 until September 4, 2009 (source: The Wall Street Journal, Baker (1996)). The three sectors of study are IT/Technologies (electronics sector), Manufacturing (manufacturing sector), and Banking/Insurance (banking sector). The first sector consists of the following five companies: Cisco (CSCO), Hewlett-Packard (HP), IBM (IBM), Intel (INTC) and United Technologies (UTX). The stock prices of these companies are presented in Figure 4.10. The second sector is about four manufacturing companies: Alcoa Inc. (AA), Caterpillar (CAT), E.I. du Pont de Nemours and Co. (DD) and 3M company (MMM). The stock prices are presented in Figure 4.11. Finally, the third sector covers four banking/insurance groups' stock prices: American Express (AXP), Bank of America Corporation (BAC), JPMorgan Chase (JPM) and Travelers insurance (TRV). The stock prices are presented in Figure 4.12.

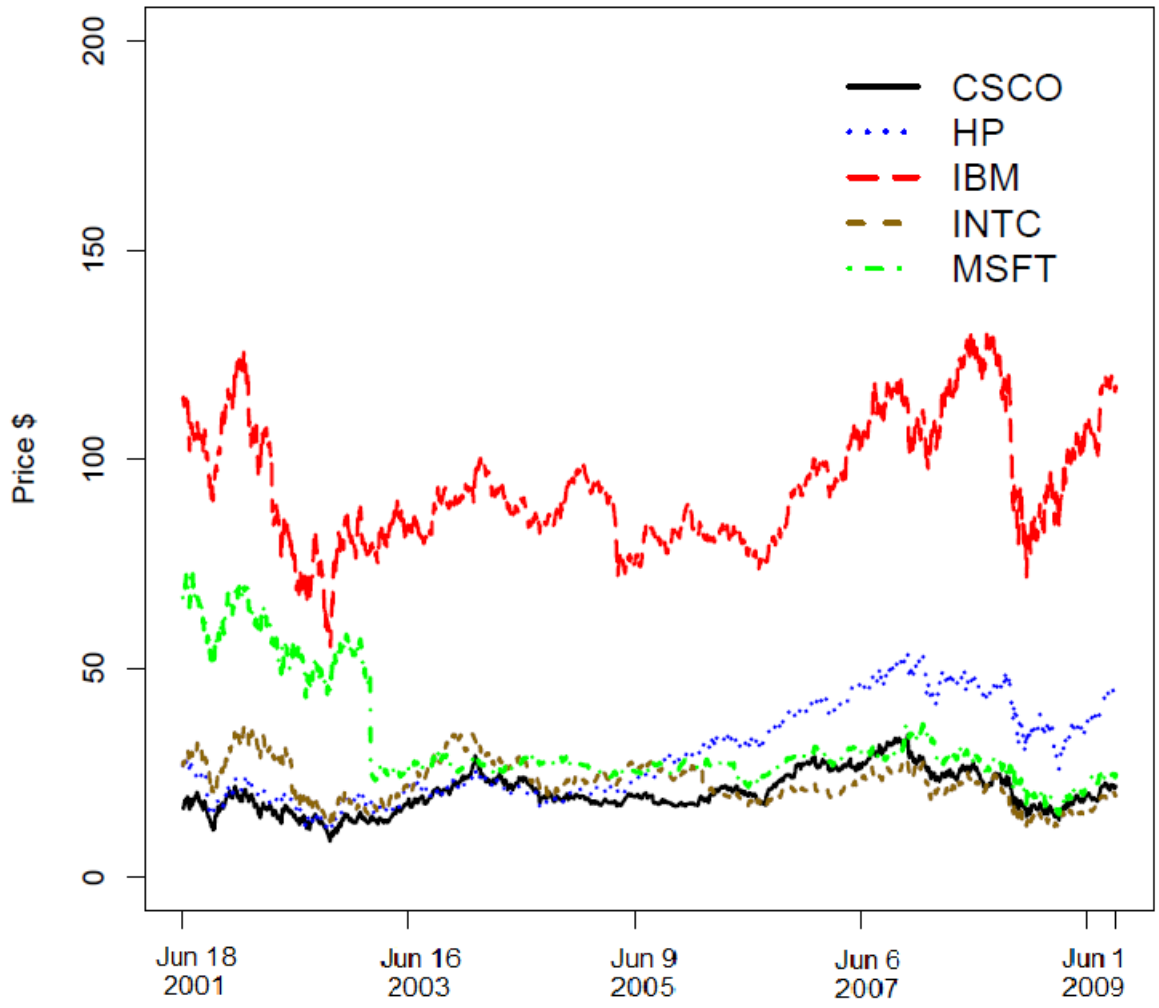

Figure 4.10: The Dow Jones stock prices for the Electronics sector: Cisco Systems (CSCO), Hewlett-Packard (HP), International Business Machines Corporation (IBM), Intel Corporation (INTC), Microsoft Corporation (MSFT). Daily data collected from the Dow Jones from 18 June 2001 to 4 September 2009. (*Source:* The Wall Street Journal, Baker (1996))

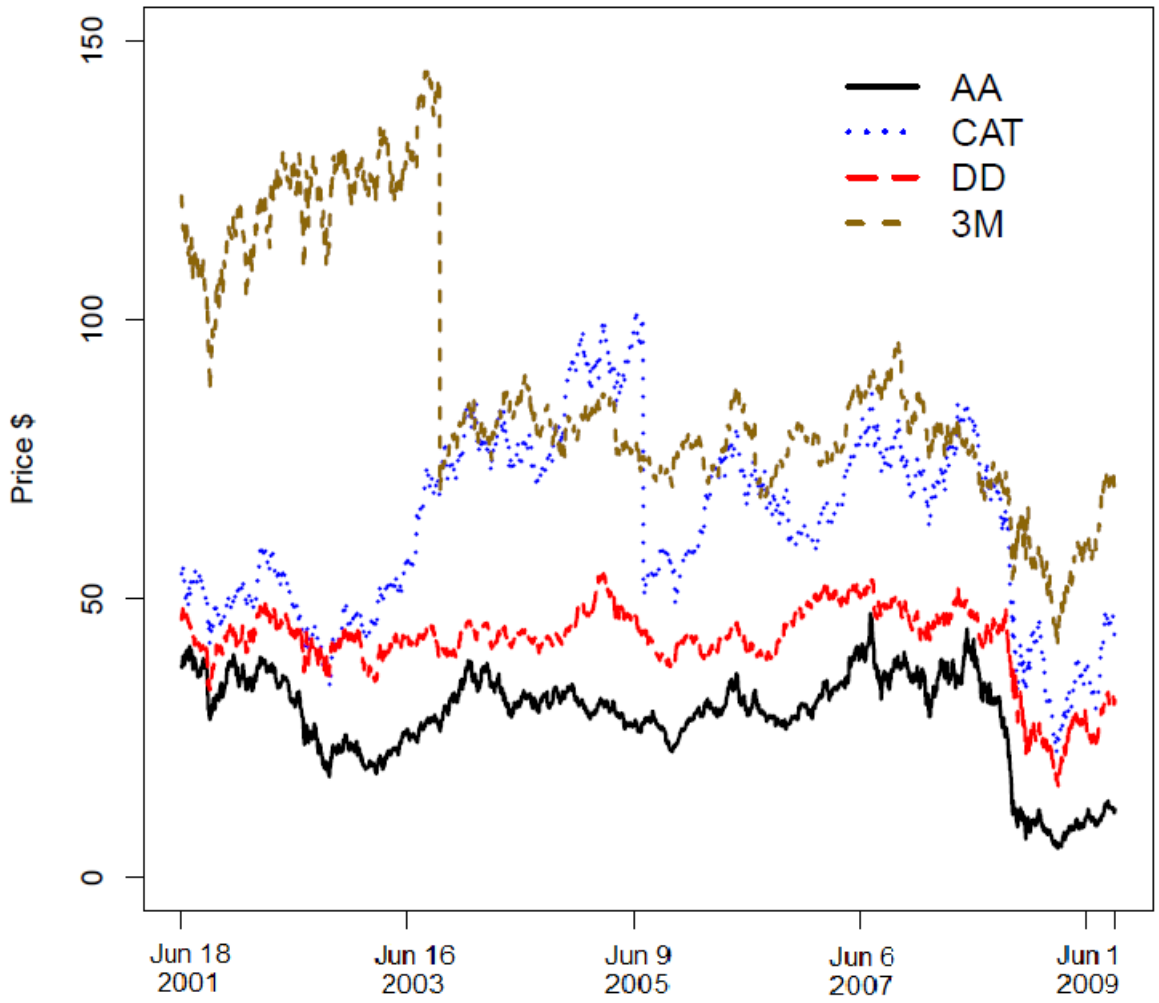

Figure 4.11: The Dow Jones stock prices for the Manufacturing sector: Alcoa Inc. (AA), Caterpillar Inc. (CAT), E.I. DuPont de Nemours & Co. (DD), Minnesota Mining and Manufacturing Company (3M). Daily data collected from the Dow Jones from 18 June 2001 to 4 September 2009. (*Source:* The Wall Street Journal, Baker (1996))

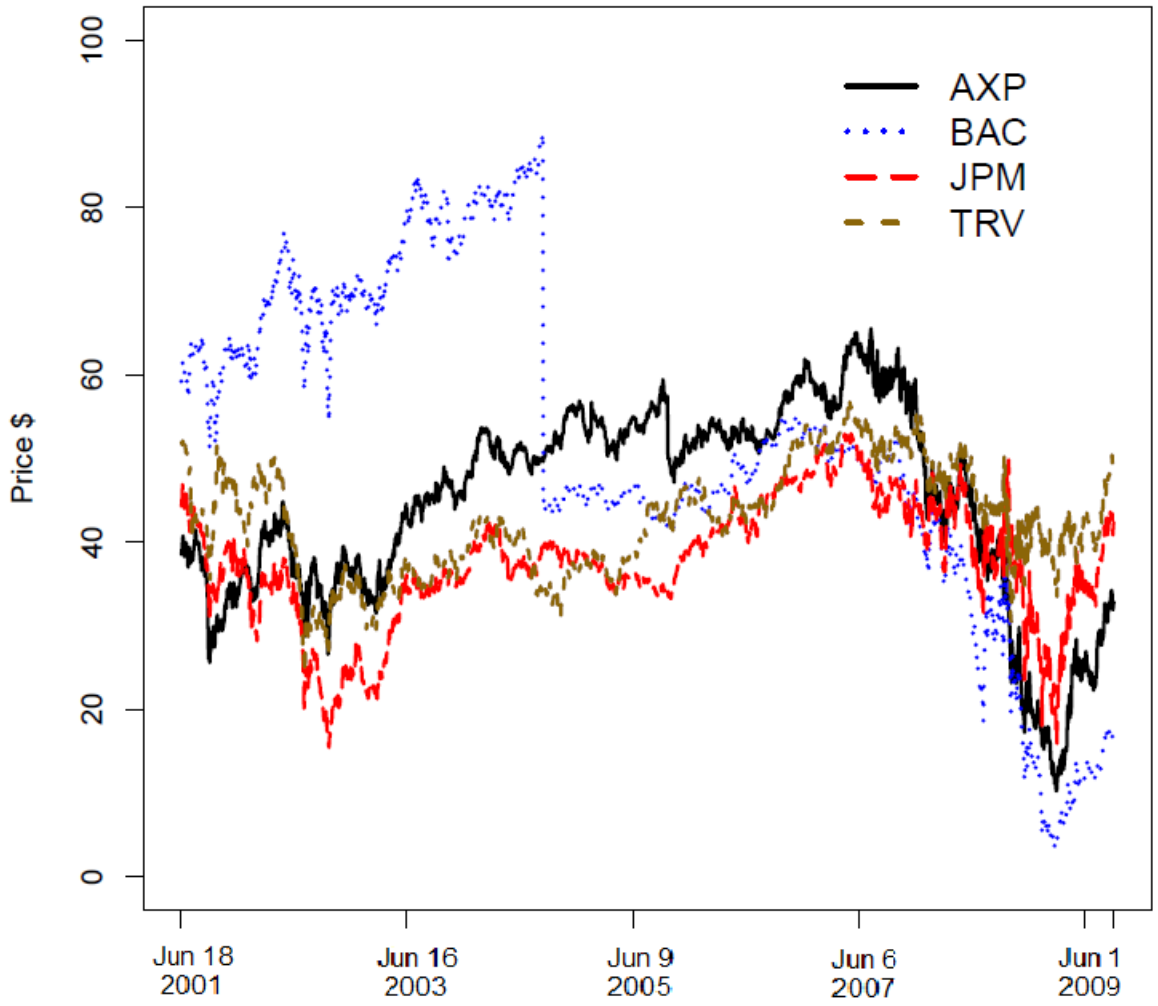

Figure 4.12: The Dow Jones stock prices for the Banking sector: American Express Company (AXP), Bank of America Corporation (BAC), JPMorgan Chase & Co (JPM), Travelers Companies Inc. (TRV). Daily data collected from the Dow Jones from 18 June 2001 to 4 September 2009. (*Source:* The Wall Street Journal, Baker (1996))

For the three sectors, we initialize the parameters of the VECM and the hyperparameters of the measurement and transition equations with the same methods as described in Section 4.7.2

for the simulated data of this chapter. We take a pre-sample of size $\tau = 50$ like for the simulated data sets. We then follow the same steps as seen in Algorithms 4, 5 and 6 for the synthetic data sets (see Section 4.7) and the European panel data set (see Section 4.8.1).

Results

The electronics sector shows a significant change in the value of the rank (see Figure 4.13), and yet, the change in the cointegration rank is only detected with an irrelevance criterion of 8%. Based on the irrelevance criterion of 8%, we have split our time period into 2 different periods for the electonics sector (see Figure 4.13). The first time period is from June 2001 until March 2005 where the cointegration rank found is 3. The second time period is from March 2005 until the last time point in June 2009 where the cointegration rank is 4.

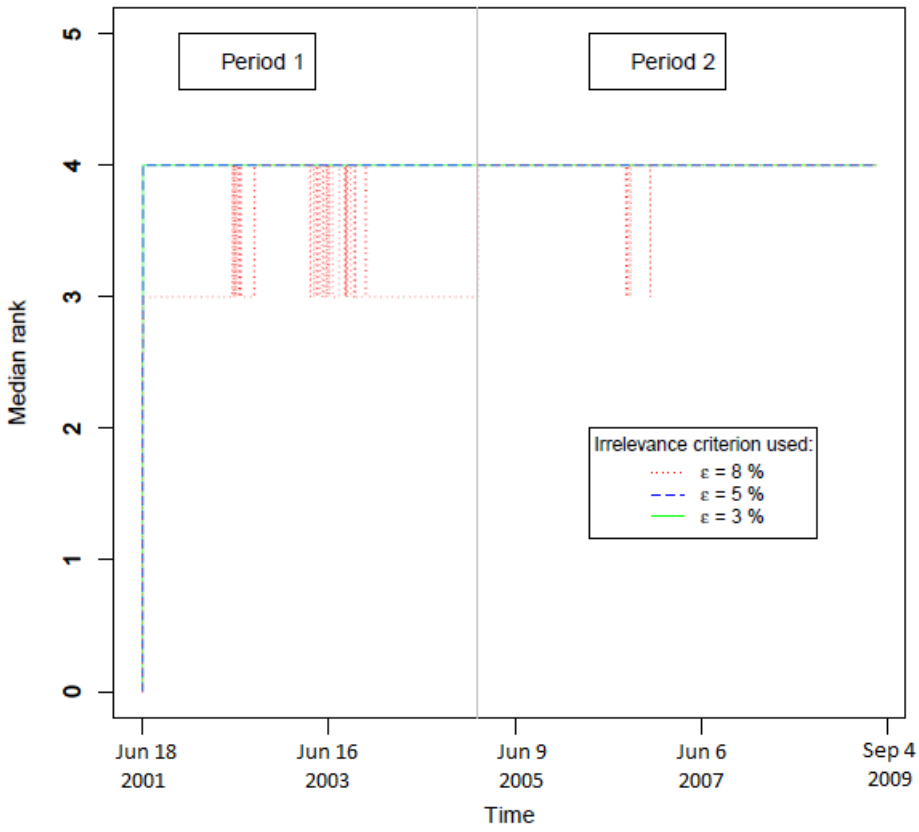

Figure 4.13: Evolution of the median cointegration rank for the Electronics Sector: CSCO-HP-IBM-INTC-UTX. Time period: 18 June 2001 to 4 September 2009.

Figure 4.14 gives the evolution of the cointegration rank for the Manufacturing sector. From that Figure we can say that the cointegration rank is roughly the same over time. Based on the irrelevance criterion $\varepsilon = 8\%$, we detect some jumps from cointegration rank 3 to 2 but it can still be considered as not significant enough to distinguish different time periods.

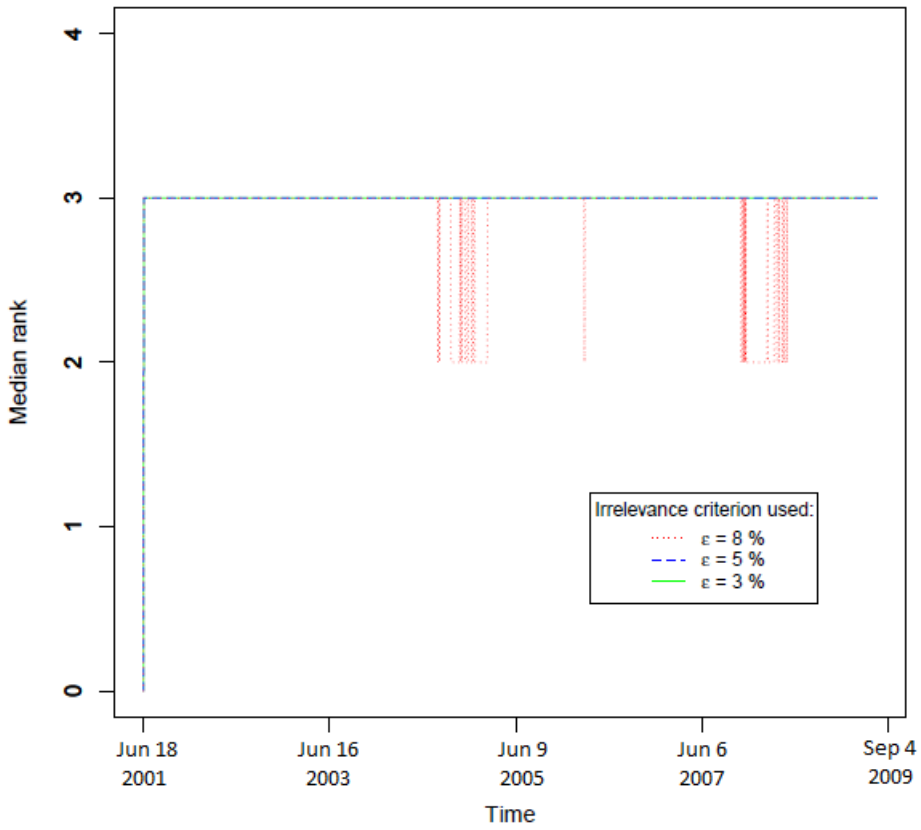

Figure 4.14: Evolution of the median cointegration rank for the Manufacturing Sector: AA-CAT-DD-MMM. Time period: 18 June 2001 to 4 September 2009.

In Figure 4.15, we have the evolution of the cointegration rank for the Banking/Insurance sector. There we can say that based on the irrelevance criterion of 8% and 5%, there is a tendency for the cointegration rank to decrease from 3 to 2 at the end of the decade.

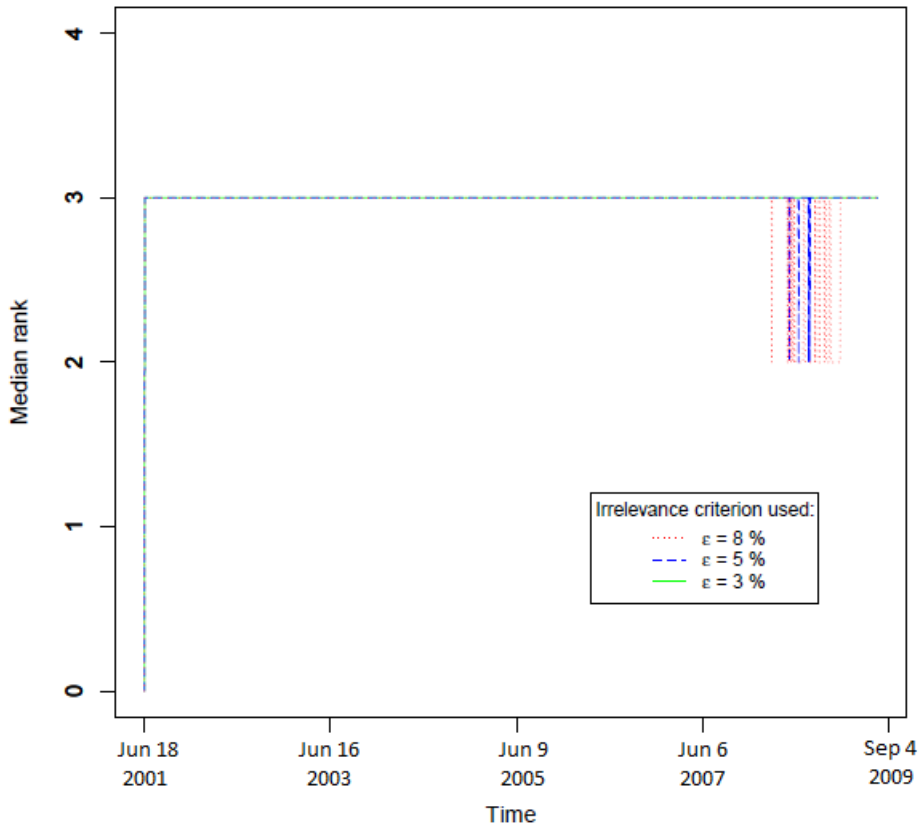

Figure 4.15: Evolution of the median cointegration rank for the Banking Sector: AXP-BAC-JPM-TRV. Time period: 18 June 2001 to 4 September 2009.

The time-varying independent cointegrating relations for the electronics sector

This section presents the independent cointegrating relations of the sector of the Dow Jones where an evolution of the cointegration rank has been observed: the Electronics sector (see Figure 4.13). The cointegration rank for the first time period (before March 2005) will be taken as 3 based on the irrelevance criterion $\varepsilon = 8\%$ whereas for the second time period (after March 2005) we will consider four independent cointegrating relations.

Table 4.12 represents three independent cointegrating relations taken in May 2004 (first time-period, rank 3). Table 4.13 represents four independent cointegrating relations taken in May 2004

(first time-period, rank 4).

Table 4.12: Independent cointegrating relations in May 2004 for the Electronics sector of the Dow Jones.

CSCO	HP	IBM	INTC	MSFT
1	0	0	-0.275	-0.167
0	1	0	-1.208	-0.536
0	0	1	0.043	-0.264

Table 4.13: Independent cointegrating relations in January 2008 for the Electronics sector of the Dow Jones.

CSCO	HP	IBM	INTC	MSFT
1	0	0	0	-0.686
0	1	0	0	-1.876
0	0	1	0	-0.363
0	0	0	1	-0.363

Like for the European panel data (see Section 4.8.1), we decided to see the smoothing effect of the FFBS algorithm by studying the coefficients taken a little later than January 2008. Tables 4.14 and 4.15 are taken respectively in February 2008 and March 2008. We can see on these tables that there is no abrupt change, but rather a smooth evolution between the values of February and March 2008 (Tables 4.14 and 4.15) of the coefficients compared with January 2008 (Table 4.13).

Table 4.14: Independent cointegrating relations in February 2008 for the Electronics sector of the Dow Jones.

CSCO	HP	IBM	INTC	MSFT
1	0	0	0	-0.739
0	1	0	0	-1.822
0	0	1	0	-0.315
0	0	0	1	-0.358

Table 4.15: Independent cointegrating relations in March 2008 for the Electronics sector of the Dow Jones.

CSCO	HP	IBM	INTC	MSFT
1	0	0	0	-0.659
0	1	0	0	-1.651
0	0	1	0	-0.285
0	0	0	1	-0.324

4.9 Discussion

The methods of time-varying cointegration allow us to see a movement of the cointegration rank over time, but can also permit to derive the cointegrating coefficients (also time-varying) over different periods. However, the methods seen in this chapter imply the use of a non-singular distribution over parameters that are in fact singular (e.g. the long-run impact matrix Π is of rank lower than p). These methods used in Chapters 3 and 4 can indeed be criticized for that aspect. The next chapter of this thesis deals with this issue of singularity of the parameters of the VECM and proposes Bayesian inference around singular distributions, but for the static model only (see Chapter 5). A discussion in Chapter 6 is brought in order to give ideas on how

we could include a singular distribution for Π in a dynamic VECM.

When looking at the simulated data (see Section 4.7.1), we have obtained reasonable accuracy in the determination of the cointegration rank, and obtained its evolution over time. The observation of the evolution of the cointegration rank is quite striking. In addition we retrieve the cointegration relations well (see Tables 4.2, 4.4 and 4.6). For that reason, we have trusted our methods and applied it to a part of the European data set seen in Chapter 3 and to some sectors of the Dow Jones.

We found a decrease of the cointegration rank like in Chapter 3 between three economies of the European panel data set (see Section 4.8.1): the variables involved are the net trading and the long-term interest rate between France, Germany and Spain over the time period (1993-2008), including therefore the two periods before and after the introduction of the Euro. That decrease is observed to be occuring two years before the official date of the adoption of the Euro (June 1997) as the evolution of the cointegration rank in Figure 4.9 suggests. The introduction of the Euro in 1999 may not be the main reason why the Eurozone countries are not converging in the post-Euro decade (1999-2008).

Finally, we applied our algorithm to three sectors of the Dow Jones from June 2001 until June 2009. We detect the presence of two time periods between the companies of the Electronics sector in the Dow Jones (see Figure 4.13), where the number of independent cointegrating relations is increasing. We use an irrelevance criterion of 8% to conclude of these two time periods. That increase would occur approximately in the middle of the decade (March 2005), and thus would suggest more convergence between the companies of the electronics sector.

As for the evolution of the independent cointegrating relations over time, we can see the smoothness of our algorithms by looking at the values of the cointegrating coefficients not changing abruptly over time (see Tables 4.7, 4.8, 4.9 and 4.10 for the European panel data set and Tables 4.13, 4.14 and 4.15 for the Electronics sector of the Dow Jones).

Chapter 5

Cointegration analysis based on singular distributions

5.1 Introduction

As we have seen in the previous chapters (see Chapters 3 and 4), Π should be a matrix of reduced rank and the prior of Π should reflect this important fact. Therefore, a more systematic way would be to start to introduce a singular prior for the long-run relations matrix, and include it in the MCMC algorithm. This singular prior, which involves the cointegration rank, entails a singular posterior distribution, which has the same lower rank. We give a detailed explanation in Section 5.6. Thus the cointegration rank cannot change over the iterations of the MCMC procedure. The method in this chapter does not include the rank in the Gibbs sampling, unlike in the two previous chapters. The cointegration rank r is estimated from the data, by using Johansen tests, see Johansen (1991), and the cointegration rank found from these tests is used as a *hyperparameter* in the singular prior distribution of the cointegrating matrix Π. The rank thus enters as a hyperparameter. However, we mention in our discussion in Section 5.10 about the possibility of finding a posterior distribution for the rank in the same way as Villani (2005).

Section 5.2 recapitulates the definition of a matrix-variate normal singular distribution es-

tablished in previous works (Díaz-García and Gutiérrez-Jáimez, 1997; Díaz-García *et al.*, 2006, 1997; Gupta and Nagar, 2000), and the definition of its density and the Hausdorff measure. In that same section, we also show how we can simulate a matrix from a matrix-variate normal singular distribution.

Section 5.4 is about obtaining the full conditional posterior distribution of Ψ: we are using Bayes' theorem with the densities of the prior and the likelihood of the parameters. The prior and the full conditional distribution of Ψ are non-singular. We are using the same prior (Gaussian) for the lag parameters as in Chapter 3. The covariance matrix Σ will follow an Inverse-Wishart prior distribution from which we obtain a full conditional Inverse-Wishart posterior distribution by conjugacy.

In Section 5.6, we retrieve a matrix-variate normal singular full conditional posterior distribution for the cointegrating matrix Π by using the joint distribution of the prior and the data. The prior mean of Π will be taken as equal to 0 (no cointegration). The covariance matrix of this prior is the Kronecker product of a positive semidefinite hyperparameter matrix S of rank r with the covariance matrix Σ of full rank.

We learn that the rank defined for the singular covariance matrix S of the prior distribution of Π is the same as the rank of the singular covariance matrix of the full conditional posterior distribution of Π. Besides, although we set a prior mean equal to 0, we retrieve a mean of rank r in the full conditional distribution of Π. As a consequence, all the matrices Π simulated in the algorithm will be singular of rank r. Therefore, the rank cannot be changed or modified throughout the iterations of the MCMC algorithm. We open a discussion at the end of the chapter about the possibility of deriving a posterior distribution of the rank r given the data.

We also learn that fixing the covariance hyperparameter S makes wrong assumptions in the prior of S. In order to infer S, we will in fact decompose the Moore-Penrose generalized inverse of S, that is S^+, into the product UU' where U has a matrix variate normal prior distribution. As the full conditional distribution of U will be of an unknown form, we will use a Metropolis-Hastings algorithm in order to infer U at each step of the MCMC procedure. A covariance

matrix S is then retrieved from the simulated matrix U. Note that the prior distribution implied for S^+ is a singular Wishart distribution, see Gupta and Nagar (2000), which would therefore imply a pseudo-Wishart distribution for S or generalized singular Inverse-Wishart distribution, see Díaz-García *et al.* (2006).

In Section 5.9, these new methods will be applied to two simulated data sets and to some real data sets: we will compare the cointegrating relations between the net tradings of the European panel data set found in Chapter 3 (see Section 3.4.4) and between the companies of the Electronics sector for the second time-period for the Dow Jones data set seen in Chapter 4 (see Section 4.8.2).

5.2 The matrix-variate normal singular distribution

5.2.1 Definition

We first recall the definition introduced by Gupta and Nagar (2000) of a matrix-variate normal singular distribution:

Definition 7. *Let $X(p\times n)$ be a random matrix with $E(X) = M$ and $cov(Vec(X)) = \Xi\otimes\Lambda$, where either $\Lambda(p \times p)$ or $\Xi(n \times n)$ or both are positive* ***semidefinite*** *with ranks $p_1(\leq p)$ and $n_1(\leq n)$ respectively $^{(\star)}$. Then X is said to have a matrix-variate normal singular distribution, denoted in this thesis as $X \sim NS_{p,n}(M, \Lambda, \Xi|p_1, n_1)$, if there exist matrices $H_1 \in \mathcal{V}_{p_1,p}$, $P_1 \in \mathcal{V}_{n_1,n}$ of ranks p_1 and n_1 respectively such that $X = H_1 Y P_1{}' + M$ for some random matrix $Y \sim N_{p_1,n_1}(0, \Delta, G)$ with $G(n_1 \times n_1) > 0$ and $\Delta(p_1 \times p_1) > 0$. H_1 and Δ are the matrices associated with the spectral decomposition of Λ: $\Lambda = H_1 \Delta H_1{}'$. P_1 and G are the matrices associated with the spectral decomposition of Ξ: $\Xi = P_1 G P_1{}'$.*

$^{(\star)}$ *In order to satisfy the singular property of the distribution of X, it suffices that only one of the covariance matrices is singular (either $p_1 < p$ or $n_1 < n$), or both of them ($p_1 < p$ and $n_1 < n$). If both of the covariance matrices are positive definite ($p_1 = p$ and $n_1 = n$), then the distribution of X becomes a non-singular matrix-variate normal distribution.*

Then we can write:

$$X \sim NS_{p,n}(M, \Lambda, \Xi | p_1, n_1) \Leftrightarrow Vec(X') \sim NS_{pn}(Vec(M'), \Lambda \otimes \Xi | p_1 n_1)$$
$$\Leftrightarrow Vec(X) \sim NS_{pn}(Vec(M), \Xi \otimes \Lambda | p_1 n_1)$$

Alternatively, we can say that $Vec(X)$ follows a vector-variate normal singular distribution with mean $Vec(M)$ and variance $\Xi \otimes \Lambda$ with rank $p_1 n_1$. In order to avoid any confusion with the non-singular normal distribution, we decide to denote the matrix-variate normal singular distribution by NS, as *Normal Singular*.

5.2.2 A probability density function for the matrix-variate normal singular distribution

Díaz-García *et al.* (2006) defined the density function of the random matrix X from Definition 7 as the following:

Definition 8. *Let $X \sim NS_{p,n}(M, \Lambda, \Xi | p_1, n_1)$ as in Definition 7. Then its density function is given by:*

$$\frac{1}{(2\pi)^{p_1 n_1/2}(\Pi_{i=1}^{p_1} \lambda_i^{n_1/2})(\Pi_{j=1}^{n_1} \xi_j^{p_1/2})} \exp\left(-\frac{1}{2} \operatorname{Tr}(\Xi^+ (X - M)' \Lambda^+ (X - M))\right) \tag{5.1}$$

where A^+ is the Moore-Penrose inverse of the positive semidefinite matrix A, and $\{\lambda_i\}_{1 \le i \le p_1}$ and $\{\xi_j\}_{1 \le j \le n_1}$ are the nonzero eigenvalues of Λ and Ξ, respectively.

This density is defined according to the Hausdorff measure (dX), that is recalled in the following proposition of Díaz-García *et al.* (2006):

Proposition 1. *For any $p \times p$ matrix X of lower rank $r < p$, there exists a Singular Value Decomposition (SVD) such that $X = H_1 D P_1'$ in which $H_1 \in \mathcal{V}_{r,p}$, $P_1 \in \mathcal{V}_{r,p}$ and $D = diag(D_{11}, D_{22}, \cdots, D_{rr})$ with $D_{11} > D_{22} > \cdots > D_{rr} > 0$. Such a decomposition is called the non-singular part of the SVD of X. If X is random, we can recall the Hausdorff measure (dX) as (see Hausdorff (1918)*

and Díaz-García et al. *(2006)):*

$$(dX) = 2^{-r}|D|^{2p-2r} \prod_{i<j}^{r} (D_{ii}^2 - D_{jj}^2)\ (dD)\ (H_1'dH1)\ (P_1'dP1) \tag{5.2}$$

where $(dD) = \bigwedge_{i=1}^{r} dD_{ii}$, $(H_1'dH1) = \bigwedge_{i=1}^{r} \bigwedge_{j=i+1}^{p} h_j'dh_i$ *and* $(P_1'dP1) = \bigwedge_{i=1}^{r} \bigwedge_{j=i+1}^{p} p_j'dp_i$.

$\bigwedge$ *denotes the exterior product, or the wedge product, see Muirhead (1982).*

With $r = \min(p_1, n_1)$, the density (5.1) of Definition 8 can be written as:

$$dF_X(X) = \frac{1}{(2\pi)^{p_1 n_1/2}(\Pi_{i=1}^{p_1}\lambda_i^{n_1/2})(\Pi_{j=1}^{n_1}\xi_j^{p_1/2})} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Xi^+(X-M)'\Lambda^+(X-M))\right)(dX)$$

where $H = (H_1|H_2) \in \mathcal{O}_p$ and $P = (P_1|P_2) \in \mathcal{O}_n$ are associated with the spectral decompositions of Λ and Ξ, respectively. (dX) represents the Hausdorff measure recalled in Equation (5.2).

5.2.3 Method to simulate a matrix-variate normal singular distribution

We can simulate a random matrix X following a matrix-variate normal distribution. If we want to simulate such a matrix, we will first need to decompose each variance matrix Λ and Ξ defining the variance of X (see Definition 7). According to Appendix A.4 and the expression of A as in (A.5), we can write any semi-definite positive matrix A as:

$$A = F_r D_r F_r'$$

where F_r are the eigenvectors associated with the non-zero eigenvalues of A, corresponding to the diagonal elements of the diagonal matrix D_r.

We can use the same decomposition for $\Lambda = H_1 \Delta H_1'$ and $\Xi = P_1 G P_1'$. For Λ, it suffices to take Δ as the diagonal matrix of size $p_1 \times p_1$ containing the p_1 non-zero eigenvalues, whereas their corresponding eigenvectors will be stored in the matrix H_1, which will be of size $p \times p_1$. We decompose Ξ by employing the same methods and obtain the diagonal matrix G containing the n_1 non-zero eigenvalues of Ξ and $P_1 \in \mathcal{V}_{n_1,n}$ containing the n_1 corresponding eigenvectors.

After that, it is straightforward to simulate $Y \sim N_{p_1,n_1}(0, \Delta, G)$, which is a non-singular distribution with positive definite matrices $\Delta > 0$ and $G > 0$. We then multiply the simulated

matrix Y on the left and right hand sides by H_1 and ${P_1}'$, respectively, to obtain $H_1 Y {P_1}'$. Finally we add the mean of the distribution of X, that is, M. We therefore obtain a value for X:

$$X = H_1 Y {P_1}' + M$$

For example, suppose we want to simulate $X \sim NS_{3,2}(M, \Xi, \Delta|2, 1)$ where:

$$M = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

and

$$\Xi = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

As we can see, the ranks of Λ and Ξ are respectively 2 and 1, which makes them positive semidefinite matrices of lower rank. Then we have that the spectral decompositions of Λ (5.3) and Ξ (5.4) are:

$$\Lambda = H_1 \Delta {H_1}' \text{ with } H_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \Delta = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \tag{5.3}$$

and

$$\Xi = P_1 G {P_1}' \text{ with } P_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } G = \begin{bmatrix} 2 \end{bmatrix} = 2 \tag{5.4}$$

Now we can simulate $Y \sim N_{2,1}(0, \Delta, G)$ from the non-singular normal distribution. We obtain:

$$Y = \begin{bmatrix} -4.38 \\ 2.73 \end{bmatrix}$$

Finally, we find the value of X by $X = H_1 Y P_1{}' + M$:

$$X = \begin{bmatrix} 1 & 2 \\ 11.76 & 4 \\ -0.46 & 6 \end{bmatrix}$$

5.3 Prior distributions and the likelihood of the model

In this section we describe the joint prior distribution of the parameters used in the model. We recapitulate the VECM model (2.4) by:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Psi_i \Delta x_{t-i} + \epsilon_t \tag{5.5}$$

where $\epsilon_t \sim N(0, \Sigma)$.

5.3.1 Prior on Π given S and Σ

In this chapter, the cointegrating matrix Π has a matrix-variate normal singular prior distribution with mean 0 and covariance $S \otimes v^{-1}\Sigma$ where S is a $p \times p$ positive semidefinite matrix of rank r and v is a fixed scalar hyperparameter. The prior of Π depends on the parameters $S \in \mathcal{S}_p^+(r)$ and $\Sigma > 0$:

$$\Pi | S, \Sigma \sim NS_{p,p}(0, v^{-1}\Sigma, S | p, r) \Leftrightarrow Vec(\Pi) | S, \Sigma \sim NS_{p^2}(0, S \otimes v^{-1}\Sigma | p \times r) \tag{5.6}$$

We will denote the prior density of Π as $f(\Pi | S, \Sigma)$. Following Definition 8 from Díaz-García *et al.* (2006), we have:

$$f(\Pi | S, \Sigma) = \frac{1}{2\pi^{pr}} |\Sigma|^{-\frac{r}{2}} \prod_{j=1}^{r} {\sigma_j}^{-\frac{p}{2}} \exp\left(-\frac{1}{2} \operatorname{Tr}(v S^+ \Pi' \Sigma^{-1} \Pi) \right) \tag{5.7}$$

where S^+ denotes the Moore-Penrose generalized inverse of the positive semidefinite parameter matrix $S \in \mathcal{S}_p^+(r)$ and $\{\sigma_j\}_{1 \le j \le r}$ consist of the eigenvalues of S.

The density (5.7) is defined under the Hausdorff measure $(d\Pi)$:

$$dF_\Pi(\Pi) = \frac{1}{2\pi^{pr}} |\Sigma|^{-\frac{r}{2}} \prod_{j=1}^{r} {\sigma_j}^{-\frac{p}{2}} \exp\left(-\frac{1}{2} \operatorname{Tr}(v S^+ \Pi' \Sigma^{-1} \Pi)) \right) (d\Pi)$$

where $(d\Pi)$ is the Hausdorff measure defined in the same way as in Proposition 1 in Díaz-García *et al.* (2006) and recalled in this chapter by Proposition 1. We can take the non-singular part of the SVD of the matrix Π. $H = (H_1|H_2) \in \mathcal{O}_p$ is associated with the spectral decomposition of $\Sigma = H\Delta H'$ with: $H_1 \in \mathcal{V}_{r,p}$ and $H_2 \in \mathcal{V}_{p-r,p}$. $P = (P_1|P_2) \in \mathcal{O}_p$, with $P_1 \in \mathcal{V}_{r,p}$ and $P_2 \in \mathcal{V}_{p-r,p}$, is associated with the spectral decomposition of $S = PGP'$ where:

$$G = \begin{bmatrix} G_r & 0 \\ 0 & 0 \end{bmatrix}$$

G_r contains the r non-zero eigenvalues of S. The non-singular part of the SVD of Π can be written as:

$$\Pi = H_1 D P_1'$$

where $D = \text{diag}(D_{11}, \cdots, D_{rr})$ and $D_{11} > \cdots > D_{rr} > 0$.

According to Díaz-García *et al.* (2006), since $r = \min(r, p)$, then the Hausdorff measure of Π is defined as:

$$(d\Pi) = 2^{-r}|D|^{2p-2r} \prod_{i<j}^{r} (D_{ii}^2 - D_{jj}^2)\ (dD)\ (H_1'dH1)\ (P_1'dP1) \tag{5.8}$$

with $(dD) = \bigwedge_{i=1}^{r} dD_{ii}$, $(H_1'dH1) = \bigwedge_{i=1}^{r} \bigwedge_{j=i+1}^{p} h_j'dh_i$ and $(P_1'dP1) = \bigwedge_{i=1}^{r} \bigwedge_{j=i+1}^{p} p_j'dp_i$.

5.3.2 Issues in fixing S

In this section, we highlight the fact that S should not be used as a fixed value. Before any analysis of a time series data set, it is difficult to define a fixed positive semidefinite matrix as a hyperparameter.

Fixing S would mean that some coefficients of Π would have different variances than others. This results in an impact in the cointegrating relations because some cointegrating coefficients would have larger variance than others. Therefore, fixing S would imply the need to have some information about the cointegrating relations before running the MCMC procedure. In dealing with the non-singular case in Chapter 3, we had taken S as the identity matrix and obtained the same variance covariance matrix $v^{-1} \times \Sigma$ in the prior of each column of Π, bringing therefore a

uniform distribution of the variance among the coefficients of Π.

By randomizing S, we allow more equality between the cointegrating coefficients and therefore more objectivity in the definition of this hyperparameter.

5.3.3 Inference on S and introduction to a Bayesian hierarchical model

S is a positive semidefinite $p \times p$ matrix of rank $r < p$ and we will actually focus on the Moore-Penrose generalized inverse of S, i.e. S^+. We will decompose it into the following form $S^+ = UU'$ where U is a $p \times r$ random matrix. This latter parameter is in fact a hyperparameter, on which we will define a hyperprior distribution. For this hyperprior, we decide to take a matrix-variate normal distribution of mean 0 and variance given by I_r and a positive definite parameter matrix $B > 0$ of size $p \times p$. Note that the matrix B enters as a hyperparameter of a hyperprior and thus should be classified as a hyperhyperparameter in this Bayesian hierarchical model:

$$S^+ = UU' \text{ with } U \sim N_{p,r}(0, B, I_r) \tag{5.9}$$

We also notice that according to the definition of Gupta and Nagar (2000) and Uhlig (1994), the product $S^+ = UU'$ follows a singular Wishart distribution. This definition introduced by Gupta and Nagar (2000) is recapitulated below:

Definition 9. *If we have r vectors u_j that are i.i.d. random p-vectors having a normal distribution with mean 0 and variance $B > 0$, with $r < p$, and if we stack the vectors u_j into a $p \times r$ matrix $U = [u_1, u_2, \cdots, u_r]$, then the product $G = UU'$ is a $p \times p$ positive semidefinite matrix that has a singular Wishart distribution (denoted as WS in this thesis) of degrees of freedom r:*

$$\forall j \in [\![1, r]\!], \;\; u_j \sim N_p(0, B) \;\; \Rightarrow G = UU' \sim WS_p(r, B) \text{ with } U = [u_1, u_2, ..., u_r] \tag{5.10}$$

As a consequence we can say that the covariance hyperparameter S is following a Pseudo-Inverse Wishart prior distribution, that is the corresponding singular Inverse-Wishart of the singular Wishart distribution, see Díaz-García *et al.* (2006).

Let us now recapitulate the Bayesian hierarchical model below:

- Stage 1: The errors of the VECM are Gaussian and are distributed with mean 0 and parameter covariance matrix Σ.

- Stage 2: The parameters Π, Ψ and Σ have a prior distribution and we derive their respective full conditional distributions from the likelihood of the VECM (constructed from the errors of the VECM).

- Stage 3: The hyperparameter U, from which S^+ and S is constructed, has a hyperprior and we derive a full conditional distribution depending on Π and Σ thanks to the prior distribution of Π given Σ and U. The likelihood of U is actually proportional to the prior distribution of Π given Σ. The hyperparameter U depends on the data through the parameters Π and Σ.

However, the full conditional distribution of U will be of unknown form and we will have to use the Metropolis-Hastings algorithm (see Metropolis *et al.* (1953) and Hastings (1970)) in order to simulate U at each step of the final MCMC procedure (see Section 5.7). While the other parameters Π, Ψ and Σ will be simulated by Gibbs steps, the hyperparameter U will be simulated by a Metropolis-Hastings step, in this context, as part of a hybrid or Metropolis-within-Gibbs algorithm. This introduction of a Metropolis-Hastings step in Gibbs sampling has been used several times in the literature, see e.g. Gilks *et al.* (1995), Martino *et al.* (2015).

5.3.4 Prior on Ψ given Σ

The prior for the lag parameter matrix will be the same prior used as in Chapter 3, that is, a Gaussian distribution depending on Σ. The prior of $\Psi|\Sigma$ is then given by:

$$Vec(\Psi)|\Sigma \sim N_{pd}(0, I_d \otimes \Sigma) \tag{5.11}$$

We will denote the prior of $\Psi|\Sigma$ as $f(\Psi|\Sigma)$ and we can write the relation:

$$f(\Psi|\Sigma) \propto |\Sigma|^{-\frac{d}{2}} \exp\left(-\frac{1}{2} Vec(\Psi)'(I_d \otimes \Sigma)^{-1} Vec(\Psi)\right) \propto |\Sigma|^{-\frac{d}{2}} \exp\left(-\frac{1}{2} \operatorname{Tr}(\Sigma^{-1}\Psi\Psi')\right) \tag{5.12}$$

5.3.5 Prior on Σ

The parameter Σ, corresponding to the variance covariance matrix of the errors ϵ_t of (5.5), remains non-singular since it is a positive definite matrix. The prior on Σ will be an Inverse-Wishart distribution with hyperparameters $A > 0$ and q estimated in the same way as in Chapter 3:

$$\Sigma \sim IW_p(A, q) \tag{5.13}$$

The prior of Σ does not depend on the other parameters of the VECM (5.5) and will have the following relation:

$$f(\Sigma) \propto |\Sigma|^{-\frac{q+p+1}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}A)\right) \tag{5.14}$$

5.3.6 The joint prior distribution and the likelihood

The joint prior distribution of the VECM (5.5) is based on the priors defined previously in Sections 5.3.1, 5.3.3, 5.3.4 and 5.3.5:

$$f(\Pi, \Psi, \Sigma, U) \propto f(\Pi|\Sigma, U)f(U)f(\Psi|\Sigma)f(\Sigma) \tag{5.15}$$

The likelihood is the same as the one used in Chapter 3 (see Section 3.3.1). We define the same matrix of the errors $E' = Y' - \Pi X' - \Psi Z'$ as in Equation (2.13) (see Section 2.3.2). We thus obtain the following distribution of the vectorized form of E' (5.16) defining the likelihood of our model (5.17):

$$Vec(E') \sim N(0, I_T \otimes \Sigma) \tag{5.16}$$

$$L(\mathcal{D}; \Pi, \Psi, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}E'E)\right) \tag{5.17}$$

We notice that neither E nor Σ contain the covariance hyperparameter S, which means that they do not contain the hyperparameter U: U will depend on its prior $f(U)$ and the prior of $\Pi|U, \Sigma$ only.

5.4 Full conditional posterior distribution of the non-singular lag parameters: Ψ

Based on the non-singular prior distribution of Ψ given Σ, we are going to find the full conditional posterior distribution of Ψ, i.e. $\Psi|\Sigma, \Pi, \mathcal{D}$ established in Result 2 below:

Result 2. *Full conditional posterior distribution of* Ψ*.*

$$Vec(\Psi)|\Sigma, \Pi, \mathcal{D} \sim N_{pd}(\mu_\Psi, \Omega_{\Psi 1} \otimes \Omega_{\Psi 2}) \tag{5.18}$$

with:

$$\mu_\Psi = (Y' - \Pi X')Z(I_d + Z'Z)^{-1} \tag{5.19}$$

$$\Omega_{\Psi 1} = (I_d + Z'Z)^{-1} \tag{5.20}$$

$$\Omega_{\Psi 2} = \Sigma \tag{5.21}$$

Proof. First of all the prior of $\Psi|\Sigma$ is given by $Vec(\Psi)|\Sigma \sim N_{pd}(0, I_d \otimes \Sigma)$, according to (5.11). We apply Bayes' theorem with the joint prior (5.15) and the likelihood (5.17):

$$\begin{aligned} f(\Psi|\Sigma, \Pi, U, \mathcal{D}) &\propto f(\Pi, \Psi, \Sigma, U)L(\mathcal{D}; \Pi, \Psi, \Sigma) \\ &\propto f(\Pi|\Sigma, U)f(U)f(\Psi|\Sigma)f(\Sigma)L(\mathcal{D}; \Pi, \Psi, \Sigma) \end{aligned}$$

We notice that Ψ depends on neither U (or S), nor the priors of $\Pi|U, \Sigma$ (5.7) and Σ (5.14):

$$f(\Psi|\Sigma, \Pi, \mathcal{D}) \propto f(\Psi|\Sigma)L(\mathcal{D}; \Pi, \Psi, \Sigma) \tag{5.22}$$

In the exponential term of the likelihood (5.17), we define $W = Y - X\Pi'$ and $G = Z \otimes \Sigma^{-1/2}$

and we have:

$$\begin{aligned}
\mathrm{Tr}(\Sigma^{-1}E'E) &= \mathrm{Tr}((\Sigma^{-1/2}W' - \Sigma^{-1/2}\Psi Z')(W\Sigma^{-1/2} - Z\Psi'\Sigma^{-1/2})) \\
&= \mathrm{Tr}[(W\Sigma^{-1/2} - Z\Psi'\Sigma^{-1/2})(\Sigma^{-1/2}W' - \Sigma^{-1/2}\Psi Z') \\
&= \mathrm{Tr}[(\Sigma^{-1/2}W' - \Sigma^{-1/2}\Psi Z')(\Sigma^{-1/2}W' - \Sigma^{-1/2}\Psi Z')] \\
&= (Vec[\Sigma^{-1/2}W'] - Vec[\Sigma^{-1/2}\Psi Z'])'(Vec[\Sigma^{-1/2}W'] - Vec[\Sigma^{-1/2}\Psi Z']) \\
&= (Vec[\Sigma^{-1/2}W'] - GVec[\Psi])'(Vec[\Sigma^{-1/2}W'] - GVec[\Psi]) \\
&= (Vec[\Sigma^{-1/2}W']' - Vec[\Psi]'G')(Vec[\Sigma^{-1/2}W'] - GVec[\Psi]) \\
&= Vec[\Sigma^{-1/2}W']'Vec[\Sigma^{-1/2}W'] - Vec[\Psi]'G'Vec[\Sigma^{-1/2}W'] - Vec[\Sigma^{-1/2}W']'GVec[\Psi] \\
&+ Vec[\Psi]'G'GVec[\Psi]
\end{aligned}$$

Then we have:

$$\begin{aligned}
\mathrm{Tr}(\Sigma^{-1}E'E) = Vec[\Psi]'G'GVec[\Psi] &- Vec[\Psi]'G'Vec[\Sigma^{-1/2}W'] \\
&- Vec[\Sigma^{-1/2}W']'GVec[\Psi] + (\textit{a term not depending on } \Psi)
\end{aligned}$$

By multiplying the likelihood (5.17) by the prior (5.12) of $\Psi|\Sigma$, we can obtain the following relation:

$$f(\Psi|\Sigma, \Pi, \mathcal{D}) \propto \exp\left(-\frac{1}{2}g_\Psi\right)$$

where:

$$\begin{aligned}
g_\Psi &= Vec(\Psi)'(I_d \otimes \Sigma^{-1})Vec(\Psi) + Vec[\Psi]'G'GVec[\Psi] \\
&- Vec[\Psi]'G'Vec[\Sigma^{-1/2}W'] - Vec[\Sigma^{-1/2}W']'GVec[\Psi] \\
&+ (\textit{a term not depending on } \Psi) \\
&= Vec(\Psi)'(I_d \otimes \Sigma^{-1} + G'G)Vec(\Psi) - Vec(\Psi)'G'Vec(\Sigma^{-1/2}W') \\
&- Vec(\Sigma^{-1/2}W')'GVec(\Psi) + (\textit{a term not depending on } \Psi)
\end{aligned}$$

Besides, we have:

$$\begin{aligned} I_d \otimes \Sigma^{-1} + G'G &= I_d \otimes \Sigma^{-1} + (Z \otimes \Sigma^{-1/2})'(Z \otimes \Sigma^{-1/2}) \\ &= I_d \otimes \Sigma^{-1} + (Z' \otimes \Sigma^{-1/2})(Z \otimes \Sigma^{-1/2}) \\ &= I_d \otimes \Sigma^{-1} + (Z'Z \otimes \Sigma^{-1}) \\ &= (I_d + Z'Z) \otimes \Sigma^{-1} \end{aligned}$$

We will then obtain:

$$\begin{aligned} g_\Psi = &Vec(\Psi)'((I_d + Z'Z) \otimes \Sigma^{-1})Vec(\Psi) - Vec(\Psi)'G'Vec(\Sigma^{-1/2}W') \\ &- Vec(\Sigma^{-1/2}W')'GVec(\Psi) + (\textit{a term not depending on } \Psi) \end{aligned} \tag{5.23}$$

Since the prior of Ψ is a matrix-variate normal non-singular distribution and that the errors are also Gaussian, then the full conditional posterior distribution of Ψ will also be a matrix-variate normal non-singular distribution by the property of conjugacy. Therefore, we only have to identify the mean μ_Ψ (5.19) and the covariance matrices $\Omega_{\Psi 1}$ (5.20) and $\Omega_{\Psi 2}$ (5.21) of the full conditional distribution of Ψ (5.18) of Result 2. We have:

$$f(\Psi|\Sigma, \Pi, \mathcal{D}) \propto \exp\left(-\frac{1}{2}(Vec(\Psi) - Vec(\mu_\Psi))'\Omega_\Psi^{-1}(Vec(\Psi) - Vec(\mu_\Psi))\right) \tag{5.24}$$

By expanding the expression in the general formula of the full conditional posterior distribution of Ψ (5.24) and by defining $\Omega_\Psi = \Omega_{\Psi 1} \otimes \Omega_{\Psi 2}$, we have:

$$\begin{aligned} &(Vec(\Psi) - Vec(\mu_\Psi))'\Omega_\Psi^{-1}(Vec(\Psi) - Vec(\mu_\Psi)) \\ &= Vec(\Psi)'\Omega_\Psi^{-1}Vec(\Psi) - Vec(\Psi)'\Omega_\Psi^{-1}Vec(\mu_\Psi) - Vec(\mu_\Psi)'\Omega_\Psi^{-1}Vec(\Psi) \\ &+ (\textit{a term not depending on } \Psi) \end{aligned} \tag{5.25}$$

The aim now is to identify $\Omega_{\Psi 1}$, $\Omega_{\Psi 2}$ and μ_Ψ from (5.23). We can already obtain the precision matrix $\Omega_\Psi^{-1} = \Omega_{\Psi 1}^{-1} \otimes \Omega_{\Psi 2}^{-1}$ by looking at the quadratic term in $Vec(\Psi)$ in (5.23):

$$\Omega_\Psi^{-1} = (I_d + Z'Z) \otimes \Sigma^{-1}$$

We then have:

$$(I_d + Z'Z)^{-1} \otimes \Sigma = \Omega_{\Psi_1} \otimes \Omega_{\Psi_2}$$

and retrieve $\Omega_{\Psi_1} = (I_d + Z'Z)^{-1}$ and $\Omega_{\Psi_2} = \Sigma$.

The mean μ_Ψ is also identified from (5.23) and from (5.25) by:

$$Vec(\Psi)'\Omega_\Psi^{-1} Vec(\mu_\Psi) = Vec(\Psi)'G'Vec(\Sigma^{-1/2}W')$$

Then we have $\Omega_\Psi^{-1} Vec(\mu_\Psi) = G'Vec(\Sigma^{-1/2}W')$, from which we obtain (since Ω_Ψ^{-1} is invertible):

$$\begin{aligned} Vec(\mu_\Psi) &= \Omega_\Psi G'Vec(\Sigma^{-1/2}W') \\ &= (I_d + Z'Z)^{-1} \otimes \Sigma(Z \otimes \Sigma^{-1/2})'Vec(\Sigma^{-1/2}W') \\ &= (I_d + Z'Z)^{-1} \otimes \Sigma(Z' \otimes \Sigma^{-1/2})Vec(\Sigma^{-1/2}W') \end{aligned}$$

Then, by using the property $(B' \otimes A)Vec(X) = Vec(AXB)$ with the sizes of A, B and X correctly compatible, we have:

$$(Z' \otimes \Sigma^{-1/2})Vec(\Sigma^{-1/2}W') = Vec(\Sigma^{-1/2}\Sigma^{-1/2}W'Z) = Vec(\Sigma^{-1}W'Z)$$

Finally:

$$((I_d + Z'Z)^{-1} \otimes \Sigma)Vec(\Sigma^{-1}W'Z) = Vec(\Sigma\Sigma^{-1}W'Z(I_d + Z'Z)^{-1}) = Vec(W'Z(I_d + Z'Z)^{-1})$$

which means that:

$$\mu_\Psi = W'Z(I_d + Z'Z)^{-1} = (Y' - \Pi X')Z(I_d + Z'Z)^{-1}$$

□

5.5 Bayesian inference on the non-singular variance matrix of the errors Σ

The parameter Σ, corresponding to the variance covariance matrix of the errors ϵ_t, is kept as non-singular and invertible. The prior of Σ is still chosen as an Inverse-Wishart distribution

with hyperparameters $A > 0$ and q (5.13) like in Chapter 3. By using Bayes' theorem, we have:

$$f(\Sigma|\Pi, \Psi, U, \mathcal{D}) \propto L(\mathcal{D}; \Pi, \Psi, \Sigma) f(\Pi|U, \Sigma) f(\Psi|\Sigma) f(\Sigma)$$

It is straightforward to obtain the full conditional posterior distribution of Σ since the Inverse-Wishart distribution of the variance matrix respects the property of conjugacy when the likelihood of the model is Gaussian. We then have:

$$\begin{aligned}
&f(\Sigma|\Pi, \Psi, U, \mathcal{D}) \propto L(\mathcal{D}; \Pi, \Psi, \Sigma) f(\Pi|U, \Sigma) f(\Psi|\Sigma) f(\Sigma) \\
&\propto |\Sigma|^{-\frac{T}{2}} |\Sigma|^{-\frac{r}{2}} |\Sigma|^{-\frac{d}{2}} |\Sigma|^{-\frac{q+p+1}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}E'E)\right) \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}v\Pi S^{+}\Pi')\right) \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}\Psi\Psi')\right) \\
&\exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}A)\right) \\
&\propto |\Sigma|^{-\frac{T+r+d+q+p+1}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}(E'E + v\Pi S^{+}\Pi' + \Psi\Psi' + A)\right)
\end{aligned}$$

We can then recognize the scale parameter matrix of the full conditional posterior Inverse-Wishart distribution of Σ, denoted as A_{post} here:

$$A_{post} = E'E + v\Pi S^{+}\Pi' + \Psi\Psi' + A > 0$$

A_{post} is the sum of positive definite and positive semidefinite matrices (the hyperparameter scale A is also positive definite). As a consequence, A_{post} is a positive definite matrix. The degrees of freedom of that posterior, denoted as q_{post}, will be identified from the following relation:

$$q_{post} + p + 1 = T + r + d + q + p + 1$$

where we can immediately derive q_{post}:

$$q_{post} = T + r + d + q$$

The full conditional distribution of Σ is then given by:

$$\Sigma|\Pi, \Psi, U, \mathcal{D} \sim IW_p(E'E + v\Pi S^{+}\Pi' + \Psi\Psi' + A, T + r + d + q) \tag{5.26}$$

5.6 Bayesian inference on the singular parameter of the VECM: Π

5.6.1 Full conditional posterior distribution of Π

In this section we will use the joint distribution of the matrix-variate singular normal prior of Π (5.6) and the likelihood of the model (5.17) to derive the full conditional posterior distribution of Π.

Let β and y be two random Gaussian vectors of size $m > 1$ and $n > 1$, respectively. We denote P and R as the variance of β and y, respectively, and we assume that these two variables depend on each other with the $m \times n$ covariance matrix C:

$$C = Cov[\beta, y]$$

Then, the joint distribution of β and y is also Gaussian and it is defined as the following:

$$\begin{bmatrix} \beta \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} \hat{\beta} \\ \hat{y} \end{bmatrix}, \begin{bmatrix} P & C \\ C' & R \end{bmatrix} \right) \tag{5.27}$$

In addition, we will also note that if at least one of the two parameters β or y has a singular distribution, then their joint distribution will also be a singular normal distribution. For that, let us define two vectors $\beta \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$. We can now define the mapping $\mathcal{F}$ from $\mathbb{R}^m \times \mathbb{R}^n$ to $\mathbb{R}^{mn}$:

$$\begin{aligned} \mathcal{F}: \ \mathbb{R}^m \times \mathbb{R}^n &\longrightarrow \mathbb{R}^{mn} \\ (\beta, y) &\longmapsto \begin{bmatrix} \beta \\ y \end{bmatrix} \end{aligned}$$

Then let $\lambda \in \mathbb{R}$, $(u_1, u_2) \in \mathbb{R}^m \times \mathbb{R}^n$ and $(v_1, v_2) \in \mathbb{R}^m \times \mathbb{R}^n$. We can verify the property of linearity:

$$\mathcal{F}(\lambda(u_1, u_2) + (v_1, v_2)) = \mathcal{F}(\lambda u_1 + v_1, \lambda u_2 + v_2)$$

$$= \begin{bmatrix} \lambda u_1 + v_1 \\ \lambda u_2 + v_2 \end{bmatrix} = \lambda \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$
$$= \lambda \mathcal{F}((u_1, u_2)) + \mathcal{F}((v_1, v_2))$$

Therefore $\mathcal{F}$ is linear. Now, by definition, any linear transformation of Gaussian random variables is also a Gaussian random variable. Therefore the joint distribution of β and y (5.27) is Gaussian. Moreover, if β has a strictly singular Gaussian distribution, that is, its covariance matrix P is strictly rank deficient (with rank strictly lower than m), then the covariance matrix of the joint distribution (5.27) is also singular, and therefore the joint distribution will be a singular normal distribution as well.

The posterior distribution of $\beta|y$ can be retrieved by using the joint distribution of the prior β and the data y (5.27):

$$\beta|y \sim N(\beta^\star, P^\star)$$

with:

$$\beta^\star = \hat{\beta} + CR^{-1}(y - \hat{y}) \tag{5.28}$$
$$P^\star = P - CR^{-1}C' \tag{5.29}$$

By using the same approach we will find the full conditional posterior distribution of Π by using the joint distribution of the prior of Π (5.7) and the likelihood of the model (5.17). We will also show that this posterior distribution will have a mean of rank r and that we retrieve a singular covariance matrix with rank $r \times p$ like in the prior distribution of Π (5.6). Result 3 below gives the full conditional posterior distribution which preserves the conjugacy and the rank of the covariance matrix:

Result 3. *Full conditional posterior distribution of* Π.

$$Vec(\Pi)|\Sigma, \Psi, \mathcal{D} \sim NS_{p^2}(Vec(\mu_\Pi), \Omega_{\Pi 1} \otimes \Omega_{\Pi 2}|r \times p) \tag{5.30}$$

with:

$$\mu_{\Pi} = Q'XS^{1/2}(vI_p + S^{1/2}X'XS^{1/2})^{-1}S^{1/2} = Q'X\Omega_{\Pi 1} \tag{5.31}$$

$$Q = Y - Z\Psi'$$

$$\Omega_{\Pi 1} = S^{1/2}(vI_p + S^{1/2}X'XS^{1/2})^{-1}S^{1/2} \tag{5.32}$$

$$\Omega_{\Pi 2} = \Sigma \tag{5.33}$$

$$Rank(\Omega_{\Pi 1}) = Rank(\mu_{\Pi}) = r \tag{5.34}$$

Proof. Firstly, we use β as $Vec(\Pi)|\Psi, S, \Sigma$ and y as $Vec(Y')|\Psi, \Sigma$ from the general example (5.27). We need to clarify the prior variance P of $Vec(\Pi)|\Psi, S, \Sigma$, the variance R of $Vec(Y')|\Psi, \Sigma$ and the covariance matrix C between $Vec(\Pi)|\Psi, S, \Sigma$ and $Vec(Y')|\Psi, \Sigma$.

We have:

$$P = Var[Vec(\Pi)|\Psi, S, \Sigma] = S \otimes v^{-1}\Sigma$$

We then have for the variance R of y:

$$\begin{aligned} R &= Var(Vec(Y')|\Psi, \Sigma) \\ &= Var(Vec(\Pi X' + \Psi Z' + E')|\Psi, \Sigma) \\ &= Var((X \otimes I_p)Vec(\Pi) + (Z \otimes I_p)Vec(\Psi) + Vec(E')|\Psi, \Sigma) \end{aligned}$$

Now we notice that the term $(Z \otimes I_p)Vec(\Psi)$ is actually a constant for $Vec(Y')|\Psi, \Sigma$ and so we have:

$$\begin{aligned} &Var((X \otimes I_p)Vec(\Pi) + (Z \otimes I_p)Vec(\Psi) + Vec(E')|\Psi, \Sigma) \\ &= Var((X \otimes I_p)Vec(\Pi) + Vec(E')|\Psi, \Sigma) \\ &= (X \otimes I_p)Var(Vec(\Pi))(X' \otimes I_p) + Var(Vec(E')) \\ &= (X \otimes I_p)(S \otimes v^{-1}\Sigma)(X' \otimes I_p) + (I_T \otimes \Sigma) \\ &= (XSX' \otimes v^{-1}\Sigma) + (I_T \otimes \Sigma) \\ &= (I_T + v^{-1}XSX') \otimes \Sigma \end{aligned}$$

As for the covariance between $Vec(\Pi)|\Psi, S, \Sigma$ and $Vec(Y')|\Psi, \Sigma$, we have:

$$\begin{aligned}
C &= Cov[Vec(\Pi), Vec(Y')|\Psi, \Sigma] \\
&= Cov[Vec(\Pi), (X \otimes I_p)Vec(\Pi)|\Psi, \Sigma] \\
&= Var[Vec(\Pi)|\Psi, \Sigma](X \otimes I_p)' \\
&= (S \otimes v^{-1}\Sigma)(X' \otimes I_p) \\
&= (v^{-1}SX' \otimes \Sigma)
\end{aligned}$$

We then retrieve the variance part of $\Pi|\Psi, S, \Sigma$, which is called $\Omega_\Pi = \Omega_{\Pi 1} \otimes \Omega_{\Pi 2}$. Recall that if $r < p$, then the variance is strictly singular of lower rank $p \times r < p^2$. We notice that since S is positive semidefinite, then S is symmetric and $S' = S$. According to Equation (5.29), the variance $P^\star = \Omega_\Pi$, is given by:

$$\begin{aligned}
\Omega_\Pi &= P - CR^{-1}C' \\
&= (v^{-1}S \otimes \Sigma) - (v^{-1}SX' \otimes \Sigma)((I_T + v^{-1}XSX') \otimes \Sigma)^{-1}(v^{-1}SX' \otimes \Sigma)' \\
&= (v^{-1}S \otimes \Sigma) - (v^{-1}SX' \otimes \Sigma)((I_T + v^{-1}XSX')^{-1} \otimes \Sigma^{-1})(v^{-1}XS \otimes \Sigma) \\
&= (v^{-1}S \otimes \Sigma) - (v^{-1}SX'(I_T + v^{-1}XSX')^{-1}v^{-1}XS \otimes \Sigma\Sigma^{-1}\Sigma) \\
&= (v^{-1}S \otimes \Sigma) - ((v^{-2}SX'(I_T + v^{-1}XSX')^{-1}XS) \otimes \Sigma)
\end{aligned}$$

Now we have:

$$\begin{aligned}
&v^{-2}SX'(I_T + v^{-1}XSX')^{-1}XS \\
&= v^{-1}S^{1/2}v^{-1/2}S^{1/2}X'(I_T + v^{-1}XSX')^{-1}v^{-1/2}XS^{1/2}S^{1/2}
\end{aligned} \tag{5.35}$$

We recall the following result for any $T \times p$ matrix H:

$$H'(I_T + HH')^{-1}H = (I_p + H'H)^{-1}H'H \tag{5.36}$$

Indeed we have:

$$\begin{aligned} H'(I_T + HH')^{-1} = (I_p + H'H)^{-1}M &\Leftrightarrow M = (I_p + H'H)H'(I_T + HH')^{-1} \\ &\Leftrightarrow M = (H' + (H'H)H')(I_T + HH')^{-1} \\ &\Leftrightarrow M = (H' + H'(HH'))(I_T + HH')^{-1} \\ &\Leftrightarrow M = H'(I_T + HH')(I_T + HH')^{-1} \\ &\Leftrightarrow M = H' \end{aligned} \tag{5.37}$$

And therefore:

$$H'(I_T + HH')^{-1} = (I_p + H'H)^{-1}H'$$

from which:

$$H'(I_T + HH')^{-1}H = (I_p + H'H)^{-1}H'H$$

We can take $H = v^{-1/2}XS^{1/2}$ in Equation (5.35) in order to obtain:

$$\begin{aligned} &v^{-1}S^{1/2}v^{-1/2}S^{1/2}X'(I_T + v^{-1}XSX')^{-1}v^{-1/2}XS^{1/2}S^{1/2} \\ &= v^{-1}S^{1/2}(I_p + v^{-1}S^{1/2}X'XS^{1/2})^{-1}v^{-1/2}S^{1/2}X'XS^{1/2}v^{-1/2}S^{1/2} \end{aligned} \tag{5.38}$$

By applying the result (5.36) with $H = v^{-1/2}XS^{1/2}$ to (5.38), we can obtain the singular posterior variance of $Vec(\Pi)|\Psi, S, \Sigma$:

$$\begin{aligned} \Omega_\Pi &= (v^{-1}S \otimes \Sigma) - ((v^{-2}SX'(I_T + v^{-1}XSX')^{-1}XS) \otimes \Sigma) \\ &= (v^{-1}S \otimes \Sigma) - (v^{-1}S^{1/2}(I_p + v^{-1}S^{1/2}X'XS^{1/2})^{-1}v^{-1/2}S^{1/2}X'XS^{1/2}v^{-1/2}S^{1/2} \otimes \Sigma) \\ &= (v^{-1}S - v^{-2}S^{1/2}(I_p + v^{-1}S^{1/2}X'XS^{1/2})^{-1}S^{1/2}X'XS) \otimes \Sigma \\ &= \Omega_{\Pi 1} \otimes \Omega_{\Pi 2} \end{aligned}$$

with $\Omega_{\Pi 2}$ as required in (5.33).

Now we simplify the form of $\Omega_{\Pi 1}$:

$$\begin{aligned}
\Omega_{\Pi 1} &= (v^{-1}S - v^{-2}S^{1/2}(I_p + v^{-1}S^{1/2}X'XS^{1/2})^{-1}S^{1/2}X'XS) \\
&= v^{-1}(S - S^{1/2}(vI_p + S^{1/2}X'XS^{1/2})^{-1}S^{1/2}X'XS) \\
&= v^{-1}(S^{1/2}S^{1/2} - S^{1/2}(vI_p + S^{1/2}X'XS^{1/2})^{-1}S^{1/2}X'XS^{1/2}S^{1/2}) \\
&= v^{-1}S^{1/2}(I_p - (vI_p + S^{1/2}X'XS^{1/2})^{-1}S^{1/2}X'XS^{1/2})S^{1/2} \\
&= v^{-1}S^{1/2}(vI_p + S^{1/2}X'XS^{1/2})^{-1}(vI_p + S^{1/2}X'XS^{1/2} - S^{1/2}X'XS^{1/2})S^{1/2} \\
&= v^{-1}S^{1/2}(vI_p + S^{1/2}X'XS^{1/2})^{-1}vI_pS^{1/2} \\
&= S^{1/2}(vI_p + S^{1/2}X'XS^{1/2})^{-1}S^{1/2}
\end{aligned}$$

as required in (5.32).

We need to prove now that the rank of $\Omega_{\Pi 1}$ is equal to r (5.34). We can recall that if A is a $p \times p$ positive semi-definite matrix of lower rank $r < p$ and B is a $p \times p$ positive definite matrix of full rank p then the rank of ABA is r. Indeed, we have $ABA = AB^{1/2}B^{1/2}A = (AB^{1/2})(AB^{1/2})'$. Then, the rank of ABA will be the same as $AB^{1/2}$, that is, the same as A since B is invertible. Now, if we replace A by the positive semidefinite matrix $S^{1/2}$ of rank r and B by the positive definite matrix $(vI_p + S^{1/2}X'XS^{1/2})^{-1}$, then we have that $\Omega_{\Pi 1}$ is of rank r, as required.

The full conditional posterior mean of Π (5.31), that we call μ_Π, is retrieved thanks to Equation (5.28). We have:

$$\begin{aligned}
Vec(\mu_\Pi) &= \hat{\beta} + CR^{-1}(y - \hat{y}) \\
&= 0 + (v^{-1}SX' \otimes \Sigma)((I_T + v^{-1}XSX') \otimes \Sigma)^{-1}(y - \hat{y}) \\
&= (v^{-1}SX' \otimes \Sigma)((I_T + v^{-1}XSX')^{-1} \otimes \Sigma^{-1})(y - \hat{y}) \\
&= (v^{-1}SX'(I_T + v^{-1}XSX')^{-1} \otimes I_p)(y - \hat{y})
\end{aligned}$$

Note that:

$$v^{-1}SX'(I_T + v^{-1}XSX')^{-1}$$

$$= (v^{-1/2}S^{1/2})(v^{-1/2}S^{1/2}X')(v^{-1/2}XS^{1/2}S^{1/2}X'v^{-1/2} + I_T)^{-1}$$
$$= (v^{-1/2}S^{1/2})(v^{-1/2}S^{1/2}X'XS^{1/2}v^{-1/2} + I_p)^{-1}v^{-1/2}S^{1/2}X'$$
$$= v^{-1}S^{1/2}(v^{-1}S^{1/2}X'XS^{1/2} + I_p)^{-1}S^{1/2}X'$$

Since $y = Vec(Y')|\Psi, \Sigma$, we have:

$$\begin{aligned}\hat{y} &= E[Vec(Y')|\Psi, \Sigma] \\ &= E[Vec(\Pi X' + \Psi Z' + E')|\Psi, \Sigma] \\ &= E[Vec(\Pi X') + Vec(\Psi Z') + Vec(E')|\Psi, \Sigma] \\ &= Vec(\Psi Z')\end{aligned}$$

Thus, given Σ and Ψ, we obtain:

$$y - \hat{y} = Vec(Y') - Vec(\Psi Z') = Vec(Y' - \Psi Z') = Vec(Q')$$

Finally, we get:

$$\begin{aligned}Vec(\mu_\Pi) &= (v^{-1}SX'(I_T + v^{-1}XSX')^{-1} \otimes I_p)Vec(Q') \\ &= Vec(I_pQ'(v^{-1}SX'(I_T + v^{-1}XSX')^{-1})') \\ &= Vec(Q'(v^{-1}S^{1/2}(v^{-1}S^{1/2}X'XS^{1/2} + I_p)^{-1}S^{1/2}X')') \\ &= Vec(v^{-1}Q'XS^{1/2}(v^{-1}S^{1/2}X'XS^{1/2} + I_p)^{-1}S^{1/2}) \\ &= Vec(Q'XS^{1/2}(vI_p + S^{1/2}X'XS^{1/2})^{-1}S^{1/2}) \\ &= Vec(Q'X\Omega_{\Pi 1})\end{aligned}$$

Hence, we retrieve Equation (5.31):

$$\mu_\Pi = Q'XS^{1/2}(vI_p + S^{1/2}X'XS^{1/2})^{-1}S^{1/2}$$

□

To complete the proof we need to show that the rank of the full conditional posterior mean of Π, i.e. the rank of μ_Π, is equal to the rank r. We have indeed:

$$\mu_\Pi = Q'XS^{1/2}(vI_p + S^{1/2}X'XS^{1/2})S^{1/2} = Q'X\Omega_{\Pi 1}$$

Then, since Q and X are of size $T \times p$, with T very large compared to p, then the product of the two matrices $Q'X$ is of full rank p. And since $\Omega_{\Pi 1}$ has rank r, then the product $Q'X\Omega_{\Pi 1}$, i.e. μ_Π, also has rank r, as in Equation (5.34).

The rank of the realisations of Π is the rank of the posterior mean

The long-run impact matrix Π is a random variable that can be simulated according to Result 3 and Definition 7 by:

$$\Pi \leftarrow \mu_\Pi + HYP_1{}' \tag{5.39}$$

where H and P_1 are retrieved by the spectral decomposition of the variance matrices $\Omega_{\Pi 1}$ and $\Omega_{\Pi 2}$, respectively:

$$\Omega_{\Pi 1} = P_1 G_r P_1{}' \tag{5.40}$$

$$\Omega_{\Pi 2} = (\Sigma =) H \Delta H' \tag{5.41}$$

In our case, Σ is of full rank, so H corresponds to the p eigenvectors associated with the p non-zero eigenvalues of Σ contained in the diagonal matrix Δ. $P_1 \in \mathcal{V}_{r,p}$ corresponds to the r eigenvectors associated with the r non-zero eigenvalues of S contained in the diagonal matrix G_r. Besides, Y is a $p \times r$ random variable following a matrix non-singular normal distribution with variance given by Δ and G_r.

We notice that $HYP_1{}'$ is the multiplication of two matrices HY and $P_1{}'$. H is a $p \times p$ invertible matrix and Y is of full rank, since it is simulated by a $p \times r$ non-singular distribution. Consequently the product of the two matrices H and Y is also a full rank matrix. As for the columns of P_1, they represent the r eigenvectors of $\Omega_{\Pi 1}$, so P_1 is a $p \times r$ full rank matrix as well. Finally, by multiplying the two full rank matrices HY with $P_1{}'$, the rank of $HYP_1{}'$ is r.

Let us denote $\mathcal{M}^r_{p,p}(\mathbb{R})$ to be the set of $p \times p$ square matrices of rank $r < p$ and let $Z = HYP_1{}'$. Then Z is a random variable taking values in $\mathcal{M}^r_{p,p}(\mathbb{R})$. We also have $\mu_\Pi \in \mathcal{M}^r_{p,p}(\mathbb{R})$ since we know according to Result 3 that $\mu_\Pi = Q'X\Omega_{\Pi 1}$ and that $Q'X$ is an invertible matrix, implying that μ_Π has the same rank as $\Omega_{\Pi 1}$, that is r (5.34).

We will show that:

$$\mu_\Pi + Z \in \mathcal{M}^r_{p,p}(\mathbb{R}) \text{ almost everywhere} \tag{5.42}$$

Suppose we denote E_{μ_Π} as the following set:

$$E_{\mu_\Pi} = \left\{ Z \in \mathcal{M}^r_{p,p}(\mathbb{R}) : \ \mu_\Pi + Z \notin \mathcal{M}^r_{p,p}(\mathbb{R}) \right\} \tag{5.43}$$

We can show that the set E_{μ_Π} is negligible compared with $\mathcal{M}^r_{p,p}(\mathbb{R})$, i.e. it is a finite set. We have indeed $Z + \mu_\Pi \notin \mathcal{M}^r_{p,p}(\mathbb{R})$ if and only if there exist a linear transformation T of μ_Π such that $T(\mu_\Pi) = Z + \mu_\Pi \notin \mathcal{M}^r_{p,p}(\mathbb{R})$. However, there is a finite number of linear transformations T such that $T(\mu_\Pi) \notin \mathcal{M}^r_{p,p}(\mathbb{R})$. Therefore there is a finite number of realisations of the random variable Z such that $Z \in E_{\mu_\Pi}$. Therefore E_{μ_Π} is negligible compared with $\mathcal{M}^r_{p,p}(\mathbb{R})$, i.e. E_{μ_Π} is a finite set, and then, if we denote P as the probability measure associated with $(\mathcal{M}^r_{p,p}(\mathbb{R}), \mathcal{F}(\mathcal{M}^r_{p,p}(\mathbb{R})))$, we have:

$$P[Z \in E_{\mu_\Pi}] = 0 \tag{5.44}$$

Therefore the simulated matrix Π is taking values in $\mathcal{M}^r_{p,p}(\mathbb{R}))$ with probability 1 almost surely. Hence the result (5.42):

$$\mu_\Pi + Z = \Pi \in \mathcal{M}^r_{p,p}(\mathbb{R}) \text{ almost everywhere} \tag{5.45}$$

We conclude that any simulation of Π will have rank r.

5.6.2 Fixed cointegration rank

In this paragraph, we highlight the fact that throughout this thesis we do not set a prior and perform Bayesian inference on the cointegration rank. Although defining a prior distribution on the cointegrating matrix Π induces a prior distribution for the rank, we prefer to infer the cointegrating matrix Π conditional on the rank. Therefore we do not put any prior on the cointegration rank. We can either use Johansen tests on the data in order to have a pre-determined cointegration rank or use the methods seen in Chapter 3 where we extract the rank from a Gibbs sampler based on a non-singular inference approach on the long-run impact matrix Π.

5.7 Metropolis-Hastings to estimate the conditional distribution of U

In this section, we recapitulate the Metropolis-Hastings step in order to be able to simulate the hyperparameter U (5.9). In many works involving Bayesian statistics, the Metropolis-Hastings algorithm is used in order to obtain samples from a distribution of unknown form, see Metropolis *et al.* (1953) and Hastings (1970). The Metropolis-Hastings algorithm is a Markov Chain Monte Carlo method and can be included as part of a hybrid or Metropolis-within-Gibbs algorithm.

Let us now describe in terms of density the prior of U defined in (5.9) (see Section 5.3.3). The density of this prior is given by:

$$f(U) = 2\pi^{-\frac{pr}{2}}|B|^{-\frac{r}{2}}\exp\left(-\frac{1}{2}\operatorname{Tr}(U'B^{-1}U)\right)$$

which is proportional to:

$$f(U) \propto \exp\left(-\frac{1}{2}\operatorname{Tr}(U'B^{-1}U)\right) \tag{5.46}$$

In terms of U, the density (5.7) of the distribution of $\Pi|U,\Sigma$ can be written as:

$$\begin{aligned} f(\Pi|U,\Sigma) &\propto \prod_{j=1}^{r}{\sigma_j}^{-\frac{p}{2}}\exp\left(-\frac{1}{2}\operatorname{Tr}(vS^{+}\Pi'\Sigma^{-1}\Pi)\right) \\ &\propto \prod_{j=1}^{r}{\sigma_j}^{-\frac{p}{2}}\exp\left(-\frac{1}{2}\operatorname{Tr}(vUU'\Pi'\Sigma^{-1}\Pi)\right) \\ &\propto \prod_{j=1}^{r}{\sigma_j}^{-\frac{p}{2}}\exp\left(-\frac{1}{2}\operatorname{Tr}(U'(v\ \Pi'\Sigma^{-1}\Pi)U)\right) \end{aligned} \tag{5.47}$$

where $\prod_{j=1}^{r}{\sigma_j}^{-\frac{p}{2}}$ depends on U because the σ_js represent the non-zero eigenvalues of S, which is a function of U: $S = (UU')^{+}$. The density (5.47) is proportional to the likelihood of U used to derive the full conditional of U. This distribution is conditional on Π and Σ. Indeed, by using

Bayes' theorem and multiplying (5.46) and (5.47) we find:

$$\begin{aligned} f(U|\Pi,\Sigma) &\propto f(\Pi|U,\Sigma)f(U) \\ &\propto \prod_{j=1}^{r} {\sigma_j}^{-\frac{p}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(U'(v\Pi'\Sigma^{-1}\Pi)U)\right) \exp\left(-\frac{1}{2}\operatorname{Tr}(U'B^{-1}U)\right) \\ &\propto \prod_{j=1}^{r} {\sigma_j}^{-\frac{p}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(U'(B^{-1}+v\Pi'\Sigma^{-1}\Pi)U)\right) \end{aligned} \tag{5.48}$$

The presence of the eigenvalues of S in the expression (5.48) above brings some difficulties in finding the full conditional distribution of U as those eigenvalues are directly linked to the values of U: $S^+ = UU'$. The full conditional distribution of U is therefore of an unknown form and we need to use the Metropolis-Hastings algorithm in order to infer U. We will therefore define a function ϕ proportional to this unknown density, simply equal to the expression obtained in (5.48):

$$\phi(U) = \prod_{j=1}^{r} {\sigma_j}^{-\frac{p}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(U'(B^{-1}+v\Pi'\Sigma^{-1}\Pi)U)\right) \tag{5.49}$$

For that, at step i, we propose a symmetric distribution centered on the previous value $U^{(i-1)}$. The symmetric distribution is a matrix-variate normal distribution with mean $U^{(i-1)}$ and variance I_r and C where C is a positive definite matrix of size $p \times p$. We will denote this proposal distribution as $g(U|V)$ where V is acting as the value of U obtained from the previous iteration, that is, if we are at step i of the MCMC algorithm, V will take the value $U^{(i-1)}$:

$$U|V \sim N_{p,r}(V, C, I_r) \tag{5.50}$$

The initial parameter $U^{(0)}$ is simulated from a matrix-variate normal distribution with mean 0 and variance $I_r \otimes C$. The Metropolis-Hastings algorithm is constructed as follows:

Algorithm 7 Metropolis-Hastings algorithm to simulate hyperparameter U at step i of the MCMC procedure.

At step i of the MCMC procedure:

Simulate U_{prop} from $g(U|U^{(i-1)})$.

Construct the acceptance ratio α as: $\alpha = \frac{\phi(U_{prop})}{\phi(U^{(i-1)})}$

IF $(\alpha \geq 1)$ THEN: $U \leftarrow U_{prop}$.

ELSE IF$(\alpha < 1)$ THEN:

Simulate h from a Uniform distribution over $[0, 1]$:

IF$(h < \alpha)$ THEN: $U \leftarrow U_{prop}$.

ELSE: $U \leftarrow U^{(i-1)}$.

After a burn-in period, the simulated $U^{(i)}$ will converge to the full conditional distribution of U, see Metropolis *et al.* (1953). This period is the same burn-in period of 20,000 iterations used for the other parameters simulated by Gibbs steps. Like in Chapter 3, we decide to take the last 10,000 iterations in order to evaluate the distribution of the parameters Π, Ψ and Σ. We will use the same period to evaluate the distribution of U.

5.8 Gibbs sampler

5.8.1 Setting of the hyperparameters

We use the same hyperparameters as in Chapter 3 so $A = (q - p + 1) \times \Sigma^{(0)}$ where $q = p + 4$ and $\Sigma^{(0)}$ is equal to $\widehat{\Sigma}$ estimated from a pre-sample as in Section 2.5. The hyperparameter v will be taken equal to 0.001 like in Chapter 3, in order to give a weakly informative prior on $\Pi|S, \Sigma$. The cointegration rank r is pre-determined either by Johansen tests or by the methods seen in Chapter 3 when comparing the same data sets. r is therefore taken as a fixed hyperparameter in our model.

The parameter C introduced in the proposal distribution $g(U|V)$ of the Metropolis-Hastings step for simulating U (5.50) has to be chosen so that the acceptance rate of the Metropolis step satisfies certain conditions (see Section 5.9.3). We discuss in Section 5.9.3 the fact that the matrix C should not have too small values because if that was the case, the proposal distribution of U would simulate almost the same value at each iteration. This would make the covariance matrix S have the same value at each iteration approximately, leading to faulty cointegrating relations.

In Section 5.3.3, we have seen that the distribution of U involves a parameter matrix B that we set as fixed. The conditional distribution of U (5.48) induces in the exponential part the addition of the matrices B^{-1} and $v\Pi'\Sigma^{-1}\Pi$ (see Section 5.7). A sensitivity analysis is needed here in order to see the effects of choosing B. In the methodology proposed in this chapter, we will first do a sensitivity analysis for the variance B of the prior distribution of U (see Section 5.9.2). Once B is specified, we will adjust the acceptance rate of the Metropolis algorithm by changing the values of C, i.e. the variance of the proposal distribution (see Section 5.9.3).

5.8.2 Algorithm of the Gibbs sampler

We recapitulate here the Gibbs sampler used in this chapter, i.e. a static VECM, but using a singular distribution over the long-run impact matrix. First of all, we can introduce a Directed Acyclic Graph (DAG) in order to summarize the techniques covered in this chapter (see Figure 5.8.2). The algorithms for the Gibbs sampler are presented below the DAG (see Algorithms 8, 9 and 10).

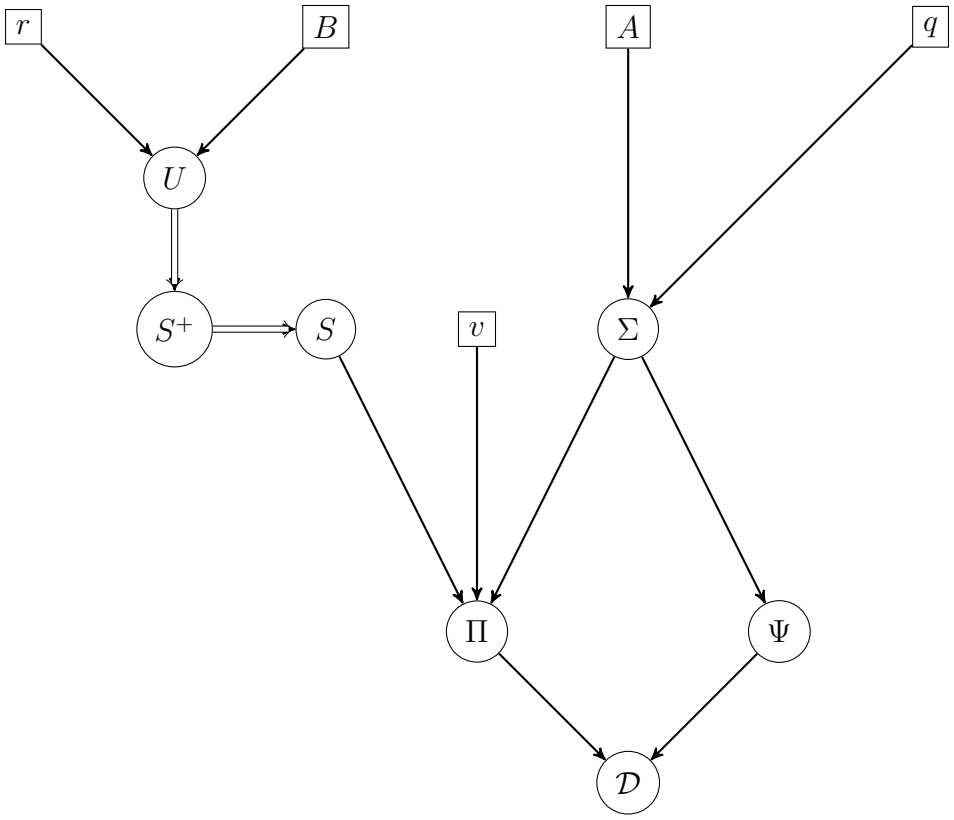

Figure 5.1: Directed Acyclic Graph of the Bayesian model studied in this chapter: Square boxes contain the fixed parameters, circles contain the randomized parameters. The cointegration rank r is a fixed parameter in this model. Double arrows mean that the first parameter is used to compute the following parameter. Simple arrows mean that the first parameter contributes to the distribution of the following parameter.

Algorithm 8 Initialization based on the pre-sample.

Set the size of the pre-sample $\tau = T/3$.

- Create matrices ΔX, X_{-1}, ΔZ following the instructions in Section 2.5.

For the pre-sample data set containing observations $[\![1, \tau]\!]$:

- Initialize $\Pi^{(0)} \leftarrow \widehat{\Pi}$, $\Psi^{(0)} \leftarrow \widehat{\Psi}$, $\Sigma^{(0)} \leftarrow \widehat{\Sigma}$ from the LS estimates (2.30) seen in Section 2.5.

Set the values of the hyperparameters:

- $v = 0.001$, $q = p + 4$, $A = (q - p - 1) \times \Sigma^{(0)}$ and B is specified in Section 5.9.2.
- C is specified according to the acceptance rate (see Section 5.9.3).
- r is estimated from the Johansen frequentist trace tests and enters as a fixed hyperparameter in the Gibbs sampler.
- Initialize $U^{(0)} \sim N(0, I_r \otimes C)$.

Algorithm 9 Gibbs sampler on the sample.

For the sample data set containing observations $[\![\tau+1, T]\!]$

- Create matrices Y, X, Z and E following the instructions in Section 2.3.2 for $t \in [\![\tau+1, T]\!]$.

We have: $\mathcal{D} = \{X, Y, Z\}$.

Gibbs Sampler:

Set the number of iterations $m = 50{,}000$.

for $i \in [\![1, m = 50,000]\!]$ **do**

- $U^{(i)}$ is constructed from the Metropolis-Hastings step recapitulated in Algorithm 7.
- $S^{+(i)} = U^{(i)}U^{(i)\prime}$ (see Section 5.3.3).
- $S^{(i)}$ is retrieved by taking the Moore-Penrose inverse of $S^{+(i)}$.
- Sample $\Pi^{(i)}$ from the singular full conditional posterior distribution of Result 3 with value $\Psi^{(i-1)}$ instead of Ψ and $\Sigma^{(i-1)}$ instead of Σ.
- Sample $\Psi^{(i)}$ from the full conditional posterior distribution (5.18) with value $\Pi^{(i)}$ instead of Π and $\Sigma^{(i-1)}$ instead of Σ.
- Sample $\Sigma^{(i)}$ from the full conditional Inverse-Wishart posterior distribution (5.26) with value $\Pi^{(i)}$ instead of Π and $\Psi^{(i)}$ instead of Ψ.

end for

Algorithm 10 Final results: Obtaining the independent cointegrating relations.

- $\Pi_{mean} \leftarrow \underset{\{m-10{,}000 \leq i \leq m\}}{\text{mean}} [\Pi^{(i)}]$
- The independent cointegrating relations (i.e. β) are then obtained from Π_{mean} and r, by using the operation of Section 3.3.7.

$$\alpha = \begin{bmatrix} \Pi_{mean\ r,r} \\ \Pi_{mean\ p-r,r} \end{bmatrix} \quad \Longrightarrow \quad \beta' = (\alpha'\alpha)^{-1}\alpha'\Pi_{mean} = \begin{bmatrix} I_r & B' \end{bmatrix} \tag{5.51}$$

5.9 Applications

5.9.1 Application to the synthetic data set of Chapter 3

We take the same two data sets as in Chapter 3 (see Section 3.4.1). The first data set (P_1) consists of the same synthetic data set from Chapter 3 consisting of seven time series $x_{1t}, x_{2t}, ..., x_{7t}$ with $T = 350$ data points each and with four independent cointegrating relations $y_{1t}, y_{2t}, y_{3t}, y_{4t}$. The process u_{it} represents a white noise process for any $i = 1, \ldots, 7$ and we have:

$$x_{1t} = \sum_{i=1}^{t} u_{1i} \sim I(1) \ , \qquad x_{2t} = \sum_{i=1}^{t} u_{2i} \sim I(1)$$

$$x_{3t} = x_{2t} + x_{1t} + u_{3t} \sim I(1) \Longrightarrow y_{1t} = x_{3t} - x_{2t} - x_{1t} \sim I(0)$$

$$x_{4t} = x_{2t} + u_{4t} \sim I(1) \Longrightarrow y_{2t} = x_{4t} - x_{2t} \sim I(0)$$

$$x_{5t} = x_{1t} + u_{5t} \sim I(1) \Longrightarrow y_{3t} = x_{5t} - x_{1t} \sim I(0)$$

$$x_{6t} = \sum_{i=1}^{t} u_{6i} \sim I(1)$$

$$x_{7t} = x_{6t} - x_{2t} + u_{7t} \sim I(1) \Longrightarrow y_{4t} = x_{7t} - x_{6t} + x_{2t} \sim I(0)$$

The second data set (P_2) consists of the same five time series $x_{1t}, x_{2t}, ..., x_{5t}$ with three independent cointegrating relations y_{1t}, y_{2t}, y_{3t} (see Section 3.4.1). Letting v_{it} be a white noise process for any $i = 1, 2, 3$, we have:

$$x_{1t} = \sum_{i=1}^{t} v_{1i} \sim I(1) \ , \qquad x_{2t} = \sum_{i=1}^{t} v_{2i} \sim I(1)$$

$$x_{3t} = x_{2t} + x_{1t} + v_{3t} \sim I(1) \Longrightarrow y_{1t} = x_{3t} - x_{2t} - x_{1t} \sim I(0)$$

$$x_{4t} = x_{2t} + v_{4t} \sim I(1) \Longrightarrow y_{2t} = x_{4t} - x_{2t} \sim I(0)$$

$$x_{5t} = x_{1t} + v_{5t} \sim I(1) \Longrightarrow y_{3t} = x_{5t} - x_{1t} \sim I(0)$$

5.9.2 Sensitivity analysis around B

As seen in Section 5.3.3, the matrix S has an Inverse-Wishart prior distribution implied by the Gaussian prior of the $p \times r$ matrix U obtained from the decomposition $S^+ = UU'$. The prior

of $Vec(U)$ is a multivariate normal distribution with mean 0 and variance $I_r \otimes B$. In this section, we will study the impact of certain values of the variance parameter B on the cointegrating relations and on the posterior distributions of some parameters of the VECM.

We decide to take the first simulated data set P_1 to do this sensitivity analysis. For each algorithm, we set C, i.e. the value for the variance of the proposal distribution of U, for which we obtain a good acceptance rate for the Metropolis-Hastings algorithm (see Algorithm 7, Section 5.7). The value of this acceptance rate is discussed in much more detail in Section 5.9.3, Table 5.2. We run Algorithms 8, 9 and 10 three times for the first simulated data set P_1: the first MCMC procedure uses large values for B ($B = 10 \times I_p$), the second MCMC procedure takes the identity matrix ($B = I_p$) and the third MCMC procedure uses an even smaller norm for B ($B = 0.1 \times I_p$).

The posterior and prior densities of some coefficients of the VECM when $B = 10 \times I_p$ are presented in Figure 5.2. We decide to display the density of a coefficient of the covariance matrix S in the top right (S_{22}). If B is too big, then the values of U can indeed become quite big and therefore the values of S will become very much concentrated around 0. The prior and the posterior will be very similar (see Figure 5.2). The other coefficients show a posterior density that is different from the prior, as expected. Note that the shape of the density of the coefficient of Π is due to the singular property of the cointegrating matrix Π.

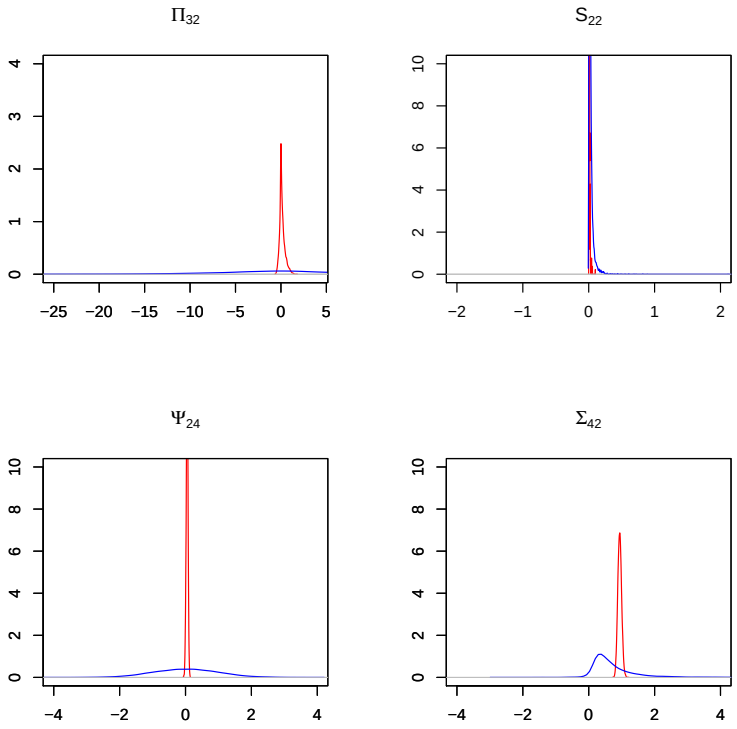

Figure 5.2: Posterior and prior densities of the coefficients Π_{32}, S_{22}, Ψ_{24} and Σ_{42} with $B = 10 \times I_p$: Posteriors in red and priors in blue.

Prior and posterior densities when using $B = I_p$ are presented in Figure 5.3 and seem to be quite different from each other for all the coefficients. In particular, the posterior density of S_{22} is sharper than its corresponding flat prior.

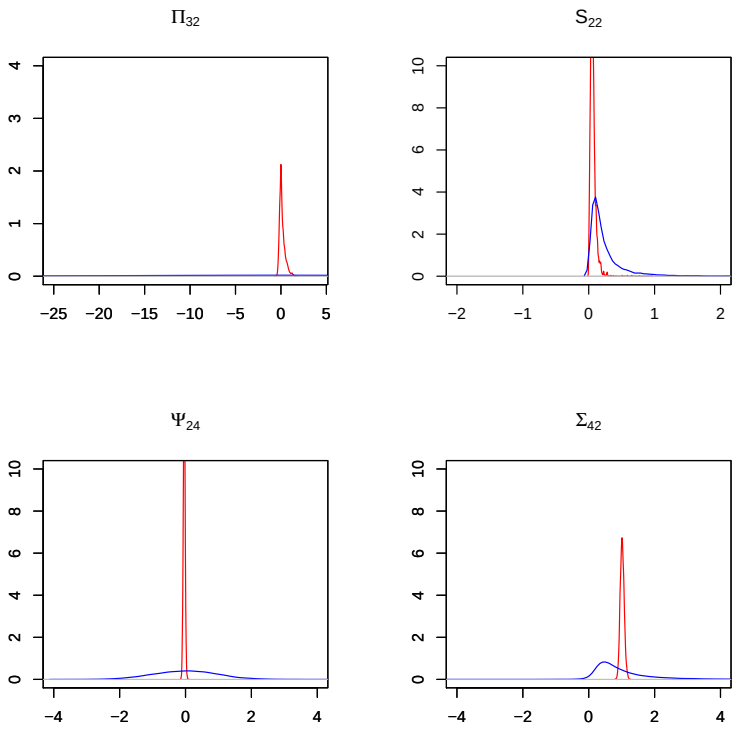

Figure 5.3: Posterior and prior densities of the coefficients Π_{32}, S_{22}, Ψ_{24} and Σ_{42} with $B = 1 \times I_p$: Posteriors in red and priors in blue.

Finally, densities shown in Figure 5.4 where we use $B = 0.1 \times I_p$ present a rather flat posterior distribution for S_{22}. From this sensitivity analysis, we think it is preferable to take B equal to the identity (see Figure 5.3) for the simulated data sets and the real data sets later on. In the next section (see Section 5.9.3), we will try to adjust the variance parameter of the proposal distribution for the Metropolis-Hastings algorithm (see Algorithm 7, Section 5.7) to be acceptable.

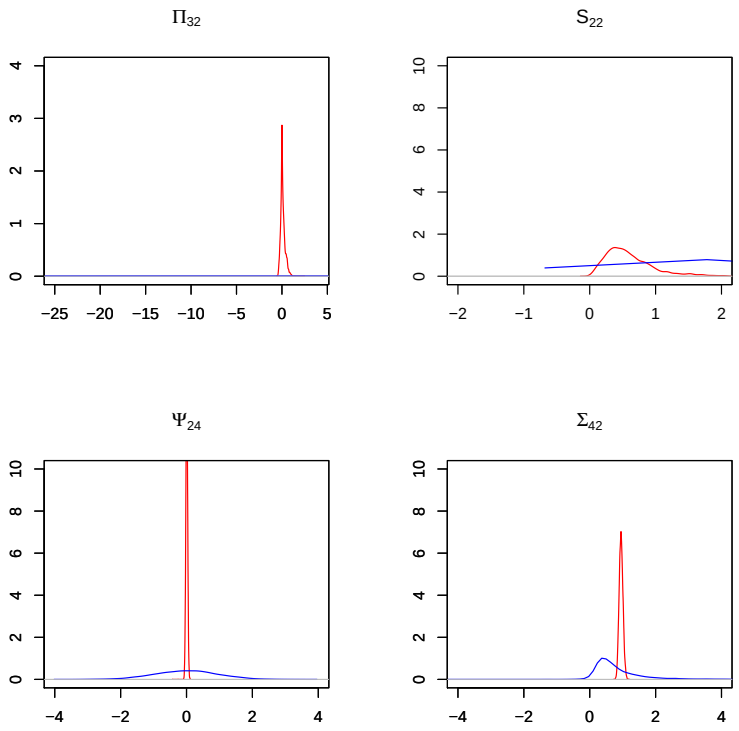

Figure 5.4: Posterior and prior densities of the coefficients Π_{32}, S_{22}, Ψ_{24} and Σ_{42} with $B = 0.1 \times I_p$: Posteriors in red and priors in blue.

5.9.3 Adjusting the acceptance rate of the Metropolis step with the variance C of the proposal distribution

In this section, we highlight the fact that we should be careful in parameterizing the variance C of the proposal distribution in the Metropolis-Hastings algorithm (see Algorithm 7). We can compute the acceptance rate, denoted as a_r, i.e. the proportion of times in the MCMC procedure for which the proposal value is accepted. If we have a multidimensional random walk Metropolis algorithm, which is the case when simulating U, then Roberts *et al.* (1997) show that the optimal acceptance rate is 0.234. However, in the literature, it is generally accepted to have an acceptance rate that lies between 0.20 and 0.40. We will try in this section to obtain the optimal acceptance rate for our simulated data sets by scaling the value of C when the variance B of the prior of U

is equal to the identity: $B = I_p$ (see Section 5.9.2).

Small values of C will reduce the variance around the previous value of U in the MCMC algorithm, and make each simulated $U^{(i)}$ very close to the mean of the proposal distribution, i.e. the previous value $U^{(i-1)}$. Therefore the acceptance ratio α (see Algorithm 7) is more likely to be equal to 1, and we will have an acceptance rate that is more likely to be close to 1. Therefore we would need to increase the size of C, so that the acceptance rate is closer to the optimal value 0.234, see Roberts *et al.* (1997). If the norm of C tends to be large, then the simulated values for $U^{(i)}$ by the proposal distribution will be more likely far from the previous value $U^{(i-1)}$, making the acceptance ratio more likely to be different from 1. We will therefore bring down the acceptance rate to smaller values.

A good acceptance rate leads to more volatility for the values of U as shown at the top of Figure 5.5. However, if the acceptance rate is too low, the simulated $U^{(i)}$s are rejected at almost every step i of the MCMC algorithm. This example is a little extreme but if we look at the bottom of Figure 5.5, we can see that the trace plots of the coefficients U_{21} and U_{32} stay the same for almost all the iterations of the MCMC algorithm. As a consequence, the covariance matrix $S = (UU')^+$ simulated at each iteration becomes almost fixed during the MCMC algorithm because the previous value is chosen. Therefore we come back to the problem where S is fixed (see Section 5.3.2). The resulting independent cointegrating relations presented in Table 5.1 below and obtained from Algorithm 10 are not correctly found.

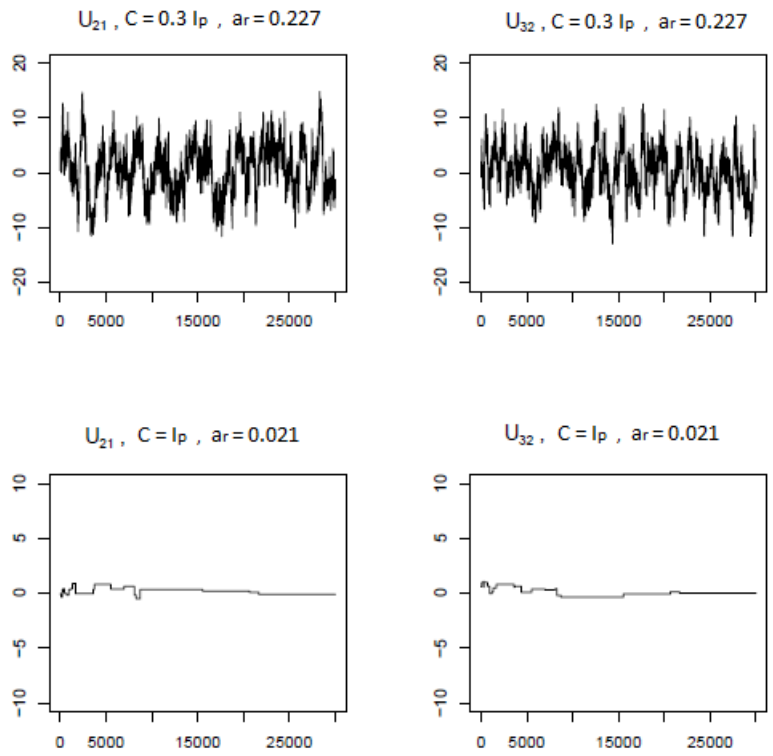

Figure 5.5: Trace plots of the coefficients U_{21}, U_{32} for the first simulated data set P_1. Trace plots on the top with $C = 0.3 \times I_p$ and $a_r = 0.227$, and trace plots on the bottom with $C = I_p$ and $a_r = 0.021$.

Table 5.1: Cointegrating relations for the first simulated data set P_1 with $B = I_p$ and $C = I_p$ and $a_r = 0.021$.

x_{7t}	x_{5t}	x_{4t}	x_{3t}	x_{6t}	x_{2t}	x_{1t}
1	0	0	0	-0.662	-0.239	-0.831
0	1	0	0	0.168	1.147	0.121
0	0	1	0	0.921	-0.601	1.032
0	0	0	1	2.709	0.481	2.281

Table 5.2 compares the values of C with the acceptance rate a_r that we obtain after running

the MCMC procedure defined by Algorithms 8 and 9 in Section 5.8.2. We use the same other hyperparameters as defined in Algorithm 8 but we change the value of C only. At the end of each MCMC procedure, we collect the corresponding acceptance rate a_r. From these tables we can see that as we reduce the amplitude of C, the acceptance rate increases towards the optimal acceptance rate desired.

Table 5.2: Acceptance ratio for the two simulated data sets P_1 and P_2

Simulated data set P_1		Simulated data set P_2	
C	a_r	C	a_r
$1 \times I_p$	0.021	$1 \times I_p$	0.135
$0.5 \times I_p$	0.112	$0.8 \times I_p$	0.157
$0.3 \times I_p$	0.227	$0.5 \times I_p$	0.254

Table 5.3 shows the independent cointegrating relations for the first simulated data set with the optimal acceptance rate found from Table 5.2, that is the closest a_r to the value 0.234, see Roberts *et al.* (1997). The value of $a_r = 0.227$ is taken and then we choose $C = 0.3 \times I_p$. After that, we run the algorithm seen in Section 5.8 with $m = 30{,}000$ iterations and a burn-in period of 20,000 iterations. We use the fixed hyperparameters defined in Algorithm 9.

Table 5.3: Cointegrating relations for the first simulated data set P_1: $C = 0.3 \times I_p$, $a_r = 0.227$.

x_{7t}	x_{5t}	x_{4t}	x_{3t}	x_{6t}	x_{2t}	x_{1t}
1	0	0	0	-0.952	0.911	0.070
0	1	0	0	0.009	0.076	-0.946
0	0	1	0	-0.043	-0.950	-0.069
0	0	0	1	-0.024	-0.946	-0.993

Table 5.4 shows the cointegrating relations obtained from the second simulated data set P_2 by specifying the value C for which we have an acceptance rate of 0.254, that is $C = 0.5 \times I_p$.

Here we can see that when C is well chosen we obtain accurate estimates of the cointegrating coefficients.

Table 5.4: Cointegrating relations for the second simulated data set P_2: $C = 0.5 \times I_p$, $a_r = 0.254$.

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	0	0	0.001	-0.991
0	1	0	-0.997	-0.002
0	0	1	-0.989	-0.995

Comparison with the static model of Chapter 3 for the simulated data sets

In this section, we recall the cointegrating relations found by applying the methods seen in Chapter 3 with Algorithms 1, 2 and 3. The cointegrating relations found with the methods described in this chapter are not very different if we compare the first simulated data set P_1 (see Table 5.3 and Table 3.3) or the second simulated data set P_2 (see Table 5.4 and Table 3.4).

Table 5.5: Cointegrating relations for the first simulated data set P_1 with the static model of Chapter 3.

x_{7t}	x_{5t}	x_{4t}	x_{3t}	x_{6t}	x_{2t}	x_{1t}
1	0	0	0	-0.941	1.009	0.043
0	1	0	0	-0.058	0.043	-0.971
0	0	1	0	-0.014	-0.965	0.047
0	0	0	1	-0.056	-0.961	-0.961

Table 5.6: Cointegrating relations for the second simulated data set P_2 with the static model of Chapter 3.

x_{5t}	x_{4t}	x_{3t}	x_{2t}	x_{1t}
1	0	0	-0.012	-1.009
0	1	0	-0.983	0.014
0	0	1	-1.019	-1.009

The advantage of the method seen in Chapter 3 is that the cointegration rank is evaluated during the MCMC procedure. However in this chapter, the novelty relies on the fact that we used singular distributions to infer the cointegrating matrix. This chapter would constitute the first step in a new approach for inferring the cointegrating matrix in the VECM.

5.9.4 Posterior summaries

In this section, we highlight posterior summaries of some parameters of the VECM for the first simulated data set P_1. The trace plots shown in Figure 5.6 are taken after running the MCMC procedure presented by Algorithms 8 and 9 and by using the parameters $C = 0.3 \times I_p$ and $B = I_p$ that we thought were appropriate in Sections 5.9.3 and 5.9.2, respectively. We also display trace plots of some coefficients of the hyperparameter U, showing convergence as well (see Figure 5.7).

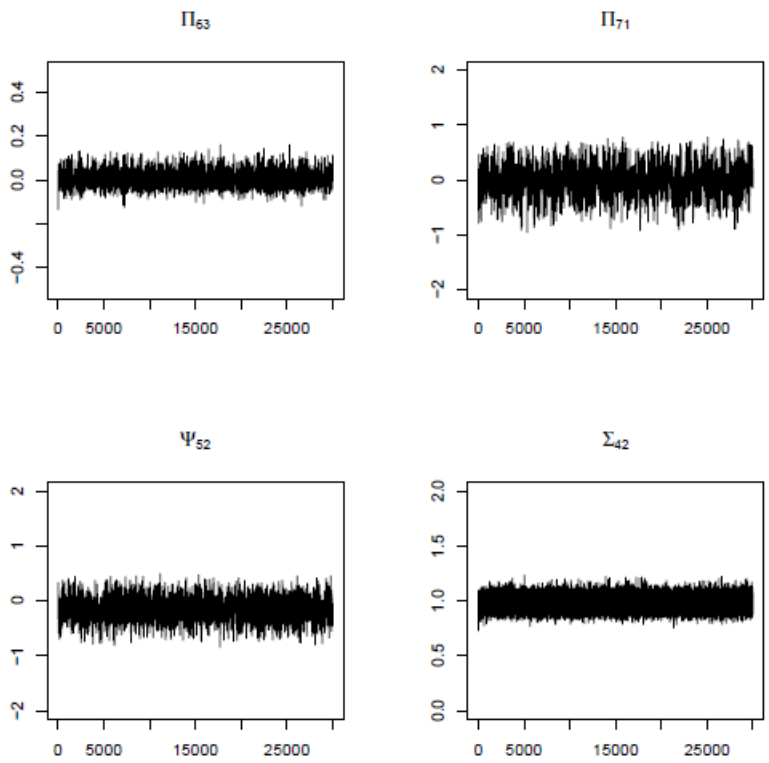

Figure 5.6: Trace plots of the coefficients Π_{63}, Π_{71}, Ψ_{52} and Σ_{42} for the first simulated data set P_1.

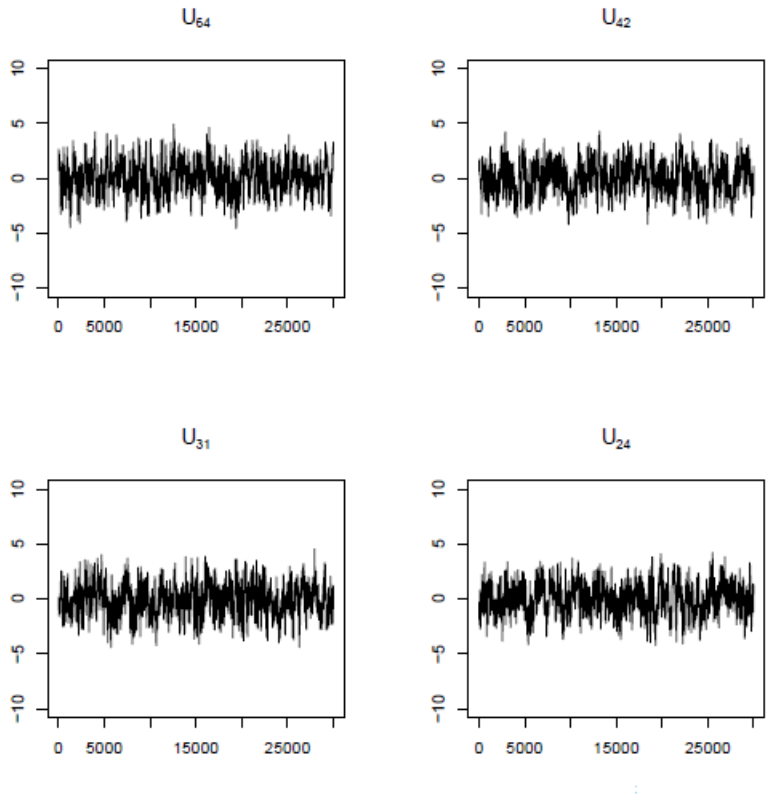

Figure 5.7: Trace plots of the coefficients U_{64}, U_{42}, U_{31} and U_{24} for the first simulated data set P_1.

We can see convergence of those parameters in these trace plots (see Figure 5.6). Their respective densities are represented in Figure 5.8. We decide to show two coefficients of the singular long-run impact matrix (Π_{63} and Π_{71}). The property of singularity for the distribution of Π makes the shape of the distribution non-Gaussian. This comes from the fact that a singular distribution is not defined under the Lebesgue measure and therefore it is not entirely correct to expect these coefficients to have a Gaussian shaped distribution. The posterior distribution of Ψ is however non-singular and Gaussian, and we should expect a symmetric distribution for its coefficients (see Ψ_{52} in Figure 5.8).

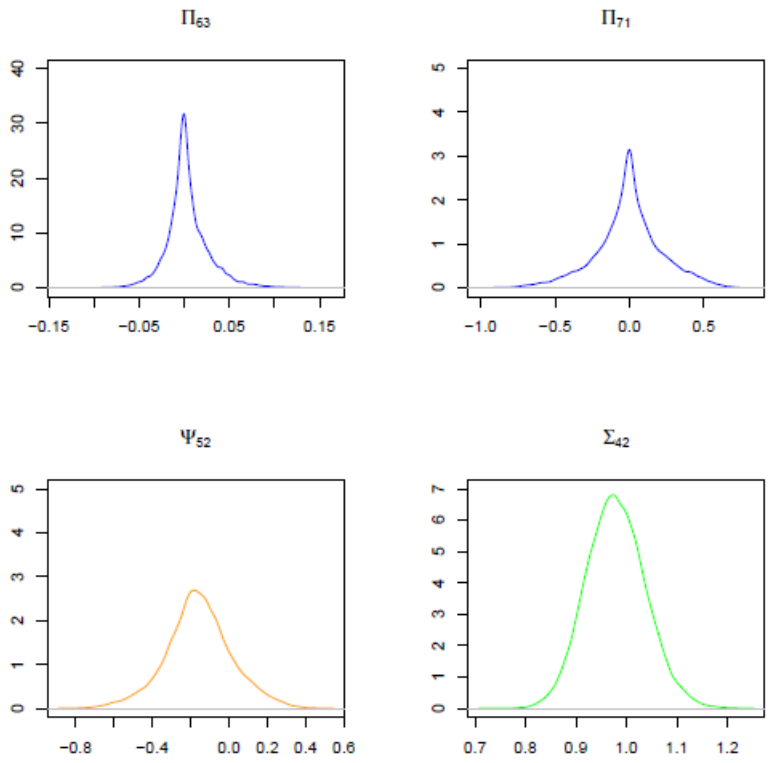

Figure 5.8: Posterior densities of the coefficients Π_{63}, Π_{71}, Ψ_{52} and Σ_{42} for the first simulated data set P_1.

5.9.5 Comparison with the static model of Chapter 3 for the European net tradings

We will compare the results we obtain from the method seen in this chapter and the method seen in Chapter 3 with non-singular posterior distributions for Π and Ψ. In Chapter 3 we compared the independent cointegrating relations between the four net tradings of France, Germany, Italy and Spain during the decade preceding the introduction of the Euro (see Table 3.8, Section 3.4.4). Table 5.7 below compares the cointegrating relations obtained for these time series previously and by using the Bayesian model of this chapter. We use the Algorithms seen in this chapter (see Algorithms 8 , 9 and 10). After having run some MCMC procedures, it was found

that with a variance matrix of the proposal distribution $C = 0.6 \times I_p$, we have an acceptance rate $a_r = 0.231$. The independent cointegrating relations presented on the left hand side of Table 5.7 are derived with $C = 0.6 \times I_p$. We can clearly see a similarity between the results of the two methods.

Table 5.7: Estimated relations between net trading, pre Euro (1991–1998): Method of Chapter 3 and method of Chapter 5.

Chapter 5				Chapter 3			
FraNX	GerNX	ItaNX	SpaNX	FraNX	GerNX	ItaNX	SpaNX
1	0	0	-1.240	1	0	0	-1.234
0	1	0	-1.647	0	1	0	-1.630
0	0	1	-1.393	0	0	1	-1.383

Table 5.8 shows the six independent cointegrating relations between the European economies before the Euro found based on the Bayesian model seen in this chapter. We run Algorithms 8, 9 and 10. In Algorithm 8, we set the cointegration rank $r = 6$. The six independent cointegrating relations presented in Table 3.9 of Section 3.4.6 are not completely the same but they share quite significant similarities. In particular, the net tradings of Germany, Italy and Spain are coevolving positively with the net trading of France (-1.59, -0.79 and -0.63) with a better coevolution between Germany and France (-1.59).

Table 5.8: Cointegrating relations for the pre-Euro time period (1991-1998) obtained with $C = 0.1 \times I_p$ and $B = I_p$. Acceptance rate $a_r = 0.278$

GerNX	ItaNX	SpaNX	ItaIR	GerUR	ItaUR	FraNX	GerIR	FraIR	SpaIR	FraUR	SpaUR
1	0	0	0	0	0	-1.59	-0.42	0.82	-0.42	-0.18	-0.14
0	1	0	0	0	0	-0.79	-0.26	1.28	-0.26	-0.84	0.09
0	0	1	0	0	0	-0.63	-0.29	-0.28	0.60	0.72	-0.55
0	0	0	1	0	0	0.68	-0.06	-0.80	-0.48	0.26	-0.58
0	0	0	0	1	0	-0.61	-0.01	-0.33	0.71	0.60	-0.56
0	0	0	0	0	1	-0.89	-0.16	0.90	-0.29	-0.69	0.32

5.9.6 Application to six major stock market indices

In this section we will study six stock market indices across the world: three stock market indices from the United States of America (NASDAQ, S&P 500, Dow Jones), one stock market index from Japan (Nikkei 225) and two stock market indices from Europe (Paris CAC 40 and Euro Next). The Japanese Nikkei 225 is one of the stock market indices of the Tokyo Stock Exchange and represents the movement of the 225 main Japanese equities' market values. It is a daily price-weighted index in Yen and has been published in the *Nihon Keizai Shimbun* newspaper since 1950. It is similar to the Dow Jones Industrial Average index used in America (see Chapter 4). As for the CAC 40 (Cotation Assistée en Continu) and Euro Next 100, they represent major European stock market indices and are a weighted measure of the most relevant companies' market values in Europe. The Dow Jones Industrial Average index (see Chapter 4), NASDAQ composite (National Association of Securities Dealers Automated Quotations) and Standard and Poor 500 are the three most commonly followed stock market indices in US stock markets.

Our time series consist of daily data collected from 1 January 2012 to 20 September 2016 (source *Yahoo*) for the six stock market indices: NASDAQ, S&P 500, Dow Jones, Nikkei 225,

CAC 40 and Euro Next. The six time series are represented in Figure 5.9.

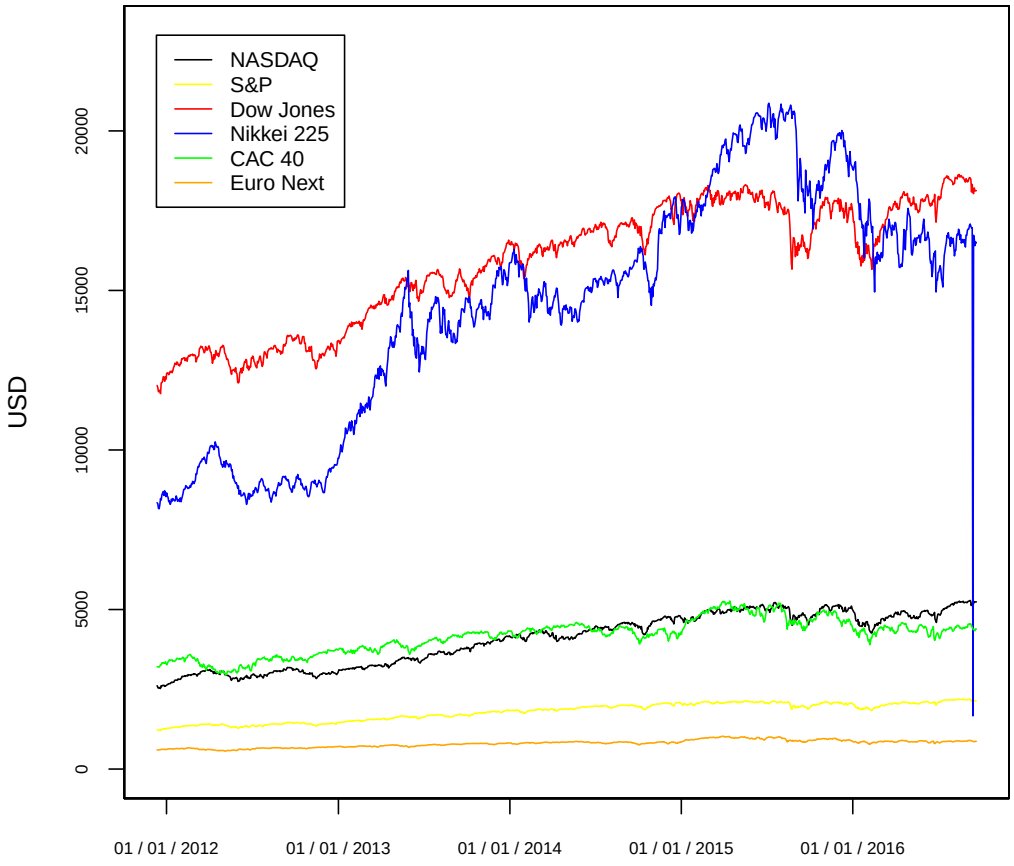

Figure 5.9: Stock market index of NASDAQ, S&P 500, Dow Jones, Nikkei 225, CAC 40 and Euro Next from 1 January 2012 to 20 September 2016.

We follow the instructions of Algorithms 8 and 9 with the number of iterations equal to 30,000. The hyperparameters and initial parameters are based on a pre-sample of size $T/3$ as Algorithm 8 describes. This pre-sample consists of the time period going from 1 January 2012 until 17 July 2013. The sample on which the Bayesian analysis is conducted goes from 18 July 2013 until 20 September 2016. Johansen tests are based on the sample data set and are presented in Table

5.9. Based on these tests, we find a cointegration rank of 2 with 95% confidence level. The test statistic of 43.98 is indeed the first value smaller than the corresponding critical value using the 5% threshold (48.28), and this corresponds to the row where the null hypothesis is $r \leq 2$. Therefore, we reject the null hypothesis that the cointegration rank is smaller than or equal to 1, but we do not reject a rank smaller than or equal to 2, hence the choice of a cointegration rank of 2. Therefore we fix the hyperparameter $r = 2$ in our algorithms.

Table 5.9: Johansen tests for the sample data set of the six stock market indices NASDAQ, S&P 500, Dow Jones, Nikkei 225, CAC 40 and Euro Next. Time Period: 18 July 2013 - 20 September 2016.

Test: $r \leq r_0$	Statistic	Critical values (90%)	Critical values (95%)	Critical values (99%)
0	208.76	85.18	90.39	104.20
1	105.11	66.49	70.60	78.87
2	43.98	45.23	48.28	55.43
3	23.66	28.71	31.52	37.22
4	11.11	15.66	17.95	23.52
5	4.07	6.50	8.18	11.65

The independent cointegrating relations presented in Table 5.10 reveal a positive relation between NASDAQ and the Japanese Nikkei 225 (-0.958) and CAC 40 (-0.361). The Euro Next index is coevolving negatively with the Japanese Nikkei 225 ($+1.614$) and positively with CAC 40 (-9.454). As for the two other American stock market indices (S&P 500 and Dow Jones), they do not seem to coevolve positively with neither NASDAQ ($+0.024$ and $+0.742$) nor Euro Next ($+0.537$ and $+0.589$).

Table 5.10: Independent cointegrating relations between NASDAQ, Euro Next, S&P 500, Dow Jones Index, Nikkei 225, CAC 40. Time period of the sample: July 2013 - September 2016

NASDAQ	Euro Next	S&P 500	Dow Jones	Nikkei 225	CAC 40
1	0	0.024	0.742	-0.958	-0.361
0	1	0.537	0.589	1.614	-9.454

5.10 Discussion

In this chapter, we described a method assuming a singular prior on the long-run impact matrix Π, which led to a singular posterior distribution for Π. Furthermore, assuming a singular prior on the long-run matrix with a certain lower rank $r < p$ implies the derivation of a singular (full conditional) posterior distribution with same rank r. Therefore the rank cannot change throughout the algorithm. We need to set a value of r for the rank before running the algorithm. In the methods given in this chapter, we decide to create a second covariance matrix S in the prior of $\Pi|\Sigma$ that is positive semidefinite of rank r. The prior variance of $Vec(\Pi)|S, \Sigma$ is then defined by $S \otimes v^{-1}\Sigma$. However, fixing S is not adequate because the prior of Π would not be uniformly distributed over the cointegrating coefficients, implying some knowledge about them before the analysis (see Section 5.3.2). This chapter introduces a Metropolis step in order to infer an a priori Gaussian $p \times r$ matrix U taken from the decomposition of the Moore-Penrose generalized inverse of S: $S^+ = UU'$. However unlike in Chapter 3 (where we fix $S = I_p$ as full rank, see Section 3.3.2), and because S is not a diagonal matrix, the method seen in this chapter considers correlation between the vectors of Π.

For the simulated data sets, the rank was already known before running the algorithm. The cointegrating relations in the simulated data sets match correctly with the reality (see Tables 5.3 and 5.4). By comparing the four net tradings before the introduction of the Euro, the cointegration rank of 3 found in Chapter 3 was used (see Section 3.4.4). The comparison with

the cointegrating relations found for the four net tradings before the introduction of the Euro in Chapter 3 are very similar (Table 5.7). We also made an application to six major stock market indices (see Section 5.9.6).

However, in this chapter the cointegration rank is considered as fixed. For future works and on the same path as Villani (2005), we could infer the cointegration rank r and find a posterior distribution depending on the data only. We can for instance give a discrete uniform prior for r: $f(r) = 1/(p+1)$ for $r \in [\![0, p]\!]$. As described in the literature review (see Section 2.4.3), we would then need to integrate out Π, Ψ, Σ and U from the joint posterior distribution of all these parameters in order to obtain the marginal likelihood of the data given the cointegration rank:

$$f(\mathcal{D}|r) = \int \cdots \int L(\mathcal{D}; \Pi, \Psi, \Sigma, r) f(\Pi, \Psi, \Sigma, U, r) \; d\Sigma \; d\Psi \; (d\Pi) \; dU$$

where $(d\Pi)$ defines the Hausdorff measure (5.8). The posterior distribution of the cointegration rank given the data is then established by the ratio (2.19) (see Section 2.4.3):

$$f(r|\mathcal{D}) = \frac{f(\mathcal{D}|r) f(r)}{\sum_{r=0}^{p} f(\mathcal{D}|r) f(r)}$$

Similar to Kleibergen and Paap (2002), we could also use Bayes factors in order to evaluate the cointegration rank. In Section 2.4.4, we saw how Kleibergen and Paap (2002) set a Bayes factor $BF[r|p]$ (2.26) for each rank r in order to test it against the full rank model:

$$BF[r|p] = \frac{f(\mathcal{D}|\text{rank} = r)}{f(\mathcal{D}|\text{rank} = p)} \tag{5.52}$$

If this Bayes factor has a value larger than 1, then the model with rank r is more likely than the full rank model. The posterior probability of the rank (2.28) based on the posterior odds ratio (2.27) would basically evaluate the rank (see Section 2.4.4). In that case, the most likely cointegration rank will be the one for which the posterior probability is the highest. The use of singular distributions for the cointegrating matrix Π could maybe simplify the computations of the Bayes factors based on different Error Correction Models (induced by different corresponding cointegration ranks).

Another idea for inferring the cointegration rank would be to use a Metropolis step for

the joint distribution of U and r. Indeed, since U is conditional on the rank, as being the number of columns in U, we cannot infer the rank r separately from it. We can use a uniform or binomial distribution as the proposal distribution $\pi(r)$ for the cointegration rank r. As for $U|r$, we can use the same Gaussian proposal distribution as seen in this chapter, called $\pi(U|r)$ here. The multiplication of both $\pi(r)$ and $\pi(U|r)$ creates our joint proposal distribution $\pi(U, r) = \pi(r)\pi(U|r)$. We also need to use the joint prior distribution $f(U, r) = f(U|r)f(r)$ in order to derive an unknown form of density for our full conditional distribution $f(U, r|\Pi, \Sigma)$ from which we will be able to define an acceptance ratio at each step of the MCMC algorithm. However, with this new approach in mind, we would have to consider a reversible-jump Markov Chain Monte Carlo methodology in order to sample this joint posterior distribution with varying dimensions, see Green (1995).

Chapter 6

Consideration for future work: A dynamic VECM including a singular distribution for the time-varying cointegrating matrix

6.1 Introduction

By continuing with the same idea of having a dynamic model as in Chapter 4, we can think about constructing a forward filtering and backward recursion algorithm, by using a singular distribution on the cointegration matrix. We provide a set of ideas for this new concept in this chapter.

In our assumptions, at each time t, the long-run impact matrix Π_t will have a singular distribution given the rank r_t and Ψ_t will have a non-singular distribution. Like in Chapters 3,4 and 5, we also assume independence between our long-run relationships matrix Π_t and the lag parameter matrix Ψ_t. We can split the transition equation into a part concerning Π_t and a part concerning Ψ_t:

$$y_t = (Z_t' \otimes I_p)\theta_t + u_t, \;\; u_t \sim N(0, \Sigma_t) \quad \textit{Measurement equation} \tag{6.1}$$

$$\pi_t = F\pi_{t-1} + \nu_{1t}, \;\; \nu_{1t} \sim NS(0, Q_{1,t}|p \times r_t) \quad \textit{Transition equation 1} \tag{6.2}$$

$$\psi_t = G\psi_{t-1} + \nu_{2t}, \;\; \nu_{2t} \sim N(0, Q_{2,t}) \quad \textit{Transition equation 2} \tag{6.3}$$

with

$$\theta_t = \begin{bmatrix} \pi_t \\ \psi_t \end{bmatrix} \tag{6.4}$$

$Q_{1,t}$ is a positive semidefinite matrix of rank $p \times r_t$ implying a vector-variate normal distribution for ν_{1t}. The lag parameter matrix Ψ_t is still considered as non-singular and therefore $Q_{2,t}$ is defined as the positive definite variance matrix of $\psi_t = Vec(\Psi_t)$.

Then for each t, and given the rank r_t, the forward filtering part of π_t would then consist of the following five steps:

$$\pi_{t|t-1} = F\pi_{t-1|t-1} \tag{6.5}$$

$$P_{t|t-1} = FP_{t-1|t-1}F' + Q_{1,t-1} \tag{6.6}$$

$$K_t = P_{t|t-1}(x_{t-1} \otimes I_p)((x_{t-1}{}' \otimes I_p)P_{t|t-1}(x_{t-1} \otimes I_p) + \Sigma_t)^{-1} \tag{6.7}$$

$$\pi_{t|t} = \pi_{t|t-1} + K_t(y_t - (x_{t-1}{}' \otimes I_p)\pi_{t|t-1}) \tag{6.8}$$

$$P_{t|t} = P_{t|t-1} - K_t(x_{t-1}{}' \otimes I_p)P_{t|t-1} \tag{6.9}$$

Since for each time t, $P_{t|t-1} \geq 0$ and $\Sigma_t > 0$, then the sum $(x_{t-1}{}' \otimes I_p)P_{t|t-1}(x_{t-1} \otimes I_p) + \Sigma_t > 0$ and is therefore invertible (step (6.7) of the forward filtering algorithm).

The backward recursion part of π_t is defined as the following:

$$\pi_{t|t+1} = \pi_{t|t} + P_{t|t}F'P^{+}_{t+1|t}(\pi_{t+1} - F\pi_{t|t}) \tag{6.10}$$

$$P_{t|t+1} = P_{t|t} - P_{t|t}F'P^{+}_{t+1|t}FP_{t|t} \tag{6.11}$$

6.2 Bayesian inference about the transition equation

Since the cointegration rank is time-varying, the variance of the transition equation depends on it, i.e. $Q_{1,t}$, is also dynamic, unlike the variance of the transition equation seen in Chapter 4.

Given the fact that there exists a unique Moore-Penrose inverse matrix of $Q_{1,t}$, which we will call Λ_{1t} here, we will try to find a prior distribution for that latter matrix.

$Q_{1,t}$ is a singular matrix in our assumptions and of rank r_t, then the prior of Λ_{1t} is also singular of rank r_t. The singular Wishart distribution introduced in Definition 9 of Section 5.3.3 will be used for the dynamic precision Λ_{1t}. Under the condition of singularity, this prior depends on the dynamic rank r_t associated with Π_t. Therefore, the prior explained in the next section will be the prior of $\Lambda_{1t}|r_t$, where r_t is also a dynamic parameter.

From Uhlig (1994) and Gupta and Nagar (2000), who explored the singular Wishart distribution (see Definition 9 in Section 5.3.3), we can also define the singular Inverse-Wishart distribution as Díaz-García *et al.* (1997) and Uhlig (1994) suggested. The idea would then be to elicit an Inverse-Wishart prior distribution on the variance parameter of the transition equation (6.2). Instead, we decide not to go too far in developing definitions about too many singular distributions and prefer to focus more on the singular Wishart distribution, which is easier to handle and simulate. The idea is now more to conduct Bayesian inference on the precision matrix Λ_{1t} of the transition equation (6.2) rather than the covariance matrix $Q_{1,t}$ itself.

Prior of the singular precision matrix given the time-varying cointegrating rank r_t

A singular Wishart distribution is used for the precision matrix $\Lambda_{1t} = Q_{1,t}^{+}$ in (6.2). This distribution involves two hyperparameters: the scale matrix B and the degrees of freedom, w_t, which is a time-varying hyperparameter involving the dynamic cointegration rank r_t. It is useful to note that B is a positive definite matrix: $B > 0$. For the singular Wishart distribution, the number of degrees of freedom is equal to the rank of the matrix we want to simulate.

The vector ν_{1t} is simulated from a singular normal distribution of which the variance is of rank $r_t \times p$. Suppose $w_t = r_t \times p$ and let us assume that the rank of the matrix Π_t is $r_t < p$, then we can state the prior of each Λ_{1t} as:

$$\Lambda_{1t} \sim WS(B, w_t) \tag{6.12}$$

The full conditional distribution of $Q_{1,t}$

We denote the density of each ν_{1t} as $f(\nu_{1t})$ from the transition equation (6.2). But we notice that ν_{1t} has a singular normal distribution and therefore the likelihood function of the transition equation will be proportional to the singular likelihood implied by the distribution of ν_{1t}. This likelihood will therefore be singular and defined on the Hausdorff measure, see Díaz-García *et al.* (2006) and Proposition 1 in Section 5.2.2.

Now we notice that in order to find the full conditional posterior distribution of Λ_{1t}, we will have to multiply the singular prior density (6.12) of rank w_t by the singular likelihood proportional to the singular normal distribution of ν_{1t}. We know that a posterior distribution for Λ_{1t} exists, but in order to apply Bayes' theorem we need to study more about the Hausdorff measure, in which our singular parameter and our likelihood do have a density.

6.3 Discussion

In this new approach, we encounter three difficulties:

- First of all, this approach would be based on the knowledge of a stochastic process representing the cointegration rank $(r_t)_{1\leq t\leq T}$ beforehand. We are not sure about what type of stochastic process could be involved in order to describe precisely the cointegration rank.

- Then, given the stochastic process r_t, the FFBR algorithm would require us to find a full conditional singular posterior distribution for the precision matrix of the transition equation (6.2) Λ_{1t}. Setting a prior and obtaining a posterior singular Wishart on the precision matrix Λ_{1t} is not a simple task. We cannot simply apply Bayes' theorem by multiplying the likelihood by the density of the Wishart prior, and then derive a conjugate posterior distribution, like in Chapter 4. The density of a singular parameter is not defined under the Lebesgue measure. The only fact that we know is that the posterior distribution of the precision matrix will have to be defined on the same space as the prior distribution, i.e. the space of square matrices of lower rank r_t.

- Finally, even if we derive a suitable stochastic process for the cointegration rank (r_t) and a suitable defined posterior distribution for the precision parameter of Π_t, then the simulations of $\Lambda_{1,t}$, would give a matrix of rank $r_t p$, as required. However, the addition with $FP_{t-1|t-1}F'$ in (6.6) would not necessarily give a matrix of rank $r_t p$. Then, following the steps (6.7), (6.8) and (6.9), the covariance matrix (6.11), of which the rank defines the rank of the singular distribution of Π_t, will not necessarily be $r_t p$.

Chapter 7

Conclusions and future work

7.1 Conclusions

7.1.1 Main findings

The goal of this thesis is to develop methods to determine the cointegration rank and the cointegrating relationships for a set of difference stationary time series. For that, we develop Bayesian methods around the cointegration matrix of the Vector Error Correction Model.

In previous studies, the long-run relationships matrix was split into two full rank matrices on which non-singular priors were set. The work of this thesis focuses on developing Bayesian inference methods on the cointegrating matrix of the VECM. The first two methods include the determination of the cointegration rank inside the Markov Chain Monte Carlo procedure or the Forward Filtering Backward Recursion algorithm (see Chapters 3 and 4). These two methods set a non-singular prior on the cointegration matrix and determines the cointegration rank based on the number of irrelevant singular values of the cointegration matrices simulated. We are then able to avoid the use of frequentist Johansen tests in order to determine the cointegration rank. The last method introduces a singular prior distribution for the cointegration matrix depending on the cointegration rank, pre-determined from the data by using Johansen tests.

In Chapter 3 a full conditional posterior distribution is used for the cointegration matrix and the lag parameters matrix, by integrating out the covariance matrix of the errors. At each step of the Gibbs sampler, a cointegration rank is estimated from the number of relevant singular values of the cointegration matrix simulated. These methods are then applied to the four most important economies of the Eurozone over a period before and a period after the introduction of the Euro. The cointegration rank is decreasing between the period before the Euro and the period after the Euro, allowing less cointegrating relations and therefore less comovements between those four economies. We also note a divergence in the net trading between the Mediterranean countries (including France) and Germany after the introduction of the single currency.

Chapter 4 offers a time-varying estimation of the parameters of the Vector Error Correction Model and sets a Forward Filtering Backward Recursion algorithm for estimating, in particular, the time-varying cointegration rank and matrix. This model that allows the parameters to change over time is called the dynamic VECM, in order to differentiate it from a static VECM, seen in the previous chapter (Chapter 3). We create simulated data sets split into two or three time periods, where the number of cointegrating relationships and the relationships themselves change over time. Results show a good similarity between the evolution of the cointegration rank over the separate time-periods, but also the nature of the cointegrating relationships, and the simulated data sets created (see Section 4.7.3). We then decide to study the evolution of the cointegration rank on real data sets. We detect a decrease of the cointegration rank in a part of the European panel data set studied in the previous chapter. We also study stock market indices of companies from three different sectors from the Dow Jones: manufacturing companies, banking/insurance and electronic companies. The dynamic VECM is used over the time period covering June 2001 to June 2009. We noted that during this time period, the cointegration rank of the Electronics sector seemed to start at 3, and then increase to 4 from 2005 (see Figure 4.13).

Chapter 5 sets a singular distribution on the cointegration matrix of the VECM conditional on the rank. The long-run relationships matrix is indeed a singular matrix in our hypothesis. We established in this chapter another way of obtaining the posterior distribution of Π than by

using Bayes' theorem. We also retrieved the important property that the rank of the singular full conditional posterior distribution of Π is equal to the rank of the prior distribution of Π. We decided to use the traditional Johansen cointegration tests on the data, or use the method seen in Chapter 3 to determine the rank before running the Gibbs sampler.

The simulated data set guarantees the efficiency of this method, with again a very good similarity between the synthetic cointegration relationships and their estimated coefficients from the MCMC procedure. The new methods are then applied to some real data sets from the previous chapters and to six major stock market indices taken from January 2012 to September 2016.

7.1.2 Advantages of the novel methods on Bayesian cointegration

The main advantage in the methods developped in this project is the fact that we are estimating the long-run relationships matrix, and thus being able to extract a rank out of it, after each simulation. We can thus identify a distribution of the cointegration rank given the data. Furthermore, the cointegration relationships can be selected by manipulating the mean, the mode or the median of the general cointegrating matrix.

This thesis elaborates Bayesian inference methods for the cointegration matrix and the cointegration rank. The methods seen in Chapters 3 and 4 do not require the use of Johansen tests in order to determine the cointegration rank. These methods also estimate the cointegrating relationships, coming from simulations of the long-run impact matrix. We can use this new approach of estimation by constructing dynamic models that highlight in particular the evolution of the cointegration rank over time, as well as the evolving cointegration relationships (see the simulated data in Chapter 4).

7.1.3 Limitations of these Bayesian estimations

One of the key issues about doing Bayesian estimations is the time required to run an algorithm. In R Core Team (2013), Gibbs sampling for large matrices often takes quite long, especially for the dynamic version of the VECM seen in Chapter 4, taking about an hour. How-

ever, by defining the main functions in C++, it is possible to re-use them in our R programs thanks to the RCPP package, see Eddelbuettel (2013) and Eddelbuettel and François (2011). The Gibbs samplers are approximately three times faster as a consequence.

Our thesis is also limited by the fact that we do not set any Bayesian inference on the cointegration rank. Once we simulate a cointegration matrix, the rank is determined from the cointegration matrix simulated. The rank is a one-to-one function of the matrix and therefore, if we set a prior and derive a posterior distribution for the matrix, then we should not set a prior on the rank. There are two ways in which we can determine the rank: either the rank is determined with the Johansen tests on the data before running the analysis (see Chapter 5), or the rank is determined from the simulations of the cointegration matrix, based on its irrelevant singular values, because the latter follows a non-singular distribution (see Chapters 3 and 4).

Another issue that occurs in our methods comes from the fact that the lag order of the original Vector Autoregressive model is not determined by Bayesian inference, but estimated by a frequentist analysis across the whole data set at the beginning, by comparing different AIC values.

7.2 Future work and other directions

For the static Vector Error Correction Models, one possibility would be to create a prior on the lag-order of the Vector Autoregressive model at the beginning, and include it in the MCMC algorithm by sampling from its full conditional posterior distribution. In this thesis, we applied a lag order of 2 for all sets of time series studied. An argument for this is given in Appendix B. In this Appendix, we reinforce the assumption that if we choose a small lag order, the lag parameter matrices would automatically compensate and we would retrieve the cointegrating relationships in Π. We finally choose a lag order of 2 in order for our simulations of the lag parameter matrices Ψ to be faster. The lag order could, for example, have a Poisson prior distribution (where the support $\mathbb{N}$ is infinite, in which case no restriction is imposed). However, this new technique would involve changing the dimensions of the lag parameters Ψ, and in the case where the simulated

lag order $k^{(i)}$ would be big, then the simulation process of $\Psi^{(i)}$ would become much slower.

In the models of Chapter 3 and Chapter 4 where we assumed the cointegrating matrix to have a non-singular prior, we did not explore the idea of having a prior distribution for the rank implied by the priors of the singular values of Π. An interesting future work would be to derive a posterior distribution for the rank and therefore, to infer it in our MCMC procedure.

The singular approach in the static Error Correction Model in Chapter 5 could be improved by also inferring the cointegration rank r with matrix U (see Section 5.3.3). In this chapter, we decompose the Moore-Penrose generalized inverse of the covariance matrix S into the product UU' where the $p \times r$ matrix U has a Gaussian prior distribution. The number of columns of U is defined to be r and therefore, U is conditional on r. We can start by setting a joint prior distribution for (U, r): $f(U, r) = f(U|r)f(r)$. We can use the same Gaussian prior for $U|r$ and a uniform prior for r, for instance. After that, we could estimate the joint conditional distribution of (U, r) associated with this prior by a Metropolis-Hastings step.

In Chapter 6 we described several theoretical approaches for a dynamic Error Correction Model that would involve a singular prior on the cointegration matrix. The aim is to try to establish a cointegration rank evolving over time along with the cointegrating relations. The cointegration rank could be defined as a stochastic process $(r_t)_{1 \leq t \leq T}$. Based on this moving cointegration rank, we could define a full conditional singular posterior distribution for the precision matrix of the transition equation (6.2), and therefore obtain singular cointegrating matrices Π_t evolving over time. More works have to be explored on this side.

Appendix A

Generalized inverse of a positive semi-definite matrix

A.1 Introduction and definition

This section recalls some properties of the generalized inverse and in particular the generalized inverse of a positive semidefinite matrix. These properties are necessary to be reminded because we will have to deal with positive semidefinite covariance matrices in the singular distribution of the long-run relations matrix Π (see Chapter 5). One of the key points of the singular distribution is that they have a singular variance.

In this section, m and n will denote two natural integers such that $m > 1$ and $n > 1$. The aim of this section is to explain how to obtain a solution of the following system:

$$Ax = b \tag{A.1}$$

where A is a $m \times n$ singular matrix (i.e. non invertible), $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$.

Following the notes from Abu-Saman (2012) we first recall the definition of a generalized inverse of a matrix.

Definition 10. *If $A \in \mathcal{M}_{m,n}(\mathbb{R})$, then G is a generalized inverse of A if $G \in \mathcal{M}_{n,m}(\mathbb{R})$ and: $AGA = A$.*

With that definition in mind it is obvious that the inverse of any $n \times n$ invertible matrix A is *a* generalized inverse of A. In addition, it is good to remember that there may be several *generalized inverse matrices* for one matrix. In this chapter, we recall and implement the method in order to obtain ***a*** generalized inverse matrix of ***a positive semidefinite*** matrix.

A.2 Solution of linearly-dependent equations

We will prove the following theorem mentioned by Abu-Saman (2012):

Theorem 1. *Let $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and assume that G is a generalized inverse of A. Then for any fixed $b \in \mathbb{R}^m$:*

1. *The system $Ax = b$ has a solution $x \in \mathbb{R}^n$ if and only if $AGb = b$*

2. *If $Ax = b$ has any solution, then:*
   x is a solution of $Ax = b$ if and only if $x = Gb + (I_n - GA)z$ for some $z \in \mathbb{R}^n$

$$\textbf{\textit{A particular solution of}}\ Ax = b\ \textit{for}\ b\ \textit{in the range of}\ A\ \textit{is}\ x = Gb \tag{A.2}$$

Proof 1. *Let us prove "$Ax = b$ has a solution $x \in \mathbb{R}^n \Leftrightarrow AGb = b$."*

1. *$\underline{Ax = b \text{ has a solution } x \in \mathbb{R}^n \Rightarrow AGb = b:}$*

   $Ax = b$ has a solution. Then $AGAx = AGb$, and since $AGA = A$ then we have $AGAx = Ax = AGb$. Finally, since $Ax = b$ we obtain $AGb = b$, as required.

2. *$\underline{AGb = b \Rightarrow Ax = b \text{ has a solution } x \in \mathbb{R}^n:}$*

   $AGb = b \Rightarrow x = Gb$ is a solution of $Ax = b$.

Proof 2. *Let us assume that $Ax = b$ has a solution.*

We will prove "x is a solution of $Ax = b \Leftrightarrow x = Gb + (I_n - GA)z$ for some $z \in \mathbb{R}^n$."

1. *$\underline{x \text{ is a solution of } Ax = b \Rightarrow x = Gb + (I_n - GA)z \text{ for some } z \in \mathbb{R}^n:}$*

x is a solution of $Ax = b$. Then, thanks to the first point of Theorem 1, $x = Gb$ is a solution and $x = Gb = Gb + (I_n - GA)0$ implies that:

$\exists z \in \mathbb{R}^n$, such that $x = Gb + (I_n - GA)z$ (by taking $z = 0$).

2. *$x = Gb + (I_n - GA)z$ for some $z \in \mathbb{R}^n \Rightarrow x$ is a solution of $Ax = b$*

If $x = Gb + (I_n - GA)z$ for some $z \in \mathbb{R}^n$, then

$$Ax = A(Gb + (I_n - GA)z) = AGb + A(I_n - GA)z = AGb + (A - AGA)z$$

and as $A = AGA$, then we have $A - AGA = 0$ and $Ax = AGb$ which means $x = Gb$ is a solution of $Ax = b$.

A.3 The unicity of the generalized inverse of a positive semidefinite matrix

This section describes a method to obtain a generalized inverse matrix for any positive semidefinite matrix. p will define a non-zero integer in this section.

Definition 11. *A $(p \times p)$ square matrix A is said to be positive semidefinite if it satisfies these two conditions:*

1. *A is symmetric, i.e. $A' = A$.*
2. *$\forall x \in \mathbb{R}^p, x'Ax \geq 0$.*

We can also recall the property of the eigenvalues for a positive semidefinite matrix.

Proposition 2. *If A is a $p \times p$ positive semidefinite matrix, then all its eigenvalues are positive or equal to 0:*

If $\lambda_1, \lambda_2, ..., \lambda_p$ represent the eigenvalues of A, then $\forall j \in \{1, ..., p\}$, $\lambda_j \geq 0$.

Now we will see how we can obtain a generalized inverse of a positive semidefinite real matrix A. If A is a $p \times p$ positive semidefinite matrix, then A will be symmetric. If, in addition, A is a real matrix, then A is diagonalizable in such a way that there exists a diagonal $p \times p$ matrix D, of which the diagonal is composed of the p eigenvalues of A (positive or null), and an *orthogonal* $p \times p$ matrix F such that $A = FDF'$. If $\lambda_1, \lambda_2, \ldots, \lambda_p$ represent the p eigenvalues of A and let us assume now that A is of rank $r < p$, then we have $\lambda_1 > 0, \lambda_2 > 0, \ldots, \lambda_r > 0$ and $\lambda_{r+1} = ... = \lambda_p = 0$ and let us call D_r the $r \times r$ invertible matrix:

$$D_r = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_r \end{bmatrix}$$

Then, we can write the diagonal matrix D as:

$$D = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$$

Now, we will show that a generalized inverse of matrix A is $G = FD^+F'$ where:

$$D^+ = \begin{bmatrix} {D_r}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

with

$${D_r}^{-1} = \begin{bmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 1/\lambda_r \end{bmatrix}$$

Firstly, note that if we expand AGA, we have:

$$\begin{aligned} AGA &= (FDF')(FD^+F')(FDF') \\ &= FDF'FD^+F'FDF' \end{aligned}$$

But since F is orthogonal, we have $F'F = FF' = I_p$ and:

$$AGA \;\; = FDD^{+}DF'$$

It is obvious to see now that:

$$DD^{+} = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} {D_r}^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} D_r {D_r}^{-1} & 0 \\ 0 & 0 \end{bmatrix} = I_{p|r}$$

where

$$I_{p|r} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

We also have that:

$$I_{p|r}D = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} D = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} = D$$

So in fact:

$$AGA = FDD^{+}DF' = FDF' = A \tag{A.3}$$

Finally, from (A.3) and Definition 10 in Section A.1, we conclude that $G = FD^{+}F'$ is indeed a generalized inverse of A.

A.4 Decomposition of a positive semidefinite matrix A with reduced diagonal matrix

Let A be a $p \times p$ positive semidefinite matrix that is of rank $r < p$ (to be taken as one of our covariance matrices). This section recapitulates the fact that there exists a decomposition of A as the product of a $p \times r$ matrix F_r with an invertible $r \times r$ matrix D_r and multiplied again by the transpose of F_r. This property will be used later in order to be able to create a function in R Core Team (2013) that can simulate a random matrix from a singular normal distribution. Let us prove then that such a decomposition exists:

Theorem 2. *Let A be a $p \times p$ positive semidefinite matrix of rank $0 < r < p$. Then there exists a $p \times r$ matrix F_r and an invertible diagonal $r \times r$ matrix D_r such that:*

$$A = F_r D_r F_r'$$

Since A is a $p \times p$ positive semidefinite matrix, A is symmetric (and real) and then there exists an orthogonal $p \times p$ matrix F and a diagonal $p \times p$ matrix D such that:

$$A = FDF'$$

We denote by λ_1, λ_2, ..., λ_p the eigenvalues of A and v_1, v_2, ..., v_p their corresponding eigenvectors. We can choose D to be the diagonal matrix of which the elements of the diagonal are the eigenvalues of A put in a descending order: the r first elements of the diagonal will correspond to the non-zero eigenvalues of A: $\lambda_1 > 0$, ... , $\lambda_r > 0$ (the rest of the eigenvalues of A being zero: $\lambda_{r+1} = \cdots = \lambda_p = 0$). We have:

$$D = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$$

with

$$D_r = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_r \end{bmatrix}$$

F is the matrix of which the rows correspond to the eigenvectors of the respective eigenvalues contained in D. Hence, we have:

$$A = FDF' = [v_1, v_2, ..., v_p] \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} [v_1', v_2', ..., v_p']'$$

Then we have:

$$FD = [v_1, v_2, ..., v_p] \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} = [\lambda_1 v_1, \lambda_2 v_2, ..., \lambda_p v_p]$$

But since $\forall j \in \{r+1, ..., p\}$, $\lambda_j = 0$, we have:

$$FD = [\lambda_1 v_1, \lambda_2 v_2, ..., \lambda_r v_r, 0, ..., 0]$$

Finally, by multiplying FD by the transpose of F, we have:

$$FDF' = [\lambda_1 v_1, \lambda_2 v_2, ..., \lambda_r v_r, 0, ..., 0] \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_p \end{bmatrix} = \lambda_1 v_1 v'_1 + \lambda_2 v_2 v'_+ \cdots + \lambda_r v_r v'_r$$

Therefore, we have:

$$A = FDF' = \sum_{j=1}^{r} \lambda_j v_j v'_j \tag{A.4}$$

Now, let us call F_r the $p \times r$ matrix of which the rows are the r first rows of F:

$$F_r = [v_1, v_2, ..., v_r]$$

We will prove that $F_r D_r F'_r = \sum_{j=1}^{r} \lambda_j v_j v'_j$.

For that we need to derive the product $F_r D_r F'_r$ in terms of the eigenvectors and eigenvalues of A:

$$F_r D_r = [v_1, v_2, ..., v_r] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_r \end{bmatrix} = [\lambda_1 v_1, \lambda_2 v_2, ..., \lambda_r v_r]$$

Then by post-multiplying $F_r D_r$ by F'_r, we have:

$$F_r D_r F_r{}' = [\lambda_1 v_1, \lambda_2 v_2, ..., \lambda_r v_r] \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_r \end{bmatrix} = \lambda_1 v_1 v'_1 + \lambda_2 v_2 v'_2 + \cdots + \lambda_r v_r v'_r$$

Therefore:

$$F_r D_r F'_r = \sum_{j=1}^{r} \lambda_j v_j v'_j$$

Hence, from equation (A.4), we have:

$$A = FDF' = F_r D_r F_r{}' \tag{A.5}$$

Therefore for any positive semidefinite matrix A we can find a decomposition as in Theorem 2. Thanks to this result, we can implement a function in R that can simulate a matrix from a normal singular distribution. During the study of this chapter, we have created a special package in R to be able to simulate some very important singular distributions: the singular matrix normal distribution and singular Wishart and Inverse-Wishart distributions (derived from Wishart and used later on in Chapter 5).

Appendix B

The choice of the lag order

In this thesis, we assumed a lag order of 2 for our vector autoregressive process, in order to ligthen our algorithms and make our programs run faster. This part of the Appendix shows that the use of a lag order of $k = 2$ in order to estimate the cointegrating relations can be used no matter what the lag order of the VAR model is. The method used in this Appendix concerns the ones described in Chapter 3, assuming a non-singular prior distribution for the long-run relations matrix Π (in a non-time varying model).

We will respectively simulate a VAR(4) model and test the algorithm of Chapter 3 of this thesis, using a non-singular prior on Π and by assuming a lag order of $k = 2$.

We create a set of $p = 5$ time series such that they are difference stationary and they depend on four lag parameter matrices Γ_j. The vector autoregressive model is then constructed as follows:

$$x_t = \Gamma_1 x_{t-1} + \Gamma_2 x_{t-2} + \Gamma_3 x_{t-3} + \Gamma_4 x_{t-4} + \epsilon_t, \;\; \epsilon_t \sim N(0,1) \tag{B.1}$$

where:

$$\Gamma_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & -0.1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.2 & 0.5 & 0 \\ 0 & 0.1 & 0 & 0 & 0 \end{bmatrix} \quad \Gamma_2 = \begin{bmatrix} 0.5 & 0 & 0 & 0 & 0 \\ 0.1 & 1 & 0 & 0 & 0.1 \\ 0 & 0 & 0.5 & 0 & 0 \\ -1.0 & 0 & 0 & -0.1 & 0 \\ 0 & 0 & 0 & 0.3 & 0 \end{bmatrix}$$

$$\Gamma_3 = \begin{bmatrix} 0 & 0 & 0 & 0.1 & 0 \\ -0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.1 & 0.2 \\ 0.5 & 0 & 0.2 & 0.1 & 0 \\ 0 & 0.6 & 0 & 0 & 0.1 \end{bmatrix} \quad \Gamma_4 = \begin{bmatrix} 0.3 & 0 & 0 & 0.1 & 0 \\ -0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.1 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The data consists of $T = 700$ data points. An ADF test has been run on our five time series to see if they are effectively I(1). The following p-values have been found: 0.43, 0.70, 0.09, 0.92 and 0.79, which are all above 0.05, retaining then the null hypothesis of non-stationary time series. On the difference of those time series, we find that all the p-values are smaller than 0.01, proving that their differences are stationary.

We have: $\Pi = -(I_p - \Gamma_1 - \Gamma_2 - \Gamma_3 - \Gamma_4)$, and the following matrix for our model represents the cointegrating relations:

$$\Pi = \begin{bmatrix} -0.20 & 0 & 0 & 0.20 & 0.05 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.50 & -0.20 & 0.20 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.70 & 0 & 0.30 & -0.90 \end{bmatrix} \tag{B.2}$$

Since the matrix is of rank 3, the cointegration rank will again be 3 for this model (B.1).

The Gibbs sampler gives an estimation of the rank that is 3 on the dot.

After applying the steps of Section 3.3.7 to the mean of the last 2,000 matrices Π simulated using the Gibbs sampling algorithm, we obtain the following independent cointegrating relations:

$$\widehat{\beta}' = \begin{bmatrix} 1 & 0 & 0 & -0.9933 & -0.2538 \\ 0 & 1 & 0 & 0.4259 & -1.2681 \\ 0 & 0 & 1 & 0.4346 & -0.4012 \end{bmatrix} \tag{B.3}$$

Then, by applying the same steps, from the actual cointegrating matrix (B.2) of our model,

we have the following independent cointegrating relations:

$$\beta' = \begin{bmatrix} 1 & 0 & 0 & -1.0000 & -0.2500 \\ 0 & 1 & 0 & 0.4286 & -1.2857 \\ 0 & 0 & 1 & 0.4000 & -0.4000 \end{bmatrix} \tag{B.4}$$

The difference between β and $\widehat{\beta}$ is very small, and the percentage of the norm of that difference is equal to:

$$||\beta - \widehat{\beta}||/||\beta|| = 0.0438/1.9357 = 2.27\%$$

This small difference obtained between the actual relations and the results we obtain with a lag order of 2 confirms that we could indeed trust the results obtained in this thesis, and proceed by using our algorithms with a lag order 2. Even though the lag order may not be the same as in the true models along this thesis, the Markov chain Monte Carlo procedure indeed adjusts the parameters estimated.

Bibliography

Abu-Saman, A. (2012). Solution of linearly-dependent equations by generalized inverse of a matrix, *International Journal of Science and Emerging Technologies*, **4** (2).

Ang, A. and Bekaert, G. (2002). Regime switches in interest rates, *Journal of Business and Economic Statistics*, **20** (2), 163–182.

Baker, G. (1996). The Wall Street Journal, http://www.wsj.com.

Banerjee, A., Dolado, J., Galbraith, J. and Hendry, D. (1993). *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press. ISBN-13: 9780198288107.

Banerjee, S. and Roy, A. (2014). *Linear Algebra and Matrix Analysis for Statistics*, Texts in Statistical Science (1st ed.), Chapman and Hall/CRC. ISBN 9781420095388.

Bauwens, L. and Giot, A. (1998). A Gibbs sampling approach to cointegration, *Computational Statistics*, **13**, 339–368.

Bauwens, L. and Lubrano, M. (1996). Identification restrictions and posterior densities in cointegrated Gaussian VAR systems, *Advances in Econometrics*, **11**, 3–28.

Bierens, H. and Martins, L. (2010). Time-varying cointegration, *Econometric Theory*, **26**, 1453–1490.

Carter, C. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika*, **82**, 339–350.

Casella, G. and George, E. (1992). Explaining the Gibbs sampler, *The American Statistician*, **46** (3), 167–174.

Chao, J. and Phillips, P. (1999). Model selection in partially nonstationary vector autoregressive processes with reduced rank structure, *Journal of Econometrics*, **91**, 227–271.

Chintrakarn, P. and Herzer, D. (2012). More inequality, more crime? A panel cointegration analysis for the United States, *Economics Letters*, **116** (3), 389–391.

Choi, I. and Saikkonen, P. (2004). Testing linearity in cointegrating smooth transition regressions, *Econometrics Journal*, **7**, 341–365.

Cogley, T. (2005). How fast can the new economy grow? A Bayesian analysis of the evolution of trend growth, *Journal of Macroeconomics*, **27** (2), 179–207.

Cogley, T. and Sargent, T. (2001). Evolving post-World War II U.S. inflation dynamics, *NBER Macroeconomics Annual*, **16**, 331–373.

Cogley, T. and Sargent, T. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US, *Review of Economic Dynamics*, **8** (2), 262–302.

Conigliani, C. and Tancredi, A. (2009). A Bayesian model averaging approach for cost-effectiveness analyses, *Health Economics*, **18** (7), 807–821.

Corander, J. and Villani, M. (2004). Bayesian assessment of dimensionality in reduced rank regression, *Statistica Neerlandica*, **58** (3), 255–270.

Damghani, B., Welch, D., O'Malley, C. and Knights, S. (2012). The misleading value of measured correlation, *Wilmott*, **2012** (62), 64–73.

DeJong, D. (1992). Co-integration and trend-stationarity in macroeconomic time series, *Journal of Econometrics*, **52**, 347–370.

Díaz-García, J. and Gutiérrez-Jáimez, R. (1997). Proof of the conjectures of H. Uhlig on the singular multivariate beta and the jacobian of a certain matrix transformation, *The Annals of Statistics*, **25** (5), 2018–2023.

Díaz-García, J., Gutiérrez-Jáimez, R. and Gutiérrez-Sánchez, R. (2006). Distribution of the generalised inverse of a random matrix and its applications, *Journal of Statistical Planning and Inference*, **136** (1), 183–192.

Díaz-García, J., Gutiérrez-Jáimez, R. and Mardia, K. (1997). Wishart and Pseudo-Wishart distributions and some applications to shape theory, *Journal of Multivariate Analysis*, **63** (1), 73–87.

Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74** (366), 427–431.

Diebold, F. and Kilian, L. (2001). Measuring predictability: theory and macroeconomic applications, *Journal of Applied Econometrics*, **16** (6), 657–669.

Dorfman, J. (1994). A numerical Bayesian test for cointegration of AR processes, *Journal of Econometrics*, **66**, 289–324.

Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*, New York: Springer. ISBN 978-1-4614-6867-7.

Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ Integration, *Journal of Statistical Software*, **40** (8), 1–18.

Engle, R. and Granger, C. (1987). Co-integration and error correction: representation, estimation and testing, *Econometrica*, **55** (2), 251–276.

Fama, E. (1970). Efficient capital markets: a review of theory and empirical work, *Journal of Finance*, **25** (2), 383–417.

Fruewirth-Schnatter, S. (1994). Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering, *Statistics and Computing*, **4**, 259–269.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics, *Journal of Econometrics*, **75** (1), 121–146.

Gilks, W. R., Best, N. and Tan, K. K. C. (1995). Adaptative rejection Metropolis sampling within Gibbs sampling, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **44** (4), 455–472.

Granger, C. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, **16** (1), 121–130.

Granger, C. and Lee, H. (1991). *Economic structural change*, chap. An Introduction to Time-Varying Parameter Cointegration, Berlin, Heidelberg: Springer, pp. 139–157.

Granger, C. and Newbold, P. (1974). Spurious regressions in econometrics, *Journal of Econometrics*, **2** (2), 111–120.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82** (4), 711–732.

Gupta, A. and Nagar, D. (2000). *Matrix Variate Distributions*, CHAPMAN and HALL/CRC.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97–109.

Hausdorff, F. (1918). Dimension und äusseres Mass, *Mathematische Annalen*, **79**, 157–179.

Huerta, G. and West, M. (1999). Priors and component structures in autoregressive time series models, *Journal of the Royal Statistical Society*, **4**, 881–899.

Johansen, S. (1988). Statistical analysis of cointegrating vectors, *Journal of Economic Dynamics and Control*, **12** (2), 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, **59** (6), 1551–1580.

Johansen, S. (1996). Likelihood-based inference for cointegration of some non-stationary time series, *Time Series Models: In Econometrics, Finance and Other Fields*, 69–100.

Johansen, S. (1997). Likelihood-based Inference in cointegrated vector autoregressive models, *The Scandinavian Journal of Economics*, **99** (2), 351–354.

Johansen, S. (2005). The interpretation of cointegrating coefficients in the cointegrated vector autoregressive model, *Oxford Bulletin of Economics and Statistics*, **67** (1), 93–104.

Johansen, S. (2006). *T.C. Mills and K. Patterson (Ed.), The Handbook of Econometrics, Cointegration: an overview, Chapter 14, Volume 1. Theoretical Econometrics, Basingstoke: Palgrave Macmillan.*

Kaufmann, R. and Stern, D. (2002). Cointegration analysis of hemispheric temperature relations, *Journal of Geophysical Research*, **107** (D2), ACL 8–1–ACL 8–10.

Keynes, J. (1936). *The General Theory of Employment, Interest and Money*, Kypros Press.

Kleibergen, F. and van Dijk, H. (1994). On the shape of the likelihood/posterior in cointegration models, *Econometric Theory*, **10** (3), 514–551.

Kleibergen, F. and Paap, R. (2002). Prior, posteriors and Bayes factors for a Bayesian analysis of cointegration, *Journal of Econometrics*, **111**, 223–249.

Koop, G. (1991). Cointegration tests in present value relationships: A Bayesian look at the bivariate properties of stock prices and dividends, *Journal of Econometrics*, **49**, 105–140.

Koop, G. (1994). An objective Bayesian analysis of common stochastic trends in international stock prices and exchange rates, *Journal of Empirical Finance*, **1**, 343–364.

Koop, G., Leon-Gonzalez, R. and Strachan, R. (2011). Bayesian inference in a time varying cointegration model, *Journal of Econometrics*, **165**, 210–220.

Koop, G., Strachan, R., van Dijk, H. and Villani, M. (2006). *T.C. Mills and K. Patterson (Ed.), Bayesian Approaches to Cointegration Volume 1 Econometric Theory 1st ed.*

Koop, G. and Tobias, J. (2006). Semiparametric Bayesian inference in smooth coefficient models, *Journal of Econometrics*, **134** (1), 283–315.

Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, **54** (1-3), 159–178.

Luetkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*, Springer.

Martino, L., Read, J. and Luengo, D. (2015). Independent doubly adaptive rejection Metropolis sampling within Gibbs sampling, *IEEE Transactions on Signal Processing*, **63** (12), 3123–3138.

Meligkotsidou, L., Tzavalis, E. and Vrontos, I. (2014). A Bayesian method of distinguishing unit root from stationary processes based on panel data models with cross-sectional dependence, *Statistics and Computing*, **24** (3), 297–315.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines, *The Journal of Chemical Physics*, **21**, 1087–1091.

Muirhead, R. (1982). *Aspects of Multivariated Statistical Therory*, New York: John Wiley & Sons.

Ostergaard, J., Rahbek, A. and Ditlevsen, S. (2017). Oscillating systems with cointegrated phase processes, *Journal of Mathematical Biology*, 1–39.

Phillips, P. (1991). Optimal inference in cointegrated systems, *Econometrica*, **59** (2), 283–306.

Phillips, P. (1996). Econometric model determination, *Econometrica*, **64** (4), 763–812.

Phillips, P. and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika*, **75** (2), 335–46.

Pinheiro, J. and Bates, D. (1996). Unconstrained parametrizations for variance-covariance matrices, *Statistics and Computing*, **11**, 289–296.

Pourahmadi, M. (2000). Maximum likelihood estimation for generalized linear models for multivariate normal covariance matrices, *Biometrika*, **87**, 425–435.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy, *The Review of Economic Studies*, **72** (3), 821–852.

Puntanen, S., Styan, G. and Isotalo, J. (2011). *Full Rank Decomposition*, Springer.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.R-project.org/`

Rad, H., Low, R. and Faff, R. (2016). The profitability of pairs trading strategies: distance, cointegration and copula methods, *Quantitative Finance*, **16** (10), 1541–1558.

Roberts, G., Gelman, A. and Gilks, W. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms, *The Annals of Applied Probability*, **7** (1), 110–120.

Said, S. and Dickey, D. (1984). Testing for unit roots in autoregressive-moving averagemodels of unknown order, *Biometrika*, **71** (3), 599–607.

Schmidt, A. D. (2008). *Pairs Trading - A Cointegration Approach*, Ph.D. thesis, University of Sydney.

Schotman, P. and van Dijk, H. (1991). A Bayesian analysis of the unit root hypothesis in real exchange rates, *Journal of Econometrics*, **49**, 195–238.

Sims, C. (1988). Bayesian skepticism on unit root econometrics, *Journal of Economic Dynamics and Control*, **12** (2-3), 463–474.

Smith, M. and Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data, *Journal of the American Statistical Association*, **97**, 1141–1153.

Stillman, R. (1986). *Dow Jones Industrial Average: History and Role in an Investment Strategy*, Irwin Professional Publishing.

Stock, J. and Watson, M. (1996). Evidence on structural instability in macroeconomic time series relations, *Journal of Business and Economic Statistics*, **14** (1), 11–30.

Strachan, R. (2003). Valid Bayesian estimation of the cointegrating error correction model, *Journal of Business and Economic Statistics*, **21** (1), 185–195.

Strachan, R. and Inder, B. (2004). Bayesian analysis of the error correction model, *Journal of Econometrics*, **123**, 307–325.

Uhlig, H. (1994). On singular Wishart and singular multivariate Beta distributions, *The Annals of Statistics*, **22** (1), 395–405.

Villani, M. (2000). *Aspects of Bayesian Cointegration*, Ph.D. thesis, Stockholm University.

Villani, M. (2005). Bayesian reference analysis of cointegration, *Econometric Theory*, **21** (2), 326–357.

Warne, A. (2006). Bayesian inference in cointegrated VAR models with applications to the demand for Euro Area M3, *Working paper series*, (692).