Department of Economics

Identification and estimation of sources of common fluctuations: new methodologies and applications

Katarzyna Maciejowska

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Florence, May 2010

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Identification and estimation of sources of common fluctuations: new methodologies and applications

Katarzyna Maciejowska

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Jury Members:

Professor Helmut Lütkepohl, EUI, Supervisor
Professor Massimiliano Marcellino, EUI
Professor Joerg Breitung, University of Bonn
Professor George Kapetanios, Queen Mary University of London

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Contents

Acknowledgement

I wish to thank my supervisor Prof. Helmut Lütkepohl for his commitment and helpful comments and Profs. Massimilano Marcellino and Anindya Banerjee for inspiration and support of my research.

I would not have completed this thesis without the continual support and devotion of my husband, Arek.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Introduction

This thesis addresses the problem of how to identify and model sources of common fluctuations of economic variables. It is an interesting question not only for researchers but also for policy makers and other authorities. The literature presents two approaches. The first one is based on an assumption that the important structural shocks can be captured by a small set of macroeconomic variables. The most popular models used in this context are structural vector autoregression models (SVAR). The second approach follows from a belief that there exists a small number of factors that affect many economic processes. Therefore, it involves analysis of large data sets, with both time and cross-sectional dimensions large enough to describe the factor structure.

We dedicate the first part of the thesis to the problem of identification and estimation of structural shocks in small SVAR models. We follow the ideas of Rigobon (2003) and Lanne and Lütkepohl (2008), which show that the statistical property of the data may provide enough information to identify the structure of the model. The papers argue that a shift in the error covariance matrix allows for the estimation of the structural parameters of interest. The literature concentrates on models in which the shift is a result of a structural brake or a mixed distribution of errors.

In the first chapter, we discuss issues associated with the estimation of a SVAR model with a mixture of two normal distributions. We show that in this class of models, the likelihood function is unbounded and there exist spurious maxima that impede the Maximum Likelihood estimation. We discuss how these problems can be solved. Moreover, we illustrate the estimation problems with a Monte Carlo experiment. We investigate which estimation method and maximization algorithm is the most efficient and robust to the existence of spurious maximizers.

The second chapter is a result of a joint work with M. Lanne and H. Lütkepohl. In this paper, we allow for a more flexible class of data generating processes. It is assumed that the reduced form errors follow the Markov switching process and the error covariance matrix varies across states. We argue that this property is sufficient to identify structural shocks. The setup of the model is formulated and discussed and we show how it can be used for testing restrictions that are considered just-identifying in a traditional SVAR framework. The approach is illustrated with two empirical examples: a small model of US economy and a model of European/US interest rate linkages.

The second part of the thesis concentrates on an alternative approach, which is based on the analysis of large data sets. This third chapter discusses a problem of generalized factor models with both time and cross sectional dimensions increasing to infinity. So far, the literature considers only models with stationary or random walk factors. As many macroeconomic variables also have deterministic time trends, we consider it important to extend the methodology and allow for different types of deterministic components and higher order processes. In the presented paper, we show an estimation method for a model with different types of factors and derive the convergence rates and limiting distributions of the estimators. We show how asymptotic theory can be applied for testing if an observable variable is a common factor. We illustrate the theory with an empirical example: an analysis of the real activity of the US economy.

The SVAR and Factor models have been successfully combined into a Factor Augmented VAR model, where a typical small set of macroeconomic variables is extended by common factors estimated from a large panel i.e. Bernanke, Boivin and Eliasz (2005). We believe that models that merge both approaches will play an important role in the future of macroeconometric analysis.

Part I

Identification and estimation of SVAR models

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Chapter 1

Comparison of estimation methods of SVAR models with mixtures of two normal distributions - a Monte Carlo analysis

1.1 Introduction

Structural vector autoregressive (SVAR) models are widely used in applied macroeconomics. They allow for the estimation of structural shocks and impulse responses from empirical data and therefore, can be used to evaluate economic theory. However, this class of models requires additional information about the theoretical setup or the data in order to identify the structural parameters. A standard approach to obtain identifiability is to impose parameter constraints that can be justified by the economic theory. Unfortunately, there is no agreement on which of the identification schemes should be used and imposing just-identifying restrictions makes it impossible to empirically evaluate some of the underlying economic assumptions. The above critique raises the question of whether there is a property of the data instead of the economic theory that can be used to identify SVAR parameters. Rigobon (2003) shows that if there is a shift in the variance of the structural shocks it can provide enough information to identify the SVAR model. Lanne and Lütkepohl (2008) generalizes this approach and develops a test for the presence of a variance shift and for the stability of the correlation structure. This paper follows the specification of Lanne and Lütkepohl (2005), which assumes nonnormality of structural shocks rather then a discrete change in the variance. The residuals are allowed

to be distributed according to the mixture of two normal distributions and it is demonstrated how this property can be used to identify the parameters.

Scientific literature provides many works that discuss the issue of mixture models. Mixture models can be found both in economics and in other disciplines such as biology, medicine, engineering and marketing, among others. They were first used by biometrician Karl Pearson (1894), who analyzed a population of crabs and proved the existence of two subspecies in the examined sample. In the 1960s economists tried to use the ML approach to estimate the model parameters (Day (1969)). However, it was the EM algorithm described by Dempster, Laird and Rubin (1977) that significantly simplified the estimation procedure and therefore helped to popularize the mixture models.

The mixture models are also special cases of Markov switching (MS) models. A Markov process simplifies to a mixture distribution if diagonal elements of its transition matrix sum to one. Markov switching models are very flexible and can account for both nonliearities in the mean and heteroscedasticity. They are extensively used in econometrics (Kim and Nelson (1999), Sims and Zha (2006), Smith, Naik and Tsai (2006)), especially in business cycle analysis (Hamilton (1989), Goodwin (1993), Diebold and Rudebusch (1996), Kim and Nelson (1998)). They were popularized by the seminal paper Hamilton (1989), which discusses the estimation issues for univariate processes. The approach was extended to a multivariate case by Krolzig (1997).

An open question that still needs to be examined are small sample properties of mixture model estimators. This issue is of special interest when mixture models are applied in macroeconomic analysis because they are associated with a usage of relatively short time series. Therefore, the main scope of the paper is to evaluate the performance of different estimation methods and maximization algorithms in the context of SVAR models with mixtures of normal distributions, as proposed by Lanne and Lütkepohl (2005), and discuss the difficulties associated with the estimation process. Since the mixture models are special cases of MS models, we believe that our research also contributes to the discussion on estimation issues of MS models, especially in the context of structural analysis.

The paper is structured as follows. In Section 1.2, SVAR model with a mixture of two normal distributions is introduced and the identification issues are discussed. Estimation methods and optimization algorithms are considered in Section 1.3. In Section 1.4, a Monte Carlo experiment is described and results for different estimation methods and optimization algorithms are presented. Finally, conclusions are provided in Section 1.5.

1.2 SVAR models with a mixture of normal distributions

1.2.1 Model description

The literature discusses different types of SVAR models: A-model, B-model and AB-model (see Lütkepohl (2005)). The classification depends on the relationships the model attempt to describe, i.e., whether we are interested in the relations between the observable variables or responses to unobservable impulses. In this paper we will focus on the B-model that describes the direct, instantaneous effect of the structural shocks on the endogenous variables. In the B-model it is assumed that the forecast error ε is a linear function of the structural shock, u. The model can be written in the following way

$$y_t = A_0 + \sum_{i=1}^{p} A_i y_{t-i} + \varepsilon_t \tag{1.1}$$

where $\varepsilon_t = Bu_t$ and the variance-covariance matrices of structural and forecast errors are $\Sigma_u = I_k$ and $\Sigma_\varepsilon = BB'$, respectively.

In the setup, y_t is a $k \times 1$ vector of endogenous variables, ε_t is a $k \times 1$ vector of forecast errors and u_t is a $k \times 1$ vector of structural shocks with an identity covariance matrix $\Sigma_u = I_k$. A_0 is a $k \times 1$ vector of constants and $A_i, i = 1, ..., p$ are $k \times k$ matrices of the autoregressive parameters. B is a $k \times k$ nonsingular matrix that describes the transition mechanisms of the structural shocks u_t.

The structural VAR model has $k + p \cdot k^2 + k^2$ unknown parameters. The reduced form of the model (1.1) allows for estimation of only $k + p \cdot k^2 + k\,(k+1)\,/2$ parameters. In order to identify all structural parameters, an additional $k\,(k-1)\,/2$ linearly independent restrictions need to be imposed.

Lanne and Lütkepohl (2005) proposes solving the identification problem by making an assumption on the distribution of shocks. It is assumed that the structural shocks vector, u_t, has a mixed normal distribution. It means that

$$u_t \sim \begin{cases} N(0,I_k) & \text{with probability } \gamma \\ N(0,\Psi) & \text{with probability } 1-\gamma \end{cases}$$

where the variance-covariance matrix Ψ is diagonal. Under this specification, the unconditional variance of the structural shock is $\Sigma_u = \gamma I_k + (1-\gamma)\,\Psi$. The matrix Σ_u is no longer identity matrix but it is still diagonal. The diagonality of the matrix Σ_u ensures that the structural shocks are uncorrelated. Lanne and Lütkepohl (2005) proves that if all diagonal elements of the matrix Ψ are distinct then the structural parameters of the model are identifiable. The issue of identifiability will be discussed in more detail in Section 1.2.3.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

1.2.2 Density function of forecast errors

In order to analyze the properties of the model we need to derive the density function for the forecast errors. Since the errors, ε_t, are a linear combination of the structural shocks, u_t, then they also have a mixed normal distribution

$$\varepsilon_t \sim \begin{cases} N(0,BB') & \text{with probability } \gamma \\ N(0,B\Psi B') & \text{with probability } 1-\gamma \end{cases}$$

Therefore, the density function $f(\varepsilon_t; B, \Psi, \gamma)$ is given by

$$\begin{aligned} f(\varepsilon_t; B, \Psi, \gamma) &= \gamma (2\pi)^{-k/2} \det(BB')^{-1/2} \exp\left(-\frac{1}{2}\varepsilon_t' (BB')^{-1} \varepsilon_t\right) \\ &\quad + (1-\gamma)(2\pi)^{-k/2} \det(B\Psi B')^{-1/2} \exp\left(-\frac{1}{2}\varepsilon_t' (B\Psi B')^{-1} \varepsilon_t\right) \end{aligned} \tag{1.2}$$

The function is a sum of two components

$$f(\varepsilon_t; B, \Psi, \gamma) = \gamma f_1(\varepsilon_t; B) + (1-\gamma) f_2(\varepsilon_t; B, \Psi) \tag{1.3}$$

where

$$f_1(\varepsilon_t; B) = (2\pi)^{-k/2} \det(BB')^{-1/2} \exp\left(-\frac{1}{2}\varepsilon_t' (BB')^{-1} \varepsilon_t\right)$$

and

$$f_2(\varepsilon_t; B, \Psi) = (2\pi)^{-k/2} \det(B\Psi B')^{-1/2} \exp\left(-\frac{1}{2}\varepsilon_t' (B\Psi B')^{-1} \varepsilon_t\right)$$

Under the assumption of no time correlation of errors, the joint density can be written as follows

$$f(\varepsilon; B, \Psi, \gamma) = \prod_{t=1}^{T} f(\varepsilon_t; B, \Psi, \gamma)$$

with $\varepsilon = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$.

In further sections, for notational simplicity, $f(\varepsilon_t; \theta, \gamma)$ is used instead of $f(\varepsilon_t; B, \Psi, \gamma)$, where $\theta' = \{vec(B)' : diag(\Psi)'\}$.

1.2.3 Identification

There is a theoretical question whether it is possible to uniquely identify the parameters of SVAR models with the mixture of two normal distributions. In the literature there are papers that address the issue of parameters identification in different kinds of models. Following Rothenberg (1971), we can distinguish between locally and globally identifiable structures. Let us denote $f(\varepsilon; \delta)$ as a density function of a random variable ε for parameters $\delta \in \Delta$.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Definition 1 *A parameter point $\delta \in \Delta$ is said to be globally identifiable if there is no other $\tilde{\delta} \in \Delta$ such that $f\left(\varepsilon;\tilde{\delta}\right) = f(\varepsilon;\delta)$ for all ε.*

Definition 2 *A parameter point $\delta \in \Delta$ is said to be locally identifiable if there exists an open neighborhood of δ containing no other $\tilde{\delta}$ such that $f\left(\varepsilon;\tilde{\delta}\right) = f(\varepsilon;\delta)$ for all ε.*

In the case of standard mixture models, it is straightforward to see that they are not globally identifiable. One can always change the order of the mixture components without changing the overall distribution. This problem is known as the "label switching". In the simple mixture model, in which the density function is described by

$$f(\varepsilon;\theta,\gamma) = \sum_{i=1}^{n} \gamma_i f_i(\varepsilon;\theta_i)$$

where $\theta = \{\theta_1, ..., \theta_n\}$ is a set of mixture components parameters and $\gamma = \{\gamma_1, ..., \gamma_n\}$ is a set of mixing proportions, such that for all $i \in \{1, .., n\}$ $\gamma_i > 0$ and $\sum_{i=1}^{n} \gamma_i = 1$, "label switching" means that for any permutation of indices $k_1, ..., k_n$

$$f\left(\varepsilon;\tilde{\theta},\tilde{\gamma}\right) = \sum_{i=1}^{n} \gamma_{k_i} f_{k_i}(\varepsilon;\theta_{k_i}) = \sum_{i=1}^{n} \gamma_i f_i(\varepsilon;\theta_i) = f(\varepsilon;\theta,\gamma)$$

where $\tilde{\theta} = \{\theta_{k_i}, ..., \theta_{k_n}\}$ and $\tilde{\gamma} = \left\{\gamma_{k_1}, ..., \gamma_{k_n}\right\}$.

In the SVAR model with the mixture of two normal distributions, the error term ε_t follows (1.3). It means that the mixture components are defined by different parameter vectors. Thus, components cannot be simply flipped around by changing their order. However, for any B, Ψ and γ, there exist $\tilde{B} = B\Psi^{0.5}$, $\tilde{\Psi} = \Psi^{-1}$ and $\tilde{\gamma} = 1 - \gamma$ such that for all $\varepsilon \in R$ there is $f\left(\varepsilon;\tilde{B},\tilde{\Psi},\tilde{\gamma}\right) = f(\varepsilon;B,\Psi,\gamma)$. The proof can be found in Appendix 1.6.1.

An additional problem that arises from the specification of SVAR models is the identifiability of the matrices B and Ψ. It can be shown that one can change the order of columns of B and corresponding diagonal elements of Ψ without influencing the values of the likelihood function. Moreover, the columns of B can be multiplied by -1 and it will not affect the values of the density function.

There are no doubts that the parameters of the SVAR models with the mixture of two normal distributions are not globally identifiable. It was shown, however, by Lanne and Lütkepohl (2005) that under some mild conditions they may be locally identifiable. The necessary and sufficient condition for the local identification is that the diagonal elements of the matrix Ψ are all mutually different.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

1.3 Estimation methods

The problem of estimating parameters of mixture models has been a subject of a large body of literature. Redner and Walker (1984) and McLachlan and Peel (2000) provide a survey of both theoretical and empirical publications discussing the properties and applications of different types of estimators. Recently, due to the increase of computational efficiency, most of the research concentrates on the application of the maximum-likelihood method. As the functional form of the residual distribution in the mixture models is usually treated as known, ML seems to be a plausible approach.

In the presented work, two estimation methods will be used. First, the standard maximum likelihood estimation will be described. Second, a two steps quasi ML estimation, which allows for the estimation of the autoregressive and mixture parameters separately, will be presented. Finally, the properties of the ML estimators will be discussed.

1.3.1 Maximum Likelihood and two steps quasi Maximum Likelihood estimators

The maximum likelihood estimation method depends on the assumed functional form of the joint error distribution. In the SVAR model with the mixture of two normal densities, the p.d.f. of the forecast errors, ε_t, for a given period t is given by (1.2). Therefore, the value of the log-likelihood function $L(\theta, \gamma|\varepsilon_t)$ for the t-th error, ε_t, is

$$\begin{aligned} L(\theta,\gamma|\varepsilon_t) &= \ln\left(f(\varepsilon_t;\theta,\gamma)\right) \\ &= -\frac{k}{2}\ln(2\pi) \\ &\quad + \ln\left(\begin{array}{c} \gamma\det(BB')^{-1/2}\exp\left(-\frac{1}{2}\varepsilon_t'(BB')^{-1}\varepsilon_t\right) + \\ (1-\gamma)\det(B\Psi B')^{-1/2}\exp\left(-\frac{1}{2}\varepsilon_t'(B\Psi B')^{-1}\varepsilon_t\right) \end{array}\right) \end{aligned}$$

A constant term $-\frac{k}{2}\log(2\pi)$ will be omitted in further analysis. The joint log-likelihood is

$$\begin{aligned} L(\theta,\gamma|\varepsilon) &= \sum_{t=1}^{T} L(\theta,\gamma|\varepsilon_t) \\ &= \ln\left(f(\varepsilon_t;\theta,\gamma)\right) \end{aligned}$$

The maximization problem

$$\max_{\theta\in\Omega,\gamma\in(0,1)} L(\theta,\gamma|\varepsilon) = \max_{\theta\in\Omega,\gamma\in(0,1)} \sum_{t=1}^{T} \ln\left(f(\varepsilon_t;\theta,\gamma)\right)$$

where θ is a vector of parameters defined as before and

$$\Omega = \{\theta : \det(B) \neq 0, diag(\Psi) > 0\}$$

is a set of all possible parameter vectors, does not have a closed form solution and therefore iterative optimization procedures have to be used.

One step Maximum Likelihood

In this method one searches for the maximum of the log-likelihood function over both the autoregressive and mixture parameters. We can rewrite the model with the lag polynomial

$$A(L)y_t - A_0 = \varepsilon_t$$

where $A(L) = I_k - \sum_{i=1}^{p} A_i L^i$ and L is a lag operator, such that $L^i y_t = y_{t-i}$. A_0 is a $k \times 1$ vector of constants. Then the estimators $\hat{A}_0, \hat{A}_1, ..., \hat{A}_p, \hat{B}, \hat{\Psi}, \hat{\gamma}$ are chosen to maximize

$$L(\theta, \gamma, A|y) = \sum_{t=p}^{T} \ln f(A(L)y_t - A_0; \theta, \gamma)$$

where $A = (A_0, A_1, ..., A_p)$, $y = (y_1, y_2, ..., y_T)$ and $f(.; \theta, \gamma)$ is defined in (1.2).

Two steps quasi Maximum Likelihood

In this method the estimation procedure consists of two steps. Firstly, the autoregressive parameters are estimated with the LS or quasi ML method. Then the estimates of the residuals are computed according to the formula

$$\hat{e}_t = y_t - \left(\hat{A}_0 + \sum_{i=1}^{p} \hat{A}_i y_{t-i} \right)$$

Finally, the mixture of two normal distributions is fitted to the estimated residuals $\hat{e}_t$ with the ML method. Then parameters $\hat{B}, \hat{\Psi}, \hat{\gamma}$ are chosen to maximize

$$L(\theta, \gamma|\hat{e}) = \sum_{t=p}^{T} \ln f(\hat{e}_t; \theta, \gamma)$$

where $\hat{e} = (\hat{e}_p, \hat{e}_{p+1}, ..., \hat{e}_T)$ and $f(.)$ is defined as in (1.2).

This is a quasi ML method because it is conditional on the estimates of the estimates of the autoregressive parameters, which in principle differs form the true ones. Thus,

$$L(\theta, \gamma|\hat{e}) \neq L(\theta, \gamma|\varepsilon)$$

Fortunately, the autoregressive parameters can be consistently estimated with the LS or quasi ML method and therefore, the estimates of the mixture parameters $\hat{B}$, $\hat{\Psi}$ and $\hat{\gamma}$ converge to the true ones. This estimation method is however less efficient than the full Maximum Likelihood approach.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

1.3.2 Numerical maximization algorithms

As mentioned before, the ML problem does not have a closed form solution. Therefore, numerical maximization algorithms need to be used to obtain the ML estimates of the parameters. There exist general iterative procedures, such as Newton's methods, two steps quasi Newton's methods and conjugate gradient methods, which can be used in this context. There are, however, other methods that are more specific and thus more suitable for the mixture distributions models. One of them is the EM algorithm. It was formalized by Dempster, Laird and Rubin (1977) and designed for estimation problems with incomplete data. McLachlan and Krishnan (1997) provides a broad review of the literature dedicated to its theoretical and empirical properties.

EM algorithm

The estimation of the SVAR models with the mixture of distributions can be analyzed from the perspective of the incomplete data problem. Let us assume that the data generating process of the shocks ε_t is

$$\varepsilon_t \sim \begin{cases} N(0,BB') & \text{if } \mathrm{Z}_t = 1 \\ N(0,B\Psi B') & \text{if } \mathrm{Z}_t = 0 \end{cases}$$

where Z_t is an indicator variable. Then the density function of ε_t conditional on Z_t could be rewritten as follows

$$f\left(\varepsilon_t | Z_t; \theta\right) = f_1\left(\varepsilon_t; \theta\right)^{Z_t} f_2\left(\varepsilon_t; \theta\right)^{1-Z_t}$$

where $f_1\left(\varepsilon_t; \theta\right)$ and $f_2\left(\varepsilon_t; \theta\right)$ are defined in Section 2.2.

In the mixture model the mixing probabilities are assumed to be constant over time. It corresponds to the assumption

$$\begin{aligned} prob\left(Z_t = 1\right) &= \gamma \\ prob\left(Z_t = 0\right) &= 1 - \gamma \end{aligned}$$

Therefore, Z_t needs to have a Bernoulli distribution

$$g\left(Z_t; \gamma\right) = \gamma^{Z_t}\left(1-\gamma\right)^{1-Z_t}$$

The joint density function of ε_t and Z_t is given by

$$f_c\left(\varepsilon_t, Z_t; \theta, \gamma\right) = f_1\left(\varepsilon_t; \theta\right)^{Z_t} f_2\left(\varepsilon_t; \theta\right)^{1-Z_t} \gamma^{Z_t}\left(1-\gamma\right)^{1-Z_t}$$

and

$$\begin{aligned} \ln\left(f_c\left(\varepsilon_t, Z_t; \theta, \gamma\right)\right) &= Z_t\left\{\ln\left(\gamma\right) + \ln\left(f_1\left(\varepsilon_t; \theta\right)\right)\right\} \\ &\quad + \left(1 - Z_t\right)\left\{\ln\left(1-\gamma\right) + \ln\left(f_2\left(\varepsilon_t; \theta\right)\right)\right\} \end{aligned}$$

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

The complete-data log likelihood $L_c(\theta,\gamma|\varepsilon)$ (meaning that both ε_t and Z_t are assumed to be observable) can be written as follows

$$\begin{aligned} L_c(\theta,\gamma|\varepsilon) &= \sum_{t=1}^{T} L_c(\theta,\gamma|\varepsilon_t) \\ &= \sum_{t=1}^{T} \ln\left(f_c(\varepsilon_t, Z_t;\theta,\gamma)\right) \end{aligned}$$

Therefore,

$$\begin{aligned} L_c(\theta,\gamma|\varepsilon) &= \sum_{t=1}^{T} Z_t\left\{\ln(\gamma)+\ln\left(f_1(\varepsilon_t;\theta)\right)\right\} \\ &\quad + \sum_{t=1}^{T}(1-Z_t)\left\{\ln(1-\gamma)+\ln\left(f_2(\varepsilon_t;\theta)\right)\right\} \end{aligned}$$

The EM algorithm consists of two steps: E (computing the expectation of $L_c(\theta,\gamma|\varepsilon)$ conditional on the the observable data ε_t) and M (maximizing the expected $L_c(\theta,\gamma|\varepsilon)$ over the parameter space $\Omega\cup(0,1)$.

E - Step In this step the expected value of the complete-data log likelihood is computed. The expected value of the $L_c(\theta,\gamma|\varepsilon)$ conditional on the the observable data ε for an initial parameters vector θ_0 and γ_0 is given by $Q(\theta,\gamma;\theta_0,\gamma_0)$

$$\begin{aligned} Q(\theta,\gamma;\theta_0,\gamma_0) &= E\left(\sum_{t=1}^{T} Z_t\left\{\log(\gamma)+\log\left(f_1(\varepsilon_t;\theta)\right)\right\}|\varepsilon;\theta_0,\gamma_0\right) \\ &\quad + E\left(\sum_{t=1}^{T}(1-Z_t)\left\{\log(1-\gamma)+\log\left(f_2(\varepsilon_t;\theta)\right)\right\}|\varepsilon;\theta_0,\gamma_0\right) \\ &= \sum_{t=1}^{T} E(Z_t|\varepsilon;\theta_0,\gamma_0)\left\{\log(\gamma)+\log\left(f_1(\varepsilon_t;\theta)\right)\right\} \\ &\quad + \sum_{t=1}^{T} E(1-Z_t|\varepsilon;\theta_0,\gamma_0)\left\{\log(1-\gamma)+\log\left(f_2(\varepsilon_t;\theta)\right)\right\} \end{aligned}$$

Let us denote by $\tau_t(\theta_0,\gamma_0)$ an expected value of the indicator variable Z_t for the initial parameters values θ_0 and γ_0

$$\begin{aligned} \tau_t(\theta_0,\gamma_0) &= E(Z_t|\varepsilon;\theta_0,\gamma_0) \\ &= 0\cdot f(Z_t=0|\varepsilon;\theta_0,\gamma_0)+1\cdot f(Z_t=1|\varepsilon;\theta_0,\gamma_0) \\ &= f_c(\varepsilon_t,Z_t=1;\theta_0,\gamma_0)/f(\varepsilon_t;\theta_0) \\ &= \gamma_0 f_1(\varepsilon_t;\theta_0)/f(\varepsilon_t;\theta_0) \end{aligned}$$

Then

$$\begin{aligned} E\left(1-Z_t|\varepsilon;\theta_0,\gamma_0\right) &= 1-\gamma_0 f_1\left(\varepsilon_t;B_0\right)/f\left(\varepsilon_t;\theta_0\right) \\ &= 1-\tau_t\left(\theta_0,\gamma_0\right) \end{aligned}$$

Thus, $Q\left(\theta,\gamma;\theta_0,\gamma_0\right)$ takes the form

$$\begin{aligned} Q\left(\theta,\gamma;\theta_0,\gamma_0\right) &= \sum_{t=1}^{T}\tau_t\left(\theta_0,\gamma_0\right)\left\{\log\left(\gamma\right)+\log\left(f_1\left(\varepsilon_t;\theta\right)\right)\right\} \\ &\quad +\sum_{t=1}^{T}\left(1-\tau_t\left(\theta_0,\gamma_0\right)\right)\left\{\log\left(1-\gamma\right)+\log\left(f_2\left(\varepsilon_t;\theta\right)\right)\right\} \end{aligned}$$

M - Step In this step the new estimates of θ and γ are chosen to maximize $Q\left(\theta,\gamma;\theta_0,\gamma_0\right)$.

$$\left(\hat{\theta},\hat{\gamma}\right)=\arg\max_{\theta\in\Omega,\gamma\in(0,1)}Q\left(\theta,\gamma;\theta_0,\gamma_0\right)$$

The $Q\left(\theta,\gamma;\theta_0,\gamma_0\right)$ function can be decomposed into two parts

$$Q\left(\theta,\gamma;\theta_0,\gamma_0\right)=Q_1\left(\gamma;\theta_0,\gamma_0\right)+Q_2\left(\theta;\theta_0,\gamma_0\right)$$

such that

$$\begin{aligned} Q_1\left(\gamma;\theta_0,\gamma_0\right) &= \log\left(\gamma\right)\sum_{t=1}^{T}\tau_t\left(\theta_0,\gamma_0\right)+\log\left(1-\gamma\right)\sum_{t=1}^{T}\left\{1-\tau_t\left(\theta_0,\gamma_0\right)\right\} \\ &= \log\left(\gamma\right)\sum_{t=1}^{T}\tau_t\left(\theta_0,\gamma_0\right)+\log\left(1-\gamma\right)\left\{T-\sum_{t=1}^{T}\tau_t\left(\theta_0,\gamma_0\right)\right\} \end{aligned}$$

$$Q_2\left(\theta;\theta_0,\gamma_0\right)=\sum_{t=1}^{T}\tau_t\left(\theta_0,\gamma_0\right)\log\left(f_1\left(\varepsilon_t;\theta\right)\right)+\left(1-\tau_t\left(\theta_0,\gamma_0\right)\right)\log\left(f_2\left(\varepsilon_t;\theta\right)\right)$$

The first component depends only on the mixing proportions γ whereas the second one depends on θ. Consequently, the maximization problem can be solved by separately estimating the proportion parameter γ and the rest of the parameters θ. It can be easily shown that the $Q_1\left(\gamma;\theta_0,\gamma_0\right)$ is maximized by

$$\hat{\gamma}=\sum_{t=1}^{T}\tau_t\left(\theta_0,\gamma_0\right)/T$$

Finally,

$$\hat{\theta}=\arg\max_{\theta\in\Omega}Q_2\left(\theta;\theta_0,\gamma_0\right)$$

Iterations of the algorithm Once the new estimates of the parameters $\hat{\theta}$ and $\hat{\gamma}$ are obtained, the two steps E and M are repeated for $\theta_0 = \hat{\theta}$ and $\gamma_0 = \hat{\gamma}$. The algorithm is terminated when a stopping condition is fulfilled. There are two popular stopping rules

1. The algorithm is stopped when the value of the log-likelihood function does not change by more then δ

$$\left|\log L\left(\hat{\theta}, \hat{\gamma}|\varepsilon\right) - \log L\left(\theta_0, \gamma_0|\varepsilon\right)\right| \leq \delta$$

2. The algorithm is stopped when the parameters do not change much. It means that for some chosen δ

$$\left\|\bar{\theta} - \bar{\theta}_0\right\| \leq \delta$$

where $\|.\|$ denotes some norm and $\bar{\theta} = \left(\hat{\theta}', \hat{\gamma}\right)'$, $\bar{\theta}_0 = \left(\theta_0', \gamma_0\right)'$.

1.3.3 Problems with Maximum Likelihood estimation

The maximum likelihood estimators suffer from two problems: the likelihood function is unbounded and the parameters are not globally identified. The second issue was discussed before and due to local identifiability does not threaten the estimation process but influences the interpretation of the estimated parameters. The first one is much more serious and some modification of the estimation procedures need to be considered.

Unbounded Likelihood function

An example of an unbounded likelihood function for a mixture model was given by Kiefer and Wolfowitz (1956). Let us consider an univariate, mixture model with a shift in a variance

$$x_t \sim \begin{cases} N(\mu,1) & \text{with probability } 0.5 \\ N(\mu, \sigma^2) & \text{with probability } 0.5 \end{cases}$$

Then the density function for x_t is given by

$$\begin{aligned} f\left(x_t; \mu, \sigma\right) &= 0.5 \frac{1}{\left(2\pi\right)^{0.5}} \exp\left(-0.5\left(x_t - \mu\right)^2\right) \\ &\quad + 0.5 \frac{1}{\left(2\pi\right)^{0.5}} \frac{1}{\sigma} \exp\left(-0.5 \frac{\left(x_t - \mu\right)^2}{\sigma^2}\right) \end{aligned}$$

Let us assume that there is a finite number of observations $\{x_t\}$ and $\max_t |x_t - \mu| = m < \infty$. Suppose we choose $\mu = x_1$ and a sequence of standard deviations $\sigma_n \to 0$. Then, for all $x_t = \mu$, the density function diverges to infinity.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

$$\begin{aligned} f(x_t;\mu,\sigma_n) &= 0.5\frac{1}{(2\pi)^{0.5}}\exp\left(-0.5\left(x_t-\mu\right)^2\right) \\ &\quad +0.5\frac{1}{(2\pi)^{0.5}}\frac{1}{\sigma_n}\exp\left(-0.5\frac{(x_t-\mu)^2}{\sigma_n^2}\right) \\ &= 0.5\frac{1}{(2\pi)^{0.5}}+0.5\frac{1}{(2\pi)^{0.5}}\frac{1}{\sigma_n}\to\infty \end{aligned}$$

The density for $x_t \neq \mu$ is bounded away from zero

$$\begin{aligned} f(x_t;\mu,\sigma_n) &= 0.5\frac{1}{(2\pi)^{0.5}}\exp\left(-0.5\left(x_t-\mu\right)^2\right) \\ &\quad +0.5\frac{1}{(2\pi)^{0.5}}\frac{1}{\sigma_n}\exp\left(-0.5\frac{(x_t-\mu)^2}{\sigma_n^2}\right) \\ &\to 0.5\frac{1}{(2\pi)^{0.5}}\exp\left(-0.5\left(x_t-x_1\right)^2\right) \\ &\geq 0.5\frac{1}{(2\pi)^{0.5}}\exp\left(-0.5m^2\right)>0 \end{aligned}$$

Thus, $L(\mu,\sigma_n|x)=\prod_{t=1}^{T} f(x_t;\mu,\sigma_n)\to\infty$

The problem seems to be equally severe for the SVAR models with a mixture of two normal distributions. The density function for an error, ε_t, is given by the following formula

$$\begin{aligned} f(\varepsilon_t) &= \gamma(2\pi)^{-k/2}\det(B)^{-1}\exp\left(-\frac{1}{2}\left(B^{-1}\varepsilon_t\right)'B^{-1}\varepsilon_t\right)+ \\ &\quad (1-\gamma)(2\pi)^{-k/2}\det(\Psi)^{-1/2}\det(B)^{-1}\exp\left(-\frac{1}{2}\left(B^{-1}\varepsilon_t\right)'\Psi^{-1}B^{-1}\varepsilon_t\right) \end{aligned}$$

We can always find a matrix B such that $\det(B)<M_1<\infty$ and there exists a time index $s\in\{1,...,T\}$ such that the ith element of $b_s=B^{-1}\varepsilon_s$ is equal to zero, $b_{is}=\left[B^{-1}\varepsilon_s\right]_i=0$, for some $i\in\{1,...,k\}$. We can choose the sequence Ψ_n of diagonal, positive definite matrices that satisfies $\Psi_{ii}^n\to 0$ and $\Psi_{jj}^n>M_2>0$ for $j\neq i$. We know that

$$-\left(B^{-1}\varepsilon_t\right)'\Psi^{-1}B^{-1}\varepsilon_t=-\sum_{j\neq i}\frac{1}{\Psi_{jj}}b_{jt}^2-\frac{1}{\Psi_{ii}}b_{it}^2$$

For $t=s$

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

$$\frac{1}{\Psi_{ii}} b_{it}^2 = 0$$

Therefore,

$$-\left(B^{-1}\varepsilon_t\right)' \Psi^{-1} B^{-1}\varepsilon_t = -\sum_{j\neq i} \frac{1}{\Psi_{jj}^n} b_{jt}^2 > -\frac{1}{M_2} \sum_{j\neq i} b_{jt}^2 > -\infty$$

and

$$\exp\left(-\frac{1}{2}\left(B^{-1}\varepsilon_t\right)' \Psi_n^{-1} B^{-1}\varepsilon_t\right) \gg 0$$

Since

$$\det\left(\Psi_n\right) \to 0$$

then

$$\det\left(\Psi_n\right)^{-1/2} \det\left(B\right)^{-1} \exp\left(-\frac{1}{2}\left(B^{-1}\varepsilon_t\right)' \Psi^{-1} B^{-1}\varepsilon_t\right) \to \infty.$$

Thus, $f\left(\varepsilon_t\right) \to \infty$.

For $t \neq s$ the value of density function $f\left(\varepsilon_t\right)$ is bounded away from zero

$$f\left(\varepsilon_t\right) > \gamma\left(2\pi\right)^{-k/2} \det\left(B\right)^{-1} \exp\left(-\frac{1}{2}\left(B^{-1}\varepsilon_t\right)' B^{-1}\varepsilon_t\right) > 0$$

So $L\left(\theta, \gamma|\varepsilon\right) = \prod_{t=1}^{T} f\left(\varepsilon_t\right) \to \infty$. Therefore the likelihood function is unbounded.

The problem of an unbounded likelihood function rises some questions about the ML estimators.

What is the ML estimator for the unbounded likelihood function? When the likelihood function is unbounded then the global maximizer of the likelihood function does not exist. Therefore one can not talk about the ML estimator in the traditional sense (see McLachlan and Peel (2000) for some discussion). It does not mean, however, that there is no sequence of local maximizers with properties of consistency, efficiency and asymptotic normality. Redner and Walker (1984) provides the regularity conditions under which, for the class of locally identifiable mixtures, such a sequence exists. Moreover, when the parameter space is compact and contains the true parameters in its interior, the MLE is a point at which the likelihood obtains its largest local maximum.

How can the ML estimation procedure be improved? Hathaway (1985) proposes imposing a set of constraints (ensuring that the parameter space is compact and does not include singularity points) that allows for the consistent estimation of the parameters. In the case of univariate time series, the constraint is $\min_{i,j}\left(\sigma_i/\sigma_j\right) \geq c$ for some constant $c > 0$. In the multivariate case,

Hathaway (1985) proposes to constrain all of the characteristic roots of $\Sigma_i\Sigma_j^{-1}$ (for any $1 \leq i \neq j \leq k$) to be greater or equal to some minimum value $c > 0$. These kind of restrictions will lead to constrained (global) maximum-likelihood formulations which are strongly consistent (if they are satisfied by the true parameters). The main disadvantage of the approach is the arbitrary choice of the value of $c > 0$. It is particularly difficult, when there is no initial intuition about the data generating process and no information to base the guess on.

Some other forms of the constraints are discussed in the literature. For example, McLachlan and Peel (2000) proposes to limit the distance between the component generalized variances by restricting the ratio $|\Sigma_i| / |\Sigma_j|$ to be greater or equal to $c > 0$.

What can we do in the case of the SVAR models with mixture of two normal densities? One may want to impose similar constraints on the parameters in the case of the SVAR model with the mixture of two normal densities. There are, however, differences between the setup presented in this paper and one discussed typically in the literature, they are associated with the components variances. In the SVAR models, the variances are composed of two matrices: B and Ψ: $\Sigma_1 = BB'$ and $\Sigma_2 = B\Psi B'$. Thus

$$\Sigma_2\Sigma_1^{-1} = B\Psi B' \cdot B'^{-1}B^{-1} = B\Psi B^{-1}$$

Let us denote by $\lambda(A)$ a set of all eigenvalues of the square matrix A. Then

$$\lambda\left(\Sigma_2\Sigma_1^{-1}\right) = \lambda\left(B\Psi B^{-1}\right) = \lambda(\Psi) = diag(\Psi)$$

So the Hathaway constraints for the two components case are equivalent to the following

$$\begin{array}{l} 0 < c \leq \min_{i\in\{1,..,K\}} \Psi_{i,i} \\ \max_{i\in\{1,..,K\}} \Psi_{i,i} \leq 1/c < \infty \end{array} \tag{1.4}$$

How to treat the obtained results? How can we evaluate the local maximum we find? The mixture models suffer not only the problem of unbounded likelihood function but also the problem of spurious maximizers. Spurious maximizers are typically generated by a small group of observations, which are located close together (Day (1969)). They are characterized by a big relative difference between components variances. Thus imposing restrictions on the parameters may reduce the number of spurious maximizers. The minimum eigenvalue of the $\Sigma_i\Sigma_j^{-1}$ can also be used to evaluate the local maximizers of the unconstrained likelihood and to choose the most interesting one.

1.4 Monte Carlo Experiment

The purpose of the Monte Carlo experiment is to investigate how a choice of an estimation method and maximization algorithm influences estimates of the

parameters. The exercise helps to answer the question what is the cost of using the two steps quasi ML instead of ML method. If there are no significant differences, then the two steps quasi ML approach will be a very attractive from the practical point of view as it allows to reduce significantly the complexity of the problem. Other interesting issues are the ability of different maximization algorithms to find the true, rather than spurious, local maximizers and the robustness to the guesses of the initial parameter values.

1.4.1 Experimental design

In the experiment, two data generating processes are considered: VAR in levels and VECM, both with the mixture of two normal distributions. The VECM process

$$\Delta y_t = A_0 + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + Bu_t$$

can be represented as a VAR process

$$y_t = A_0 + \sum_{j=1}^{p} A_j y_{t-j} + Bu_t$$

where the relationship between the VECM and VAR parameters is described as follow:

$$\begin{aligned} A_1 &= \alpha\beta' + \Gamma_1 + I_k \\ A_2 &= \Gamma_2 - \Gamma_1 \\ &\vdots \\ A_{p-1} &= \Gamma_{p-1} - \Gamma_{p-2} \\ A_p &= -\Gamma_{p-1} \end{aligned}$$

Therefore, in both cases the data sets used in the research can be generated according to the VAR specification. It is assumed that u_t follows a mixture of two normal distributions $N(0, I)$ and $N(0, \Psi)$ with mixing proportions γ and $1-\gamma$ ($\gamma \in (0,1)$), respectively

For each type of data generating process, the Monte Carlo experiment consists of 1000 replications. In each replication ($i = 1, \ldots 1000$), a time series is generated according to the following algorithm :

1. For each replication i and time period t a variable Z_{it} is generated from the binomial distribution with $prob(Z_{it} = 1) = \gamma$ and $prob(Z_{it} = 0) = 1 - \gamma$. Firstly, we draw randomly v_{it} form the uniform distribution on the interval $[0, 1]$. Then the value of Z_{it} is assigned $Z_{it} = 1 \Leftrightarrow v_{it} \leq \gamma$, $Z_{it} = 0 \Leftrightarrow v_{it} > \gamma$.

2. Structural shocks u_{it} are generated according to the distribution $N(0, I)$ if $Z_{it} = 1$ and $N(0, \Psi)$ if $Z_{it} = 0$ for each time period t (or alternatively $u_{it} \sim N(0, I) \Leftrightarrow v_{it} \leq \gamma$, $u_{it} \sim N(0, \Psi) \Leftrightarrow v_{it} > \gamma$).

3. Time series $\{y_{it}\}$ are generated from the formula

$$y_{it} = A_0 + \sum_{j=1}^{p} A_p y_{i,t-j} + B u_{it}$$

under the assumption $y_{i0} = 0$.

4. The first 100 observation of y_{it} are dismissed to reduce the influence of the choice of the initial observations on the outcome.

Finally, parameters of the SVAR or SVECM model are estimated with two methods: ML and two steps quasi ML. In both estimation methods, four algorithms are used to search for the parameter values that maximize the likelihood function: three general maximization algorithms (BFGS, NEWTON and BHHH provided in the CML library in GAUSS) and the EM algorithm.

The outcomes, for each of the estimation methods and the maximization algorithms, are evaluated on the basis of:

- number of successful estimates (algorithm converges)
- ratio of estimates that satisfy the conditions (1.4) for $c = 0.01$
- mean and variance of the estimated parameters
- convergence to the true parameter values for increasing sample size
- sensitivity to choice of the initial values

1.4.2 Choice of parameters values

The Monte Carlo experiment was performed for three different lengths of the time series $T = 50, 150, 500$. Time dimensions $T = 50, 150$ correspond to lengths of time series used in the empirical analysis, whereas $T = 500$ captures the asymptotic behavior of examined estimators and maximization algorithms.

In both data generating processes, the residuals Bu_t were distributed according to the mixture of two normal distributions with the following parameter values:

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Psi = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} \tag{1.5}$$

Two different proportion parameters were considered. Firstly, the mixture proportion was set to $\gamma = 0.5$, thus Bu_t was equally often distributed according to $N(0, BB')$ as to $N(0, B\Psi B')$. Finally $\gamma = 0.8$, which means that the second

component was much more rarely observable. It was expected that the choice of γ would influence the small sample properties of the estimators in three ways: by effecting a rate of successful estimates, a frequency of choosing the true, rather then spurious, maximizers and efficiency (measured by estimator variance).

Structural Vector Autoregressive Model (SVAR)

In the first part of the experiment data was generated according to the VAR model with the order of autoregression $p = 1$.

$$y_t = A_0 + A_1 y_{t-1} + Bu_t \tag{1.6}$$

The autoregressive parameters were chosen to ensure that the process y_t was stationary

$$A_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, A = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \tag{1.7}$$

Structural Vector Error Correction Model (SVECM)

The order of autoregression is set as $p = 2$ and the model takes the following form

$$\Delta y_t = A_0 + \alpha\beta' y_{t-1} + \Gamma \Delta y_{t-1} + Bu_t \tag{1.8}$$

The parameters of the SVECM model were chosen to ensure that the process is well defined[1]

$$\begin{aligned} \alpha &= \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}, \beta = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ A_0 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Gamma = \begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix} \end{aligned} \tag{1.9}$$

1.4.3 Results

Ratios of successful estimates Tables 1.1 and 1.2 present ratios of successful estimates for the VAR model, which are computed as the number of the outcomes with nonsingular covariance matrix and $0 < \gamma < 1$, divided by

[1] Let us denote by $C(z)$ the following polynomial

$$C(z) = (1 - z) I_k - \alpha\beta' z - \sum_{i=1}^{p-1} (1 - z) z\Gamma_i$$

Then, the VECM process is well defined if the following conditions hold

1. $\det(C(z)) = 0 \Rightarrow |z| \geq 1$
2. The number of unit roots $z = 1$, is exactly $k - r$, where $r = rk(\alpha) = rk(\beta)$

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

the total number of Monte Carlo iterations. Results indicate that the general maximization algorithms suffer many problems when estimating model parameters. More frequently, the parameters converge to singularity points or end up on the boundaries ($\gamma = 0$ or 1). This unwanted behavior is the strongest for the short time series ($T = 50$), when the ratios vary between $10\% - 60\%$ for algorithms that start with the true parameters values and $5\% - 35\%$ when they begin with false parameters values. For long time series ($T = 500$), the ratios are $70\% - 90\%$ and $30\% - 80\%$ respectively. In practice, we can expect the second case to occur more often and therefore, the results question the usage of this kind of algorithms. The EM algorithm outperforms the rest of algorithms in terms of the number of successful estimates. It converges to local maxima in almost all cases. Its disadvantage is, however, a very slow rate of convergence and lengthy time of computation (for more details see Redner and Walker (1984), McLachlan and Krishnan (1997)).

Tables 1.11 and 1.12 summarize the ratios of successful estimates[2] for the VECM model. For the two steps quasi ML method, they are qualitatively similar to those obtained in the VAR experiment. When the estimation procedures are initiated at true parameter values, the general maximization algorithms (NEWTON and BFGS) converge in $15-60\%$ cases for short time series $T = 50$, compared with 90% for the EM algorithm. As the time dimension increases, differences between algorithms decrease and the ratios for general maximization algorithms reach almost 100%. When the estimation begins with parameter values that differ from the true ones, the ratios of successful estimates for the BFGS do not exceed 35% for all time lengths ($T = 50, 500$), whereas the NEWTON algorithm converges in $30 - 90\%$ cases depending on the time dimension. Both general maximization algorithms perform significantly worse then the EM algorithm, for which the rate of convergence is close to 100%.

When the ML method is considered, there appears to be more differences between the VAR and VECM experiments. The general maximization algorithms converge in around $20-30\%$ of the cases for $T = 50$ and $80-100\%$ of the cases for $T = 500$. The EM algorithm, however, does not perform significantly better and converges only in 40% of cases for $T = 50$ and 95% of cases for $T = 500$. These results indicate that the complexity of the estimation problem influences significantly the chances of successful convergence.

Finally, comparisons of different maximization algorithms bring two conclusions. Firstly, there are algorithms, such as BHHH[3], very sensitive to the length of the time series. For $T = 50$, it falls far behind the BFGS and NEWTON algorithms. Secondly, BFGS is more frequently successful than the NEWTON algorithm when the initial guesses are close to the true parameters. The difference seems significant especially for very short time series. The results show, however, that the NEWTON algorithm is much more robust to the initial guesses of the parameters. Thirdly, the ratios of successful estimates and the true local maximizers hardly depend on the number of observations.

[2] As in the VAR experiment the BHHH algorithm performes much worse then other algorithms, it is ommited in futher research.

[3] Comparison based on the VAR experiment

It is interesting to compare the results of ML and two steps quasi ML methods. It appears that the two steps quasi ML method leads more often to the successful estimates and to the true maximizers rather then the spurious ones. These preliminary results can not fully support the choice of this method in empirical applications, as the precision of estimates needs to be taken into account. However, it already indicates the advantages of simplifying the estimation problem.

Autoregressive (VAR and VECM) parameters The comparison of the parameter estimates is based on the outcomes of the BFGS algorithm[4]. For all the two steps procedures, regardless of the maximization algorithm, the autoregressive parameters were estimated in the same way. Therefore, there is no need to compare results between the algorithms. Tables 1.5, 1.6, 1.15 and 1.16 present the means and the variances of the estimators for VAR and VECM models respectively. The outcomes satisfy condition (1.4) and are presented for the ML and two steps quasi ML separately. It is worth emphasizing that both methods produce very similar results. They confirm the consistency of the estimators, hence in all considered cases the mean converges to true parameter values and the variance decreases[5].

Mixture parameters Firstly, the estimates of the mixing parameters are compared on the basis of a ML with a BFGS maximization algorithm. Their properties (mean and the variance) are summarized in the Tables 1.7 and 1.17. The outcomes are less satisfying then in the autoregressive parameters case, but still show the consistency as the mean converges to the true parameter values and the variance decreases. It may be noticed that most of the problems arise while estimating the matrix Ψ. The biggest of the diagonal elements is estimated very imprecisely (its variance across Monte Carlo iterations reaches 313.04 for $T = 50$ for VAR and 331.26 for VECM model) and thereby influences the estimates of the rest of parameters.

Secondly, the results for three estimation procedures: a ML with BFGS (called M1), a two steps quasi ML with BFGS (called M2) and a two steps quasi ML with an EM (called EM2) are compared. The outcomes for the mixing proportion $\gamma = 0.5$ are illustrated in the Figures 1.1 and 1.2. It shows that the two steps quasi ML method with EM algorithm is the most precise in estimating the crucial Ψ matrix (when both the mean and the variance of the estimators are taken into account). For other mixture parameters, the outcomes are comparable across all three procedures (for more details see Tables 1.8-1.9 and Tables 1.18-1.19).

[4]Results for other maximization algorithms are very simmilar and therefore they are not discussed in details.

[5]The t-ratios mean and variacne were also computed and they confirm good properties of the estimators (converge to the first two moments of N(0,1)). Tables that summarize the t-ratios are available upon request

Spurious maximizers The importance of the spurious maximizers problem is illustrated by the results in Table 1.10. It summarizes the mean and the variance of the VAR and mixture parameters estimators for the cases in which the condition (1.4) is not satisfied. For $T = 50$, the mean of Ψ_2 estimators reaches almost 5000 and decreases to 2936 for $T = 150$. It means that in some cases the estimation procedures produce very unrealistic results which are characterized by high values of $\hat{\Psi}$ and low values of mixing proportion estimators (mean of $\hat{\gamma}$ was 0.193 and 0.117 for $T = 50, 150$ respectively).

The autoregressive parameters estimators were not affected by the existence of the spurious maximizers. Even when the mixing parameters were estimated incorrectly, they were still similar to the results for cases in which (1.4) is satisfied and converged to true parameter values as the sample size increases. It suggests that the estimators of autoregressive parameters are robust to the choice of a local maximizer.

As previously discussed algorithms may converge to the spurious maximizers rather then to the true ones. To disregard these cases, the condition (1.4) was checked for every estimate. Tables 1.3, 1.4, 1.13 and 1.14 summarize the ratios of the number of true local maximizers to the number of successful estimates. The results show that the ratio increases with the length of the times series. For $T = 50$, it starts from 66% to 84%, whereas for $T = 500$ all the results exceed 99%. Unfortunately, the low ratio for short time series means that when the macroeconomic time series are used it may be expected that the spurious maximizers will arise quite often.

1.5 Conclusions

In this paper, we describe and discuss issues associated with an estimation of structural VAR models with mixtures of two normal distributions. The main theoretical difficulties that arise are a lack of global identifiability of parameters and an unbounded likelihood function. The first issue can be easily overcome because, under some mild restrictions, the parameters are locally identifiable and therefore, a ML estimation method can be applied. The second problem requires a new definition of a ML estimator because a global maximum of a likelihood function does not exist. Moreover, the likelihood function has many spurious local maxima, which make it difficult to find the proper ML estimates. We present how the issue is solved in the literature and adopt this approach to the SVAR models with a mixture distribution.

Finally, we perform a Monte Carlo experiment that compares different estimation methods and maximization algorithms. The outcomes indicate that there are no significant differences in the efficiency between the two discussed estimation methods: ML and two steps quasi ML. This result favours the two steps method as it is simpler and less computationally demanding. Next, we compare the properties of different maximization algorithms. The general maximization algorithms seem to perform worse then the EM algorithm. It is more frequent that they are not able to produce any results or lead to spurious maxi-

mizers. Estimates based on these methods vary more across the MC iterations, particularly for short time series. The differences between these two types of algorithms become negligible for long time series $T = 500$, when the ratio of successful estimations and the moments of the obtained estimators equalize. The main disadvantages of the EM algorithm are difficulties with computing the variance of the estimators[6] and the lengthy time of computations.

The experiment confirms that spurious maximizers are one of the crucial problems when estimating the parameters of SVAR models with the mixture of normal distributions. It happens that the estimates, which constitute local maxima of the likelihood function, are produced by a small group of observations with a low variance. Therefore, they give a high value of the likelihood function but do not represent a ML estimate with its statistical properties. The existence of spurious maximizers threatens the estimates of the mixing parameters but does not affect the estimates of the autoregressive parameters.

1.6 Appendix

1.6.1 "Label Switching"

We will show that for $\tilde{B} = B\Psi^{0.5}$, $\tilde{\Psi} = \Psi^{-1}$ and $\tilde{\gamma} = 1 - \gamma$ and any $\varepsilon \in R$ the following equality holds

$$f\left(\varepsilon; \tilde{B}, \tilde{\Psi}, \tilde{\gamma}\right) = f\left(\varepsilon; B, \Psi, \gamma\right)$$

The density function $f\left(\varepsilon; B, \Psi, \gamma\right)$ consists of two components

$$\begin{aligned} f\left(\varepsilon; B, \Psi, \gamma\right) &= \gamma \det\left(BB'\right)^{-0.5} \exp\left(-0.5\varepsilon'\left(BB'\right)^{-1}\varepsilon\right) \\ &\quad + (1-\gamma)\det\left(B\Psi B'\right)^{-0.5} \exp\left(-0.5\varepsilon'\left(B\Psi B'\right)^{-1}\varepsilon\right) \\ &= f_1\left(\varepsilon\right) + f_2\left(\varepsilon\right) \end{aligned}$$

Lets $\tilde{f}_1\left(\varepsilon\right)$ and $\tilde{f}_2\left(\varepsilon\right)$ denote the components of the density function computed for the new parameters vectors $\tilde{B}$, $\tilde{\Psi}$ and $\tilde{\gamma}$. Then the first component $\tilde{f}_1\left(\varepsilon\right) = f_2\left(\varepsilon\right)$

$$\begin{aligned} \tilde{f}_1\left(\varepsilon\right) &= \tilde{\gamma}\det\left(\tilde{B}\tilde{B}'\right)^{-0.5} \exp\left(-0.5\varepsilon'\left(\tilde{B}\tilde{B}'\right)^{-1}\varepsilon\right) \\ &= (1-\gamma)\det\left(B\Psi^{0.5}\Psi'^{0.5}B'\right)^{-0.5} \exp\left(-0.5\varepsilon'\left(B\Psi^{0.5}\Psi'^{0.5}B'\right)^{-1}\varepsilon\right) \\ &= (1-\gamma)\det\left(B\Psi B'\right)^{-0.5} \exp\left(-0.5\varepsilon'\left(B\Psi B'\right)^{-1}\varepsilon\right) \\ &= f_2\left(\varepsilon\right) \end{aligned}$$

[6] To estimate asymptotic variance of the parameters some modification of the algorithm need to be introduced.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

and the second one $\tilde{f}_2(\varepsilon) = f_1(\varepsilon)$

$$\begin{aligned}
\tilde{f}_2(\varepsilon) &= (1-\tilde{\gamma}) \det\left(\tilde{B}\tilde{\Psi}\tilde{B}'\right)^{-0.5} \exp\left(-0.5\varepsilon'\left(\tilde{B}\tilde{\Psi}\tilde{B}'\right)^{-1}\varepsilon\right) \\
&= \gamma \det\left(B\Psi^{0.5}\Psi^{-1}\Psi'^{0.5}B'\right)^{-0.5} \exp\left(-0.5\varepsilon'\left(B\Psi^{0.5}\Psi^{-1}\Psi'^{0.5}B'\right)^{-1}\varepsilon\right) \\
&= \gamma \det\left(BB'\right)^{-0.5} \exp\left(-0.5\varepsilon'\left(BB'\right)^{-1}\varepsilon\right) \\
&= f_1(\varepsilon)
\end{aligned}$$

Finally,

$$\begin{aligned}
f\left(\varepsilon; \tilde{B}, \tilde{\Psi}, \tilde{\gamma}\right) &= \tilde{f}_1(\varepsilon) + \tilde{f}_2(\varepsilon) \\
&= f_2(\varepsilon) + f_1(\varepsilon) \\
&= f(\varepsilon; B, \Psi, \gamma)
\end{aligned}$$

1.7 Results: SVAR

Table 1.1: VAR. Ratio of successful estimates, algorithms initiated with the true parameters values.

		ML				two steps quasi ML			
γ	T	BFGS	NEWTON	BHHH	EM	BFGS	NEWTON	BHHH	EM
0.5	50	0.592	0.262	0.102	1.000	0.625	0.334	0.17	1.00
	150	0.896	0.515	0.611	0.996	0.882	0.498	0.583	0.994
	500	0.995	0.734	0.972	0.992	0.992	0.735	0.969	0.992
0.8	50	0.384	0.165	0.016	0.998	0.410	0.186	0.023	1.00
	150	0.758	0.403	0.230	0.993	0.733	0.415	0.228	0.992
	500	0.979	0.635	0.749	0.990	0.976	0.647	0.757	0.919

NOTE: Two methods are considered: Maximum Likelihood and two steps quasi Maximum Likelihood. For each estimation method four maximization algorithms are evaluated: BFGS, NEWTON, BHHH and EM. The data generating process is described by (1.5), (1.6) and (1.7). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Table 1.2: VAR. Ratio of successful estimates, algorithms not initiated with the true parameters values.

		ML				two steps quasi ML			
γ	T	BFGS	NEWTON	BHHH	EM	BFGS	NEWTON	BHHH	EM
0.5	50	0.212	0.287	0.060	1.000	0.242	0.344	0.123	0.999
	500	0.306	0.727	0.768	0.992	0.275	0.728	0.660	0.989
0.8	50	0.161	0.272	0.058	0.998	0.160	0.371	0.046	0.998
	500	0.584	0.822	0.628	0.994	0.385	0.820		0.990

NOTE: Two methods are considered: Maximum Likelihood and two steps quasi Maximum Likelihood. For each estimation method four maximization algorithms are evaluated: BFGS, NEWTON, BHHH and EM. The data generating process is described by (1.5), (1.6) and (1.7). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.3: VAR. Ratio of successful estimates that satisfy condition (1.4) for $c = 0.01$ to all successful estimates, algorithms initiated with the true parameters values.

		ML				two steps quasi ML			
γ	T	BFGS	NEWTON	BHHH	EM	BFGS	NEWTON	BHHH	EM
0.5	50	0.775	0.786	0.863	0.812	0.913	0.904	0.935	0.946
	150	0.948	0.940	0.957	0.948	0.984	0.970	0.992	0.989
	500	0.998	0.997	0.998	0.998	0.999	0.999	0.998	1.00
0.8	50	0.930	0.915	0.875	0.903	0.971	0.962	0.956	0.919
	150	0.991	0.985	1.00	0.75	0.999	0.995	1.00	0.986
	500	1.00	1.00	1.00	0.999	1.00	1.00	1.00	1.00

NOTE: Two methods are considered: Maximum Likelihood and two steps quasi Maximum Likelihood. For each estimation method four maximization algorithms are evaluated: BFGS, NEWTON, BHHH and EM. The data generating process is described by (1.5), (1.6) and (1.7). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Table 1.4: VAR. Ratio of successful estimates that satisfy condition (1.4) for $c = 0.01$ to all successful estimates, algorithms not initiated with the true parameters values.

		ML				two steps quasi ML			
γ	T	BFGS	NEWTON	BHHH	EM	BFGS	NEWTON	BHHH	EM
0.5	50	0.901	0.840	0.800	0.749	0.967	0.936	1.00	0.930
	500	0.997	1.00	0.996	0.999	0.996	1.00	0.997	0.999
0.8	50	0.969	0.893	0.810	0.850	0.962	0.921	0.956	0.944
	500	1.00	0.998	0.995	0.999	1.00	0.999		1.00

NOTE: Two methods are considered: Maximum Likelihood and two steps quasi Maximum Likelihood. For each estimation method four maximization algorithms are evaluated: BFGS, NEWTON, BHHH and EM. The data generating process is described by (1.5), (1.6) and (1.7). T and γ denote the length of the sample and a mixing proportion parameter, respectively.

Table 1.5: VAR. The mean and the variance of the autoregressive parameters estimates for the two steps quasi ML method (BFGS algorithm iniciated with the true parameters values).

Parameters		$A_1^{(0)}$	$A_2^{(0)}$	A_{11}	A_{21}	A_{12}	A_{22}
True values		0	0	0.5	0	0	0.5
γ	T	Mean					
0.5	50	−0.0009	0.0090	0.4385	0.0066	−0.0056	0.4440
	150	−0.0051	−0.0035	0.4794	−0.0015	−0.0022	0.4803
	500	−0.0014	−0.0010	0.4937	−0.0029	−0.0007	0.4940
0.8	50	−0.0058	−0.0085	0.4434	−0.0122	−0.0026	0.4510
	150	−0.0006	0.0009	0.4774	−0.0023	−0.0001	0.4782
	500	0.0006	0.0021	0.4921	−0.0006	−0.0002	0.4935
		Variance					
0.5	50	0.0247	0.0855	0.0152	0.0529	0.0055	0.0156
	150	0.0076	0.0215	0.0046	0.0168	0.0018	0.0049
	500	0.0021	0.0059	0.0015	0.0045	0.0005	0.0016
0.8	50	0.0269	0.0481	0.0171	0.0366	0.0099	0.0129
	150	0.0071	0.0132	0.0053	0.0102	0.0029	0.0052
	500	0.0018	0.0039	0.0015	0.0027	0.0008	0.0016

NOTE: The data generating process is described by (1.5), (1.6) and (1.7). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.6: VAR. The mean and the variance of the autoregressive parameters estimates for the ML method (BFGS algorithm iniciated with the true parameters values).

Parameters		$A_1^{(0)}$	$A_2^{(0)}$	A_{11}	A_{21}	A_{12}	A_{22}
True values		0	0	0.5	0	0	0.5
γ	T	Mean					
0.5	50	0.0025	−0.0056	0.4492	−0.0176	−0.0054	0.4535
	150	−0.0062	−0.0010	0.4806	0.0000	−0.0030	0.4851
	500	−0.0014	−0.0007	0.4936	−0.0043	−0.0006	0.4953
0.8	50	0.0030	−0.0080	0.4429	−0.0196	−0.0069	0.4613
	150	0.0000	0.0003	0.4783	0.0010	−0.0003	0.4832
	500	0.0008	0.0015	0.4919	−0.0010	−0.0001	0.4947
		Variance					
0.5	50	0.0285	0.099	0.0175	0.0602	0.0061	0.0190
	150	0.0082	0.0186	0.0048	0.0157	0.0018	0.0049
	500	0.0021	0.0051	0.0016	0.0038	0.0005	0.0014
0.8	50	0.0358	0.0448	0.0184	0.0410	0.0110	0.0164
	150	0.0074	0.0114	0.0053	0.0084	0.0030	0.0040
	500	0.0019	0.0032	0.0015	0.0021	0.0009	0.0013

NOTE: The data generating process is described by (1.5), (1.6) and (1.7). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Table 1.7: VAR. The mean and the variance of the mixing parameter estimates for the ML method (BFGS algorithm iniciated with the true parameters values).

Parameters		B_{11}	B_{21}	B_{12}	B_{22}	Ψ_1	Ψ_2	γ
True values		1	0	0	1	1	5	0.5/0.8
γ	T	Mean						
0.5	50	0.9696	0.0041	−0.0008	0.6739	1.1309	18.541	0.5183
	150	1.0072	−0.0006	0.0000	0.8503	1.0355	10.008	0.4939
	500	1.0054	−0.0011	−0.0005	0.9607	0.9791	6.0943	0.5031
0.8	50	0.9413	−0.0131	−0.0071	0.6328	1.2300	15.284	0.6339
	150	0.9816	−0.0222	0.0090	0.8626	0.9604	7.2953	0.7134
	500	0.9958	−0.0010	0.0016	0.9635	0.9544	5.4723	0.7717
		Variance						
0.5	50	0.0716	0.3132	0.0306	0.0862	4.2206	340.36	0.0353
	150	0.0284	0.1302	0.0147	0.0711	0.4660	120.44	0.0356
	500	0.0008	0.0388	0.0037	0.0314	0.1051	12.934	0.0181
0.8	50	0.0540	0.1640	0.0337	0.0533	6.7184	223.14	0.0346
	150	0.0153	0.0679	0.0187	0.0343	0.4133	19.281	0.0299
	500	0.0030	0.0181	0.0067	0.0088	0.1301	1.7017	0.0110

NOTE: The data generating process is described by (1.5), (1.6) and (1.7). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.8: VAR. The mean and the variance of the mixing parameter estimates for the two steps quasi ML method (BFGS algorithm iniciated with the true parameters values).

Parameters		B_{11}	B_{21}	B_{12}	B_{22}	Ψ_1	Ψ_2	γ
True values		1	0	0	1	1	5	0.5/0.8
γ	T	Mean						
0.5	50	0.9791	0.0151	−0.0016	0.7006	1.0384	16.008	0.4562
	150	1.0021	0.0064	−0.003	0.8605	0.9686	9.4433	0.4798
	500	1.0047	−0.0001	−0.0008	0.9739	0.9808	5.8277	0.5047
0.8	50	0.9435	0.0127	−0.0101	0.6837	0.9831	11.008	0.6201
	150	0.9778	−0.0161	0.0080	0.8734	0.9754	6.8268	0.7101
	500	0.9954	−0.0007	0.0012	0.9689	0.9565	5.3703	0.7727
		Variance						
0.5	50	0.0688	0.3102	0.0277	0.0972	7.5682	313.04	0.0488
	150	0.0312	0.1436	0.0171	0.0759	0.4699	111.72	0.0393
	500	0.0078	0.0313	0.0039	0.0300	0.1029	9.9939	0.0181
0.8	50	0.0382	0.1795	0.0361	0.0548	1.8765	83.093	0.0417
	150	0.0153	0.0656	0.0192	0.0344	0.5007	16.762	0.0311
	500	0.0029	0.0183	0.0069	0.0083	0.1258	1.5437	0.0107

NOTE: The data generating process is described by (1.5), (1.6) and (1.7). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.9: VAR. The mean and the variance of the mixing parameter estimates for the two steps quasi ML method (EM algorithm iniciated with the true parameters values).

Parameters		B_{11}	B_{21}	B_{12}	B_{22}	Ψ_1	Ψ_2	γ
True values		1	0	0	1	1	5	0.5/0.8
γ	T	Mean						
0.5	50	0.943	0.0260	−0.0045	0.8453	0.8848	11.443	0.5333
	500	1.0049	0.0020	−0.0011	0.9765	0.9780	5.7325	0.5046
0.8	50	0.9669	−0.0258	0.0124	0.8504	0.7101	8.3221	0.7634
	500	0.9968	0.0003	0.0011	0.9735	0.9471	5.3650	0.7780
		Variance						
0.5	50	0.0582	0.3278	0.0334	0.1064	6.8999	189.44	0.0467
	500	0.0074	0.0304	0.0039	0.0275	0.0964	8.2750	0.0163
0.8	50	0.0280	0.1730	0.0546	0.0495	0.4517	65.635	0.0395
	500	0.0029	0.0173	0.0067	0.0077	0.1339	1.4941	0.0094

NOTE: The data generating process is described by (1.5), (1.6) and (1.7). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Table 1.10: VAR, The mean and the variance of estimators for the ML method, the mixing proportion $\gamma = 0.5$. The data generating process is described by (1.6), (1.7) and (1.5).

T	50		150	
	Mean	Var	Mean	Var
$A_1^{(0)}$	-0.0275	0.0367	0.0046	0.0097
$A_2^{(0)}$	-0.0689	0.1363	0.0269	0.0406
$A_{1,1}$	0.4370	0.0180	0.4845	0.0064
$A_{2,1}$	0.01741	0.0957	-0.0348	0.0281
$A_{1,2}$	-0.0034	0.0073	0.0072	0.0020
$A_{2,2}$	0.4480	0.0294	0.4817	0.0061
$B_{1,1}$	1.1507	0.2199	1.0997	0.1557
$B_{2,1}$	0.0588	0.3988	-0.0215	0.2382
$B_{1,2}$	-0.0008	0.0008	0.0036	0.0009
$B_{2,2}$	0.0724	0.0027	0.0869	0.0028
Ψ_1	40.557	77630	40.050	60582
Ψ_2	4988.49	$1.77e+008$	2936.21	55389027
γ	0.1929	0.0032	0.1168	0.0013

1.8 Results: SVECM

Table 1.11: VECM. Ratio of successful estimates, algorithms initiated with the true parameters values.

		ML			two steps quasi ML		
γ	T	BFGS	NEWTON	EM	BFGS	NEWTON	EM
0.5	50	0.366	0.218	0.403	0.627	0.336	0.976
	150	0.882	0.588	0.801	0.919	0.566	0.996
	500	0.987	0.843	0.970	0.993	0.750	0.988
0.8	50	0.245	0.174	0.350	0.343	0.166	0.949
	150	0.726	0.501	0.706	0.740	0.381	0.987
	500	0.969	0.785	0.935	0.975	0.667	0.993

NOTE: Two methods are considered: Maximum Likelihood and two steps quasi Maximum Likelihood. For each estimation method four maximization algorithms are evaluated: BFGS, NEWTON, BHHH and EM. The data generating process is described by (1.5), (1.8) and (1.9). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.12: VECM. Ratio of successful estimates, algorithms not initiated with the true parameters values.

		ML			two steps quasi ML		
γ	T	BFGS	NEWTON	EM	BFGS	NEWTON	EM
0.5	50	0.130	0.144	0.268	0.241	0.336	0.999
	500	0.349	0.716	0.891	0.293	0.739	0.987
0.8	50	0.112	0.140	0.299	0.172	0.366	1.000
	500	0.287	0.759	0.931	0.329	0.847	0.988

NOTE: Two methods are considered: Maximum Likelihood and two steps quasi Maximum Likelihood. For each estimation method four maximization algorithms are evaluated: BFGS, NEWTON, BHHH and EM. The data generating process is described by (1.5), (1.8) and (1.9). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Table 1.13: VECM. Ratio of successful estimates that satisfy condition (1.4) for $c = 0.01$ to all sucesfull estimates, algorithms initiated with the true parameters values.

		ML			two steps quasi ML		
γ	T	BFGS	NEWTON	EM	BFGS	NEWTON	EM
0.5	50	0.839	0.972	0.990	0.907	0.881	0.944
	150	0.926	0.995	0.999	0.979	1	0.987
	500	0.998	1	1	0.999	0.999	0.999
0.8	50	0.894	0.977	0.991	0.983	0.952	0.969
	150	0.983	1	1	0.996	1	0.985
	500	1	1	1	1	1	0.999

NOTE: Two methods are considered: Maximum Likelihood and two steps quasi Maximum Likelihood. For each estimation method four maximization algorithms are evaluated: BFGS, NEWTON, BHHH and EM. The data generating process is described by (1.5), (1.8) and (1.9). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.14: VECM. Ratio of successful estimates that satisfy condition (1.4) for $c = 0.01$ to all sucesfull estimates, algorithms not initiated with the true parameters values.

		ML			two steps quasi ML		
γ	T	BFGS	NEWTON	EM	BFGS	NEWTON	EM
0.5	50	0.854	0.951	1	0.975	0.881	0.913
	500	0.997	1	1	0.997	0.999	0.999
0.8	50	0.866	0.971	1	0.994	0.937	0.950
	500	1	1	1	1	0.999	0.999

NOTE: Two methods are considered: Maximum Likelihood and two steps quasi Maximum Likelihood. For each estimation method four maximization algorithms are evaluated: BFGS, NEWTON, BHHH and EM. The data generating process is described by (1.5), (1.8) and (1.9). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.15: VECM. The mean and the variance of the parameters estimates for the two steps quasi ML method (BFGS algorithm initiated with the true parameters values).

Parameters		β_2	α_1	α_2	$A_1^{(0)}$	$A_2^{(0)}$	Γ_{11}	Γ_{21}	Γ_{12}	Γ_{22}
True values		-1	-0.1	0.1	0	0	0.2	0.5	0.5	0.2
γ	T	Mean								
0.5	50	-1.628	-0.172	0.174	0.105	0.087	0.177	0.429	0.428	0.216
	150	-1.004	-0.125	0.132	-0.012	0.019	0.191	0.476	0.472	0.210
	500	-1.000	-0.108	0.111	-0.002	-0.001	0.197	0.490	0.491	0.205
0.8	50	-1.287	-0.180	0.157	-0.042	-0.062	0.178	0.426	0.427	0.190
	150	-1.007	-0.124	0.127	-0.017	-0.004	0.192	0.469	0.476	0.202
	500	-1.001	-0.107	0.110	-0.002	0.000	0.197	0.491	0.493	0.204
		Variance								
0.5	50	225.420	0.015	0.040	2.471	5.720	0.011	0.031	0.015	0.039
	150	0.039	0.002	0.006	0.125	0.242	0.003	0.009	0.003	0.010
	500	0.000	0.000	0.001	0.012	0.019	0.001	0.002	0.001	0.003
0.8	50	1595.93	0.016	0.027	1.575	3.309	0.013	0.024	0.017	0.029
	150	0.036	0.003	0.005	0.125	0.206	0.004	0.007	0.004	0.008
	500	0.001	0.001	0.001	0.010	0.013	0.001	0.002	0.001	0.002

NOTE: The data generating process is described by (1.5), (1.8) and (1.9). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.16: VECM. The mean and the variance of the parameters estimates for the ML method (BFGS algorithm initiated with the true parameters values).

Parameters		β_2	α_1	α_2	$A_1^{(0)}$	$A_2^{(0)}$	Γ_{11}	Γ_{21}	Γ_{12}	Γ_{22}
True values		-1	-0.1	0.1	0	0	0.2	0.5	0.5	0.2
γ	T	Mean								
0.5	50	-0.999	-0.193	0.180	0.025	0.179	0.182	0.441	0.407	0.222
	150	-1.002	-0.125	0.126	0.000	0.006	0.192	0.480	0.472	0.208
	500	-1.000	-0.108	0.110	0.000	-0.003	0.197	0.492	0.491	0.205
0.8	50	-0.979	-0.196	0.162	0.116	-0.116	0.182	0.453	0.414	0.207
	150	-0.999	-0.125	0.123	-0.011	-0.019	0.196	0.478	0.474	0.203
	500	-1.001	-0.107	0.108	0.000	-0.001	0.197	0.493	0.493	0.203
		Variance								
0.5	50	0.045	0.016	0.035	1.653	2.048	0.014	0.036	0.016	0.040
	150	0.009	0.002	0.006	0.128	0.180	0.003	0.008	0.003	0.009
	500	0.000	0.000	0.001	0.012	0.018	0.001	0.002	0.001	0.002
0.8	50	0.060	0.015	0.021	1.064	0.999	0.017	0.029	0.017	0.029
	150	0.010	0.003	0.004	0.109	0.118	0.004	0.006	0.004	0.007
	500	0.000	0.001	0.001	0.009	0.011	0.001	0.002	0.001	0.002

NOTE: The data generating process is described by (1.5), (1.8) and (1.9). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.17: VECM. The mean and the variance of the mixing parameters estimates for the ML method (BFGS algorithm initiated with the true parameters values).

Parameters		B_{11}	B_{21}	B_{12}	B_{22}	Ψ_1	Ψ_2	γ
True values		1	0	0	1	1	5	0.5/0.8
γ	T	Mean						
0.5	50	0.953	−0.018	0.011	0.648	1.059	19.615	0.551
	150	0.987	−0.006	−0.001	0.828	1.021	11.012	0.494
	500	0.997	−0.009	0.001	0.953	0.990	6.183	0.499
0.8	50	0.902	0.078	−0.018	0.612	1.202	16.464	0.649
	150	0.979	−0.002	0.000	0.849	0.949	7.593	0.710
	500	0.994	0.001	0.000	0.964	0.945	5.432	0.773
		Variance						
0.5	50	0.071	0.342	0.032	0.076	2.599	331.26	0.029
	150	0.031	0.144	0.016	0.079	1.524	169.026	0.036
	500	0.008	0.034	0.004	0.033	0.107	15.931	0.019
0.8	50	0.053	0.172	0.026	0.052	2.768	225.49	0.031
	150	0.016	0.073	0.020	0.036	0.781	39.406	0.033
	500	0.003	0.019	0.007	0.009	0.115	1.576	0.011

NOTE: The data generating process is described by (1.5), (1.8) and (1.9). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.18: VECM. The mean and the variance of the mixing parameters estimates for the two steps quasi ML method (BFGS algorithm initiated with the true parameters values).

Parameters		B_{11}	B_{21}	B_{12}	B_{22}	Ψ_1	Ψ_2	γ
True values		1	0	0	1	1	5	0.5/0.8
γ	T	Mean						
0.5	50	0.945	−0.003	−0.002	0.680	1.192	14.223	0.442
	150	0.985	−0.017	−0.001	0.856	1.097	9.458	0.475
	500	0.997	−0.012	0.002	0.970	0.992	5.750	0.500
0.8	50	0.936	0.008	−0.008	0.671	0.966	11.231	0.593
	150	0.973	0.002	−0.002	0.876	0.930	6.969	0.711
	500	0.994	−0.002	0.001	0.971	0.946	5.328	0.773
		Variance						
0.5	50	0.078	0.317	0.030	0.091	0.924	241.47	0.048
	150	0.032	0.151	0.018	0.082	0.638	122.55	0.044
	500	0.008	0.033	0.004	0.030	0.102	6.488	0.018
0.8	50	0.043	0.161	0.029	0.054	1.539	162.63	0.045
	150	0.014	0.072	0.021	0.037	0.344	32.394	0.037
	500	0.003	0.019	0.007	0.008	0.114	4.015	0.011

NOTE: The data generating process is described by (1.5), (1.8) and (1.9). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

Table 1.19: VECM. The mean and the variance of the mixing parameters estimates for the two steps quasi ML method (EM algorithm initiated with the true parameters values).

Parameters		B_{11}	B_{21}	B_{12}	B_{22}	Ψ_1	Ψ_2	γ
True values		1	0	0	1	1	5	0.5/0.8
γ	T	Mean						
0.5	50	0.957	−0.001	−0.003	0.839	0.828	10.781	0.531
	150	0.989	−0.015	0.003	0.898	0.940	8.398	0.496
	500	0.997	−0.014	0.003	0.972	0.991	5.688	0.499
0.8	50	0.943	−0.009	0.004	0.864	0.767	7.235	0.787
	150	0.980	−0.004	0.002	0.905	0.788	6.782	0.739
	500	0.994	−0.002	0.002	0.973	0.939	5.266	0.775
		Variance						
0.5	50	0.059	0.352	0.039	0.103	0.739	178.21	0.050
	150	0.029	0.152	0.021	0.077	0.435	90.99	0.040
	500	0.008	0.033	0.005	0.027	0.101	6.139	0.016
0.8	50	0.027	0.149	0.052	0.046	0.685	55.824	0.037
	150	0.013	0.081	0.028	0.036	0.342	29.583	0.037
	500	0.003	0.020	0.007	0.008	0.115	1.538	0.010

NOTE: The data generating process is described by (1.5), (1.8) and (1.9). We denote by T and γ the length of the sample and a mixing proportion parameter, respectively.

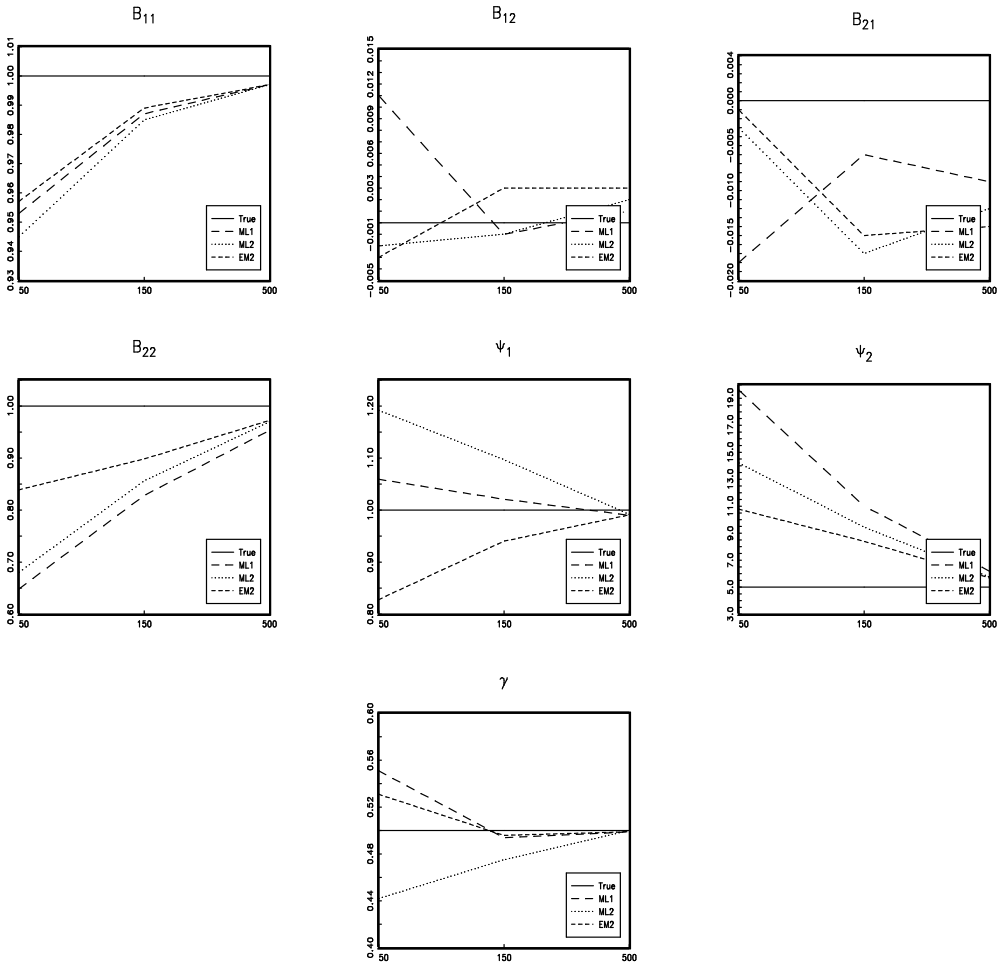

Figure 1.1: The mean of the estimates of mixture parameters for VECM conditional on the sample length. "True" describes the true parameter values whereas ML1, ML2 and EM2 present the results for the ML method with BFGS algorithm, two steps quasi ML method with BFGS algorithm and two steps quasi ML method with EM algorithm, respectively. The data generating process is described by (1.5), (1.8) and (1.9).

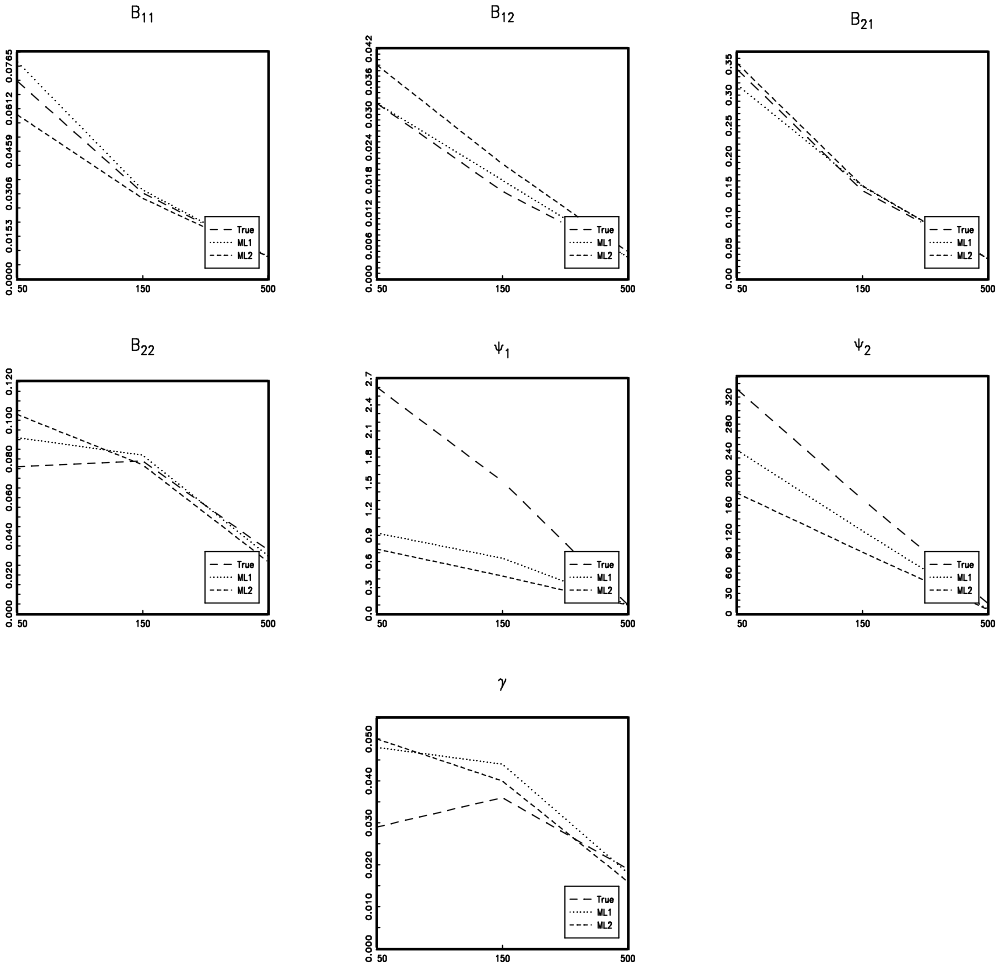

Figure 1.2: The variance of the estimates of mixture parameters for VECM conditional on the sample length. ML1, ML2 and EM2 present the results for the ML method with BFGS algorithm, two steps quasi ML method with BFGS algorithm and two steps quasi ML method with EM algorithm, respectively. The data generating process is described by (1.5), (1.8) and (1.9).

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Chapter 2

Structural vector autoregressions with Markov switching[1]

2.1 Introduction

In structural vector autoregressive (SVAR) modelling a major problem is to find convincingly identified shocks which are informative about the actual reactions of a set of variables to unexpected exogenous innovations. Although economic theories and models often provide some information which can be used for identification, this is not always sufficient to fully identify the shocks of interest. Different cures for this problem have been proposed over the years. In the earlier VAR literature, a triangular orthogonalization of the shocks which results in a recursive structure was quite popular (e.g., Sims (1980)). This kind of identification was often based on some ad hoc reasoning and it sometimes is proposed that different orderings are used, which would result in different recursive structures and check the robustness of the main results (Amisano and Giannini (1997), Lütkepohl (2005, Section 2.3.2)). Another proposal is to identify only some of the shocks (see Christiano, Eichenbaum and Evans (1999)). This approach works well as long as there is information to identify the shocks of primary interest. Unfortunately, this is not always possible (see again Christiano, Eichenbaum and Evans (1999)). Other approaches use restrictions for the long-run effects of the shocks (Blanchard and Quah (1989), King, Plosser, Stock and Watson (1991), Pagan and Pesaran (2008)), inequality restrictions (Uhlig (2005), Canova and De Nicoló (2002), Faust (1998)), Bayesian methods (Koop (1992)) or statistical properties of the data; the residual distribution (Lanne and Lütkepohl (2008)), structural breaks or heteroskedasticity (Rigobon

[1] This is a joint article with M. Lanne and H. Lütkepohl published in EUI Working Paper ECO 2009/06

(2003), Lanne and Lütkepohl (2008)).

In this study, we will consider the latter type of identifying information. In other words, we will use specific properties of a statistical model to achieve identification. More precisely, we will consider special features of Markov regime switching (MS) models to identify structural shocks. These models were introduced by Hamilton (1989) as tools for time series econometrics. They were extended to the VAR case by Krolzig (1997) and they have been considered for SVAR analysis, e.g., by Sims and Zha (2006) and Rubio-Ramirez, Waggoner and Zha (2005). Sims, Waggoner and Zha (2008) presents Bayesian methodology for handling general versions of MS-SVAR models. They were found to be useful, for instance, in business cycle analysis. Thus, they are potentially suitable models in many situations where SVAR models have been used traditionally. In contrast to other MS-VAR studies, we will argue that in these models shocks can be identified by the assumption that they are orthogonal across different regimes. Conditions will be given which ensure identification of the shocks under this assumption. A crucial condition is that the residual covariance matrices of the VAR model vary across regimes. In fact, since identification will hinge on MS in the residual covariance matrix, we will focus on a model where the other parameters are constant across regimes. Such models were found to be particularly useful in applications reported by Sims and Zha (2006) and Sims, Waggoner and Zha (2008).

An important advantage of our approach is that some crucial assumptions necessary for the identification of the shocks can be checked with statistical methods. We will also discuss an extension of the setup to systems with integrated and cointegrated variables. In that case, we will consider vector error correction models (VECMs) which makes it easy to accommodate long-run restrictions for the effects of the shocks in a way proposed by King, Plosser, Stock and Watson (1991) and others.

To illustrate our approach, we apply it to two examples from Lanne and Lütkepohl (2009). The first one considers a stationary system consisting of US gross domestic product (GDP), an interest rate and stock prices. It was previously used to investigate the impact of fundamental shocks on stock prices. The second example is based on a VECM and analyzes the relation between European and US interest rates.

The paper is structured as follows. In the next section, our model setup is presented, identification is discussed and the associated estimation strategy is considered. In Section 2.3, the empirical applications are presented and conclusions are provided in Section 2.4. A theoretical result regarding matrix decompositions is given in the Appendix.

2.2 The model

2.2.1 General setup

We consider a K-dimensional reduced form VAR(p) model of the type

$$y_t = Dd_t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{2.1}$$

where $y_t = (y_{1t}, \ldots, y_{Kt})'$ is a K-dimensional vector of observable time series variables, d_t is a deterministic term with coefficient matrix D, the A_j's ($j = 1, \ldots, p$) are $(K \times K)$ coefficient matrices and u_t is a K-dimensional white noise error term with mean zero and positive definite covariance matrix Σ_u, that is, $u_t \sim (0, \Sigma_u)$. If some of the variables are cointegrated, the VECM form may be more convenient,

$$\Delta y_t = D^* d_t^* + \alpha\beta' y_{t-1}^* + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t, \tag{2.2}$$

where Δ denotes the differencing operator, defined such that $\Delta y_t = y_t - y_{t-1}$, $\Gamma_j = -(A_{j+1} + \cdots + A_p)$ ($j = 1, \ldots, p-1$) are $(K \times K)$ coefficient matrices, α is a $(K \times r)$ loading matrix of rank r, β is the $(K^* \times r)$ cointegration matrix which may include parameters associated with deterministic terms and y_{t-1}^* is y_{t-1} augmented by deterministic terms in the cointegration relations. The rank r is the cointegrating rank of the system. The term d_t^* represents unrestricted deterministic components and its parameter matrix is denoted by D^*.

In the standard SVAR approach, a transformation of the reduced form residuals u_t is used to obtain the structural shocks, ε_t. A transformation matrix B is chosen such that $\varepsilon_t = B^{-1}u_t \sim (0, I_K)$ has identity covariance matrix, that is, the structural shocks are assumed to be orthogonal and typically their variances are normalized to one. Hence, $\Sigma_u = BB'$. To obtain identified, unique structural shocks, some restrictions have to be imposed on B. Often zero restrictions or long-run constraints are used in this context. A zero restriction on B implies that a certain shock does not have an instantaneous effect on one of the variables, whereas long-run restrictions exclude permanent effects of shocks on some or all of the variables. Specific examples will be considered in our applications in Section 2.3. It is also straightforward to extend the models considered so far as to allow for restrictions to be placed on the instantaneous relations of the variables rather than the shocks. This is most easily done in the context of the so-called AB model, the VAR version of which has the form

$$Ay_t = Dd_t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + B\varepsilon_t. \tag{2.3}$$

For this model the reduced form covariance matrix is $\Sigma_u = A^{-1}BB'^{-1\prime}$. More restrictions are needed to identify both A and B. However, often one of the two matrices is the identity matrix and it is just a matter of convenience to place the restrictions on the other matrix.

Notice that, although normality of the u_t's is often assumed for convenience, such an assumption is usually not backed by theoretical considerations nor is it necessarily required for asymptotic inference. Moreover, VAR residuals are

often found to be nonnormal in applied work. In the following, we will specify a Markov switching structure on the residuals which implies a more general distribution class for the u_t's and we will discuss how that can be used for the identification of shocks.

2.2.2 Markov regime switching residuals

We assume that the distribution of the error term, u_t, depends on a Markov process s_t. More precisely, it is assumed that s_t ($t = 0, \pm 1, \pm 2, \dots$) is a discrete Markov process with two different regimes, 0 and 1. We focus on a two regime case here for convenience to simplify the following notation and discussion. An extension to more than two regimes is straightforward. The case of two regimes only is also considered in the applications in Section 2.3 and it is therefore preferable here to simplify the discussion of the identification of shocks.

The *transition probabilities* are

$$p_{ij} = \Pr(s_t = j | s_{t-1} = i), \quad i, j = 0, 1.$$

The *conditional* distribution of u_t given s_t is assumed to be normal,

$$u_t | s_t \sim N(0, \Sigma_{s_t}). \tag{2.4}$$

Although the conditional normality assumption is made for convenience, it should be clear that it opens up a much wider class of distributions than just the unconditional normal. We will discuss this issue further below. The distributional assumption will be used for setting up the likelihood function. If normality of the conditional distribution does not hold, the estimators will only be pseudo maximum likelihood (ML) estimators. The normality assumption in (2.4) is not essential for our identification of shocks.

Note that in our model the transition probabilities are the same in all periods. They can be conveniently summarized in the (2×2) *transition matrix*

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}.$$

This matrix contains all necessary conditional probabilities to reconstruct the distributions of the stochastic process s_t. For example, the unconditional distribution of s_t can be derived from the conditional probabilities in P (see, e.g., Hamilton (1994)). For later reference, we mention that the unconditional probabilities of the states of an ergodic Markov chain are $\Pr(s_t = 0) = 1 - \Pr(s_t = 1) = (1 - p_{11})/(2 - p_{00} - p_{11})$.

Moreover, $p_{10} = 1 - p_{00}$ and $p_{01} = 1 - p_{11}$. If $p_{00} = p_{01}$ and $p_{11} = p_{10}$, the conditional distributions of the states are independent of the previous state, that is,

$$\Pr(s_t = j) = \Pr(s_t = j | s_{t-1} = 0) = \Pr(s_t = j | s_{t-1} = 1), \quad j = 0, 1.$$

Hence, the MS model reduces to a model with mixed normal (MN) errors,

$$u_t \sim \begin{cases} N(0, \Sigma_0) \text{ with probability } \gamma = p_{00}, \\ N(0, \Sigma_1) \text{ with probability } 1 - \gamma = p_{11}. \end{cases}$$

In that case, the transition matrix has the form

$$P = \begin{bmatrix} \gamma & \gamma \\ 1-\gamma & 1-\gamma \end{bmatrix}. \tag{2.5}$$

Given that mixed normal distributions constitute a very large and flexible class of distributions, this shows that assuming a conditional normal distribution in (2.4) results in a very rich distribution class for the error terms. The case of mixed normal errors in the context of SVAR analysis was considered by Lanne and Lütkepohl (2009).

Identification of shocks in the MS model can be achieved by the assumption that the shocks are orthogonal across regimes and only the variances of the shocks change across regimes while the impulse responses are not affected. In particular, the instantaneous effects are the same in all regimes. Note that the assumption of time invariant impulse responses throughout the sample period is common in standard SVAR analysis and hence, not a particularly restrictive element in our setup.

A well-known result of matrix algebra establishes that there exists a $(K \times K)$ matrix B such that $\Sigma_0 = BB'$ and $\Sigma_1 = B\Lambda B'$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K)$ is a diagonal matrix (e.g., Lütkepohl (1996, Section 6.1.2)). From $\Sigma_0 = BB'$ and $\Sigma_1 = B\Lambda B'$ we get a total of $K(K+1)$ equations which can be solved uniquely for the K^2 elements of B and the K diagonal elements of Λ under mild conditions. In the Appendix, we give a result which implies that the matrix B is unique up to changes in sign if all diagonal elements of Λ are distinct and ordered in some prespecified way. For example, they may be ordered from smallest to largest or largest to smallest. The result in the Appendix is formulated in such a way so that it can be used for models with more than two regimes. For the case of two regimes, the important point to note here is that our setup delivers shocks $\varepsilon_t = B^{-1}u_t$ which are orthogonal in both regimes. Since B is unique (up to sign changes), the model is in fact identified by the assumption that the shocks have to be orthogonal and the instantaneous effects are identical across regimes. Thus, any restrictions imposed on B in a conventional SVAR framework become over-identifying in our setup and hence, can be tested against the data.

The nonuniqueness of B with respect to sign in our framework causes no problems for our purposes. The precise condition is that all signs in any of the columns of B can be reversed. This corresponds to considering negative shocks instead of positive shocks or vice versa. Usually it will not be a problem for the analyst to decide on whether positive or negative shocks are of interest. Also, from the point of view of asymptotic inference, local identification of this kind is sufficient for the usual results to hold. Provided no sign restrictions are used, this kind of nonuniqueness of the shocks with respect to sign changes is also common in standard SVAR analyses although this is not always emphasized.

Rubio-Ramirez, Waggoner and Zha (2005) also discusses identification in MS-SVAR models. However, they allow all parameters to differ across regimes. In their setup, assuming the same impulse responses in all regimes is not plausible and therefore assuming the same instantaneous effects of the shocks would be restrictive. Hence, they use alternative identification conditions. In our setup, MS is confined to the error covariance matrix only and no MS is assumed in other parameters because the latter is not needed for the identification of the shocks and we try to remain as close as possible to the standard SVAR approach which assumes time invariant impulse responses for the full sample period. Allowing for MS in the residuals only means that we basically remain within a standard SVAR model. In fact, this feature of a model was diagnosed but not used for identification in Sims and Zha (2006), for example.

2.2.3 Estimation

Under our assumption of conditional normality given a particular state in (2.4), maximum likelihood (ML) estimation is a plausible estimation method. If the conditional normality does not hold it will deliver pseudo ML estimators. Hence, for a 2-state MS-VAR model the parameters are estimated by maximizing the (pseudo) log likelihood function

$$l_T = \sum_{t=1}^{T} \log f(y_t|Y_{t-1}),$$

where $Y_{t-1} = (y_{-p+1}, \ldots, y_{t-1})$,

$$f(y_t|Y_{t-1}) = \sum_{j=0}^{1} \Pr(s_t = j|Y_{t-1}) f(y_t|s_t = j, Y_{t-1})$$

and

$$f(y_t|s_t = j, Y_{t-1}) = (2\pi)^{-K/2} \det(\Sigma_j)^{-1/2} \exp\left\{-\frac{1}{2} u_t' \Sigma_j^{-1} u_t\right\}, \quad j = 0, 1.$$

Here $\Sigma_0 = BB'$, $\Sigma_1 = B\Lambda B'$ and the u_t are the reduced form residuals. Moreover,

$$\begin{bmatrix} \Pr(s_t = 0|Y_{t-1}) \\ \Pr(s_t = 1|Y_{t-1}) \end{bmatrix} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \begin{bmatrix} \Pr(s_{t-1} = 0|Y_{t-1}) \\ \Pr(s_{t-1} = 1|Y_{t-1}) \end{bmatrix},$$

where

$$\Pr(s_t = j|Y_t)$$

$$= \Pr(s_t = j|Y_{t-1}) f(y_t|s_t = j, Y_{t-1}) \Big/ \sum_{i=0}^{1} \Pr(s_t = i|Y_{t-1}) f(y_t|s_t = i, Y_{t-1}),$$

$$j = 0, 1.$$

The optimization of l_T may be done with a suitable extension of the EM algorithm described in Hamilton (1994). A blockwise algorithm for computing the ML or, in a Bayesian framework, the posterior mode estimates was proposed by Sims, Waggoner and Zha (2008) which may be more suitable for models with many free parameters, e.g., when many variables are considered and the number of regimes is larger than 2 or 3.

The properties of Gaussian ML estimation in a univariate model of type (2.4) (that is, the process is white noise conditional on a given state of the Markov chain) were discussed by Francq and Roussignol (1997). Very general asymptotic estimation results for stationary processes are available in Douc, Moulines and Rydén (2004). The case of cointegrated VARs seems less well explored. If the cointegration relations are known, there is no problem because the results for stationary processes can be used. For the situation where the cointegration relations are unknown, we propose to use a two-step estimation procedure. In the first step, the cointegration relation is estimated by Johansen's (1995) reduced rank regression. Then an ML estimation conditional on the first-step cointegration relation is performed. Although there is no apparent reason why this procedure should not result in estimators with standard asymptotic properties, we admit that we do not know of a formal proof if the cointegration matrix is unknown and has to be estimated. In the application in Section 2.3.2, where cointegrated variables are considered, assuming known cointegration relations turns out to be reasonable.

2.3 Illustrations

2.3.1 US Model

Our first example uses a small system of US macro variables from Binswanger (2004) which has also been used by Lanne and Lütkepohl (2009). The purpose of Binswanger's analysis was to determine the impact of fundamental shocks on the stock market. The issue has been discussed previously in the literature. For example, in an SVAR analysis Rapach (2001) finds that macro shocks have an important effect on real stock prices. On the other hand, Binswanger (2004) uses US data from 1983 to 2006 and concludes that real activity shocks explain only a small fraction of the real stock price variability. It is not uncommon in SVAR analyses that the specification of the shocks is essential for the outcome.

We use quarterly US data for the period $1983Q1 - 2006Q3$ for the three variables $(gdp_t, r_t, sp_t)'$, where gdp_t denotes log real gross domestic product, r_t is the 3-months Treasury bill rate and sp_t stands for log real stock prices, as in Lanne and Lütkepohl (2009). More details on the data are given in Appendix B of the latter paper. Binswanger's objective was to assess the importance of fundamental shocks for stock price movements. Fundamental shocks in this context are shocks which have a long-term impact on economic growth and the interest rate.

Binswanger (2004) assumes that there is just one nonfundamental shock. It

is specified by the requirement that it may have a long-term impact on stock prices, sp_t, but not on gdp_t and r_t. The three structural shocks are identified by imposing zero restrictions on the matrix of long-run effects of the shocks as follows:

$$A(1)^{-1}B = \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}. \tag{2.6}$$

Here, an asterisk denotes an unrestricted element. Hence, the matrix of long-run effects, $A(1)^{-1}B$, is lower-triangular. The assumption that the last shock is nonfundamental and in particular, does not have a long-term impact on gdp_t and r_t, implies the two zeros in the last column of $A(1)^{-1}B$. The additional zero restriction in the second column has however little justification. It is to some extent arbitrary and is imposed to obtain identified shocks in the classical SVAR framework.

Lanne and Lütkepohl (2009) argues that identifying the shocks without such a restriction is desirable. They use a VAR(4) model in first differences for $y_t = (\Delta gdp_t, \Delta r_t, \Delta sp_t)'$ because the variables have unit roots but are not cointegrated. The residuals are found to be nonnormal. Therefore, they fit a model with mixed normal residuals and use this data feature to identify shocks and to check the structural restrictions imposed by Binswanger (2004). As mentioned in Section 2.2.2, a model with mixed normal residuals is just a special case of our MS model. The mixed normal model assumes that the regimes have no persistence and are assigned at random in each period. For the present example, allowing for some persistence in the regimes may be plausible for different reasons. For example, volatility changes could be linked to business cycle fluctuations and hence, may derive persistence from the fact that periods of positive and negative growth tend to last for several periods. Alternatively, the MS structure may just capture conditional heteroskedasticity which may arise from other sources than the business cycle.

Therefore, we have fitted VAR models with MS residuals, assuming that $\Sigma_1 \neq \Sigma_2$.[2] As in Lanne and Lütkepohl (2008), we have estimated an unrestricted model as well as one with the structural restrictions specified in (2.6). In addition, we have also estimated a model with only two zero restrictions on the last column of the matrix of long-run effects,

$$A(1)^{-1}B = \begin{bmatrix} * & * & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}. \tag{2.7}$$

Some estimation results and a range of tests which will be discussed in the following are given in Tables 2.1-2.3.

The first question of interest is whether the MS model is preferable to the model with mixed normal residuals that was used by Lanne and Lütkepohl

[2] The computations were done with GAUSS programs using EM iterations to get close to the optimum and then switching to the Newton-Raphson algorithm from the CML library for optimizing the likelihood function.

Table 2.1: Estimates of Structural Parameters of MS Models for $(\Delta gdp_t, \Delta r_t, \Delta sp_t)'$ with Lag Order $p = 4$ and Intercept Term (Sample Period: $1983Q2 - 2006Q3$)

	unrestricted		(2.6)		(2.7)	
Parameters	Estimates	Std.Dev.	Estimates	Std.Dev.	Estimates	Std.Dev.
λ_1	0.567	0.222	0.627	0.270	1.284	0.673
λ_2	0.931	0.402	1.643	1.245	0.614	0.255
λ_3	12.71	4.492	11.96	4.720	12.55	4.617
p_{00}	0.951	0.036	0.950	0.040	0.950	0.038
p_{11}	0.876	0.076	0.877	0.091	0.870	0.085
uncond.	0.716		0.714		0.723	
state prob.s	0.284		0.286		0.277	
$\log L/T$	7.5860		7.5666		7.5668	

NOTE: Standard errors are obtained from the inverse Hessian of the log-likelihood function.

Table 2.2: Wald Tests for Equality of λ_i's for Models from Table 2.1

	unrestricted		(2.6)		(2.7)	
H_0	test value	p-value	test value	p-value	test value	p-value
$\lambda_1 = \lambda_2 = \lambda_3$	7.974	0.019	7.739	0.021	7.677	0.022
$\lambda_1 = \lambda_2$	0.611	0.434	0.630	0.427	0.969	0.325
$\lambda_1 = \lambda_3$	7.284	0.007	5.645	0.018	6.608	0.010
$\lambda_2 = \lambda_3$	6.756	0.009	3.869	0.049	5.732	0.017

Table 2.3: LR Tests of Models for $(\Delta gdp_t, \Delta r_t, \Delta sp_t)'$

H_0	H_1	LR statistic	Assumed distribution	p-value
MN	MS	2.959	$\chi^2(1)$	0.085
(2.6)	unrestricted	3.491	$\chi^2(3)$	0.322
(2.7)	unrestricted	3.456	$\chi^2(2)$	0.178
(2.6)	(2.7)	0.036	$\chi^2(1)$	0.850

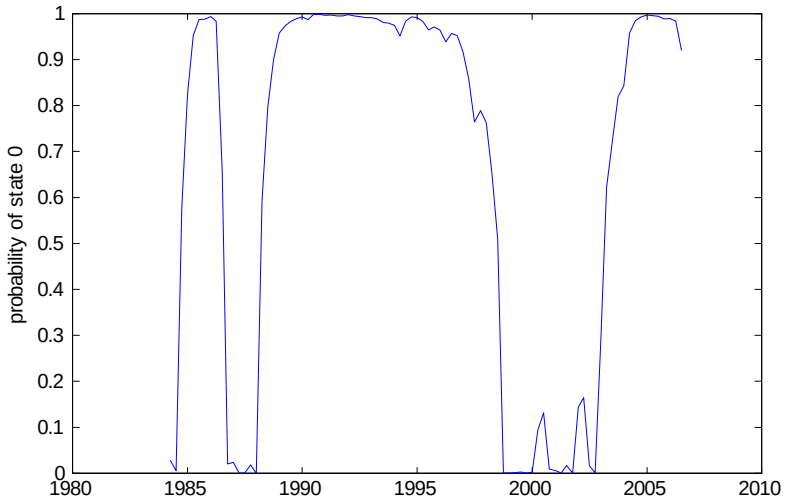

Figure 2.1: Probabilities of State 0 ($\Pr(s_t = 0|Y_T)$) for the unrestricted model for $(\Delta gdp_t, \Delta r_t, \Delta sp_t)'$ from Table 2.1.

(2009). Looking at the estimated state probabilities of the unrestricted model in Table 2.1, they are both larger than 0.8 and hence, the states appear to have some persistence. Still, it is desirable to check the MS model against the MN model more formally. Therefore, we have performed a likelihood ratio (LR) test of the restriction on the transition matrix specified in (2.5). In other words, we test the restriction that the probabilities in each row of P are constant. For this purpose, we have reestimated the unrestricted MN model from Lanne and Lütkepohl (2008) and compare the maximum of the likelihood with that of the unrestricted MS model given in Table 2.1.[3] The resulting LR test is reported in Table 2.3 together with some other LR tests which will be discussed later. The corresponding p-value turns out to be 8.5%. Thus, we can reject the MN model at a 10% level but not at the 5% level. In other words, there is weak evidence in favor of the MS model.

Further evidence is provided by the probabilities of being in State 0, which are plotted in Figure 2.1. More precisely, in Figure 2.1 we see the state probabilities conditional on the full sample information, $\Pr(s_t = 0|Y_T)$, based on the estimated unrestricted model. Obviously, these probabilities are quite persis-

[3] Since the likelihood is highly nonlinear and has multiple local maxima, it is not uncommon to obtain slightly different results with another estimation algorithm. Therefore it was necessary to reestimate the MN model with our estimation algorithm to ensure strict comparability of the results which is important for a proper comparison of the likelihood maxima. The results in Lanne and Lütkepohl (2009) are qualitatively similar to our estimation results for the MN model although they differ slightly numerically.

tent. However, they do not correspond strictly to the phases of the official US business cycle. Since one of the λ_i's of the unrestricted model in Table 2.1 is quite large ($\lambda_3 = 12.71$) while the other two are around one or a little smaller, the second state is one where at least one of the shocks has a substantially larger volatility than in the first regime. Thus, the state probabilities plotted in Figure 2.1 correspond to a regime of lower volatility at least in one of the shocks. The corresponding state appears to represent periods when the stock market had a tendency to increase. Notice that the probability of being in this state is low around the stock market crash in 1987 and during the adjustment period after the technology bubble in the first years of the new millennium. In any case, there appears to be some persistence in the state which implies that the MS model may describe the data better than the MN model. Therefore, we will now consider the previously used identifying restrictions within our MS model.

As mentioned earlier, the zero restrictions in (2.6) and (2.7) are over-identifying if the λ_i's are distinct. Hence, it is instructive to look at the estimates in Table 2.1. Clearly, the estimated λ_i's of the unrestricted model are quite different. However, their standard errors are also quite large. Therefore, we have performed Wald tests of equality of these quantities and present them in Table 2.2. These tests have asymptotic χ^2-distributions because the estimators have the usual normal limiting distributions under our assumptions. The p-values reported in Table 2.2 are based on these χ^2-distributions. In this context it may be worth noting that, in contrast to the matrix B, the λ_i's are identified even if they are identical. Thus, testing their equality makes sense. The test that all three λ_i's are equal has a p-value of 1.9% and hence, clearly rejects at a 5% level. The null hypotheses $H_0 : \lambda_1 = \lambda_3$ and $H_0 : \lambda_2 = \lambda_3$ are even rejected at the 1% level. On the other hand, at common significance levels, it cannot be rejected that $\lambda_1 = \lambda_2$. Similar results are also obtained if the restrictions in (2.6) and (2.7) are imposed. Thus, there is strong evidence that at least two of the three λ_i's are distinct.

Let us for the moment still pretend that the three λ_i's in the unrestricted model are distinct and hence, all three shocks are identified without further restrictions on B. In that case, the zero restrictions imposed in (2.6) and (2.7) are overidentifying and can be tested by LR tests. These test results are also given in Table 2.3. It turns out that none of the zero restrictions can be rejected at conventional significance levels. This result is also obtained when only the additional restriction in the second column of the matrix of long-run effects in (2.6) is tested which was not backed by theoretical considerations (see the last row in Table 2.3). The resulting p-value is 0.850 and hence, the data clearly do not object to this restriction. Although this means that we end up with the same model which was used by Binswanger (2004), the advantage of our approach is that the restrictions can be backed by statistical tests.

Of course, these conclusions are based on the assumption that all three λ_i's are distinct which does not have strong support from the data. Therefore, it is worth reflecting on the implications of some λ_i's being identical. This would mean that some of the restrictions imposed on the matrix of long-run effects in (2.6) and (2.7) may in fact not be overidentifying and hence, the LR tests

may have fewer degrees of freedom than assumed in Table 2.3. In that case, the p-values would be smaller than the ones reported in the table. However, in the absence of further information, we have no basis for rejecting the restrictions in (2.6).

2.3.2 European/US interest rate linkages

Our next example is also from Lanne and Lütkepohl (2009). It considers euro area and US interest rate linkages to investigate the relation between European and US monetary policy. It is based on an earlier study by Brüggemann and Lütkepohl (2005) which performs a standard SVAR analysis for cointegrated variables and concludes that European monetary policy depends to some extent on US monetary policy, whereas the reverse is not apparent from the data.

The paper considers monthly data for a euro area three months money market rate r_t^{EU}, a euro area 10-year bond rate R_t^{EU}, a US three months money market rate r_t^{US} and a US 10-year bond rate R_t^{US}. Thus, $y_t = (R_t^{US}, r_t^{US}, R_t^{EU}, r_t^{EU})'$. The sampling period is $1985M1 - 2004M12$. Details on the data construction and their sources are also given in Appendix B of Lanne and Lütkepohl (2009). Brüggemann and Lütkepohl (2005) finds that all four variables are $I(1)$ and that both the expectations hypothesis of the term structure and the uncovered interest rate parity hold. Hence, stationarity of the two spreads $R_t^{US} - r_t^{US}$ and $R_t^{EU} - r_t^{EU}$ as well as the two parities $R_t^{US} - R_t^{EU}$ and $r_t^{US} - r_t^{EU}$ is supported. These four relations represent three linearly independent cointegration relations from which the fourth one can be derived by a linear transformation. Therefore, Lanne and Lütkepohl (2008) considers a four-dimensional system with three known cointegration relations.

That paper uses a VECM for y_t with a constant term, three lags of Δy_t (i.e., $p = 4$), a cointegrating rank of $r = 3$ and MN residuals to investigate the impact of monetary shocks in the US and in Europe. Again it is easy to think of arguments for a more general MS specification of the residuals and hence, we have estimated the corresponding MS model and we have tested it against an MN model. Estimation and test results are given in Tables 2.4-2.6. They will be discussed in the following paragraph.

A test of our unrestricted MS model against an unrestricted MN model, i.e., of the restriction in (2.5) is reported in Table 2.6. The p-value is extremely small so that the MN model is rejected at any reasonable significance level. Hence, there is strong evidence that the MS model is preferable to the MN model for the present data set. This result is not surprising given that the estimated transition probabilities p_{00} and p_{11} are both larger than 90%, which indicates that the states have considerable persistence. The estimated probabilities of State 0, $\Pr(s_t = 0|Y_T)$, are plotted in Figure 2.2. Three of the λ_i's associated with the unrestricted model in Table 2.4 are considerably larger than one, while the other is not much smaller than one. Hence, the overall volatility in the second state (State 1) is considerably larger than in the first state. The probabilities in Figure 2.2 are those of the low volatility state. Apparently, the second half of the sample is characterized by lower volatility of shocks to the system. Indeed, the

Table 2.4: Estimates of MS-VECM for $(R_t^{US}, r_t^{US}, R_t^{EU}, r_t^{EU})'$ with Lag Order $p = 4$, Cointegrating Rank $r = 3$ and Intercept Term (Sample Period: $1985M1 - 2004M12$)

	unrestricted		one trans. shock		two trans. shocks	
Parameters	Estimates	Std.Dev.	Estimates	Std.Dev.	Estimates	Std.Dev.
λ_1	0.812	0.169	0.811	0.168	0.849	0.186
λ_2	15.87	3.433	15.90	3.332	14.73	3.232
λ_3	3.499	0.807	3.487	0.796	3.386	0.901
λ_4	8.422	1.818	8.445	1.802	8.233	1.757
p_{00}	0.904	0.036	0.905	0.036	0.911	0.038
p_{11}	0.919	0.033	0.919	0.033	0.928	0.035
uncond.	0.459		0.458		0.448	
state prob.s	0.541		0.542		0.552	
$\log L/T$	1.65305		1.65294		1.63751	

NOTE: Standard errors are obtained from the inverse Hessian of the log-likelihood function.

Table 2.5: Wald Tests for Equality of λ_i's for Models from Table 2.4

	unrestricted		one trans. shock		two trans. shocks	
H_0	test value	p-value	test value	p-value	test value	p-value
$\lambda_1 = \lambda_2$	19.22	0.000	20.46	0.000	18.58	0.000
$\lambda_1 = \lambda_3$	10.70	0.001	10.87	0.001	7.251	0.007
$\lambda_1 = \lambda_4$	17.40	0.000	17.82	0.000	17.59	0.000
$\lambda_2 = \lambda_3$	12.34	0.000	13.10	0.000	10.76	0.001
$\lambda_2 = \lambda_4$	3.808	0.051	3.940	0.047	3.206	0.073
$\lambda_3 = \lambda_4$	5.996	0.014	6.230	0.013	5.991	0.014

first differences notably of the short-term interest rate series appear to have an overall smaller variability in the second part of the sample, except for the period around the year 2000 (see Figure 2.3). The lower volatility periods correspond to the high probabilities of State 0 in Figure 2.2. Thus, the states reflect the change in volatility. For our purposes it is important to note that the MS model describes the data better than previous SVAR counterparts. Hence, it is of interest to study its implications for structural analysis.

The estimated λ_i's of all the models in Table 2.4 are quite different. One-standard error intervals around the estimates do not overlap. Again we have performed Wald tests to check equality of the λ_i's. The results of pairwise tests are presented in Table 2.5 and confirm distinct λ_i's. The p-values of all pairwise tests are smaller than 10% and most are even smaller than 1%. Thus, there is evidence that the λ_i's are distinct and hence, the shocks can be identified by assuming that they are orthogonal and have identical instantaneous impacts in both states. Consequently, we can check some of the structural assumptions that were used by Brüggemann and Lütkepohl (2005) and Lanne and Lütkepohl

Table 2.6: LR Tests of Models for $(R_t^{US}, r_t^{US}, R_t^{EU}, r_t^{EU})'$

H_0	H_1	LR statistic	Assumed distribution	p-value
MN	MS	45.39	$\chi^2(1)$	0.000
one trans. shock	unrestricted	0.051	$\chi^2(1)$	0.822
two trans. shocks	unrestricted	7.335	$\chi^2(2)$	0.026

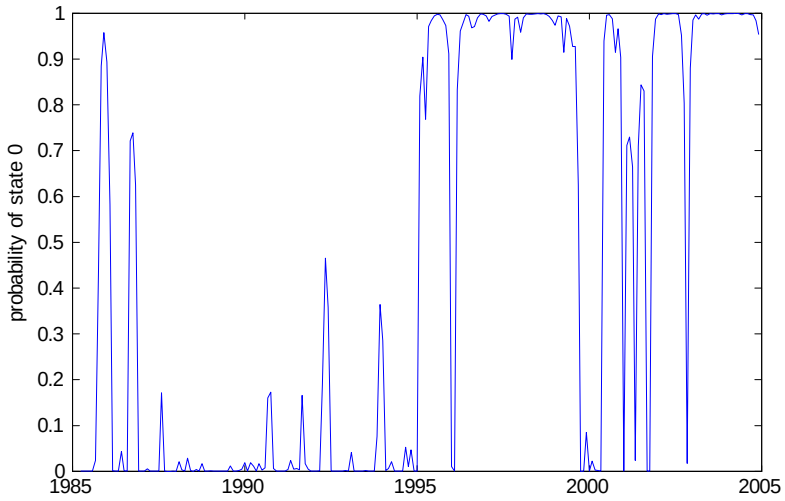

Figure 2.2: Probabilities of State 0 ($\Pr(s_t = 0|Y_T)$) for the unrestricted model for $(R_t^{US}, r_t^{US}, R_t^{EU}, r_t^{EU})'$ from Table 2.4.

(2009).

One important conclusion of the previous studies was that there are two transitory shocks that were viewed as candidates for monetary shocks. Since there are three cointegration relations, there can be up to three transitory shocks and Lanne and Lütkepohl (2008) finds that the data actually only support two such shocks in their MN framework. This issue is investigated by testing suitable zero restrictions on the matrix of long-term effects of the shocks. This matrix is known to be of the form ΞB, where $\Xi = \beta_\perp[\alpha_\perp'(I_K - \sum_{i=1}^{p-1}\Gamma_i)\beta_\perp]^{-1}\alpha_\perp'$ and $\alpha_\perp$ and $\beta_\perp$ signify orthogonal complements of α and β, respectively (e.g., Lütkepohl (2005, Section 9.2)). A shock is transitory if the corresponding column of this matrix consists of zeros. Such restrictions become testable in our MS models because the shocks are identified via the MS structure.

LR tests are presented in Table 2.6. The restrictions associated with one transitory shock (one zero column in ΞB) are not rejected at conventional sig-

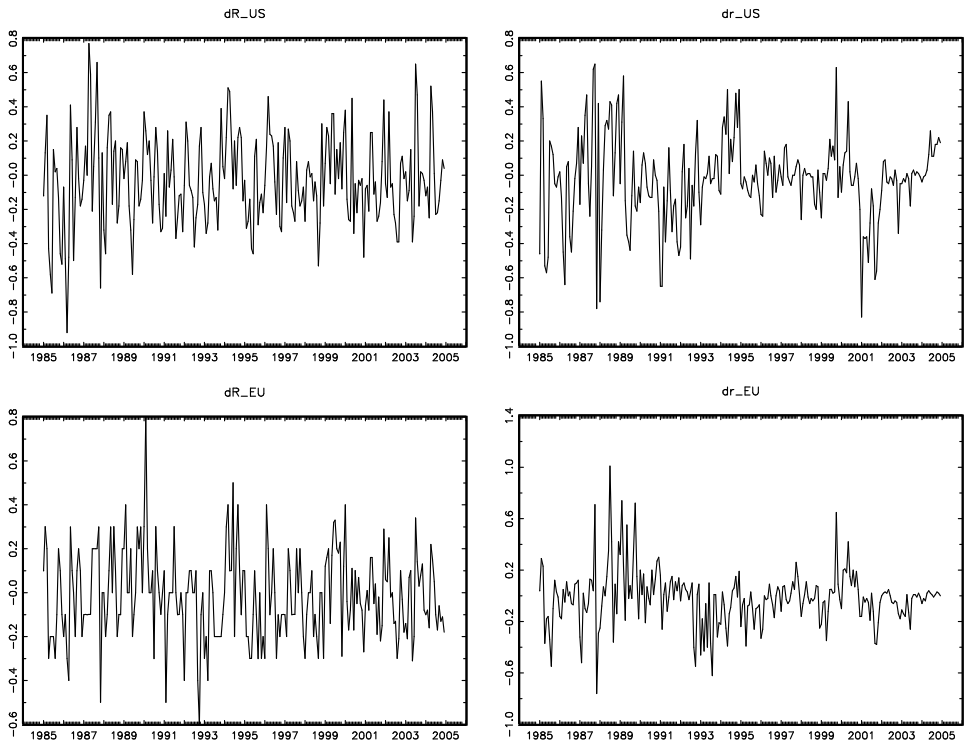

Figure 2.3: Changes in interest rates.

nificance levels whereas two transitory shocks are rejected at the 5% level, the p-value being 0.026. Note that the number of degrees of freedom of the asymptotic χ^2 distribution of the LR statistic implied by restricting a column of ΞB to zero take into account the reduced rank of the matrix of long-run effects. In particular, since the cointegrating rank is three, the (4×4) matrix ΞB has rank one so that a zero column of ΞB stands for a single restriction. Thus, in our model, the data do not support the existence of two transitory shocks.

Assuming one transitory shock only, we have also tested a number of alternative restrictions on its effects which did not help in determining a specific interpretation of this shock. In particular, we cannot identify it as a US or European monetary policy shock. Thus, we find little support for assumptions that allow us to explore the relation between US and European monetary shocks. Consequently, we do not find evidence for the hypothesis that US monetary policy has a more important impact on European monetary policy than vice versa. Thus, using the MS framework sheds doubt on whether the matter can be settled within a simple model of this type.

2.4 Conclusions

In this study we have augmented VAR models by Markov switching to obtain identified shocks. We have shown that under general conditions it is enough to assume orthogonality of the shocks and invariance of the impulse responses across regimes to obtain identification. A main advantage of this setup is that the data are informative with respect to the additional conditions needed for identification. Moreover, other assumptions which are typically used in SVAR analysis become overidentifying in our framework and hence, are testable.

We have applied these ideas to two SVAR models from the literature where a MS structure in the residual volatility is plausible. In the first example, a US macro system consisting of GDP, an interest rate and a stock price index is analyzed and it is found that in our framework previously assumed identifying restrictions can be confirmed. In the second example, the interest rate linkage between the US and the euro area is investigated. The MS model is found to be a better description of the data than previous SVAR models. Thus, it makes sense to use our framework for testing previously made identifying assumptions against the data. It turns out that a crucial restriction cannot be confirmed in our framework. Overall, our setup appears to be a useful tool to extract more information on identifying assumptions in SVAR analysis from the data.

The limited knowledge on the statistical inference procedures in particular when cointegrated variables are considered offer directions for further research. Moreover, the numerical challenges in estimating the models are nonnegligible if larger models with many variables and states are of interest. The algorithms proposed by Sims, Waggoner and Zha (2008) may be useful in this context and may help to overcome numerical problems in difficult situations. Further investigations in this direction are also left for the future.

Appendix. A Uniqueness Result for Covariance Matrix Decomposition

Proposition A. Let $\Sigma_0 = BB'$ and $\Sigma_i = B\Lambda_i B'$, where $\Lambda_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{iK})$, $i = 1, \ldots, M$, be nonsingular $(K \times K)$ covariance matrices. Then the $(K \times K)$ matrix B in the decomposition $\Sigma_0 = BB'$ is unique apart from sign reversal of its columns if for all $k \neq j \in \{1, \ldots, K\}$ there exists an $i \in \{1, \ldots, M\}$ such that $\lambda_{ik} \neq \lambda_{ij}$. □

Proof: Suppose $Q = [q_{ij}]$ is a $(K \times K)$ matrix such that

$$\Sigma_0 = BB' = BQQ'B' \tag{A.1}$$

and

$$\Sigma_i = B\Lambda_i B' = BQ\Lambda_i Q'B', \quad i = 1, \ldots, M. \tag{A.2}$$

To show the uniqueness of B up to multiplication of its columns by -1, we have to show that the only feasible matrix Q is a diagonal matrix with ± 1 on the main diagonal.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Pre- and post-multiplying $(A.1)$ by B^{-1} and its transpose, respectively, implies that $QQ' = I_K$ and hence, Q must be orthogonal. Similarly, it follows from $(A.2)$ that $Q\Lambda_i Q' = \lambda_i$ or $Q\Lambda_i = \Lambda_i Q$, $i = 1, \ldots, M$. Consequently, $\lambda_{ik} q_{kl} = \lambda_{il} q_{kl}$ for all $i = 1, \ldots, M$. Thus, $q_{kl} = 0$ for $k \neq l$ because $\lambda_{ik} \neq \lambda_{il}$ for at least one $i \in \{1, \ldots, M\}$. In other words, Q is an orthogonal diagonal matrix and hence, all diagonal elements of Q are ± 1 because the diagonal elements of a diagonal matrix are its eigenvalues and the eigenvalues of a diagonal real orthogonal matrix are all ± 1. This proves the proposition.

Part II

Generalized factor models

Chapter 3

Generalized factor model - estimation and distribution theory

3.1 Motivation

In the last decade, one could observe a growing interest in models that can extract and use information from large sets of variables. One approach is based on an assumption that there exist common factors, which can explain the variables' comovement. The factor models have been shown useful in econometric modeling. There is a series of articles that demonstrate advantages of using factors in forecasting (Stock and Watson (2002a), Stock and Watson (2002b)) and impulse response analysis (Bernanke, Boivin and Eliasz (2005), Kapetanios and Marcellino (2006))).

Recently, Stock and Watson (2005) adopts factor models for structural analysis. This article, together with other papers (Kapetanios and Marcellino (2006) and Forni, Giannone, Lippi and Reichlin (2007)) discusses the possibility of integrating the factor methods into the SVAR framework. There is empirical evidence that factors can contribute to classical VAR analysis (see Bernanke, Boivin and Eliasz (2005), Kapetanios and Marcellino (2006), Eickmeier (2009) and Forni and Gambetti (2008)).

So far, most of the research concentrates on modeling stationary panel data. Breitung and Eickmeier (2005) provides a comprehensive literature review of stationary dynamic factors models and their applications. There are, however, few articles that discuss the issue of common nonstationary trends. Bai (2004), Bai and Ng (2004) and Gonzalo and Granger (1995) describe estimation methods of nonstationary common components. Bai (2004) proposes information criteria, IPC, that allow for consistent estimation of the number of common trends and derives limiting distributions of estimated factors and common components.

Banerjee and Marcellino (2008) discusses cointegration issues related to the existence of common trends and shows how the factor analysis can contribute to the existing literature. Eickmeier (2009) uses nonstationary factors in structural analysis of economic development of euro area countries.

The literature discusses two approaches in modeling nonstationary panels. The first one is based on the differenced data and was proposed by Bai and Ng (2004). This method allows for consistent estimation of nonstationary static factors and is independent from an integration order of the idiosyncratic component[1]. The second approach uses the data in levels and was introduced by Bai (2004). It is suitable for structural analysis because it directly estimates the dynamic nonstationary factors. The concept can also be easily integrated into the generalized dynamic factor models framework. Unfortunately, the results rely on the stationarity assumption of idiosyncratic errors, which is sometimes difficult to verify.

In this paper, we follow the idea of Bai (2004) and extract factors from data in levels. We contribute to the existing literature by allowing for higher order dynamics in the data generating processes. We show that ignoring the time trend or $I(2)$ processes[2] leads to inconsistent estimation of factors and factors loadings. It has important implications for structural analysis and impulse responses. If it is not taken into consideration then some of the factor loadings grows to infinity and the relative importance of some shock increases unproportionaly. Moreover, we derive the convergence rates, the asymptotic distribution of factors, factor loadings and common components for a general model. The dynamics of the factors are summarized by a scaling matrix. It is chosen to ensure the convergence of the factors second moments. The results allow for the assessment of the accuracy of estimation procedure and for constructing confidence intervals around a rotation of true factors used in empirical analysis.

The theory is illustrated with an empirical example. We analyze a panel of 69 real variables describing the U.S. economy. We show that the data fluctuation can be summarized by a small number of common factors. Since most of the variables have a deterministic trend, then it is relevant to assume an existence of a factor with the time trend. The limiting distributions allows for testing if an interest rate, investments, a personal consumption and government spendings are the driving forces of the economy.

This paper is organized as follows: Section 3.2 describes the model and discusses the estimation issues. In Section 3.3, we derive the convergence rates and asymptotic distributions of estimates for a general model. Section 3.4 analyzes in more detail the model with $I(1)$ factors with a deterministic trend. In Section 3.5, we apply the approach to the panel measuring the real activity of U.S. economy. Finally, in Section 3.6, we summarize and conclude. The description of the data and proofs are provided in Appendix.

[1] The modeling strategy cannot be directly applied for structural analysis because it deals only with the static representation of the factor model. In order to recover dynamic factors, some additional steps have to be introduced, as in Eickmeier (2009).

[2] A process X is $I(d)$ (integrated of order d) if d is a smallest number such that $(1-L)^d X$ is stationary. Here, L denotes a lag operator.

3.2 Model description and estimation

3.2.1 Model setup

Let us denote by X a $N \times T$ panel of observable variables. We use F_t^0, λ_i^0 and r to describe the true common factors, factor loadings and number of factors, respectively. Then for any $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ it is assumed that

$$X_{it} = \lambda_i^{0\prime} F_t^0 + e_{it} \tag{3.1}$$

The residuals e_{it} are $I(0)$ error processes that can be serially correlated. F_t^0 is a $r \times 1$ vector of common factors and λ_i^0 is a $r \times 1$ vector of factor loadings.

Let X_i be $T \times 1$ vector of observations of the ith cross-section unit. Then

$$X_i = F^0 \lambda_i^0 + e_i$$

where $X_i = (X_{i1}, X_{i2}, \ldots, X_{iT})'$, $F^0 = \left(F_1^0, F_2^0, \ldots, F_T^0\right)'$ and $e_i = (e_{i1}, e_{i2}, \ldots, e_{iT})'$.

When it is needed, we will use the following notation

$$X = F^0 \Lambda_0' + e$$

where $\Lambda_0 = \left(\lambda_1^0, \lambda_2^0, \ldots, \lambda_N^0\right)'$ and e is a $N \times T$ matrix, $e = (e_1, ..., e_N)$.

In the model, we distinguish between common factors, F^0, and a common component, denoted by C. The common component is a $T \times N$ matrix that summarizes the total impact of the factors on the panel, defined as a product of factors and factor loadings

$$C = F^0 \Lambda_0'$$

The model setup is similar to the one described in Bai (2003) and Bai (2004). We do not assume any particular type of common factors. Thus, we allow for stationary, $I(1)$ or $I(2)$ factors with or without a deterministic time trend. It is assumed that a kth factor is generated by the following process

$$(1 - L)^d F_{kt}^0 = a_{kt} + u_{kt}$$

where L denotes the lag operator and d takes values $d \in \{0, 1, 2\}$. When $d = 0$ and $a_{kt} = a$ then the process is stationary, whereas for $d = 1$ or 2 the factors are nonstationary $I(1)$ or $I(2)$ processes, respectively. The a_{kt} denotes a deterministic component and u_{kt} is a stationary process. We define by u_t a $r \times 1$ vector of common shocks $u_t = (u_{1t}, ..., u_{rt})'$.

In this article, we are particularly interested in models with nonstationary factors of order not higher than one and a linear time trend. It this case either

$$(1 - L) F_t^0 = a + u_t$$

or

$$F_t^0 = at + u_t$$

Following Bai (2003), we assume that both dimensions of the panel increase to infinity $N, T \to \infty$. Throughout the paper the norm of a matrix is defined as $\|A\| = tr\left(A'A\right)^{1/2}$. We use I_r for a $r \times r$ identity matrix, $\lambda_i\left(A\right)$ for the ith largest eigenvalue of the square matrix A and $v_i\left(A\right)$ for the orthonormal eigenvector of the matrix A associated with the ith largest eigenvalue. Moreover, $[c]$ is a ceiling of the scalar c (it is the smallest integer number, such that $c \leq [c]$). We denote by $\to^p$ and $\to^d$ convergence in probability and distribution, respectively.

3.2.2 Assumptions

The following assumptions are used to derive the asymptotic properties of the estimators. Assumptions B-D are the same as in Bai (2003) and Bai (2004) and are discussed there in detail. We change Assumption A and Assumptions G-F in order to allow for factors with different dynamics.

Assumption A (Common factors)

1. $E\left\|u_t\right\|^{4+\delta} \leqslant M$ for some $\delta > 0$ and all $t \leqslant T$
2. $E\left\|F_1^0\right\|^4 \leqslant M$
3. The nonstationary $I\left(1\right)$ and $I\left(2\right)$ factors are not cointegrated.
4. There exists a diagonal scaling matrix D, which elements are functions of the time dimension T, such that for $T \to \infty$

$$D^{-1}F^{0\prime}F^0D^{-1} \to^d \Sigma$$

where Σ is a random matrix, which is positive definite with probability 1. Moreover, there exists $M \in \Re$ such that for all T

$$T\left\|D^{-2}\right\| \leq M$$

5. The maximum expected value of the normalized factors is bounded

$$\max_t E\left\|\sqrt{T}D^{-1}F_t^0\right\| \leq M$$

6. There exists a limit $\sqrt{T}D^{-1}F_t^0 \to^d F_\tau$ for $t/T = \tau$.

Assumption B (Heterogeneous factor loadings) The loading λ_i^0 is either deterministic, such that $\left\|\lambda_i^0\right\| \leqslant M$, or it is stochastic, such that $E\left\|\lambda_i^0\right\|^4 \leqslant M$. In both cases

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime} = \Lambda_0'\Lambda_0/N \to^p \Sigma_\Lambda$$

as $N \to \infty$ for some nonrandom, positive definite matrix Σ_Λ. Moreover, the matrix $\Sigma_\Lambda\Sigma$ has distinct eigenvalues with probability one.

Assumption C (Idiosyncratic component)

1. $E(e_{it}) = 0$ and $E|e_{it}|^8 \leqslant M$
2. $E(e_s' e_t/N) = \gamma_{NT}(s,t)$ with $|\gamma_{NT}(s,s)| \leqslant M$ for all s, and
$$\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_{NT}(s,t)| \leqslant M$$
3. $E(e_{is}e_{jt}) = \pi_{ij,st}$ with $|\pi_{ij,tt}| \leqslant |\pi_{ij}|$ for some π_{ij} and for all t.
$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\pi_{ij}| \leqslant M$$
4. $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}|\pi_{ij,st}| \leqslant M$

Assumption D $\{\lambda_i\}$, $\{e_t\}$, $\{u_t\}$ are mutually independent stochastic variables.

Assumptions A-D are necessary to prove the consistency of the estimators. Assumption A allows for factors with different dynamics. If all factors are stationary then the scaling matrix $D = \sqrt{T}I_r$, whereas if there are both stationary and nonstationary $I(1)$ and $I(2)$ factors without a time trend then D can be defined as follows

$$D = \begin{bmatrix} T^2 I_{r_2} & 0 & 0 \\ 0 & T I_{r_1} & 0 \\ 0 & 0 & \sqrt{T} I_{r_0} \end{bmatrix} \tag{3.2}$$

where r_k denotes the number of $I(k)$ factors. In Bai (2004), there are only $I(0)$ and $I(1)$ factors and the scaling matrix takes the form

$$D = \begin{bmatrix} T I_r & 0 \\ 0 & \sqrt{T} I_q \end{bmatrix} \tag{3.3}$$

where r and q denotes the number of nonstationary and stationary common factors, respectively.

Remark 3 *If we allow for deterministic time trends then the scaling matrix D needs to be adjusted. Suppose the factors have a linear trend. Then, an element scaled by $T^{3/2}$ needs to be added to the diagonal of D. An exception is a model in which only the $I(2)$ factors have a linear (not quadratic) trend. In this case the scaling matrix remains unchanged as in (3.2). A model with $I(1)$ factors and the linear trend is discussed in detail in Section 3.4.*

In order to identify the number of nonstationary factors, we need to assume that they are not cointegrated. Otherwise, the space spanned by the factors could be described by the lower number of common trends G^0 and a stationary component. Hence, we would be able to reduce the number of nonstationary factors by substituting the corresponding vectors of F^0 by G^0 and the stationary term.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Assumption B is standard and is introduced to ensure that the factors load to infinitely many variables. It allows us to distinguish between a common component that is pervasive and an idiosyncratic component that has a limited effect. Hence, it ensures that the factor structure is identifiable. Assumption C describes a possible time and cross-sectional dependence of the idiosyncratic components. It is extensively discussed in Bai (2004). Assumption D excludes the correlation between the idiosyncratic and common shocks. It is not restrictive because in further analysis we allow for a dynamic structure of the factors.

In order to show a stronger result, we need to impose an additional Assumption E. It restricts the correlation of the idiosyncratic errors.

Assumption E

Let us denote $\bar{\gamma}_N(t,s) = E(|e_s' e_t/N|)$. Then there exists $M \leqslant \infty$ such that

1. For each t, $\sum_{s=1}^{T} |\bar{\gamma}_N(t,s)| \leqslant M$
2. For each i, $\sum_{j=1}^{N} |\pi_{ij}| \leqslant M$

Some moment conditions are introduced in Assumption F. The first two conditions F.1 and F.2 are needed to prove consistency and to compute the convergence rates. Finally, deriving the asymptotic distributions of estimators requires additional information about the limiting distribution of $N^{-1/2}\sum_{i=1}^{N}\lambda_i^0 e_{it}$ and $D^{-1}\sum F_t^0 e_{it}$. It is provided by Assumptions F.3 and F.4. If the loadings are deterministic then the Assumption F.3 follows from the Central Limit Theorem and the fact that the loadings are bounded. Otherwise, we assume, as in Bai (2004), that the limiting distribution of the first sum is normal.

Assumption F (Moments and Central Limit Theorem)

1. There exists $M < \infty$ such for every pair (s,t),

$$E\left|N^{-1/2}\sum_{i=1}^{N}\left[e_{it}e_{is} - E\left(e_{it}e_{is}\right)\right]\right|^4 \leq M$$

2. There exists $M < \infty$ such that for any T

$$E\left|\frac{1}{T^{1/2}}\sum_{t=1}^{T} D^{-1}F_t^0\Lambda_0' e_t\right|^2 \leq M$$

3. For each t as $N \to \infty$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it} \to^d N(0,\Gamma_t)$$

where $\Gamma_t = \lim_{N\to\infty}(1/N)\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i^0\lambda_j^{0\prime}E(e_{it}e_{jt})$

4. For each i as $T \to \infty$ there exists a random variable W_i, such that

$$D^{-1} \sum_{t=1}^{T} F_t^0 e_{it} \to^d W_i$$

The distribution of the random variable W_i depends on the dynamics of the factors. If the kth factor is stationary or $I(1)$ with a time trend, then W_{ki} has a normal distribution, whereas if F_{kt} is $I(1)$ without deterministic trend then the distribution of W_{ki} is a functional of a Brownian motion, as in Bai and Ng (2004).

3.2.3 Estimation

Estimates of Λ and F are obtained by solving the optimization problem

$$\begin{aligned}\left(\tilde{\Lambda}, \tilde{F}\right) &= \arg\min_{\Lambda,F} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i F_t)^2 \\ &= \arg\min_{\Lambda,F} tr\left(\left(X - F\Lambda'\right)'\left(X - F\Lambda'\right)\right)\end{aligned}$$

where $X = \left(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_N\right)$ and $F = (F_1, F_2, \ldots, F_T)'$. For any non-zero F the optimal loading matrix is

$$\tilde{\Lambda}' = \left(FF'\right)^{-1} F'X \tag{3.4}$$

and

$$X - F\tilde{\Lambda}' = \left(I_T - F\left(F'F\right)^{-1} F'\right) X$$

Define $P_F = F\left(F'F\right)^{-1} F'$. Then the optimal vector of factors F is

$$\begin{aligned}\tilde{F} &= \arg\min_{F} tr\left(\left(X - F\tilde{\Lambda}'\right)'\left(X - F\tilde{\Lambda}'\right)\right) \\ &= \arg\min_{F} tr\left(X'\left(I_T - P_F\right)'\left(I_T - P_F\right)X\right) \\ &= \arg\min_{F} tr\left(X'\left(I_T - P_F\right)X\right) \\ &= \arg\max_{F} tr\left(X'P_F X\right)\end{aligned}$$

In order to solve the above problem, we need to impose some normalization of the factors. It is standard to assume that the product of scaled factors gives the identity matrix,

$$D^{-1}F'FD^{-1} = I_r$$

Then

$$\begin{aligned} P_F &= F\left(F'F\right)^{-1\prime}F' \\ &= FD^{-2}F' \end{aligned}$$

and the problem is equivalent to maximizing

$$tr\left(X'FD^{-2}F'X\right) = tr\left(D^{-1}F'XX'FD^{-1}\right)$$

Thus, the estimated common factors $\tilde{F}$ are proportional to the eigenvectors v corresponding with the r largest eigenvalues of the $T \times T$ matrix XX'.

$$\tilde{F} = B \cdot v$$

The scaling matrix B is diagonal and is chosen to satisfy the normalization condition

$$I_r = D^{-1}F'FD^{-1} = D^{-1}Bv'vBD^{-1} = D^{-1}BBD^{-1}$$

Thus, $B = D$ and $\tilde{F}$ is D times the eigenvectors v

$$\tilde{F} = vD \tag{3.5}$$

The estimate of the loading matrix is obtained on the basis of (3.4) and is equal to

$$\tilde{\Lambda}' = D^{-2}\tilde{F}'X \tag{3.6}$$

The results correspond with the outcomes of Bai and Ng (2002) and Bai (2004) with $D = \sqrt{T}I_r$ or $D = TI_r$, respectively. In the first case, the estimated factors are the eigenvectors v multiplied by $\sqrt{T}$. In a model with $I(1)$ factors without drift, the estimators are $\tilde{F} = vT$. In the Generalized Factor Model (GFM) presented by Bai (2004), the scaling matrix is (3.3). Thus, the estimates of the nonstationary factors are the eigenvectors corresponding with the r largest eigenvalues multiplied by T, whereas the estimates of the stationary factors are the eigenvectors corresponding with the $r+1 : r+q$ largest eigenvalues multiplied by $\sqrt{T}$.

In further analysis, we consider also another normalization of factors and factor loadings. The following lemma defines so called normalized factors, $\hat{F}$, and normalized loadings, $\hat{\Lambda}$.

Lemma 4 *Define normalized factors $\hat{F} = N^{-1}X\tilde{\Lambda}$ and a normalized loading matrix $\hat{\Lambda}$ such that $\hat{F}\hat{\Lambda}' = \tilde{F}\tilde{\Lambda}$. Then*

$$\begin{aligned} \hat{\Lambda} &= \tilde{\Lambda}V_{NT}^{-1} \\ \hat{F} &= \tilde{F}V_{NT} \end{aligned}$$

where $V_{NT} = \tilde{V}_{NT}D^{-2}/N$ and $\tilde{V}_{NT}$ is the diagonal matrix consisting of the r largest eigenvalues of the matrix XX'.

This lemma shows how the two different estimators $\hat{F}$ and $\tilde{F}$ are related to each other. It is used to derive the asymptotic distribution of $\tilde{F}$ and to construct the confidence intervals around a rotation of the true factors.

3.3 Distribution theory

In this section, we present an asymptotic theory of estimated factors, factor loadings and a common component. Firstly, we discuss the consistency issue and derive the asymptotic distribution. Finally, we show how the confidence intervals of a rotation of the true factors can be constructed.

3.3.1 Consistency

Bai (2003) and Bai (2004) prove consistency of the estimators of stationary and random walk factors. They show that the mean squared errors of the estimated factors are $O_p\left(\max\left\{N^{-1}, T^{-1}\right\}\right)$ and $O_p\left(\max\left\{N^{-1}, T^{-2}\right\}\right)$, respectively. Using similar arguments, we show that the MSE of an estimated factors with a scaling matrix D is $O_p\left(\max\left\{N^{-1}, \left\|D^{-2}\right\|\right\}\right)$. Moreover, for a given time period t the error $\hat{F}_t - H'F_t^0$ is $O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$.

Consider firstly the MSE of estimated factor.

Proposition 5 *Assume Assumptions A-D hold. There exists a nonsingular matrix* $\tilde{H}$ *and* $\delta_{NT}^{-1} = \max\left\{N^{-1/2}, \left\|D^{-1}\right\|\right\}$ *such that*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - \tilde{H}'F_t^0\right\|^2 = O_p\left(\delta_{NT}^{-2}\right)$$

The proposition states that the time average of a squared deviation between the estimated factors and the rotation of the true factors converges to zero with a growing sample size $N, T \to \infty$. The proposition is very important because it shows that the factors can be consistently estimated with a principle component method. The convergence rates are used to derive the asymptotic distribution of the estimators.

The result is in line with the existing literature. In a case of a model with stationary factors, the norm of the scaling matrix $\left\|D^{-1}\right\| = O_p\left(T^{-1/2}\right)$. Therefore, the convergence rate is $\delta_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$, as in Bai (2003). If we assume that all the factors are random walks without a drift then $\left\|D^{-1}\right\| = O_p\left(T^{-1}\right)$ and $\delta_{NT} = \min\left\{\sqrt{N}, T\right\}$. The convergence rate corresponds with the outcome presented in Bai (2004).

Finally, it is shown that for a given time period t the error converges to zero with a growing cross-sectional and time dimension. To prove the convergence rates we need to impose the more restrictive Assumption E.

Proposition 6 *Under Assumptions A-E the following holds for each* t*,*

$$\tilde{F}_t - \tilde{H}'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$$

The convergence rate is the same as in Bai and Ng (2002) for stationary factors. Since we allow for different types of factors then the rate is lower then

in Bai (2004), where only $I(1)$ factors without a trend are considered. It is, however, sufficient to derive the limiting distribution of factors.

Remark 7 *If we allow for only one type of nonstationary factors, for example $I(1)$ or $I(2)$ factors, then it is shown by Lemma 25 that*

$$\tilde{F}_t - \tilde{H}' F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)$$

This is in line with the results of Bai (2004) for the $I(1)$ factors without a time trend, where

$$\tilde{F}_t - \tilde{H}' F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-3/2}\right)$$

3.3.2 Asymptotic distributions

We investigate the asymptotic distribution of the estimated factors, the factor loadings and the common component. Firstly, we describe a limiting behavior of V_{NT} and $D^{-2}\tilde{F}' F^0$.

Lemma 8 *Under assumptions A-E, as $N, T \to \infty$*

1. *There exists a random, diagonal, full rank with probability 1 matrix V such that $V_{NT} \to^d V$*
2. *There exists a random, positive definite, with probability 1 matrix Q such that*

$$D^{-2}\tilde{F}' F^0 = Q_{NT} \to^d Q$$

The lemma defines two matrices, V and Q, used to describe the asymptotic distribution of factors and factors loadings.

Limiting distribution of estimated common factors

The following proposition shows that the factor estimates are asymptotically normal. This property is used to construct the confidence intervals around the rotation of the true factors.

Proposition 9 *Under Assumptions A-F, as $N, T \to \infty$ and $N^{1/2}\left\|D^{-1}\right\| \to 0$ we have for each t*

$$\sqrt{N}\left(\tilde{F}_t - \tilde{H}' F_t^0\right) \to^d \Sigma_\Lambda^{-1} N\left(0, \Gamma_t\right)$$

where Σ_Λ and Γ_t are defined as in the Assumptions B and F.

The proposition requires restrictions on the relation between the cross-sectional and the time dimensions. If there are stationary factors the conditions say that $N/T \to 0$. In a case of a model with only nonstationary factors without the deterministic trend, the condition is $N/T^2 \to 0$. If there is only one type of factor, it can be shown that the condition is $N^{1/2}T^{-1/2}\left\|D^{-1}\right\|$ as in Bai (2004)[3].

The results will be used to construct the confidence intervals around a rotation of true factors.

[3] The condition follows directly from Lemma 25 and the proof of Proposition 9.

Limiting distribution of estimated factors loadings

In this section, we show that the estimated factor loadings converges to some random variable.

Proposition 10 *Under the Assumptions A-F, for each i, as $N,T \to \infty$ we have*

$$D\left(\check{\lambda}_i - \check{H}^{-1}\lambda_i^0\right) \to^d \left(\bar{H}\right)^{-1}\Sigma^{-1}W_i$$

with $\bar{H}$ is defined by Lemma 28. Σ and W_i are defined by Assumption A and F, respectively.

The actual limiting distribution of factor loadings depends on the dynamics of the factors. As shown in Bai (2003), if the factors are stationary then the matrix Σ converges to the factors variance-covariance matrix. On the other hand, if all factors are random walks without a drift then Σ is defined by a Brownian motion. Moreover, if we allow for other types of factors then the elements of the random matrix Σ may take different forms.

Limiting distribution of estimated common components

Let us denote the true and estimated common components[4] by $C_{it}^0 = F_t^0\lambda_i^0$ and $\hat{C}_{it} = \hat{F}_t\hat{\lambda}_i$, respectively. The limiting distribution of the estimates of the common component depends on the relation between the cross-sectional and time dimensions T/N.

Proposition 11 *Under Assumptions A-G as $N,T \to \infty$ it holds that*

1. *If $N/T \to 0$ then for each pair (i,t)*

$$\sqrt{N}\left(\hat{C}_{it} - C_{it}^0\right) \to^d \lambda_i^{0\prime}H^{-1\prime}QN\left(0,\Gamma_t\right)$$

where Γ_t is defined in Assumption F and Q is introduced in Lemma 8.

2. *If $T/N \to 0$ then for each pair (i,t) and $t = [\tau T]$*

$$\sqrt{T}\left(\hat{C}_{it} - C_{it}^0\right) \to^d F_\tau'\Sigma^{-1}W_i$$

where Σ and F_τ are defined in Assumption A and W_i is defined in Assumption F.

3. *If $N/T \to \pi$ then for each pair (i,t) and $t = [\tau T]$*

$$\sqrt{N}\left(\hat{C}_{it} - C_{it}^0\right) \to^d \lambda_i^{0\prime}H^{-1\prime}QN\left(0,\Gamma_t\right) + \sqrt{\pi}F_\tau'\Sigma^{-1}W_i$$

where Q, Γ_t, F_τ,Σ and W_i are defined as above.

[4] The estimated common component $\hat{C}_{it}$ does not depend on the normalization of common factors.

As noted by Bai (2004), the third case is the most useful in practice, because π can be estimated by the sample ratio N/T. Moreover, the distribution of the common components in cases (2) and (3) depends on the limiting distribution of F_τ. When the factor $\bar{F}_i^0$ is stationary then $F_{\tau i}$ is normally distributed. However, if the factor $\bar{F}_i^0$ is a $I(1)$ process without a deterministic drift then $F_{\tau i}$ is a Brownian motion process with a variance described by Bai and Ng (2004).

Confidence intervals

In the article, we interpret a scalar, observable variable R_t as a common factor if it is a linear combination of the true factors plus a constant.

$$R_t = \alpha + \beta' F_t^0$$

where α is a shift parameter and β is a $r \times 1$ vector that summarize the relation between R_t and F_t^0. We allow for both a rotation and a shift of the factors because neither R_t nor F_t^0 have to be zero mean processes and they may have different levels and scalings.

Consider the rotation of $\tilde{F}$ toward R_t described by the regression

$$\begin{aligned} R_t &= \alpha + \beta' \left(\tilde{H}^{-1\prime} \tilde{F}_t \right) + u_t \\ &= \alpha + \delta' \tilde{F}_t + u_t \end{aligned}$$

Let $\left(\hat{\alpha}, \hat{\beta}\right)$ be the least-square estimator of (α, β) and $\hat{R}_t = \hat{\alpha} + \hat{\beta}' \left(\tilde{H}^{-1\prime} \tilde{F}_t \right)$. From the identity $\delta' = \beta' \tilde{H}^{-1\prime}$ it follows that $\hat{\delta}' = \hat{\beta}' \tilde{H}^{-1\prime}$. If R_t is a common factor then the following proposition holds.

Proposition 12 *Under the Assumptions A-F and no cross-section correlation for the idiosyncratic errors, as* $N, T \to \infty$ *and* $N^{1/2} \left\| D^{-1} \right\| \to 0$

$$\sqrt{N} \left(\hat{R}_t - \alpha - \beta' F_t^0 \right) \to^d \hat{\delta} V^{-1} Q N(0, \Gamma_t)$$

where V, Q *are defined in Lemma 8 and* Γ_t *is introduced in Assumption F.*

Following Bai (2004), we will approximate the 95% confidence intervals as follows

$$\left(\hat{R}_t - 1.96 \sqrt{\tilde{S}_t^2 / N}, \hat{R}_t + 1.96 \sqrt{\tilde{S}_t^2 / N} \right) \tag{3.7}$$

where $\tilde{S}_t^2 = \left(\hat{\delta} V^{-1} Q \right) \Gamma_t \left(\hat{\delta} V^{-1} Q \right)'$.

Remark 13 *As stated in Bai (2003), the matrix*

$$\hat{\delta} V^{-1} D^{-2} \tilde{F}' F^0 \Gamma_t F^{0\prime} \tilde{F} D^{-2} V^{-1} \hat{\delta}'$$

involves the product of $F^0\Lambda_0$*, which can be consistently estimated with* $\tilde{F}\tilde{\Lambda}$*. Hence, it can be substituted by*

$$\hat{\delta}V^{-1}D^{-2}\tilde{F}'\tilde{F}\tilde{\Gamma}_t\tilde{F}'\tilde{F}D^{-2}V^{-1}\hat{\delta}' = \hat{\delta}V^{-1}\tilde{\Gamma}_tV^{-1}\hat{\delta}'$$

where

$$\tilde{\Gamma}_t = \lim_{N\to\infty}(1/N)\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_j'\hat{E}\left(e_{it}e_{jt}\right)$$

Remark 14 *Bai and Ng (2006) propose two types of estimators of the matrix* $\tilde{\Gamma}_t$ *that can be used for cross sectionally uncorrelated idiosyncratic errors* e_{it}

1. $\tilde{\Gamma}_t = \frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\tilde{\lambda}_i\tilde{\lambda}_i'$
2. $\tilde{\Gamma} = \tilde{\sigma}_\varepsilon^2\frac{1}{N}\sum_{i=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_i'$*, where* $\tilde{\sigma}_\varepsilon^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{e}_{it}^2$ *for errors with equal variances* $\sigma_{\varepsilon_i}^2 = \sigma_\varepsilon^2$*.*

$\tilde{\lambda}_i$ *and* $\tilde{e}_{it}$ *correspond to the estimates of* λ_i *and* e_{it}*.*

Remark 15 *If the observable variable* R_t *belongs to the panel* ($R_t = X_{it}$) *then the parameters* $\left(\hat{\alpha},\hat{\delta}\right)$ *can be replaced with* $\left(0,\hat{\lambda}_i\right)$*, where* $\hat{\lambda}_i$ *are estimated factor loadings.*

Remark 16 *In order to compute the confidence intervals, we need to ensure that the idiosyncratic errors have zero mean*[5]*. Otherwise* $E\left(\hat{R}_t - \alpha - \beta F_t^0\right) \neq 0$ *and* $\tilde{\Gamma}_t$ *will not be a consistent estimator of the variance-covariance matrix* Γ_t*.*

3.4 Model with $I\left(1\right)$ factors with a deterministic trend

So far, the literature considers only models with either stationary factors or common trends without deterministic drift. Since most of time series have both stochastic and deterministic trends, the theory does not match the needs of macroeconomic modeling. Thus, we believe that the model that allows for a deterministic trend is interesting, especially from an empirical point of view.

In this section, we discuss in more detail issues associated with an estimation of a factor model with a linear time trend. We address the problem of determining a number of common trends with a drift. We show the convergence rates, limiting distributions of factors, factor loadings and common components. Finally, we present the results in the context of a generalized factor model, as in Bai (2004).

[5] One possible way to construct idiosyncratic errors with zero mean is to remove the mean form the orginal data set.

3.4.1 Modeling the time trend vs. detrending the data

Once we decide, on the basis of analysis of the variables in panel, that the deterministic trend plays an important role in the model, we may consider two strategies. The first approach leaves the data unchanged and models the trend together with other factors. It is discussed in detail in the following sections. The second approach consists of two steps. Firstly, the data is detrended and secondly, the factors without trend are estimated as in Bai (2004). Its main disadvantage is that it requires either a precise parametrization of the time trend or a usage of some nonlinear filtering procedures. There is no agreement on which of the detrending methods should be used in the context. Therefore, we believe that our approach is a competitive alternative.

3.4.2 Number of common factors with a drift

The first issue is the number of identifiable common trends with a deterministic drift. We show that a model with $n > 1$ common factors with time trends can be represented as a model with only one factor with time trend and $n-1$ factors without a deterministic drift. Consider a system with n factors, $F_t = (F_{1t}, ..., F_{nt})'$, that depends both on the time trend and a stochastic, zero mean variable $\omega_t = (\omega_{1t}, ..., \omega_{nt})'$

$$\begin{aligned} F_t &= At + B\omega_t \\ &= [A_{n\times 1} : B_{n\times n}] \begin{bmatrix} t \\ \omega_t \end{bmatrix} \end{aligned}$$

The matrix $[A_{n\times 1} : B_{n\times n}]$ needs to have a rank n in order to ensure that all the factors are identifiable. Since the factors are assumed to follow a deterministic time trend, the vector A has to be non-zero. Then the system can be rewritten as follow

$$F_t = C\,[I_{n\times n} : D_{n\times 1}] \begin{bmatrix} t \\ \omega_t \end{bmatrix}$$

where C is a $n \times n$ full rank matrix and $I_{n\times n}$ is an identity matrix. Let us construct a new set of factors $\tilde{F}_t = C^{-1}F_t$. Then

$$\tilde{F}_t = [I_{n\times n} : D_{n\times 1}] \begin{bmatrix} t \\ \omega_t \end{bmatrix}$$

and

$$\begin{aligned} \tilde{F}_{1t} &= t + D_{11}\omega_{nt} \\ \tilde{F}_{2t} &= \omega_{1t} + D_{21}\omega_{nt} \\ &\vdots \\ \tilde{F}_{nt} &= \omega_{n-1t} + D_{n1}\omega_{nt} \end{aligned}$$

Thus, among the factors $\tilde{F}_t$ only the first one has a time trend. Moreover, if all the factors are nonstationary and noncointegrated then at least $n-1$ of the ω_t elements have to be $I(1)$ processes. We can order the elements of ω_t in such a way that only the last component ω_{nt} is allowed to be stationary. Depending on integration order of ω_{nt} the first factor $\tilde{F}_{1t}$ will be trend stationary (when ω_{nt} is $I(0)$) or a random walk with a drift (when ω_{nt} is $I(1)$).

We have shown that the factors F_t are a linear combination of $\tilde{F}_t$, where only one factor $\tilde{F}_{1t}$ has a time trend. Without loss of generality we can replace F_t with $\tilde{F}_t$. Therefore, in further analysis, we assume that there is only one common factor with a deterministic linear trend.

Remark 17 *The arguments are valid if the trend is not linear but is a function of time $f(t)$ and loads with weights A to the factors. Then, the factors F_t can be replaced with $\tilde{F}_t$, where only one of the elements of $\tilde{F}_t$ has a deterministic component and other elements have a zero mean.*

3.4.3 Static factor model

Let us first consider a static factor model with a single nonstationary factor with a deterministic time trend. Some of the restrictive assumptions on the total number of factors and the relation between factors and observable variables will be relaxed in the Section 3.4.4, where a generalized dynamic factor model will be discussed.

Define by F_t a common nonstationary factor with a deterministic trend such that it is either $I(1)$ with a drift

$$F_t = a + F_{t-1} + u_t \tag{3.8}$$

or trend stationary.

$$F_t = at + u_t$$

with $a \neq 0$.

Since the factor has a time trend then it needs to be scaled by $T^{3/2}$. Hence, the scaling matrix $D = T^{3/2}$ and the limit of $D^{-1}F^{0\prime}F^0D^{-1} = T^{-3}\sum_{t=1}^{T}\left(F_t^0\right)^2$ equals a scalar $\Sigma = a^2/3$

$$\begin{aligned} T^{-3}\sum_{t=1}^{T}\left(F_t^0\right)^2 &= T^{-3}\sum_{t=1}^{T}a^2t^2 + o_p(1) \\ &= \sum_{t=1}^{T}a^2\left(\frac{t}{T}\right)^2\frac{1}{T} + o_p(1) \\ &\to \int_0^1 a^2x^2dx = a^2/3 \end{aligned}$$

For $t = [\tau T]$ the limit of F_t^0/T is $F_\tau = a\tau$. For a $I(1)$ factor

$$\begin{aligned} \frac{1}{T}F_t^0 &= a\frac{t}{T} + \frac{t}{T}\frac{1}{t}\sum_{s=1}^{t} u_s \\ &\rightarrow {}^p a\tau + \tau E u_t = a\tau \end{aligned}$$

and for a trend stationary factor

$$\begin{aligned} \frac{1}{T}F_t^0 &= a\frac{t}{T} + \frac{1}{T}u_t \\ &\rightarrow {}^p a\tau \end{aligned}$$

Moreover, it can be assumed that for each i, as $T \rightarrow \infty$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{T}F_t^0 e_{it} \rightarrow^d N(0, \Omega_i)$$

where $\Omega_i = \lim_{N\rightarrow\infty}(1/T)\sum_{t=1}^{T}\sum_{s=1}^{T} a^2 \frac{ts}{T^2} E(e_{it}e_{is})$. Thus, the variable W_i has a normal distribution.

Remark 18 *Suppose the deterministic trend is not linear and is described by a function $f(t)$. Then as long as*

$$0 < \lim_{T\rightarrow\infty} T^{-3}\sum_{t=1}^{T}(f(t))^2 < M$$

and

$$0 < \lim_{T\rightarrow\infty}\frac{1}{T}f(\tau T) < M$$

then the results hold and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{T}F_t^0 e_{it} \rightarrow^d N(0, \Omega_i)$$

The matrix Ω_i takes the following form form

$$\Omega_i = \lim_{N\rightarrow\infty}(1/T)\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{f(t)f(s)}{T^2}E(e_{it}e_{is})$$

Estimation

In Section 3.2, we derived estimators of the factor and factor loadings. Since $D = T^{3/2}$ then

$$\begin{aligned} \tilde{F} &= T^{3/2}v \\ \tilde{\Lambda}' &= T^{-3}\tilde{F}'X \end{aligned}$$

where $v = v_1(XX')$ is the eigenvector corresponding with the largest eigenvalue of the matrix XX'. Hence, the normalized factor and loadings can be computed as in Lemma 4, with V_{NT} being the largest eigenvalue of the matrix $XX'/(NT^3)$.

Convergence rates

The convergence rates can be computed on the basis of Proposition 5 and Lemma 25. Since $\left\|D^{-1}\right\| = T^{-3/2}$ then

$$\delta_{NT}^{-1} = \max\left\{N^{-1/2}, T^{-3/2}\right\}$$

and

$$\hat{F}_t - H'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-2}\right)$$

The convergence rates are higher than in the model with only stationary factors or common trends without a drift.

3.4.4 Generalized dynamic factor model

Finally, consider the generalized dynamic factor model with both stationary and nonstationary factors

$$X_{it} = \lambda_i^r(L)F_t^r + \lambda_i^q(L)F_t^q + e_{it} \tag{3.9}$$

where $\lambda_i^r(L)$ and $\lambda_i^q(L)$ are lag polynomials corresponding to different types of factors: F_t^r is a $r \times 1$ vector of common nonstationary factors with the first factor having a time trend and F_t^q is a $q \times 1$ vector of stationary factors. Hence, in the generalized dynamic factor model we allow for more then one factor: there are r nonstationary and q stationary dynamic factors.

$$\begin{aligned} F_t^r &= A + F_{t-1}^r + u_t^r \\ F_t^q &= u_t^q \end{aligned}$$

with $A = (a, 0, \ldots, 0)'$. Following Bai (2004) and Forni, Hallin, Lippi and Reichlin (2003), we assume

$$\begin{aligned} \lambda_i^r(L) &= \sum_{j=0}^{\infty} \lambda_{ij}^r L^j \\ \lambda_i^q(L) &= \sum_{j=0}^{\infty} \lambda_{ij}^q L^j \end{aligned}$$

with $\sum_{j=0}^{\infty} j\left|\lambda_{ij}^r\right| < \infty$ and $\sum_{j=0}^{\infty} j\left|\lambda_{ij}^q\right| < \infty$.

Since there are three types of factors the scaling matrix takes the form

$$D = \begin{bmatrix} T^{3/2} & 0 & 0 \\ 0 & T \cdot I_{r-1} & 0 \\ 0 & 0 & T^{1/2} \cdot I_q \end{bmatrix} \tag{3.10}$$

Static representation

The dynamic representation of the model (3.9) can not be directly estimated. In order to construct the estimators, we need to rewrite the model in the static form. Let us notice that (3.9) can be expressed as follows

$$\begin{aligned} X_{it} &= \lambda_i^r(L) F_t^r + \lambda_i^q(L) F_t^q + e_{it} \\ &= \varphi F_t^r + \phi^r(L) \Delta F_t^r + \lambda_i^q(L) F_t^q + e_{it} \end{aligned}$$

where the factors ΔF_t^r and F_t^q are stationary. In order to derive the asymptotic distributions, we need to approximate the model with finite order lag polynomials. Let us assume that $\phi^r(L)$, $\lambda_i^q(L)$ have an order p. Then the model can be written as

$$X_{it} = \varphi F_t^r + \Phi G_t \tag{3.11}$$

where $G_t = \left(\Delta F_t^r, \ldots, \Delta F_{t-p}^r, F_t^q, \ldots, F_{t-p}^q\right)'$ summarizes the stationary factors. Thus, the model has the static form that uniquely identifies the dynamic nonstationary factors[6] F_t^r. The representation (3.11) will be used in further analysis.

Estimation of the number of factors

In order to estimate the total number of factors, Bai (2004) proposes to use the data in first differences[7]. If the data are $I(1)$ then

$$\Delta X_{it} = \lambda_i^r(L) \Delta F_t^r + \lambda_i^q(L) \Delta F_t^q + \Delta e_{it} \tag{3.12}$$

and both ΔX_{it} and factors ΔF_t^r, ΔF_t^q are stationary. Therefore, the information criteria PC introduced by Bai and Ng (2002) can be applied. As stated in Bai (2004) the procedure allows for consistent estimation of the total number of factors (both stationary and nonstationary).

The second issue is determining the number of stationary and nonstationary factors separately. Bai (2004) shows that the number of nonstationary, dynamic factors can be estimated directly form the data in levels on the basis of representation (3.11). Bai and Ng (2004) constructs the information criteria IPC and proves their consistency for panels without a deterministic trend. In the paper, it is stated that the same information criteria can be used to estimate the total number of nonstationary factors regardless of the existence of the deterministic components and the order of integration. The number of stationary static factors, G_t, can be computed as the difference between the total number of factors and the number of nonstationary dynamic factors as in Bai (2004).

[6] The identification is achieved under the assumption of no cointegration between the nonstationary factors. See Bai (2004) for a discussion.

[7] The aim of the differencing is to ensure that the common factors are stationary. Therefore, the order of differencing should equal to the integration order of the data.

Estimation and convergence rates

Since the number of factors can be consistently estimated with the information criteria as in Bai and Ng (2002) and Bai (2004), then we assume that the true number of both stationary and nonstationary factors is known. The common factors can be estimated as follow

$$\hat{F} = vD$$

where v are the eigenvectors corresponding with the $(r+q)$ largest eigenvalues of a matrix XX' and D is given by (3.10). Thus,

1. A nonstationary common trend with a drift is estimated as the eigenvector corresponding to the largest eigenvalue of the matrix XX' multiplied by $T^{3/2}$.

2. Nonstationary common trends without a drift are estimated as the eigenvectors corresponding to $2:r$ largest eigenvalues of the matrix XX' multiplied by T.

3. Stationary common trends are estimated as the eigenvectors corresponding to $(r+1):(r+q)$ largest eigenvalues of the matrix XX' multiplied by $T^{1/2}$.

Let V_{NT} be a diagonal matrix defined in Lemma 4. It has diagonal elements V_i such that

1. V_1 is the largest eigenvalue of the matrix XX'/NT^3.

2. $V_2, \ldots, V_r$ are the $2:r$ largest eigenvalues of the matrix XX'/NT^2.

3. $V_{(1+r)}, \ldots, V_{(r+q)}$ are the $(r+1):(r+q)$ largest eigenvalues of the matrix XX'/NT.

Finally, we present the convergence rates. Since $\left\|D^{-1}\right\| = O_p\left(T^{-1/2}\right)$ then the convergence rates are $\delta_{NT}^{-1} = \min\left\{N^{-1/2}, T^{-1/2}\right\}$ and

$$\hat{F}_t - H'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\right)$$

3.5 Empirical example

In the paper, we study the behavior of 69 variables describing the real activity of US economy (an industrial production, components of the real GDP, two measures of the labor productivity and interest rates). The data are quarterly, spanning the period from January 1961 to September 2008. The description of the data is provided in the Appendix. Most of the variables in the panel are nonstationary and have both deterministic and stochastic trends.

3.5.1 Normalization

The literature on stationary panels underlines the need for data normalization. Usually, variables in panels are divided by their standard deviations. This approach ensures that all variables have equal input to the total variability of the panel. Therefore, the estimation method does not favour any of them. Moreover, the normalization does not change the theoretical results because it is associated with multiplying the data by a diagonal matrix that converges to an invertible matrix of asymptotic standard deviations.

This method cannot be directly applied for nonstationary panels because the standard deviations diverge to infinity. Thus, it will affect the asymptotic results of the estimation method. In order to normalize the data, we propose dividing them by

$$\sigma_i = \left(\sum_{t=1}^{T} \left(X_{it} - \mu_i \right)^2 / T^{n_i} \right)^{1/2}$$

where μ_i denotes the mean of the variable X_i and n_i is chosen to ensure that $\sigma_i = O_p(1)$ and that σ_i has a limit. For example, if a variable X_i is stationary then $n_i = 1$ and if X_i is an $I(1)$ process without a deterministic drift then $n_i = 1.5$. Finally, for a $I(1)$ variable X_i with a time trend there is $n_i = 2$.

The normalization ensures that the variables with the same type of dynamics have the same volatility. It has an intuitive interpretation for processes without time trends because it corresponds to a standard deviation. For data with a deterministic trend, the normalization guarantees that in the limit the slope of the trend equalize across the panel variables. Thus, it standardizes the main source of the volatility.

3.5.2 The number of factors

Firstly, we estimate the number of nonstationary factors using the IPC information criteria described by Bai (2004) and applied for data in levels. We assume that there are not more then ten common trends. Thus, we consider cases, in which $k_{\max} \leq 10$. The results are presented in Table 3.2 and indicate that there are either two or three nonstationary factors.

Finally, we estimate the number of factors from differenced data with the PC criteria described in Bai (2003). The criteria do not give conclusive results because they always choose the maximum permitted number of factors. It may indicate that either the model has a long lag structure or the cross sectional sample size is too small to provide correct estimates.

The literature discusses some alternative approaches that can be used to select the number of factors. Child (2006) provides a review of less formal, graphical methods that can be applied in this context. They are based on the eigenvalues of the panel correlation matrix. It can be seen that the sum of these eigenvalues equals the cross sectional dimension N. Therefore, the first approach is to look at the number of the eigenvalues larger then one and hence,

above the average. This criterion indicates 18 common factors, which explains 83.25% of the total variability. As stated by Child (2006), the large cross sectional dimension leads to overestimation of the number of factors. Hence, we analyze the plot of the correlation matrix eigenvalues and use a Scree test[8]. The eigenvalues are presented in Figure 3.1 and indicate that there are around ten common factors. The plot starting from the eleventh eigenvalue is almost linear and decreases steadily to zero. The first ten common factors explain 67.85% of the total variability of the panel. The result is in line with the outcome of Stock and Watson (2005), which indicates the existence of nine static factors in the stationary panel describing US economy.

Since we cannot choose the total number of factors consistently, we check the robustness of the results with respect to the number of stationary factors. We will use, as a benchmark, a model with ten factors (three common trends and seven stationary factors).

3.5.3 Macroeconomic factors

Finally, we check whether some observable variables can be interpreted as common factors. Since the unobserved factors are consistently estimated then we can use a formal test described in Section 3.3. In order to construct the confidence intervals, we need to estimate the variance-covariance matrix Γ_t. We use the estimator applied in Bai (2004). It is constructed as follow

$$\Gamma_t = \frac{1}{N}\sum_{i=1}^{N} \tilde{e}_{it}^2 \tilde{\lambda}_i \tilde{\lambda}_i'$$

where $\tilde{\lambda}_i$ are the principle components estimates of the loadings matrices and $\tilde{e}_{it} = X_{it} - \tilde{\lambda}_i \tilde{F}_t$ are the idiosyncratic residuals.

Interest rate

In most of the macroeconomics literature, interest rates are one of the driving forces of the economy. In the analysis, we focus on the interest rate measured by Federal Funds rate (FF). We rotate the estimated factors toward FF by running the regression $FF_t = \alpha + \delta \tilde{F}_t + \varepsilon_t$. Next, we compute confidence intervals around fitted values (3.7) and the percentage of FF observations that remain outside the intervals. The results for different number factors are presented in Table 3.4. The outcomes indicate that for models with at least ten factors, all observations of FF remain inside the confidence intervals. Therefore, we cannot reject the hypothesis that the FF is one of the factors driving the economy. Figure 3.2 presents the observations of FF and the estimated confidence intervals for the benchmark model.

[8] The Scree test was introduced by Cattell (1966) and is based on the observation that the plot of correlation matrix eigenvalues for uncorrelated variables is almost flat and linearly converges to zero.

Private fixed investments vs. personal consumption expenditures

Next, we consider the hypothesis that investments play an important role in the economic development. Therefore, we examine if two measures of investments; real private fixed investments in nonresidential structures and residential permanent site structures, can be considered as common factors. We proceed as before and regress the variables on the estimated common factors. Next, we construct the confidence intervals as in (3.7) and compute the percentage of observations that remain outside the confidence intervals. The results are presented in the Table 3.4. They indicate that for sufficient number of factors both variables can be interpreted as common trends.

Unfortunately, for a benchmark model with ten common factors, around 22% of observations of the investments in nonresidential structures lay outside the confidence intervals. The variable and the confidence intervals are presented in Figure 3.3. Therefore, we consider another measure of nonresidential investments: the real private fixed investments in nonresidential commercial structures. For models with at least eight factors we can not reject the null that the variable is a common factor. Moreover, for models with at least eleven factors, we could not reject the hypothesis that both measures of investments in nonresidential structures are common trends. Thus, we conclude that they are the driving forces of the economy.

The outcomes for the investments in residential permanent site structures are more clear. For all considered models, at least 90% of observations stay inside the confidence intervals. Moreover, for a benchmark model only 6.28% of observations fall outside the intervals (Figure 3.4). Hence, we interpret the investments in residential site structure as a common factor.

Finally, we analyze whether different measures of real personal consumption expenditures can be interpreted as common trends. The outcomes indicate that the null hypothesis can be reject for all model setups. Thus, we do not find any results supporting the view that the personal consumption is a main driving force of the whole economy.

Government spendings

Since we do not find any arguments in favor of a hypothesis that the private real consumption expenditure can be interpreted as common factors, we test whether government spendings have an important effect on the economy. We consider two measurements of government spendings: real federal consumption expenditures and gross investments in national defence and nondefense sectors. We proceed as before and construct the confidence intervals. The percentage of variable observations that lay outside of the intervals are presented in Table 3.4. The results indicate that for a model with at least nine factors both variables can be interpreted as common factors. Figure 3.5 shows federal expenditures in national defence and the confidence intervals for the benchmark model. It can be noticed that almost all observations stay inside the intervals (only less

then 2% are outside). Similar results are obtained for federal expenditures in nondefense sectors (Figure 3.6). The outcomes support the hypothesis that government spending have an impact on the whole economy.

3.6 Conclusions

This paper discusses the estimation methods of common factors with different types of dynamics. We generalize the existing methodology by allowing for other types of factors apart from stationary factors and common trends without a deterministic drift. In particular, we focus on nonstationary factors with a time trend. We believe that it is an important issue because most of the macroeconomic variables are subjected to a time trend. Thus, the data should be either detrended or the existence of a drift needs to be explicitly modeled. The model setup is similar to the generalized factor model presented in Bai (2004). Under some standard assumptions, we show that the common factors can be consistently estimated with a principal component method (under the assumption that both time and cross-sectional dimensions increase to infinity). Additionally, we derive convergence rates and asymptotic distributions of factors, factors loadings and common components. It allows us to construct the confidence intervals of a rotation of true factors and hence, to construct a formal test to verify if an observable variable can be interpreted as a common factor. We link the theory to the existing literature and present it as an extension to the work of Bai (2003) and Bai (2004).

The theory is illustrated with an empirical example. We analyze 69 macroeconomic variables describing the real part of the U.S. economy. We show that an interest rate, investments and government spendings can be interpreted as common factors, thus they are the driving forces of the economy. The results are in line with a macroeconomic literature. We do not find any arguments in favor of a hypothesis that personal consumption is also one of the common trends.

3.7 Appendix: Data description and estimation results

The appendix lists the variables used in the empirical analysis and describes the applied transformation (column A in the following table). All variables are in levels and all but the Federal Funds rate are expressed in logarithms

Nr	Variable
1	Real Gross Domestic Product, Quantity Indexes; (2000=100,SA)
2	Real final sales to domestic purchasers; (2000=100,SA)
3	Real personal consumption expenditures; (2000=100, SA)
4	Real personal consumption expenditures: Durable goods; (2000=100, SA)
5	Real personal consumption expenditures: Motor vehicles and parts;(2000=100, SA)
6	Real personal consumption expenditures: Household equipment; (2000=100, SA)
7	Real personal consumption expenditures: Nondurable goods; (2000=100, SA)
8	Real personal consumption expenditures: Food; (2000=100, SA)
9	Real personal consumption expenditures: Clothing and shoes; (2000=100, SA)
10	Real personal consumption expenditures: Energy goods; (2000=100, SA)
11	Real personal consumption expenditures: Services; (2000=100, SA)
12	Real personal consumption expenditures: Housing; (2000=100, SA)
13	Real personal consumption expenditures: Household operation; (2000=100, SA)
14	Real personal consumption expenditures: Electricity and gas; (2000=100, SA)
15	Real personal consumption expenditures: Transportation; (2000=100, SA)
16	Real personal consumption expenditures: Medical care; (2000=100, SA)
17	Real personal consumption expenditures: Recreation;(2000=100, SA)
18	Real gross private domestic investment; (2000=100, SA)
19	Real private fixed investment; (2000=100, SA)
20	Real private fixed investment: Nonresidential: Structures; (2000=100, SA)
21	Real private fixed investment: Nonresidential: Commercial struct.;(2000=100, SA)
22	Real private fixed investment: Nonresidential: Manufacturing struct.; (2000=100,SA)
23	Real private fixed investment: Nonresidential: Power & communic. struct.; (2000=100, SA)
24	Real private fixed investment: Nonresidential: Mining struct.; (2000=100, SA)
25	Real private fixed investment: Nonresidential: Equipment and software; (2000=100, SA)
26	Real private fixed investment: Nonresidential: Information processing equipment and software; (2000=100, SA)
27	Real private fixed investment: Nonresidential: Software; (2000=100, SA)
28	Real private fixed investment: Nonresidential: Equipment and software: Industrial equipment;(2000=100, SA)
29	Real private fixed investment: Nonresidential: Equipment and software: Transportation equipment; (2000=100, SA)
30	Real private fixed investment: Residential: Structures; (2000=100, SA)
31	Real private fixed investment: Residential: Structures: Permanent site; (2000=100, SA)
32	Real private fixed investment: Residential: Structures: Permanent site: Single family; (2000=100, SA)
33	Real private fixed investment: Residential: Structures: Other structures; (2000=100, SA)

Nr	Variable
34	Real private fixed investment: Residential: Equipment; (2000=100, SA)
35	Real Exports; (2000=100, SA)
36	Real Exports: Goods; (2000=100, SA)
37	Real Exports: Services; (2000=100, SA)
38	Real Imports; (2000=100, SA)
39	Real Imports: Goods; (2000=100, SA)
40	Real Imports: Services; (2000=100, SA)
41	Real government consumption expenditures and gross investment; (2000=100, SA)
42	Real government consumption expenditures and gross investment: Federal; (2000=100, SA)
43	Real government consumption expenditures and gross investment: Federal: National defense; (2000=100, SA)
44	Real government consumption expenditures and gross investment: Federal: National defense: Consumption expenditures; (2000=100, SA)
45	Real government consumption expenditures and gross investment: Federal: National defense: Gross investment; (2000=100, SA)
46	Real government consumption expenditures and gross investment: Federal: Nondefense; (2000=100, SA)
47	Real government consumption expenditures and gross investment: Federal: Nondefense: Consumption expenditures; (2000=100, SA)
48	Real government consumption expenditures and gross investment: Federal: Nondefense: Gross investment; (2000=100, SA)
49	Real government consumption expenditures and gross investment: State and local;
50	Real government consumption expenditures and gross investment: State and local: Consumption expenditures; (2000=100, SA)
51	Real government consumption expenditures and gross investment: State and local: Gross investment, (2000=100, SA)
52	Industrial Production Index: Total index; (2000=100, SA)
53	Industrial Production Index: Final products and nonindustrial supplies;(2000=100, SA)
54	Industrial Production Index: Consumer goods; (2000=100, SA)
55	Industrial Production Index: Durable consumer goods; (2000=100, SA)
56	Industrial Production Index: Nondurable consumer goods; (2000=100, SA)
57	Industrial Production Index: Business equipment; (2000=100, SA)
58	Industrial Production Index: Defense and space equipment; (2000=100, SA)
59	Industrial Production Index: Materials; (2000=100, SA)
60	Industrial Production Index: Construction supplies; (2000=100, SA)
61	Industrial Production Index: Business supplies; (2000=100, SA)
62	Industrial Production Index: Mining NAICS=21; (2000=100, SA)
63	Industrial Production Index: Manufacturing (SIC); (2000=100, SA)
64	Output Per Hour of All Persons: Nonfarm Business Sector; Index (1992=100,SA)
65	Output Per Hour of All Persons: Business Sector; Index (1992=100,SA)
66	Federal Fund rate
67	1-Year Treasury Constant Maturity Rate
68	3-Year Treasury Constant Maturity Rate
69	5-Year Treasury Constant Maturity Rate

Table 3.2: Choice of the number of nonstationary dynamic factors, information criteria IPC

Inf. Criteria$\backslash k_{\max}$	2	3	4	5	6	7	8	9	10
IPC_1	2	2	2	3	3	3	3	3	4
IPC_2	2	2	2	3	3	3	3	3	4
IPC_3	2	2	2	3	3	3	3	4	4

Table 3.3: Variable names and description

Name	Nr	Description
Con	3	Real personal consumption expenditures;
ConD	4	Real personal consumption expenditures: Durable goods;
ConND	7	Real personal consumption expenditures: Nondurable goods;
ConS	11	Real personal consumption expenditures: Services;
InvS	20	Real private fixed investment: Nonresidential: Structures;
InvCS	21	Real private fixed investment: Nonresidential: Commercial struct.;
InvRS	31	Real private fixed investment: Residential: Structures: Permanent site;
GovD	43	Real government consumption expenditures and gross investment: Federal: National defense;
GovND	46	Real government consumption expenditures and gross investment: Federal: Nondefense;
FF	66	Federal Fund rate

NOTE: Variable number corresponds with the ordering defined in the data description.

Table 3.4: Percentage of observations that remain outside confidence intervals for models with different number of factors

Variable	Number of factors								
Name	6	7	8	9	10	11	12	13	14
Con	51.83	53.40	18.85	24.61	31.41	25.13	26.70	31.94	28.27
ConD	72.25	72.77	72.77	71.73	74.35	63.87	49.74	38.22	27.23
ConND	80.10	65.97	34.56	38.22	42.93	39.27	45.55	48.17	45.55
ConS	14.14	17.80	15.18	17.23	18.32	23.04	36.70	30.37	32.46
InvS	78.53	44.50	52.88	23.56	21.99	0.00	0.00	0.00	0.00
InvCS	88.48	21.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00
InvRS	6.28	7.33	8.90	4.71	6.28	0.00	0.00	0.00	0.00
GovD	3.67	9.95	12.04	1.57	1.57	4.71	3.14	5.76	7.33
GovND	82.72	71.22	81.67	4.19	3.66	1.57	1.57	0.00	0.00
FF	38.22	46.60	56.54	60.21	0.00	0.00	0.00	0.00	0.00

NOTE: Variable name corresponds with the description presented in Table 3.3.

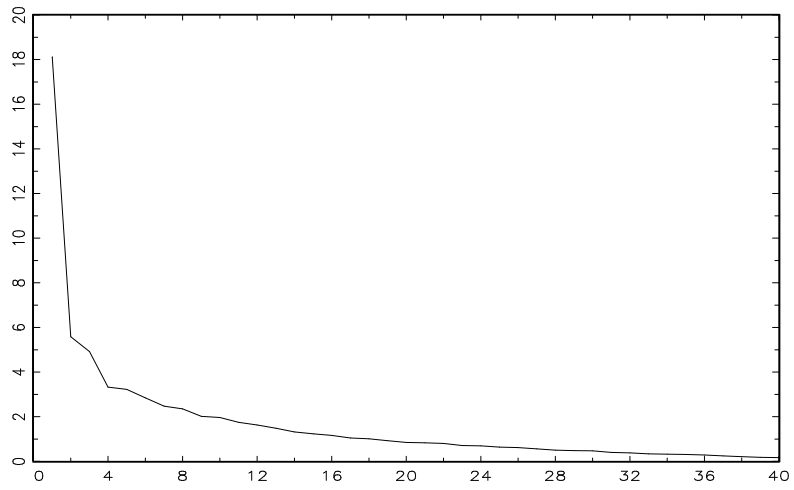

Figure 3.1: First largest eigenvalues of the panel correlation matrix.

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

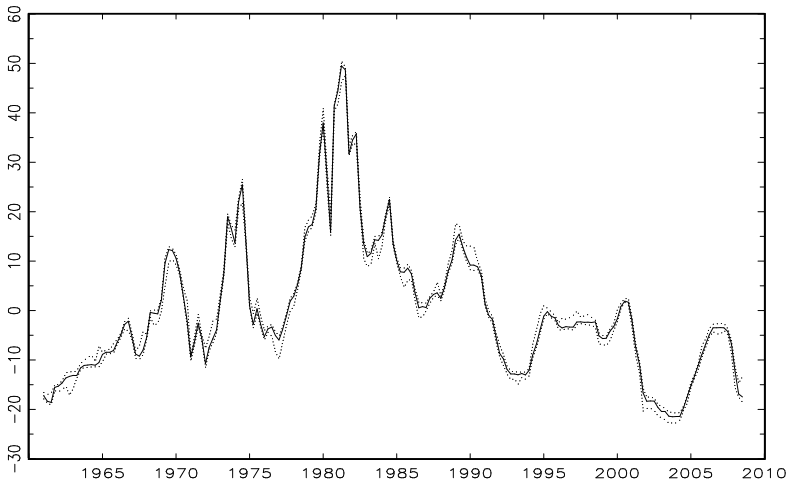

Figure 3.2: Federal Funds rate (solid line) and the confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data

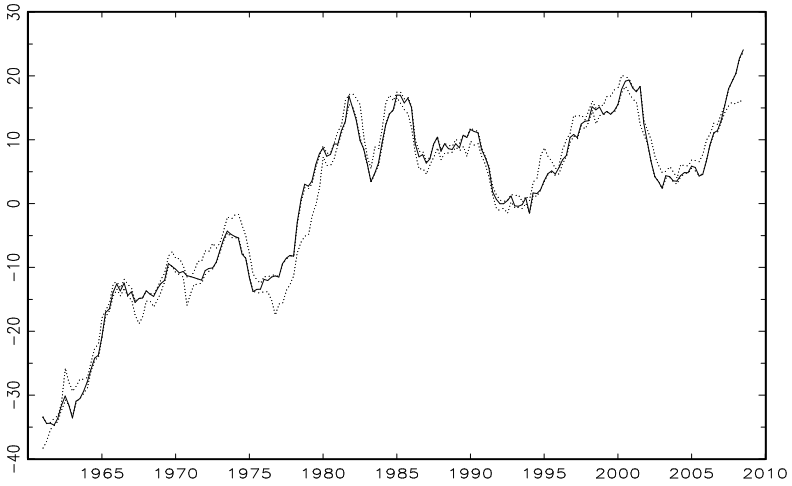

Figure 3.3: Real private fixed investments in nonresidential structures (solid line) and confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data

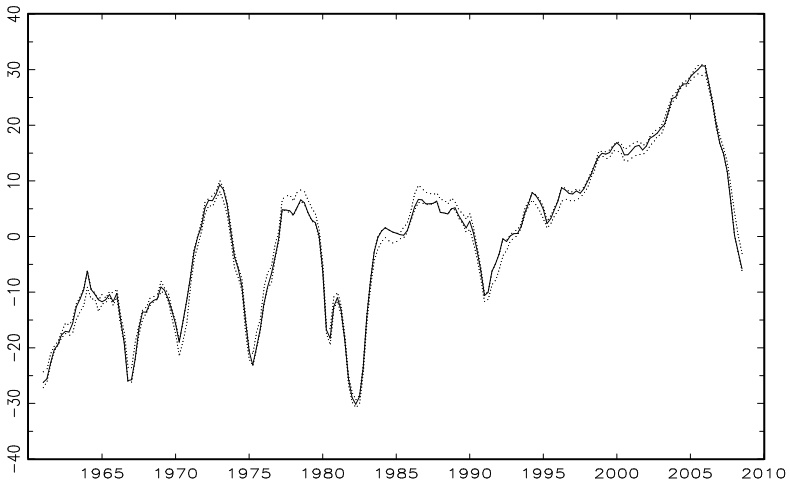

Figure 3.4: Real private fixed investments in residential permanent site structures (solid line) and confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data

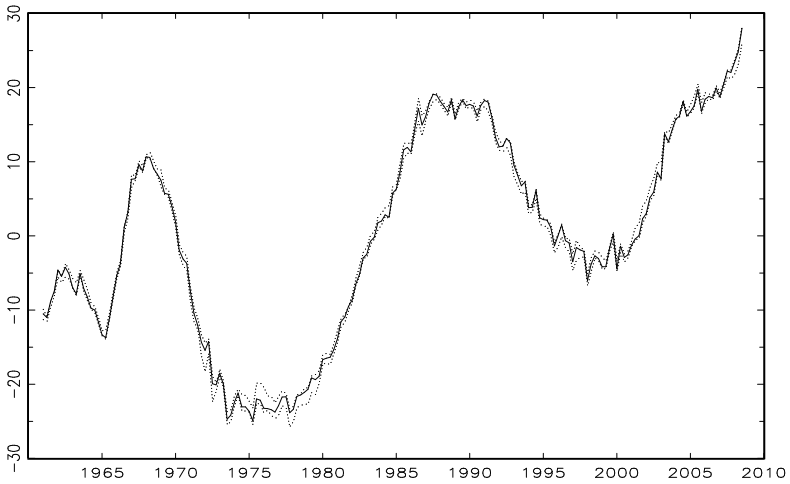

Figure 3.5: Real federal government consumption expenditures and gross investments in national defense (solid lines) and confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data

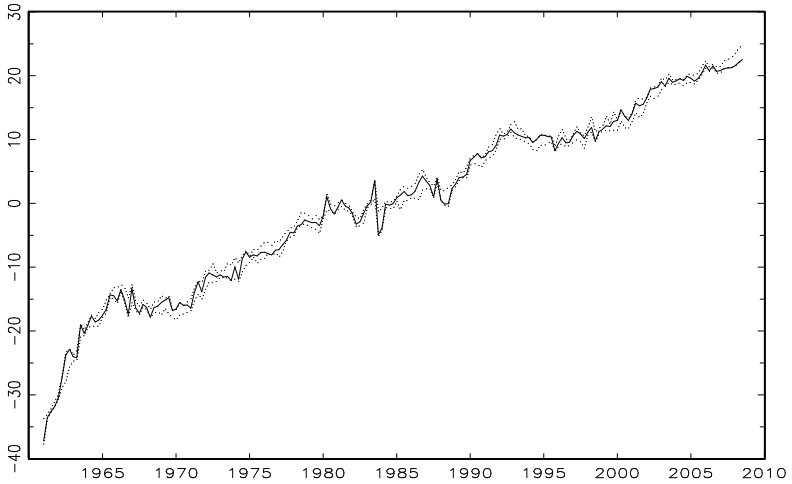

Figure 3.6: Real federal government consumption expenditures and gross investments in nondefence sectors (solid line) and confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data

3.8 Appendix: Proofs

3.8.1 General algebra results

In the following sections we use some general properties of the Euclidean norm

$$\|A\|^2 = tr\left(A'A\right)$$

The results can be found in Lütkepohl (1996).

1. $\|A\| = \|A'\|$
2. $\|cA\| = |c| \|A\|$
3. Cauchy-Schwarz inequality

$$\|AB\| \leq \|A\| \|B'\| = \|A\| \|B\|$$

4. Parallelogram identity

$$\|A+B\|^2 + \|A-B\|^2 \leq 2\left(\|A\|^2 + \|B\|^2\right)$$

Thus,

$$\begin{aligned} \|A+B\|^2 &\leq 2\left(\|A\|^2 + \|B\|^2\right) \\ &\leq 2\left(\|A\| + \|B\|\right)^2 \end{aligned}$$

and therefore,

$$\|A+B\| \leq \sqrt{2}\left(\|A\| + \|B\|\right)$$

Lemma 19 *(Eigenvalues and singular values results) Let us define by $\sigma_i(A)$ the ith largest singular value of a matrix A and by $\lambda_i(B)$ the ith largest eigenvalue of a square matrix B. Then, for any real $m \times n$ matrix A the following results holds*

1. *The matrices $A'A$ and AA' are square, symmetric and positive semidefinite*
2. *If $m \geq n$ then for $i \leq n$ there is $\lambda_i(AA') = \lambda_i(A'A)$*
3. *A and B are $m \times n$ matrices, with $r = \min\{m,n\}$ then for $1 \leq i,j,i+j-1 \leq r$*

$$\sigma_{i+j-1}(AB') \leq \sigma_i(A)\sigma_j(B)$$

4. *A is a $m \times n$ matrix, with $m \geq n$, B is a $n \times n$ square matrix then for $1 \leq i,j,i+j-1 \leq n$*

$$\sigma_{i+j-1}(AB') \leq \sigma_i(A)\sigma_j(B)$$

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Proof. The results (1) and (3) are presented in Lütkepohl (1996). Consider (2). Since the matrices AA' and $A'A$ are symmetric and positive definite then $\lambda_i(AA') \geq 0$ and $\lambda_i(A'A) \geq 0$. Moreover $rk(AA') = rk(A'A) = r$ and r equals the number of the non-zero eigenvalues of the matrices AA' and $A'A$. Therefore, for all $i = 1, ..., r$ there is $\lambda_i(AA') > 0$ and $\lambda_i(AA') = \lambda_i(A'A)$ (see Lütkepohl (1996)). For $i > r$ we have $\lambda_i(AA') = \lambda_i(A'A) = 0$. Thus, $\lambda_i(AA') = \lambda_i(A'A)$.

Consider (4). It follows directly from the part (3). We can construct a $m \times n$ matrix $\bar{B}$ such that

$$\bar{B} = \begin{bmatrix} B \\ 0_{(m-n)\times n} \end{bmatrix}$$

and

$$\left(A\bar{B}'\right)'\left(A\bar{B}'\right) = \begin{bmatrix} (AB')'(AB') & 0 \\ 0 & 0 \end{bmatrix}$$

Then $\sigma_j(B) = \sigma_j\left(\bar{B}\right)$ and $\sigma_i(AB') = \sigma_i\left(A\bar{B}'\right)$ for any $i, j \leq n$. Therefore,

$$\begin{aligned} \sigma_{i+j-1}(AB') &= \sigma_{i+j-1}\left(A\bar{B}'\right) \\ &\leq \sigma_i(A)\,\sigma_j\left(\bar{B}\right) \\ &= \sigma_i(A)\,\sigma_j(B) \end{aligned}$$

■

3.8.2 Estimation

Proof of Lemma 4. The loadings matrix $\hat{\Lambda}$ satisfies the condition

$$\hat{F}\hat{\Lambda}' = \tilde{F}\tilde{\Lambda}'$$

Moreover, we know that

$$\tilde{\Lambda}' = D^{-2}\tilde{F}'X$$

and therefore

$$\hat{F}\hat{\Lambda}' = \tilde{F}\tilde{\Lambda}' = \frac{1}{T^3}\tilde{F}D^{-2}\tilde{F}'X$$

Thus,

$$\hat{F}'\hat{F}\hat{\Lambda}' = \hat{F}'\tilde{F}D^{-2}\tilde{F}'X$$

and

$$\hat{\Lambda}' = \left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'\tilde{F}D^{-2}\tilde{F}'X \tag{3.13}$$

From definition of the normalized factor $\hat{F} = N^{-1}X\tilde{\Lambda}$ and the loadings $\tilde{\Lambda}' = D^{-2}\tilde{F}'X$ it follows that

$$\hat{F} = \frac{1}{N}X\tilde{\Lambda} = \frac{1}{N}(XX')\,\tilde{F}D^{-2}$$

Let us denote by $\tilde{V}_{NT}$ the diagonal matrix consisting of the first r largest eigenvalues of the matrix XX' and $V_{NT} = D^{-2}\tilde{V}_{NT}/N$. Then by the fact that both V_{NT} and D are diagonal there is $\hat{F}'\tilde{F} = V_{NT}D^2$ and $\hat{F}'\hat{F} = V_{NT}^2 D^2$

$$\begin{aligned}\hat{F}'\tilde{F} &= \frac{1}{N}D^{-2}\tilde{F}'\left(XX'\right)\tilde{F} = \frac{D^{-2}}{N}\tilde{V}_{NT}D^2 = V_{NT}D^2 \\ \hat{F}'\hat{F} &= \left(\frac{1}{N}\right)^2 D^{-2}\tilde{F}'\left(XX'\right)\left(XX'\right)\tilde{F}D^{-2} = \frac{D^{-2}}{N}\tilde{V}_{NT}\frac{D^{-2}}{N}D^2 = V_{NT}^2 D^2\end{aligned}$$

Finally, from equation (3.13) the normalized loadings are $\hat{\Lambda} = \tilde{V}_{NT}^{-1}\tilde{\Lambda}$

$$\begin{aligned}\hat{\Lambda}' &= \left(V_{NT}^2 D^2\right)^{-1}\left(V_{NT}D^2\right)D^{-2}\tilde{F}'X = V_{NT}^{-1}D^{-2}\tilde{F}'X \\ &= V_{NT}^{-1}\tilde{\Lambda}'\end{aligned}$$

Since $\hat{F}\hat{\Lambda}' = \tilde{F}\tilde{\Lambda}'$ then

$$\hat{F} = V_{NT}\tilde{F}$$

■

The following Lemma 20-21 discuss issues associated with the eigenvalues of matrix V_{NT}. They show that the matrix $V_{NT} = O_p(1)$.

Lemma 20 *Let us denote V_{NT}^* the diagonal matrix consisting of the first r largest eigenvalues of the matrix $F^0\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}$ in the descending order multiplied by D^{-2}. Then $V_{NT}^* = O_p(1)$ and $\lim_{T,N\to\infty} V_{NT,i}^* > 0$, where $V_{NT,i}^*$ denotes the ith diagonal element of V_{NT}^*.*

Proof. The ith diagonal element of the matrix V_{NT}^* is the ith largest eigenvalue of the matrix

$$V_{NT,i}^* = \lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right)$$

where $d_i = D_{ii}$. We show that

$$\lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right) = O_p(1)$$

Let us first notice that since $i \leq r$. Then by Lemma 19

$$\begin{aligned}\lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right) &= \lambda_i\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\frac{F^{0\prime}}{d_i}\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right) \\ &= \sigma_i^2\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right)\end{aligned}$$

where $\sigma_i(A)$ denotes a ith largest singular value of a matrix A. From Lemma 19 it follows that

$$\begin{aligned}\sigma_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right) &\leq \sigma_i\left(\frac{F^0}{d_i}\right)\sigma_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right) \\ &= \sigma_i\left(\frac{F^0}{d_i}\right)\lambda_1\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\end{aligned}$$

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

We show that $\sigma_i\left(\frac{F^0}{d_i}\right) = O_p(1)$. Let us first notice that $\sigma_i\left(d_i^{-1}D\right) = 1$. Then by Lemma 19

$$\begin{aligned}\sigma_i\left(\frac{F^0}{d_i}\right) &\leq \sigma_1\left(F^0D^{-1}\right)\sigma_i\left(d_i^{-1}D\right)\\ &= \sigma_1\left(F^0D^{-1}\right)\\ &\to {}^d\lambda_1(\Sigma)\end{aligned}$$

and $\lambda_1(\Sigma) < M$ with probability 1. Since

$$\lambda_1\left(\frac{\Lambda_0'\Lambda_0}{N}\right) \to^p \lambda_1(\Sigma_\Lambda) < M$$

then

$$\lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right) = O_p(1)$$

Finally, we show that $\lim_{T\to\infty}\lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right) > 0$. By Lemma 19

$$\sigma_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right)\sigma_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{-1/2}\right) \geq \sigma_i\left(\frac{F^0}{d_i}\right)$$

Moreover,

$$\sigma_i\left(\frac{F^0}{d_i}\right)\sigma_{r-i+1}\left(d_iD^{-1}\right) \geq \sigma_r\left(F^0D^{-1}\right)$$

where $\sigma_{r-i+1}\left(d_iD^{-1}\right) = 1$ and $\sigma_r\left(F^0D^{-1}\right) \to^p \lambda_r(\Sigma) > 0$. Thus,

$$\lim_{T\to\infty}\sigma_i\left(\frac{F^0}{d_i}\right) \geq \lambda_r(\Sigma) > 0$$

From Assumption B it follows that $\Lambda_0'\Lambda_0/N \to^p \Sigma_\Lambda$ and Σ_Λ is symmetric, positive definite. Thus

$$\begin{aligned}\sigma_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{-1/2}\right) &= \lambda_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{-1}\right)\\ &= \lambda_r\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\right)\\ &\to {}^p\lambda_r(\Sigma_\Lambda)\end{aligned}$$

where $0 < \lambda_r(\Sigma_\Lambda) < M$. Therefore,

$$\begin{aligned}\lim_{N,T\to\infty}\sigma_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right) &\geq \lim_{N,T\to\infty}\frac{\sigma_i\left(\frac{F^0}{d_i}\right)}{\sigma_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{-1/2}\right)}\\ &\geq \frac{\lambda_r(\Sigma)}{\lambda_r(\Sigma_\Lambda)} > 0\end{aligned}$$

and

$$\lim_{T,N\to\infty} V_{NT,i}^{*} > 0$$

■

Lemma 21 *Under Assumptions A and F and N, $T \to \infty$ the matrix $V_{NT} = O_p(1)$.*

Proof. From the model setup it follows that

$$\frac{XX'}{N} = F^0\left(\frac{\Lambda_0'\Lambda_0}{N}\right)F^{0\prime} + C$$

where C is a symmetric matrix

$$C = \frac{F^0\Lambda_0'e'}{N} + \frac{e\Lambda_0F^{0\prime}}{N}$$

Let us denote $\lambda_i(A)$ the ith largest eigenvalue of the matrix A. By Lütkepohl (1996)

$$\lambda_i(A+B) \leq \lambda_i(A) + \lambda_{\max}(B)$$

for symmetric matrices A and B. Therefore,

$$\begin{aligned}\lambda_i\left(\frac{XX'}{N}\right) &\leq \lambda_i\left(F^0\left(\frac{\Lambda_0'\Lambda_0}{N}\right)F^{0\prime}\right) + \lambda_{\max}(C)\\ &\leq \lambda_i\left(F^0\left(\frac{\Lambda_0'\Lambda_0}{N}\right)F^{0\prime}\right) + tr(C)\end{aligned}$$

Thus,

$$V_{NT} \leq V_{NT}^{*} + tr(C)D^{-2}$$

From the definition of the trace operator and its properties (see Lütkepohl (1996))

$$\begin{aligned}tr(C) &= 2tr\left(\frac{e\Lambda_0F^{0\prime}}{N}\right)\\ &= 2\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}e_{it}\lambda_i'\right)F_t^0\end{aligned}$$

Thus, by Assumptions A and F

$$\begin{aligned}\left\|tr(C)D^{-2}\right\| &= \left\|2\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}e_{it}\lambda_i'\right)F_t^0D^{-2}\right\|\\ &= \left\|2\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}e_{it}\lambda_i'\right)\left(\sqrt{T}F_t^0D^{-1}\right)\sqrt{T}D^{-1}\right\|\\ &\leq 2\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}e_{it}\lambda_i'\right)\left(\sqrt{T}F_t^0D^{-1}\right)\right\|\left\|\sqrt{T}D^{-1}\right\|\\ &= O_p(1)\end{aligned}$$

Hence, by Lemma 20

$$V_{NT} \leq V_{NT}^{*} + O_p(1) = O_p(1)$$

■

3.8.3 Consistency

In this section, we prove two important propositions: Propositions 5 and 6. They show that the factors can be consistently estimated and derive the corresponding convergence rates.

The following Lemmas 22 and 23 are needed to prove Proposition 5.

Lemma 22 *Under the assumptions A-C for all T and N there exists some $M < \infty$ such that*

1. $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_N(s,t)^2 \leqslant M$
2. $E\left\{\left(N^{-1/2}e_t'\Lambda_0\right)^2\right\} \leqslant M$
3. $E\left\|(NT)^{-1/2}\sum_{t=1}^{T}e_t'\Lambda_0\right\| \leqslant M$

Proof. Points (1) - (3) are proved in Bai(2004). ■

Lemma 23 *Under Assumptions A-C and $N, T \to \infty$*

1. $\left\|F^{0\prime}\tilde{F}D^{-2}\right\| = O_p(1)$
2. $\left\|e\Lambda_0 F^{0\prime}\tilde{F}D^{-2}\right\| = O_p\left(\sqrt{NT}\right)$
3. *Define a symmetric $T\times T$ matrix Φ by $\Phi_{ts} = \gamma_N(t,s)$, then* $\left\|\Phi\tilde{F}D^{-2}\right\| = O_p\left(\sqrt{T}\left\|D^{-1}\right\|\right)$
4. *Define a symmetric $T\times T$ matrix Υ as $\Upsilon = ee' - \Phi$, then* $\left\|\Upsilon\tilde{F}D^{-2}\right\| = O_p\left(\frac{T}{\sqrt{N}}\left\|D^{-1}\right\|\right)$

Proof. Consider (1). Let us denote

$$H = \frac{\Lambda_0'X'\tilde{F}D^{-2}}{N}$$

Then by Lemma 21 and Assumption B $\|H\| = O_p(1)$ because

$$\begin{aligned}\|H\| &= \left\| \frac{\Lambda_0' X \tilde{F}' D^{-2}}{N} \right\| \\ &\leq \left\| \frac{\Lambda_0'}{\sqrt{N}} \right\| \left\| \frac{X' \tilde{F} D^{-2}}{\sqrt{N}} \right\| \\ &= O_p(1)\, tr\left(\frac{D^{-2} \tilde{F}' X X' \tilde{F} D^{-2}}{N} \right) \\ &= O_p(1)\, tr\left(V_{NT}\right) = O_p(1)\end{aligned}$$

Moreover,

$$H = \frac{\Lambda_0' \Lambda_0 F^{0\prime} \tilde{F} D^{-2}}{N} + \frac{\Lambda_0' e' \tilde{F} D^{-2}}{N}$$

Then

$$\left\| \frac{\Lambda_0' \Lambda_0 F^{0\prime} \tilde{F} D^{-2}}{N} \right\| \leq \sqrt{2} \left(\|H\| + \left\| \frac{\Lambda_0' e' \tilde{F} D^{-2}}{N} \right\| \right)$$

We show that $\left\| \Lambda_0' e' \tilde{F} D^{-2}/N \right\|^2 = O_p(1)$. By Lemma 22

$$\begin{aligned}\left\| \Lambda_0' e' \tilde{F} D^{-2}/N \right\| &\leq \left\| \frac{\Lambda_0' e'}{\sqrt{NT}} \right\| \left\| \tilde{F} D^{-1} \right\| \left\| T^{1/2} D^{-1} N^{-1/2} \right\| \\ &= o_p(1)\end{aligned}$$

Thus, $\Lambda_0' \Lambda_0 F^{0\prime} \tilde{F} D^{-2}/N = O_p(1)$. Since $\Lambda_0' \Lambda_0/N$ converges to a positive definite matrix then it must be that $F^0 \tilde{F} D^{-2} = O_p(1)$.

Consider (2). From the first part of the lemma it follows that

$$\begin{aligned}\left\| e \Lambda_0 F^{0\prime} \tilde{F} D^{-2} \right\| &\leq \|e \Lambda_0\| \left\| F^{0\prime} \tilde{F} D^{-2} \right\| \\ &= O_p\left(\sqrt{NT}\right)\end{aligned}$$

Consider (3)

$$\left\| \Phi \tilde{F} D^{-2} \right\|^2 \leq \|\Phi\| \left\| \tilde{F} D^{-1} \right\| \left\| D^{-1} \right\|$$

By Lemma 22, the last component is $\|\Phi\|^2 = O_p(T)$ because

$$\begin{aligned}\|\Phi\|^2 &= \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\gamma_N(t,s)\right)^2 \\ &= T \left\{ \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\gamma_N(t,s)\right)^2 \right\} \\ &= T O_p(1) = O_p(T)\end{aligned}$$

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Thus,

$$\begin{aligned}\left\|\Phi\tilde{F}D^{-2}\right\| &= O_p\left(\left\|D^{-1}\right\|^2\right)O_p(1)\,O_p(T)\\ &= O_p\left(\sqrt{T}\left\|D^{-1}\right\|\right)\end{aligned}$$

Finally consider (4).

$$\left\|\Upsilon\tilde{F}D^{-2}\right\| \leq \|\Upsilon\|\left\|\tilde{F}D^{-1}\right\|\left\|D^{-1}\right\|$$

Under the Assumption F.1 the last component is $\|\Upsilon\|^2 = O_p\left(T^2/N\right)$ because

$$\begin{aligned}\|\Upsilon\|^2 &= \sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{e_t'e_s}{N}-\gamma_N(t,s)\right)^2\\ &= \frac{T^2}{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\frac{1}{N^{1/2}}\left(e_t'e_s-E\left(e_t'e_s\right)\right)\right\}^2\\ &= \frac{T^2}{N}O_p(1)=O_p\left(\frac{T^2}{N}\right)\end{aligned}$$

Thus,

$$\begin{aligned}\left\|\Upsilon\tilde{F}D^{-2}\right\| &= O_p\left(\left\|D^{-1}\right\|\right)O_p(1)\,O_p\left(\frac{T}{\sqrt{N}}\right)\\ &= O_p\left(\frac{T}{\sqrt{N}}\left\|D^{-1}\right\|\right)\end{aligned}$$

■

Proof of Proposition 5. Let us define a matrix H as in Lemma 23. The matrix H takes the form

$$H=\frac{\Lambda_0'X'\tilde{F}D^{-2}}{N}$$

Then it was shown that $\|H\|=O_p(1)$ and thus the matrix is well defined. The difference between the estimated factors and a rotation of the true factors can be expressed as follow

$$\begin{aligned}\hat{F}-F^0H &= \frac{1}{N}XX'\tilde{F}D^{-2}-\frac{1}{N}F^0\Lambda_0'X'\tilde{F}D^{-2} &(3.14)\\ &= \frac{1}{N}\left\{\left(F^0\Lambda_0'+e\right)X'-F^0\Lambda_0'X\right\}\tilde{F}D^{-2}\\ &= \frac{1}{N}\left\{e\Lambda_0F^{0\prime}+ee'\right\}\tilde{F}D^{-2}\\ &= \frac{1}{N}\left\{e\Lambda_0F^{0\prime}+N\Upsilon+N\Phi\right\}\tilde{F}D^{-2} &(3.15)\end{aligned}$$

where Υ and Φ are defined as in Lemma 23.

$$\begin{aligned}\frac{1}{4T}\left\|\hat{F}-F^0H\right\|^2 &\leq \frac{1}{4TN^2}\left\|\left\{e\Lambda_0F^{0\prime}+N\Upsilon+N\Phi\right\}\tilde{F}D^{-2}\right\|^2\\ &\leq \frac{1}{TN^2}\left\|e\Lambda_0F^{0\prime}\tilde{F}D^{-2}\right\|^2+\frac{N^2}{TN^2}\left\|\Phi\tilde{F}D^{-2}\right\|^2\\ &\quad+\frac{N^2}{TN^2}\left\|\Upsilon\tilde{F}D^{-2}\right\|^2\end{aligned}$$

From Lemma 23 it follows that

$$\begin{aligned}\frac{1}{4T}\left\|\hat{F}-F^0H\right\|^2 &= \frac{1}{TN^2}O_p\left(NT\right)+\frac{1}{T}O_p\left(T\left\|D\right\|^{-2}\right)+\frac{1}{T}O_p\left(\frac{T^2}{N}\left\|D\right\|^{-2}\right)\\ &= O_p\left(N^{-1}\right)+O_p\left(\left\|D\right\|^{-2}\right)+O_p\left(\frac{T}{N}\left\|D\right\|^{-2}\right)\end{aligned}$$

Under the assumption $T\left\|D\right\|^{-2}=O_p\left(1\right)$ we get

$$\frac{1}{T}\left\|\hat{F}-F^0H\right\|^2=O_p\left(N^{-1}\right)+O_p\left(\left\|D\right\|^{-2}\right)$$

and

$$\begin{aligned}\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2 &= \frac{1}{T}\left\|\left(\hat{F}-F^0H\right)V_{NT}^{-1}\right\|^2\\ &\leq \frac{1}{T}\left\|\left(\hat{F}-F^0H\right)V_{NT}^{-1}\right\|\left\|V_{NT}^{-1}\right\|^2\\ &= O_p\left(\delta_{NT}^{-2}\right)\end{aligned}$$

■

Next, we show Lemma 24 and a proof of Proposition 6.

Lemma 24 *Under Assumptions A-E, N, $T\to\infty$ and $T\left\|D^{-1}\right\|^2=O_p\left(1\right)$ for all t it holds*

1. $\left\|D^{-2}\tilde{F}'F^0\Lambda_0'e_t/N\right\|=O_p\left(N^{-1/2}\right)$
2. $\left\|e_t'e'\tilde{F}D^{-2}/N\right\|=O_p\left(\left\|D^{-1}\right\|\right)$

Proof. Consider (1). By Lemma 22 and Lemma 23

$$\begin{aligned}\left\|\frac{D^{-2}\tilde{F}'F^0\Lambda_0'e_t}{N}\right\| &\leq \left\|D^{-2}\tilde{F}'F^0\right\|\left\|\frac{\Lambda_0'e_t}{\sqrt{N}}\right\|N^{-1/2}\\ &= O_p\left(N^{-1/2}\right)\end{aligned}$$

Let us consider (2).

$$\left\|\frac{e_t'e'\tilde{F}D^{-2}}{N}\right\|\leq\left\|\frac{e_t'e'}{N}\right\|\left\|\tilde{F}D^{-1}\right\|\left\|D^{-1}\right\|$$

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

The second component by definition is $O_p(1)$. It is now shown that the first part is $\|e_t'e'/N\| = O_p(1)$.

$$\begin{aligned}\|e_t'e'/N\|^2 &= \frac{1}{N^2}\sum_{s=1}^{T}\left(e_t'e_s\right)^2 \\ &\leq \left(\frac{1}{N}\sum_{s=1}^{T}|e_t'e_s|\right)^2\end{aligned}$$

Moreover, by Assumption E

$$\begin{aligned}E\left|\frac{1}{N}\sum_{s=1}^{T}|e_t'e_s|\right| &= \sum_{s=1}^{T}E\left|\frac{e_t'e_s}{N}\right| \\ &= \sum_{s=1}^{T}\bar{\gamma}_N(t,s) \\ &= O_p(1)\end{aligned}$$

Therefore,

$$\begin{aligned}\left\|\frac{e_t'e'\tilde{F}D^{-2}}{N}\right\| &= O_p(1)\left\|D^{-1}\right\| \\ &= O_p\left(\left\|D^{-1}\right\|\right)\end{aligned}$$

■

Proof of Proposition 6. Form equation (3.14) it follows that

$$\hat{F}_t - H'F_t^0 = \frac{1}{N}\left\{e_t\Lambda_0F^{0\prime} + e_t'e'\right\}\tilde{F}D^{-2}$$

Thus, from Lemma 24 we get

$$\hat{F}_t - H'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$$

Since $\tilde{F}_t - \tilde{H}'F_t^0 = V_{NT}^{-1}\left(\hat{F}_t - \tilde{H}'F_t^0\right)$ then also

$$\tilde{F}_t - \tilde{H}'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$$

■

The following Lemma 25 is a counterpart of the Proposition 6 for models with only one type of nonstationary factors.

Lemma 25 *If there is only one type of factors (hence, $D = T^dI_r$) and $d \geq 1$ then for $N, T \to \infty$*

1. $\left\|e_t'e'\tilde{F}D^{-2}/N\right\| = O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right) + O_p\left(N^{-1/2}\right)$

2. $\tilde{F}_t - \tilde{H}' F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)$

Proof. Consider (1). Since $D = T^d I_r$ then $e_t' e' \tilde{F} D^{-2}/N = e_t' e' \tilde{F}/\left(T^{2d}N\right)$ and

$$e_t' e' \tilde{F}/\left(T^{2d}N\right) = N\Upsilon_t \tilde{F}/\left(T^{2d}N\right) + N\Phi_t \tilde{F}/\left(T^{2d}N\right)$$

where $\Phi_t = \left(\gamma_N(t,1), ..., \gamma_N(t,T)\right)$ and $\Upsilon_t = e_t e'/N - \Phi_t$.

We show that the first component $N\Upsilon_t \tilde{F}/\left(T^{2d}N\right) = O_p\left(N^{-1/2}T^{3/2-2d}\delta_{NT}^{-1}\right) + O_p\left(T^{1/2-d}N^{-1/2}\right)$.

$$\begin{aligned}
\frac{N\Upsilon_t \tilde{F}}{NT^{2d}} &= \frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t' e_s}{N} - \gamma_N(s,t)\right)\tilde{F}_s \\
&= \frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t' e_s}{N} - \gamma_N(s,t)\right)\left(\tilde{F}_s - \tilde{H}' F_s^0\right) + \frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t' e_s}{N} - \gamma_N(s,t)\right)F_s^0
\end{aligned}$$

The first part is $O_p\left(N^{-1/2}T^{3/2-2d}\delta_{NT}^{-1}\right)$ by Assumption F and Proposition 5 because

$$\begin{aligned}
\frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t' e_s}{N} - \gamma_N(s,t)\right)\left(\tilde{F}_s - \tilde{H}' F_s^0\right) &\leqslant \frac{1}{N^{1/2}T^{2d-3/2}}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - \tilde{H}' F_s^0\right)^2\right)^{1/2} \\
&\quad \times \frac{1}{T}\sum_{s=1}^{T}\left|N^{-1/2}\sum_{i=1}^{N}\left[e_{it}e_{is} - E\left(e_{it}e_{is}\right)\right]\right| \\
&= \frac{1}{N^{1/2}T^{2d-1/2}} O_p\left(\delta_{NT}^{-1}\right) O_p(1) \\
&= O_p\left(N^{-1/2}T^{3/2-2d}\delta_{NT}^{-1}\right)
\end{aligned}$$

Since for all t, $E\left|F_t^0/T^{d-1/2}\right| = O_p(1)$, it follows that

$$\begin{aligned}
E\left(\frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t' e_s}{N} - \gamma_N(s,t)\right)F_s^0\right) &\leqslant \frac{1}{T^{d-1/2}N^{1/2}}\max_{1\leqslant s\leqslant T} E\left|\frac{F_s^0}{T^{d-1/2}}\right| \\
&\quad \times E\left(\frac{1}{T}\sum_{s=1}^{T}\left|\frac{1}{N^{1/2}}\sum_{i=1}^{N}\left[e_{it}e_{is} - E\left(e_{it}e_{is}\right)\right]\right|\right) \\
&= \frac{1}{T^{d-1/2}N^{1/2}} O_p(1) O_p(1) \\
&= O_p\left(T^{1/2-d}N^{-1/2}\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{N\Upsilon_t \tilde{F}}{NT^3} &= O_p\left(N^{-1/2}T^{3/2-2d}\delta_{NT}^{-1}\right) + O_p\left(T^{1/2-d}N^{-1/2}\right) \\
&= O_p\left(N^{-1/2}\right) + O_p\left(T^{1/2-d}N^{-1/2}\right)
\end{aligned}$$

Next, we prove that $N\Phi_t\tilde{F}/\left(T^{2d}N\right) = O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)$.

$$\begin{aligned}\frac{N\Phi_t\tilde{F}}{NT^3} &= \frac{1}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\,\tilde{F}_s \\ &= \frac{1}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\left(\tilde{F}_s - \tilde{H}'F_s^0\right) + \frac{\tilde{H}'}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\,F_s^0\end{aligned}$$

The first expression is $O_p\left(T^{1/2-2d}\right)$ by Assumption E.1 and Proposition 5

$$\begin{aligned}\frac{1}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\left(\tilde{F}_s - \tilde{H}'F_s^0\right) &\leqslant \frac{1}{T^{2d-1/2}}\sum_{s=1}^{T}\left|\bar{\gamma}_{NT}(t,s)\right|\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - hF_s^0\right)^2\right)^{1/2} \\ &= \frac{1}{T^{2d-1/2}}O_p(1)\,O_p\left(\delta_{NT}^{-1}\right) \\ &= O_p\left(T^{1/2-2d}\right)\end{aligned}$$

The second expression is $O_p\left(T^{-1/2-d}\right)$ because

$$\frac{1}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\,F_s^0 \leqslant \frac{1}{T^{d+1/2}}\sum_{s=1}^{T}\left|\frac{F_s^0}{T^{d-1/2}}\right|\left|\bar{\gamma}_{NT}(t,s)\right|$$

Since for all t, $E\left|F_t^0/T\right| = O_p(1)$ then by Assumption E.1

$$\begin{aligned}E\left(\frac{1}{T^{d+1/2}}\sum_{s=1}^{T}\left|\frac{F_s^0}{T^{d-1/2}}\right|\left|\gamma_{NT}(t,s)\right|\right) &\leqslant \frac{1}{T^{d+1/2}}\max_{1\leqslant s\leqslant T}E\left|\frac{F_s^0}{T^{d-1/2}}\right|\sum_{s=1}^{T}\left|\gamma_{NT}(t,s)\right| \\ &= \frac{1}{T^{d+1/2}}O_p(1)\,O_p(1) \\ &= O_p\left(T^{-1/2-d}\right)\end{aligned}$$

Thus,

$$\begin{aligned}\frac{N\Phi_t\tilde{F}}{NT^3} &= O_p\left(T^{1/2-2d}\right) + O_p\left(T^{-1/2-d}\right) \\ &= O_p\left(T^{-1/2-d}\right) \\ &= O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)\end{aligned}$$

Therefore,

$$\begin{aligned}e_t'e'\tilde{F}D^{-2}/N &= O_p\left(N^{-1/2}\right) + O_p\left(T^{1/2-d}N^{-1/2}\right) + T^{-1/2}\left\|D^{-1}\right\| \\ &= O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)\end{aligned}$$

Consider (2). From Lemma 24 and the above point it follows that

$$\begin{aligned}\hat{F}_t - H'F_t^0 &= \frac{1}{N}\left\{e_t\Lambda_0 F^{0\prime} + e_t'e'\right\}\tilde{F}D^{-2} \\ &= O_p\left(N^{-1/2}\right) + O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right) \\ &= O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)\end{aligned}$$

Since $\tilde{F}_t - \tilde{H}'F_t^0 = V_{NT}^{-1}\left(\hat{F}_t - \tilde{H}'F_t^0\right)$ then also

$$\tilde{F}_t - \tilde{H}'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)$$

■

3.8.4 Asymptotic distribution

In this section, we derive the limiting distribution of the discussed estimators. Firstly, we show some general results and prove Lemma 8. Next, we discuss separately the issues associated with derivation of asymptotic distributions of the estimators of factors, factor loadings and common components.

Lemma 26 *Under Assumptions A-F, as $N, T \to \infty$,*

$$\left\|N^{-1}D^{-2}\tilde{F}'\left(XX'\right)\tilde{F}D^{-2} - N^{-1}D^{-2}\tilde{F}'F^0\left(\Lambda_0'\Lambda_0\right)F^{0\prime}\tilde{F}D^{-2}\right\|^2 = o_p\left(1\right)$$

Proof. Let us denote

$$b_{NT} = N^{-1}D^{-2}\tilde{F}'\left(XX'\right)\tilde{F}D^{-2} - N^{-1}D^{-2}\tilde{F}'F^0\left(\Lambda_0'\Lambda_0\right)F^{0\prime}\tilde{F}D^{-2}$$

Then

$$\begin{aligned}b_{NT} &= N^{-1}D^{-2}\tilde{F}'e\Lambda_0F^{0\prime}\tilde{F}D^{-2} + N^{-1}D^{-2}\tilde{F}'F^0\Lambda_0'e'\tilde{F}D^{-2} + N^{-1}D^{-2}\tilde{F}'ee'\tilde{F}D^{-2} \\ &= D^{-2}\tilde{F}'\left(e\Lambda_0F^{0\prime}\tilde{F}D^{-2}/N + F^0\Lambda_0'e'\tilde{F}D^{-2}/N + ee'\tilde{F}D^{-2}/N\right) \\ &= D^{-2}\tilde{F}'\left(\hat{F} - F^0H\right) + D^{-2}\tilde{F}'e\Lambda_0F^{0\prime}\tilde{F}D^{-2}/N\end{aligned}$$

Thus, by Proposition 5

$$\begin{aligned}\left\|b_{NT}\right\|/\sqrt{2} &\leq \left\|D^{-2}\tilde{F}'\left(\hat{F} - F^0H\right)\right\| + \left\|D^{-2}\tilde{F}'e\Lambda_0F^{0\prime}\tilde{F}D^{-2}/N\right\| \\ &\leq \sqrt{T}\left\|D^{-1}\right\|\left\|D^{-1}\tilde{F}'\right\|\left(\frac{1}{T}\left\|\hat{F} - F^0H\right\|\right)^{1/2} + \left\|D^{-1}\right\|\left\|D^{-1}\tilde{F}'\right\|\left\|e\Lambda_0F^{0\prime}\tilde{F}D^{-2}/N\right\| \\ &\leq O_p\left(1\right)O_p\left(\delta_{NT}^{-1}\right) + O_p\left(\left\|D^{-1}\right\|\right)O_p\left(1\right)O_p\left(N^{-1/2}\right) \\ &= O_p\left(\delta_{NT}^{-1}\right)\end{aligned}$$

Hence

$$\left\| N^{-1}D^{-2}\tilde{F}'\left(XX'\right)\tilde{F}D^{-2} - N^{-1}D^{-2}\tilde{F}'F^{0}\left(\Lambda_0'\Lambda_0\right)F^{0\prime}\tilde{F}D^{-2} \right\| = o_p\left(1\right)$$

■

Proof of Lemma 8. Consider (1). From Lemma 26 it follows that

$$\left\| D^{-2}\tilde{F}'\left(XX'/N\right)\tilde{F}D^{-2} - D^{-2}\tilde{F}'F^{0}\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}\tilde{F}D^{-2} \right\|^2 = o_p\left(1\right)$$

Let us denote V_{NT}^{*} the diagonal matrix consisting of the r largest eigenvalues of the matrix $F^{0}\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}$ multiplied by D^{-2} and F^{*}, the corresponding eigenvectors. Let us assume that $D^{-1}F^{*\prime}F^{*}D^{-1} = I$. Then

$$\left\| D^{-2}\tilde{F}'F^{0}\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}\tilde{F}D^{-2} - D^{-2}F^{*\prime}F^{0}\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}F^{*}D^{-2} \right\|^2 = o_p\left(1\right)$$

and $V_{NT} = V_{NT}^{*} + o_p\left(1\right)$. Moreover, the diagonal elements of V_{NT}^{*} are equal to the eigenvalues of the matrix $\left(F^{0\prime}F^{0}\right)\left(\Lambda_0'\Lambda_0/N\right)$ divided by $D^{-2}$and V_{NT}^{*} converges to V, where $V_{ii} = \lim_{N,T\to\infty} V_{NT,i}^{*} > 0$ by Lemma 20.

Consider (2). It can be shown that

$$\begin{aligned} D^{-1}\tilde{H}'F^{0\prime}F^{0}\tilde{H}D^{-1} &= D^{-1}\tilde{F}'\tilde{F}D^{-1} + o_p\left(1\right) \\ &= I + o_p\left(1\right) \end{aligned}$$

Since $\tilde{H} = \left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}\tilde{F}D^{-2}V_{NT}^{-1} + o_p\left(1\right)$, it holds that

$$D^{-3}V_{NT}^{-1}\tilde{F}'F^{0\prime}\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}F^{0}\left(\Lambda_0'\Lambda_0/N\right)^{0\prime}F^{0\prime}\tilde{F}V_{NT}^{-1}D^{-3} = I + o_p\left(1\right)$$

and

$$D^{-3}V_{NT}^{-1/2}\tilde{F}'F^{0\prime}\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}F^{0}\left(\Lambda_0'\Lambda_0/N\right)^{0\prime}F^{0\prime}\tilde{F}V_{NT}^{-1/2}D^{-3} = V_{NT} + o_p\left(1\right)$$

Let us denote

$$R_{NT} = \left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2} Q_{NT}V_{NT}^{-1/2}$$

From the definition of Q_{NT} and Lemma 26 it follows that $R_{NT}'R_{NT} = I + o_p\left(1\right)$. Then the equation can be transformed into

$$D^{-1}R_{NT}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}F^{0\prime}F^{0}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}R_{NT}D^{-1} = V_{NT} + o_p\left(1\right)$$

If the matrix D has all diagonal elements equal then it is straightforward that

$$R_{NT}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}D^{-1}F^{0\prime}F^{0}D^{-1}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}R_{NT} = V_{NT} + o_p\left(1\right)$$

and R_{NT} converges in distribution to the eigenvectors of the matrix $\Sigma_\Lambda^{1/2}\Sigma\Sigma_\Lambda^{1/2}$. Since the eigenvalues of the matrix $\Sigma_\Lambda^{1/2}\Sigma\Sigma_\Lambda^{1/2}$ are distinct then R is unique. Thus $Q = \Sigma_\Lambda^{-1/2} R V^{1/2}$ and Q is positive definite with probability 1.

If D has different elements on the diagonal then

$$R_i = \lim_{N,T\to\infty} v_i \left(\left(F^{0\prime} F^0 / D_{ii}^2 \right) \left(\Lambda_0' \Lambda_0 / N \right) \right)$$

where $v_i(A)$ denotes the eigenvector of matrix A corresponding with the ith largest eigenvalue. ■

Limiting distribution of estimated common factors

The following Lemma 27 is used in the proof of Proposition 9.

Lemma 27 *Under Assumptions A-F, for N, $T \to \infty$*

$$\sqrt{N}\left(\hat{F}_t - H' F_t^0 \right) \to^d Q' N(0, \Gamma_t)$$

Proof. Under the assumption $N \left\| D^{-2} \right\| \to 0$ by Proposition 6, we have

$$\sqrt{N}\left(\hat{F}_t - H' F_t^0 \right) = O_p(1) + O_p\left(\left\| D^{-1} \right\| N^{1/2} \right)$$

Thus, the limiting distribution is defined by the first term $e_t \Lambda_0 F^{0\prime} \tilde{F} D^{-2}/N$ and

$$\begin{aligned} \sqrt{N}\left(\hat{F}_t - H' F_t^0 \right) &= \frac{D^{-2} \tilde{F}' F^0 \Lambda_0' e_t}{\sqrt{N}} + o_p(1) \\ &= D^{-2} \tilde{F}' F^0 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} + o_p(1) \end{aligned}$$

By Assumption F and Lemma 8

$$\sqrt{N}\left(\hat{F}_t - H' F_t^0 \right) \to^d Q' N(0, \Gamma_t)$$

where Q is independent of $N(0, \Gamma_t)$ since it depends only on the common components that are independent from idiosyncratic disturbances. ■

Proof of Proposition 9. Under the Lemma 4 and Lemma 27

$$\begin{aligned} \sqrt{N}\left(\tilde{F}_t - \tilde{H}' F_t^0 \right) &= V_{NT}^{-1} \sqrt{N}\left(\hat{F}_t - H' F_t^0 \right) \\ &\to {}^d V^{-1} Q N(0, \Gamma_t) \end{aligned}$$

■

Limiting distribution of estimated factors loadings

Firstly, in Lemmas 28-30, we present some general results that are needed to prove Proposition 10. Then we present a proof of Proposition 10.

Lemma 28 *Under the assumption A − F for N, T → ∞ ,*

$$\bar{H} = D\tilde{H}D^{-1} = O_p(1)$$

and

$$\bar{H}\bar{H}' \to^d \Sigma^{-1}$$

Proof. Let us first notice that

$$\begin{aligned} D^{-1}\tilde{F}'\tilde{F}D^{-1} &= D^{-1}\tilde{H}'F^{0\prime}F^0\tilde{H}D^{-1} + D\tilde{H}'F^{0\prime}\left(\tilde{F} - F^0\tilde{H}\right)D^{-1} \\ &\quad + D^{-1}\left(\tilde{F} - F^0\tilde{H}\right)'F^0\tilde{H}D^{-1} \\ &\quad + TD^{-1}\frac{1}{T}\left(\tilde{F} - F^0\tilde{H}\right)'\left(\tilde{F} - F^0\tilde{H}\right)D^{-1} \end{aligned}$$

By Proposition 5 and Assumption A

$$\begin{aligned} \left\| TD^{-1}\frac{1}{T}\left(\tilde{F} - F^0\tilde{H}\right)'\left(\tilde{F} - F^0\tilde{H}\right)D^{-1}\right\| &\leq \frac{1}{T}\left\|\tilde{F} - F^0\tilde{H}\right\|^2 T\left\|D^{-2}\right\| \\ &= O_p\left(\delta_{NT}^{-2}\right) = o_p(1) \end{aligned}$$

Since $D^{-1}\tilde{F}'\tilde{F}D^{-1} = I_r = O_p(1)$, then

$$\bar{H}'\Sigma_{NT}\bar{H} + \bar{H}'B + B'\bar{H} = O_p(1) \tag{3.16}$$

where $\bar{H} = DHD^{-1}$, $\Sigma_{NT} = D^{-1}F^{0\prime}F^0D^{-1}$ and $B = D^{-1}F^{0\prime}\left(\tilde{F} - F^0\tilde{H}\right)D^{-1}$. Firstly, we show that $\|B\| = O_p(1)$. By Proposition 5 and Assumption A we have

$$\begin{aligned} \|B\| &= \left\|D^{-1}F^{0\prime}\left(\tilde{F} - F^0\tilde{H}\right)D^{-1}\right\| \\ &\leq \left\|D^{-1}F^{0\prime}\right\|\left(\frac{1}{T}\left\|\tilde{F} - F^0\tilde{H}\right\|^2\right)^{1/2}\left\|D^{-1}\right\|\sqrt{T} \\ &= O_p\left(\delta_{NT}^{-2}\right) = o_p(1) \end{aligned}$$

Since $\Sigma_{NT} = O_p(1)$ and $B = o_p(1)$, then from the properties of the quadratic form (3.16) it follows that $\bar{H}' = O_p(1)$. Then

$$\bar{H}'B + B'\bar{H} = o_p(1)$$

and

$$\bar{H}'\Sigma_{NT}\bar{H} = I + o_p(1)$$

Thus, by Assumption A

$$\begin{aligned}\bar{H}\bar{H}' &= \Sigma_{NT}^{-1} + o_p(1) \\ &\to {}^{d}\Sigma^{-1}\end{aligned}$$

■

Lemma 29 *Under Assumptions A-F, for $N, T \to \infty$*

1. $D^{-1}F^{0\prime}\left(\tilde{F} - F^0\tilde{H}\right) = O_p\left(\delta_{NT}^{-1}\right)$
2. $D^{-1}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right) = O_p\left(\delta_{NT}^{-1}\right)$

Proof. Consider (1). As noted by Bai (2004)

$$\begin{aligned}D^{-1}F^{0\prime}\left(\tilde{F} - F^0\tilde{H}\right) &= \sum_{t=1}^{T} D^{-1}F_t^0\left(\tilde{F}_t - \tilde{H}'F_t^0\right)' \\ &\leq \max_t\left(\sqrt{T}D^{-1}F_t^0\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left|\tilde{F}_t - \tilde{H}'F_t^0\right|'\end{aligned}$$

Moreover,

$$\left(\sum_{t=1}^{T}\left|\tilde{F}_t - \tilde{H}'F_t^0\right|'\right)\left(\sum_{t=1}^{T}\left|\tilde{F}_t - \tilde{H}'F_t^0\right|'\right)' \leq 2\sum_{t=1}^{T}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)'\left(\tilde{F}_t - \tilde{H}'F_t^0\right)$$

Thus, by Proposition 5

$$\left\|\sum_{t=1}^{T}\left|\tilde{F}_t - \tilde{H}'F_t^0\right|'\right\| \leq \left\|\tilde{F} - F^0\tilde{H}\right\| = O_p\left(\delta_{NT}^{-1}\sqrt{T}\right)$$

and under Assumption A

$$D^{-1}F^{0\prime}\left(\tilde{F} - F^0\tilde{H}\right) = O_p\left(\delta_{NT}^{-1}\right)$$

Part (2) follows directly from (1)

$$\begin{aligned}D^{-1}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right) &= D^{-1}\left(\tilde{F} - F^0\tilde{H}\right)'\left(\tilde{F} - F^0\tilde{H}\right) + D^{-1}\tilde{H}F^{0\prime}\left(\tilde{F} - F^0\tilde{H}\right) \\ &= TD^{-1}O_p\left(\delta_{NT}^{-2}\right) + O_p\left(\delta_{NT}^{-1}\right) \\ &= O_p\left(\delta_{NT}^{-1}\right)\end{aligned}$$

■

Lemma 30 *Under Assumptions A-E, for $N, T \to \infty$, we have for each i*

$$\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) = O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right) + O_p\left(\left\|D^{-1}\right\|\right)$$

Proof. Let us consider an expression for $\tilde{\lambda}_i$. Form the definition of $\tilde{\Lambda}' = D^{-2}\tilde{F}'X$ it follows that

$$\begin{aligned}\tilde{\lambda}_i &= D^{-2}\tilde{F}'\bar{X}_i \\ &= D^{-2}\tilde{F}'\left(F_0\lambda_i^0 + \bar{e}_i\right) \\ &= D^{-2}\left(\tilde{F}'F_0\right)\lambda_i^0 + D^{-2}\left(\tilde{F}'\bar{e}_i\right)\end{aligned}$$

Since $D^{-2}\tilde{F}'\tilde{F} = I$ and $F^0 = F^0 + \tilde{F}\tilde{H}^{-1} - \tilde{F}\tilde{H}^{-1}$ it follows

$$\begin{aligned}\tilde{\lambda}_i &= D^{-2}\tilde{F}'\tilde{F}\tilde{H}^{-1}\lambda_i^0 + D^{-2}\tilde{F}'\left(F^0 - \tilde{F}\tilde{H}^{-1}\right)\lambda_i^0 + D^{-2}\left(\tilde{F}'\bar{e}_i\right) \\ &= \tilde{H}^{-1}\lambda_i^0 + D^{-2}\tilde{F}'\left(F^0 - \tilde{F}\tilde{H}^{-1}\right)\lambda_i^0 + D^{-2}\left(\tilde{F}'\bar{e}_i\right)\end{aligned}$$

Hence,

$$\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0 = D^{-2}\tilde{F}'\left(F^0\tilde{H} - \tilde{F}\right)\tilde{H}^{-1}\lambda_i^0 + D^{-2}\tilde{F}'\bar{e}_i$$

The first part is $O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)$. By Lemma 29

$$\begin{aligned}\left\|D^{-2}\tilde{F}'\left(F^0\tilde{H} - \tilde{F}\right)\tilde{H}^{-1}\right\| &= \left\|D^{-2}\tilde{F}'\left(F^0\tilde{H} - \tilde{F}\right)\right\|\left\|\tilde{H}^{-1}\right\| \\ &= O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)\end{aligned}$$

From Assumption B it follows that $\lambda_i^0 = O_p(1)$. Therefore,

$$D^{-2}\tilde{F}'\left(F^0\tilde{H} - \tilde{F}\right)\tilde{H}^{-1}\lambda_i^0 = O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)$$

The second part can be decomposed as follows

$$D^{-2}\tilde{F}'\bar{e}_i = D^{-2}\left(\tilde{F} - F^0\tilde{H}\right)'\bar{e}_i + D^{-2}\tilde{H}'F^{0\prime}\bar{e}_i$$

By Proposition 5 the first expression $\left\|D^{-2}\left(\tilde{F} - F^0\tilde{H}\right)'\right\| = O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)$ because

$$\begin{aligned}\left\|D^{-2}\left(\tilde{F} - F^0\tilde{H}\right)'\right\| &= \left\|D^{-2}\right\|\sqrt{T}\left(\frac{1}{T}\left\|\tilde{F} - F^0\tilde{H}\right\|^2\right)^{1/2} \\ &= O_p\left(\left\|D^{-2}\right\|\sqrt{T}\right)O_p\left(\delta_{NT}^{-1}\right)O_p(1) \\ &= O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)\end{aligned}$$

Since $\bar{e}_i = O_p(1)$ then $D^{-2}\left(\tilde{F} - F^0\tilde{H}\right)'\bar{e}_i = O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)$. The second expression is $D^{-2}\tilde{H}'F^{0\prime}\bar{e}_i = O_p\left(\left\|D^{-1}\right\|\right)$

$$\begin{aligned}D^{-2}\sum_{i=1}^{T}F_t^0 e_{it} &\leqslant D^{-1}\max\left\|\sqrt{T}D^{-1}F_t^0\right\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|e_{it}| \\ &= O_p\left(\left\|D^{-1}\right\|\right)\end{aligned}$$

Thus,

$$\begin{aligned}\frac{\tilde{F}'\bar{e}_i}{T^3} &= O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)+O_p\left(\left\|D^{-1}\right\|\right)\\ &= O_p\left(\left\|D^{-1}\right\|\right)\end{aligned}$$

Finally,

$$\hat{\lambda}_i - H^{-1}\lambda_i^0 = O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\sqrt{N}\right)+O_p\left(\left\|D^{-1}\right\|\right)$$

■

Proof of Proposition 10. By Lemma 30, we have

$$D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) = O_p\left(\delta_{NT}^{-1}\right)+O_p\left(1\right)$$

Thus, the limiting distribution of $D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right)$ is determined by the last term $F^{0\prime}\bar{e}_i$. Therefore,

$$\begin{aligned}D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) &= DD^{-2}\tilde{H}'F^{0\prime}\bar{e}_i + o_p\left(1\right)\\ &= \bar{H}'\frac{1}{\sqrt{T}}\sum_{t=1}^{N}\sqrt{T}D^{-1}F_t^0 e_{it} + o_p\left(1\right)\end{aligned}$$

As discussed in Bai (2003), by Lemma 28

$$\bar{H}\bar{H}' \to^d \Sigma^{-1}$$

Thus

$$\bar{H}' \to^d \bar{H}^{-1}\Sigma^{-1}$$

where $\bar{H}$ is defined in Lemma 28. Therefore, by Assumption G there is

$$D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) \to^d \bar{H}^{-1}\Sigma^{-1}N\left(0,\Omega_i\right)$$

■

Corollary 31 *Under the Assumption A-F, for N, $T \to \infty$*

$$D\left(\hat{\lambda}_i - H^{-1}\lambda_i^0\right) \to^d V^{-1}\bar{H}^{-1}\Sigma^{-1}N\left(0,\Omega_i\right)$$

Proof. By Lemma 4 and Proposition 10

$$\begin{aligned}D\left(\hat{\lambda}_i - H^{-1}\lambda_i^0\right) &= V_{NT}^{-1}D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right)\\ &\to {}^dV^{-1}\bar{H}^{-1}\Sigma^{-1}N\left(0,\Omega_i\right)\end{aligned}$$

■

Maciejowska, Katarzyna(2010), Identification and Estimation of Sources of Common Fluctuations: New methodologies and applications
European University Institute DOI: 10.2870/19341

Limited distribution of estimated common components

Let us denote $C_{it}^0 = F_t^{0\prime}\lambda_i^0$ and $\hat{C}_{it} = \hat{F}_t'\hat{\lambda}_i$. The asymptotic distribution of common components follows from the above Proposition 9 and 10.

Proof of Proposition 11. From the definition of $\hat{C}_{it}$ and C_{it}^0, we get

$$\hat{C}_{it} - C_{it}^0 = \left(\tilde{F}_t - \tilde{H}'F_t^0\right)' \tilde{H}^{-1}\lambda_i^0 + \tilde{F}_t'\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right)$$

By Proposition 6 and Assumption B we have that

$$\left(\tilde{F}_t - \tilde{H}'F_t^0\right)' H^{-1}\lambda_i^0 = O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$$

Finally, by Proposition 10 and Lemma 30

$$\begin{aligned}\tilde{F}_t'\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) &= \tilde{F}_t'D^{-1}\sqrt{T}D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right)T^{-1/2} \\ &= O_p\left(T^{-1/2}\right) + O_p\left(\delta_{NT}^{-1}T^{-1/2}\right) = O_p\left(T^{-1/2}\right)\end{aligned}$$

1. If $N/T \to 0$ then $N^{1/2}\left\|D^{-1}\right\| \to 0$ and

$$\begin{aligned}\sqrt{N}\left(\hat{C}_{it} - C_{it}^0\right) &= O_p(1) + O_p\left(N^{1/2}T^{-1/2}\right) \\ &= O_p(1) + o_p(1)\end{aligned}$$

Thus, by Proposition 6

$$\begin{aligned}\sqrt{N}\left(\hat{C}_{it} - C_{it}^0\right) &= \lambda_i^{0\prime}\tilde{H}^{-1\prime}\sqrt{N}\left(\tilde{F}_t - \tilde{H}'F_t^0\right) + o_p(1) \\ &\to {}^d\lambda_i^{0\prime}\left(V^{-1}Q\Sigma_\Lambda\right)^{-1}V^{-1}QN(0,\Gamma_t) \\ &= \lambda_i^{0\prime}\Sigma_\Lambda N(0,\Gamma_t)\end{aligned}$$

2. If $T/N \to 0$ then

$$\begin{aligned}\sqrt{T}\left(\hat{C}_{it} - C_{it}^0\right) &= O_p\left(T^{1/2}N^{-1/2}\right) + O_p(1) \\ &= o_p(1) + O_p(1)\end{aligned}$$

By Proposition 10 and under assumption $t/T = \tau$

$$\begin{aligned}\sqrt{T}\left(\hat{C}_{it} - C_{it}^0\right) &= \tilde{F}_t'\sqrt{T}D^{-1}D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) + o_p(1) \\ &= F_t^{0\prime}\tilde{H}\sqrt{T}D^{-1}D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) + o_p(1) \\ &= F_t^{0\prime}\sqrt{T}D^{-1}\bar{H}D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) + o_p(1) \\ &\to {}^dF_\tau'\bar{H}\left(\bar{H}\right)^{-1}\Sigma^{-1}W_i \\ &= F_\tau'\Sigma^{-1}W_i\end{aligned}$$

3. If $N/T \to \pi$ and $t/T = \tau$

$$\begin{aligned}\sqrt{N}\left(\hat{C}_{it} - C_{it}^0\right) &= O_p(1) + \sqrt{\pi}O_p(1) \\ &= \lambda_i^{0\prime} H^{-1\prime}\sqrt{N}\left(\hat{F}_t - H'F_t^0\right) + \sqrt{\pi}F_t^{0\prime}\sqrt{T}D^{-1}D\left(\hat{\lambda}_i - H^{-1}\lambda_i^0\right) + o_p(1) \\ &\to {}^d\lambda_i^{0\prime}\Sigma_\Lambda N(0,\Gamma_t) + \sqrt{\pi}F_\tau'\Sigma^{-1}W_i\end{aligned}$$

■

Confidence intervals

Consider the rotation of $\tilde{F}$ towards an observable variable R_t described by the regression

$$R_t = \alpha + \beta\left(\tilde{H}^{-1}\tilde{F}_t\right) + error$$

Let $\left(\hat{\alpha}, \hat{\beta}\right)$ be the least-squares estimator of (α, β) and $\hat{R}_t = \hat{\alpha} + \hat{\beta}\left(\tilde{H}^{-1}\tilde{F}_t\right)$.

In Lemma 32 we show some properties of the factor estimators that are used in the proof of Proposition 12.

Lemma 32 *Under Assumptions A-E and* $T\left\|D^{-2}\right\| \leq M$ *we have for* $N, T \to \infty$

1. *If* $N^{1/2}T^{-1/2}\left\|D^{-1}\right\| \to 0$ *then*

$$\left\|N^{1/2}T^{-1/2}D^{-1}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\right\| = N^{1/2}T^{-1/2}O_p\left(\delta_{NT}^{-1}\right) = o_p(1)$$

2. *If* $N^{1/2}T^{-1/2}\left\|D^{-1}\right\| \to 0$ *then*

$$\left\|N^{1/2}T^{-1}\sum_{t=1}^{T}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)\right\| = N^{1/2}T^{-1/2}O_p\left(\delta_{NT}^{-1}\right) = o_p(1)$$

3. $\left\|D^{-1}T^{1/2}\tilde{F}_t\right\| = O_p(1)$

Proof. Consider (1). Let us notice that

$$\left\|N^{1/2}T^{-1/2}D^{-1}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\right\| = \|D\|\,T^{-1/2}\left\|N^{1/2}D^{-2}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\right\|$$

By Lemma 29

$$\begin{aligned}\left\|N^{1/2}D^{-2}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\right\| &= O_p\left(N^{1/2}\right)\left(O_p\left(\delta_{NT}^{-2}T\left\|D^{-2}\right\|\right) + O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)\right) \\ &= O_p\left(N^{1/2}\delta_{NT}^{-2}T\left\|D^{-2}\right\|\right) + O_p\left(N^{1/2}\left\|D^{-1}\right\|\delta_{NT}^{-1}\right) \\ &= O_p\left(N^{1/2}\delta_{NT}^{-2}\right) + O_p\left(N^{1/2}\left\|D^{-1}\right\|\delta_{NT}^{-1}\right) \\ &= O_p\left(\delta_{NT}^{-1}\right) + O_p\left(\left\|D^{-1}\right\|\right) = o_p(1)\end{aligned}$$

Thus,

$$\begin{aligned}\left\|N^{1/2}T^{-1/2}D^{-1}\tilde{F}'\left(\tilde{F}-F^0\tilde{H}\right)\right\| &= \|D\|\,T^{-1/2}o_p(1)\\ &= o_p(1)\end{aligned}$$

Consider (2).

$$\begin{aligned}\left\|\sum_{t=1}^{T}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)\right\|^2 &= tr\left(\left(\sum_{t=1}^{T}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)\right)\left(\sum_{t=1}^{T}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)\right)'\right)\\ &\leq tr\left(2\sum_{t=1}^{T}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)\left(\tilde{F}_t-\tilde{H}'F_t^0\right)'\right)\\ &= 2tr\left(\left(\tilde{F}-F^0\tilde{H}\right)'\left(\tilde{F}-F^0\tilde{H}\right)\right)\\ &= 2\left\|\tilde{F}-F^0\tilde{H}\right\|^2=O_p\left(T\delta_{NT}^{-2}\right)\end{aligned}$$

Thus,

$$\left\|N^{1/2}T^{-1}\sum_{t=1}^{T}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)\right\|=N^{1/2}T^{-1/2}O_p\left(\delta_{NT}^{-1}\right)$$

Consider (3). $D^{-1}T^{1/2}\tilde{F}_t$ can be decomposed into two parts

$$\left\|D^{-1}T^{1/2}\tilde{F}_t\right\|=\left\|D^{-1}T^{1/2}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)\right\|+\left\|D^{-1}T^{1/2}\tilde{H}'F_t^0\right\|$$

By Proposition 6, $\left\|\tilde{F}_t-\tilde{H}'F_t^0\right\|=o_p(1)$. Moreover, by Assumption A

$$\begin{aligned}\left\|D^{-1}T^{1/2}\tilde{F}_t\right\| &= o_p(1)+\left\|\bar{H}'\right\|\left\|D^{-1}T^{1/2}F_t^0\right\|\\ &= o_p(1)+O_p(1)=O_p(1)\end{aligned}$$

■

Proof of Proposition 12. One can express $\hat{R}_t-\alpha-\beta F_t^0$ as follows

$$\begin{aligned}\hat{R}_t-\alpha-\beta F_t^0 &= \hat{\alpha}+\hat{\beta}\left(\tilde{H}^{-1\prime}\tilde{F}_t\right)-\alpha-\beta F_t^0\\ &= (\hat{\alpha}-\alpha)+\left(\hat{\beta}-\beta\right)\left(\tilde{H}^{-1\prime}\tilde{F}_t\right)+\beta\tilde{H}^{-1\prime}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)\end{aligned}$$

Thus,

$$\sqrt{N}\left(\hat{R}_t-\alpha-\beta F_t^0\right)=\sqrt{N}(\hat{\alpha}-\alpha)+\sqrt{N}\left(\hat{\beta}-\beta\right)\left(\tilde{H}^{-1\prime}\tilde{F}_t\right)+\sqrt{N}\beta\tilde{H}^{-1\prime}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)$$

It can be shown that the first two terms are $o_p(1)$. Let us denote $Z_t=\left[1,\left(\tilde{H}^{-1}\tilde{F}_t\right)'\right]$ and a $T\times(1+r)$ matrix $Z'=[Z_1',\ldots,Z_T']$. We write ι to

describe a $T \times 1$ vector $\iota' = [1, ..., 1]$. The parameter vector $\psi = (\alpha, \beta)'$ is estimated with the least-squares method. Thus, $\hat{\psi} = (Z'Z)^{-1} Z'R$. Under the null $R_t = \alpha + \beta F_t^0 = \alpha + \beta \left(\tilde{H}^{-1\prime} \tilde{F}_t\right) + \beta \tilde{H}^{-1\prime} \left(\tilde{H}' F_t^0 - \tilde{F}_t\right)$ and in matrix notation $R = Z\psi + \left(F^0 \tilde{H}' - \tilde{F}\right) \tilde{H}^{-1} \beta'$. Therefore,

$$\begin{aligned} \hat{\psi} &= (Z'Z)^{-1} Z'R \\ &= (Z'Z)^{-1} Z'Z\psi + (Z'Z)^{-1} Z' \left(F^0 \tilde{H}' - \tilde{F}\right) \tilde{H}^{-1} \beta' \\ &= \psi + (Z'Z)^{-1} Z' \left(F^0 \tilde{H}' - \tilde{F}\right) \tilde{H}^{-1} \beta' \end{aligned}$$

So

$$\hat{\psi} - \psi = (Z'Z)^{-1} Z' \left(F^0 \tilde{H}' - \tilde{F}\right) \tilde{H}^{-1} \beta'$$

Let us define a $(1 + r) \times (1 + r)$ diagonal matrix

$$D_T = \begin{bmatrix} T^{1/2} & 0 \\ 0 & D \end{bmatrix}$$

where D_T is the scaling matrix. Then

$$\left(\hat{\psi} - \psi\right) = D_T^{-1} M D_T^{-1} Z' \left(F^0 \tilde{H}' - \tilde{F}\right) \tilde{H}^{-1} \beta'$$

with $M = \left(D_T^{-1} Z'Z D_T^{-1}\right)^{-1} = O_p(1)$. Let us denote the blocks of the matrix M as follow

$$M = \begin{bmatrix} M_{11} & M_{1F} \\ M_{F1} & M_{FF} \end{bmatrix}$$

where M_{11} is a 1×1 matrix and M_{FF} is a $r \times r$ matrix.

This implies that by Lemma 29 and Lemma 32 $\left\|\sqrt{N}(\hat{\alpha} - \alpha)\right\| = o_p(1)$

$$\begin{aligned} \left\|\sqrt{N}(\hat{\alpha} - \alpha)\right\| / \sqrt{2} &= \left\|N^{1/2} T^{-1/2} M D_T^{-1} Z' \left(F^0 \tilde{H}' - \tilde{F}\right) \tilde{H}^{-1} \beta'\right\| / \sqrt{2} \\ &\leq \left\|N^{1/2} T^{-1/2} M_{11} T^{-1/2} \iota' \left(F^0 \tilde{H}' - \tilde{F}\right) \tilde{H}^{-1} \beta'\right\| \\ &\quad + \left\|N^{1/2} T^{-1/2} M_{1F} D^{-1} \tilde{H}^{-1\prime} \tilde{F}' \left(F^0 \tilde{H}' - \tilde{F}\right) \tilde{H}^{-1} \beta'\right\| \\ &= \|M_{11}\| \left\|\frac{N^{1/2}}{T} \iota' \left(F^0 \tilde{H}' - \tilde{F}\right)\right\| \left\|\tilde{H}^{-2} \beta'\right\| \\ &\quad + \|M_{1F}\| \left\|\frac{N^{1/2}}{T^{1/2}} D^{-1} \tilde{F}' \left(F^0 \tilde{H}' - \tilde{F}\right)\right\| \left\|\tilde{H}^{-2} \beta'\right\| \\ &= O_p(1)\, o_p(1)\, O_p(1) + O_p(1)\, o_p(1)\, O_p(1) \\ &= o_p(1) \end{aligned}$$

By Lemma 32 $\sqrt{N}\left(\hat{\beta}-\beta\right)\left(\tilde{H}'\tilde{F}_t\right)=o_p(1)$

$$\begin{aligned}
\left\|\sqrt{N}\left(\hat{\beta}-\beta\right)\left(\tilde{H}'\tilde{F}_t\right)\right\|/\sqrt{2} &= \left\|N^{1/2}D^{-1/2}MD_T^{-1}Z'\left(F^0\tilde{H}'-\tilde{F}\right)\tilde{H}^{-1}\beta'\tilde{H}'\tilde{F}_t\right\|/\sqrt{2} \\
&\leq \left\|N^{1/2}D^{-1}M_{F1}T^{-1/2}\iota'\left(F^0\tilde{H}'-\tilde{F}\right)\tilde{H}^{-1}\beta'\tilde{H}'\tilde{F}_t\right\| \\
&\quad+\left\|N^{1/2}D^{-1}M_{FF}D^{-1}\tilde{H}^{-1\prime}\tilde{F}'\left(F^0\tilde{H}'-\tilde{F}\right)\tilde{H}^{-1}\beta'\tilde{H}'\tilde{F}_t\right\| \\
&= \|M_{F1}\|\left\|\frac{N^{1/2}}{T}\iota'\left(F^0\tilde{H}'-\tilde{F}\right)\right\|\left\|D^{-1}T^{1/2}\tilde{F}_t\right\|\left\|\tilde{H}^{-3}\beta'\right\| \\
&\quad+\|M_{FF}\|\left\|\frac{N^{1/2}}{T^{-1/2}}D^{-1}\tilde{F}'\left(F^0\tilde{H}'-\tilde{F}\right)\right\|\left\|D^{-1}T^{1/2}\tilde{F}_t\right\|\left\|\tilde{H}^{-2}\beta'\right\| \\
&= o_p(1)
\end{aligned}$$

Therefore,

$$\sqrt{N}\left(\hat{R}_t-\alpha-\beta F_t^0\right)=o_p(1)+\sqrt{N}\beta\tilde{H}^{-1\prime}\left(\tilde{H}'F_t^0-\tilde{F}_t\right)$$

Since $\left(\hat{\beta}-\beta\right)=o_p(1)$ and $\sqrt{N}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)=O_p(1)$, then β can be replaced with $\hat{\beta}$ and

$$\sqrt{N}\left(\hat{R}_t-\alpha-\beta F_t^0\right)=o_p(1)+\sqrt{N}\hat{\beta}\tilde{H}^{-1\prime}\left(\tilde{H}'F_t^0-\tilde{F}_t\right)$$

Finally, by Proposition 9

$$\begin{aligned}
\sqrt{N}\left(\hat{R}_t-\alpha-\beta F_t^0\right) &\rightarrow {}^d\hat{\beta}\tilde{H}^{-1\prime}V^{-1}QN(0,\Gamma_t) \\
&= \hat{\delta}V^{-1}QN(0,\Gamma_t)
\end{aligned}$$

■

Bibliography

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

Bai, J. (2003). Inferential theory for factor models of large dimentions, *Econometrica* **71**(1): 135–171.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data, *Journal of Econometrics* **122**: 137–183.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**(1): 191–221.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration, *Econometrica* **72**(4): 1127–1177.

Bai, J. and Ng, S. (2006). Confidential intervals for diffusion index forecasts and inference for factor-augmented regressions, *Econometrica* **74**(4): 1133–1150.

Banerjee, A. and Marcellino, M. (2008). Factor-augmented error correction models, EUI Working Pepers ECO 2008/15.

Bernanke, B. S., Boivin, J. and Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach, *The Quarterly Journal of Economics* pp. 387–422.

Binswanger, M. (2004). How do stock prices respond to fundamental shocks?, *Finance Research Letters* **1**: 90–99.

Blanchard, O. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances, *American Economic Review* **79**: 655–673.

Breitung, J. and Eickmeier, S. (2005). Dynamic factor models, Deutsche Bundesbank Discussion Paper, Series 1: Economic Studies, No 38/2005.

Brüggemann, R. and Lütkepohl, H. (2005). Uncovered interest rate parity and the expectations hypothesis of the term structure: Empirical results for the U.S. and europe, *Applied Economics Quarterly* **51**: 143–154.

Canova, F. and De Nicoló, G. (2002). Monetary disturbances matter for business fluctuations in the G-7, *Journal of Monetary Economics* **34**: 1131–1159.

Cattell, R. B. (1966). The Scree test for the number of factors, *Multivariate Behavioral Research* **1**: 245–276.

Child, D. (2006). *The essentials of factor analysis*, Continuum, New York.

Christiano, L. J., Eichenbaum, M. and Evans, C. L. (1999). *Handbook of Macroeconomics*, Elsevier.

Day, N. E. (1969). Estimating the components of a mixture of normal distributions, *Biometrica* **56**(3): 463–474.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum-likelihood from incomplete data via the EM algorithm, *Journal of Royal Statistics Society Ser. B (methodological)* **39**: 1–38.

Diebold, F. X. and Rudebusch, G. D. (1996). Measuring business cycle: A modern perspective, *Review of Economics and Statistics* **78**: 67–77.

Douc, R., Moulines, E. and Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, *Annals of Statistics* **32**: 2254–2304.

Eickmeier, S. (2009). Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model, *Journal of Applied Economietrics* .

Faust, J. (1998). The robustness of identified VAR conclusions about money, *Carnegie-Rochester Conference Series in Publilc Policy* **49**: 207–244.

Forni, M. and Gambetti, L. (2008). The dynamic efects of monetary policy: A structural factor model approach, RECent Working Paper No. 26.

Forni, M., Giannone, D., Lippi, M. and Reichlin, L. (2007). Opening the black box - structural factor models woth large cross-sections, Working Paper Series No. 712, European Central Bank.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2003). The generalized dynamic factor model: Identification and estimation, LEM Papers Series 2003/13.

Francq, C. and Roussignol, M. (1997). On the white noise driven by the hidden Markov chains, *Journal of Time Series Analysis* **18**: 553–578.

Gonzalo, J. and Granger, C. (1995). Estimation of common long-memory components in cointegrated systems, *Journal of Business & Economic Statistics* **13**(1): 27–35.

Goodwin, T. H. (1993). Business-cycle analysis with a markov-switching model, *Journal of Business & Economic Statistics* **11**(3): 331–339.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**: 357–384.

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

Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distribution, *The Annals of Statistics* **18**(2): 795–800.

Kapetanios, G. and Marcellino, M. (2006). Impulse response functions from structural dynamic factor models: A monte carlo evaluation, CEPR Discussion Paper No. 5621.

Kiefer, A. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Ann. Math. Statist* pp. 887–906.

Kim, C.-J. and Nelson, C. R. (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching, *The Review of Economics and Statistics* **80**(2): 188–201.

Kim, C.-J. and Nelson, C. R. (1999). Has the U.S. economy become more stable? a bayesian approach based on a markow-switching model of the business cycle, *The Review of Economics and Statistics* **81**(4): 608–616.

King, R. G., Plosser, C. I., Stock, J. H. and Watson, M. W. (1991). Stochastic trends and economic fluctuations, *American Economic Review* **81**: 819–840.

Koop, G. (1992). Aggregate shocks and macroeconomic fluctuations: A bayesian approach, *Journal of Applied Econometrics* **7**: 395–411.

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*, Springer-Verlag, Berlin.

Lanne, M. and Lütkepohl, H. (2005). Structural vector autoregressions with nonnormal residuals, CESinfo Working Paper No. 330.

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

Lanne, M. and Lütkepohl, H. (2009). Structural vector autoregressions with nonnormal residuals, *Journal of Business & Economic Statistics* . forthcoming.

Lütkepohl, H. (1996). *Handbook of Matrices*, John Wiley & Sons, Chichester.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

McLachlan, G. J. and Krishnan, T. (1997). *The EM Algorithm and Extensions*, Wiley.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*, Wiley.

Pagan, A. R. and Pesaran, M. H. (2008). Econometric analysis of structural systems with permanent and transitory shocks, *Journal of Economic Dynamics and Control* **32**: 3376–3395.

Rapach, D. E. (2001). Macro shocks and real stock prices, *Journal of Economics and Business* **53**: 5–26.

Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm, *Society for Industrial and Appied Mathematics* **26**: 195–239.

Rigobon, R. (2003). Identification through heteroscedasticity, *Review of Economics and Statistics* **85**: 777–792.

Rothenberg, T. J. (1971). Identification in parametric models, *Econometrica* **39**(3): 577–591.

Rubio-Ramirez, J. F., Waggoner, D. and Zha, T. (2005). Markov-switching structural vector autoregressions: Theory and applications, Discussion Paper, Federal Reserve Bank of Atlanta.

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**: 1–48.

Sims, C. A., Waggoner, D. F. and Zha, T. (2008). Methods for inference in large multiple-equation Markov-switching models, *Journal of Econometrics* **146**: 255–274.

Sims, C. A. and Zha, T. (2006). Were there regime switches in U.S. monetary policy?, *American Economic Review* **96**: 54–81.

Smith, A., Naik, P. A. and Tsai, C.-L. (2006). Markov-switching model selection using kullback-leibler divergence, *Journal of Econometrics* **134**: 553–577.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a larg number of predictors, *Journal of the American Statistical Association* **97**(460): 1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics* **20**(2): 147–162.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factors models for VAR analysis, NBER Working Papers No.11467.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from agnostic identification procedure, *Journal of Monetary Economics* **52**: 381–419.