Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Non-stationary Dynamic Factor Models

Autor:

FRANCISCO DE JESÚS CORONA VILLAVICENCIO

Director/es:

ESTHER RUIZ ORTEGA

PILAR PONCELA BLANCO

DEPARTAMENTO DE ESTADÍSTICA

Getafe, julio de 2017

Universidad
Carlos III de Madrid
www.uc3m.es

(a entregar en la Oficina de Posgrado, una vez nombrado el Tribunal evaluador , para preparar el documento para la defensa de la tesis)

TESIS DOCTORAL

Non-stationary Dynamic Factor Models

Autor: *Francisco de Jesús Corona Villavicencio*

Director/es: Esther Ruiz Ortega

Pilar Poncela Blanco

Firma del Tribunal Calificador:

Firma

Presidente: (Nombre y apellidos)

Vocal: (Nombre y apellidos)

Secretario: (Nombre y apellidos)

Calificación:

Getafe, 14 de julio de 2017

Universidad Carlos III de Madrid

PH.D. THESIS

Non-stationary Dynamic Factor Models

Author:

Francisco Corona

Advisor:

Esther Ruiz & Pilar Poncela

DEPARTMENT OF STATISTICS

Getafe, Madrid, June 5, 2017

A mi familia,

donde todo siempre es genuina y hermosamente igual

Agradecimientos

Quisiera agradecer principalmente a mis asesoras Esther Ruiz y Pilar Poncela por el apoyo brindado durante estos más de 3 años en los cuales pude aprenderles muchas cosas que valoro demasiado. De verdad, sus sugerencias, orientación, impulso y tiempo dedicado han sido más que imprescindibles para poder concluir la tesis. También agradezco el apoyo que me dieron para todos los trámites que necesitaba. No me queda más que agradecerles muchísimo por todo.

Al Consejo Nacional de Ciencia y Tecnología de México (CONACYT) y al Instituto de Innovación y Transferencia de Tecnología de Nuevo León (I2T2), porque a pesar de las dificultades económicas del país, sigue facilitando e impulsando a que varios de nosotros podamos realizar lo que nos gusta. Hago mención al proyecto CONACYT 252996 CB 2016-2018.

A la Fundación Carolina, especialmente a Juan Torres por su seguimiento y apoyo brindado durante mi estancia en España.

A Graciela González quien fungió como mi tutora en el seguimiento de beca CONACYT-I2T2 en México, pero más que eso, por su incondicional ayuda y apoyo en todo momento. También a Manuel Lecuanda por su orientación durante muchos años. A Bertín Ramírez y Nelson Muriel por leer siempre mis avances de la tesis y darme sus comentarios de valiosa ayuda. A Antonio Costilla por sus múltiples consejos que siempre me hace.

A Juan de Dios Tena y Mike Wiper, con quienes dentro del Departamento pude compartir el gusto por la Estadística en un tema que tenemos en común (el fútbol) y que de muchas formas me estimula a poder realizar otras cosas que también me agradan.

A Susana Linares quien muy eficazmente siempre ha resuelto mis dudas administrativas.

Mi agradecimiento muy especial a María, amiga y compañera de despacho y con que siempre puedo contar para cualquier cosa. También a Daniel, Javier y Raúl, que a pesar de que pasó rápido el tiempo, seguro que en algún momento nos encontraremos de vuelta.

A Velvet y *Machete*, mis primeros amigos en la Universidad y en Getafe, y con quien sigo compartiendo muchísimas cosas. También a Fátima, por la última etapa en Madrid.

A mis amigos del bar Alex (mejor conocido como el bar de Julián) de mi barrio Juan de la Cierva, por haberse preocupado siempre por mí y haberme hecho estar como en mi casa. Sin duda alguna y ustedes lo saben, conocí una de las más hermosas rutinas que ahora se convierten en experiencias que las llevo conmigo a donde quiera que vaya... *"Es lo que hay ..."*

A los "del Encuentro", *donde siempre me imaginaba estar*, prolongando mis horas de trabajo, estudio y *algo más*, y que puedo decir ahora, y sin darme cuenta, hice amigos para toda la vida.

A mis amigos de Monterrey por su apoyo incondicional pese a la distancia y también por su hospitalidad durante mis estancias en los veranos de investigación en la ciudad.

A mis amigos de la Ciudad de México, *Guanajuato 226-E*, por haberme acogido durante mi verano de investigación en el Banco de México y *por muchas cosas más...*

También a mis amigos de Baja California, con quienes de alguna forma empecé algo de esto hace ya algunos años. Especialmente agradezco a Pedro Orraca, amigo y colega de toda la vida.

A Lau, porque aunque suene como un relato de Cantinflas, *no sabíamos lo que estábamos haciendo pero era algo lo que hacíamos, aunque no sabíamos muy bien qué*. Quizá se llama aprender a vivir, pero ahora puedo ver que las horas de pláticas, las comidas y todo lo que vivimos en ese espacio de menos de 60 metros cuadrados que siempre echo de menos, todo, fue de gran ayuda ... ¿Quién lo iba a decir ese febrero de 2011? Estábamos el lugar *donde siempre me ha gustado estar...*

Y en especial a mi familia. No tengo mucho qué decir porque saben todo lo que les agradezco. A mi mamá Elvia y a mi papá *el Comandante Corona*; a mis hermanos Brenda y Víctor; a mis sobrinos Nadia, David, Itzel, Manel y Daniel; a mi cuñada Anna y *al* Andrei. Aunque uno puede pensar que la distancia nos aleja casi siempre, con ustedes es como si nunca pasara esto, como si todo fuese siempre igual, y la mejor sensación que experimento es saberlo con total certeza.

Abstract

This dissertation focuses on studying two topics of large non-stationary Dynamic Factor Models (DFMs). A very common practice when extracting factors from non-stationary multivariate time series is to differentiate each variable in the system. As a consequence, the ratio between variances and the dynamic dependence of the common and idiosyncratic differentiated components may change with respect to the original components. In the first step, we analyze the effects of these changes on the finite sample properties of several procedures to determine the number of factors. In particular, we consider the information criteria of Bai and Ng (2002), the edge distribution of Onatski (2010) and the ratios of eigenvalues proposed by Ahn and Horenstein (2013). The performance of these procedures when implemented to differentiated variables depends on both the ratios between variances and dependencies of the differentiated factor and idiosyncratic noises. Furthermore, we also analyze the role of the number of factors in the original non-stationary system as well as of its temporal and cross-sectional dimensions. Finally, we implement the different procedures to determine the number of common factors in a system of inflation rates in 15 euro area countries.

In the second step, we analyze and compare the finite sample properties of alternative factor extraction procedures in the context of non-stationary DFMs. On top of considering procedures already available in the literature, we extend the hybrid method based on the combination of Principal Components and Kalman filter and smoothing algorithms to non-stationary models. We show that, unless the idiosyncratic noise is non-stationary, procedures based on extracting the factors using the non-stationary original series work better than those based on differenced variables. The results are illustrated in an empirical application fitting non-stationary DFM to

aggregate GDP and consumption of the set of 21 OECD industrialized countries. The goal is to check international risk sharing is a short or long-run issue.

Resumen

Esta tesis se centra en estudiar dos tópicos de Modelos de Factores Dinámicos (DFMs) cuando el número de series de tiempo y su temporalidad es grande. Una práctica común cuando se extraen factores en un sistema no estacionario, es diferenciar cada una de sus variables. Como consecuencia, la relación entre las varianzas y la dinámica de dependencias tanto la parte común como la del componente idiosincrático, ambos diferenciados, puede cambiar respecto a la de los componentes originales. Primero, se analizarán los efectos muestrales que tiene esta práctica para algunos procedimientos de determinación del número de factores. En particular, nos centramos en analizar el funcionamiento de los criterios de información de Bai and Ng (2002), la diferencia de valores propios de Onatski (2010) y las razones de valores propios propuesta por Ahn and Horenstein (2013). Podemos concluir que cuando diferenciamos las variables, el funcionamiento de estos procedimientos depende de las razones de varianza y de las dependencias de los factores y errores idiosincráticos diferenciados. También analizamos el funcionamiento de los procedimientos cuando se usan las observaciones originalmente no estacionarias. Finalmente, implementamos estos procedimientos para determinar el número de factores en un sistema de tasas de inflación de 15 países de la zona euro.

Posteriormente analizamos y comparamos el funcionamiento en muestras finitas de algunas alternativas disponibles en la literatura para extraer factores comunes dentro del contexto de DFMs no estacionarios. Adicionalmente, extendemos el método híbrido basado en combinar Componentes Principales y el suavizamiento de Kalman en modelos no estacionarios. Mostramos que, al menos que los errores idiosincráticos sean no estacionarios, los procedimientos basados en extraer los factores comunes usando series originalmente no estacionarias, funcionan mejor que

los métodos que extraen los factores usando variables diferenciadas. Los resultados obtenidos anteriormente son ilustrados para estimar un DFM no estacionario usando las variables del PIB y consumo agregado de 21 países industrializados de la OCDE. El objetivo es determinar si el riesgo internacional compartido es un fenómeno de corto o de largo plazo.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Motivation

Dynamic Factor Models (DFMs) are useful for representing the dynamics of a group of N correlated time series through a small number of underlying common factors. Nowadays in economics, the availability of large amount of information collected during extended periods, T, grants the opportunity to understand several economic phenomena in a better way. Hence, researchers have the econometric tools which allow for the adequately modelling of large amounts of data, while taking into account several features of that modelling, for instance: the objectives, the economic theory which supports the inclusion of certain time series into a multivariate system, the stochastic nature of the observations, the dimensionality, etc. Since a large amount of time series data is presently available through statistical agencies, and over an increasingly long timespan, we focus on factor models in which the number of time series and the time dimension are large.

In economics, large DFMs have been mainly used with one of the following two objectives: i) estimating the few underlying factors for macroeconomic policy making (for example: obtaining the business cycle, lagging, coincident and leading indicators, instrumental variables, among many other applications) and ii) forecasting macroeconomic variables. For instance, the goal of estimation is pursued in Bernanke and Boivin (2003), Artis et al. (2004), Lahiri and Yao (2004), Bernanke et al. (2005), Favero et al. (2005) and Stock and Watson (2005) whereas forecasting is explored by Stock and Watson (2002a,b), Marcellino et al. (2003), Schumacher (2005) and Boivin and Ng (2006). Applications in other areas are given by Alonso et al. (2011), García-Martos et al. (2011), Mestekemper et al. (2013), Alonso et al. (2016), among many others.

A distinctive feature of macroeconomic time series, as is well known and documented in, for example, Engle and Granger (1987), Kunst and Neusser (1997) and Johansen (1988, 1991), is non-stationarity which entails the possibility of cointegration in multidimensional systems. Now, cointegration may be regarded as a representation in terms of common factors as is done in Stock and Watson (1988), Vahid and Engle (1993), Escribano and Peña (1994) and Gonzalo and Granger (1995). This analogy suggests that the use of DFMs is, at least conceptually, not restricted to stationary time series, and that it may prove fruitful when dealing with non-stationary vector time series. However, the most popular way of dealing with large systems of non-stationary macroeconomic time series is by differencing the variables in a univariate fashion. For some recent references, see Breitung and Eickmeier (2011), Stock and Watson (2012a,b), Barhoumi et al. (2013), Costantini (2013), Moench et al. (2013), Bräuning and Koopman (2014), Buch et al. (2014), Poncela et al. (2014), Jungbacker and Koopman (2015) and Lahiri et al. (2015).

The main reason for this extended practice is that the factors are most commonly estimated by Principal Components (PC), a technique which yields a robust asymptotic theory for this case. Bai (2003) develops the inferential theory for large factor models, considering the PC estimator, deriving the rate of convergence and the limiting distributions of the estimated factors, factor loadings and common components. He shows that the factor estimates extracted using PC are asymptotically equivalent to the maximum likelihood estimator. The results are consistent under heteroscedasticity, and under weak serial and cross-sectional correlation in the idiosyncratic errors. This allows us to carry out the corresponding and appropriate inference including confidence intervals, hypothesis tests, etc.

In this dissertation, we study the effects of differencing non-stationary DFMs when determining the number of common factors and estimating the factor space.

Modelling the levels of non-stationary time series through DFM representation, would involve a model in which

$$Y_t = PF_t + \varepsilon_t,$$

$$F_t = \Phi F_{t-1} + \eta_t,$$

$$\varepsilon_t = \Gamma \varepsilon_{t-1} + a_t.$$

where $Y_t = (y_{1t}, \ldots, y_{Nt})'$, $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ and $a_t = (a_{1t}, \ldots, a_{Nt})'$ are $N \times 1$ vectors, $P = (p_1, \ldots, p_N)'$ is an $N \times r$ matrix, $F_t = (F_{1t}, \ldots, F_{rt})'$ and $\eta_t = (\eta_{1t}, \ldots, \eta_{rt})'$, are $r \times 1$ vectors. Finally $\Phi = \text{diag}(\phi_1, \ldots, \phi_r)$ and $\Gamma = \text{diag}(\gamma_1, \ldots, \gamma_r)$ are $r \times r$ and $N \times N$ matrices respectively. In our case, we focus on the case in which the main diagonals of Φ and/or Γ can contain 1's, i.e., when the common factors and/or the idiosyncratic components may be non-stationary. Note that the DFM given by the previous equations is not identified because for any $r \times r$ non-singular matrix H, the series Y_t can be expressed in terms of a new loading matrix and a new set of common factors, $Y_t = P^* F_t^* + \varepsilon_t$, where $P^* = PH$ and $F_t^* = H^{-1} F_t$. Then, the DFM with factor loading matrix P^* and factors F_t^* is observationally equivalent to the original model. To solve this indentification problem, a normalization is necessary to uniquely define the factors. For the restrictions used to solve this identification problem see Bai (2004), Bai and Ng (2013) and Bai and Wang (2014).

It is interesting to observe that differencing a non-stationary system can introduce a trade-off between the variance of the common component, PF_t, and the variance of the idiosyncratic component. This fact can affect the identification of the number of common factors. Furthermore, differencing a cointegrated system may distort the inference and we can lose information by removing the long-run relationship (Sims, 2012), which is crucial to understand the comovements among variables.

In this context, the main objectives of this dissertation are the following:

1. Analyze the effects of stationary univariate transformations when determining the number

of common factors

2. Analyze the finite sample performance of the non-stationary common factors estimation approaches

3. Implement empirical applications for non-stationary systems

For the first objective, we analytically derive the eigenvalues of the covariance matrix, analyzing the effects of the stationary univariate transformations. Furthermore, for the first two objectives, we carry out Monte Carlo experiments considering several designs selected to represent different situations that can potentially be encountered when dealing with the empirical analysis of real macroeconomic variables. We consider different sample sizes and structure dependence on Φ, Γ, Σ_η, Σ_a, where Σ_η and Σ_a are the covariance matrix of the disturbances of the common factors and the idiosyncratic components respectively.

For the third objective, the empirical applications are based on a system of prices of 15 countries from the euro area with the goal of determining the number of common factors and how these results can be affected by the treatment applied (differencing or not) to the data. Furthermore, we study the risk sharing for a dataset from the Organisation for Economic Co-operation and Development (OECD) for industrialized countries. We focus on income and consumption variables. The goal of this exercise is to conclude whether international risk sharing is a short or long-run phenomenon. This is a novel application given that risk sharing is frequently studied as a short-run issue.

1.2. Determining the number of factors after stationary univariate transformations

The appropriate determination of the number of factors, r, is crucial to obtain a consistent estimation of the space spanned by the factors. In the context of large DFMs, there are several alternative procedures designed to determine r. Bai and Ng (2008), Stock and Watson (2011), Breitung and Choi (2013), Bai and Wang (2016), among many others, give a review of procedures for determining the number of factors, which basically consist in visual diagnostics and formal tests based on the behavior of eigenvalues from the sample covariance matrix.

The most popular procedures for determining r are the criteria proposed by Bai and Ng (2002); see Andersen et al. (2015), Everaert et al. (2015) and Panopoulou and Vrontos (2015) for some recent applications. The Bai and Ng (2002) criteria are based on modifications of the Akaike (AIC) and Bayesian (BIC) information criteria, taking into account the cross-sectional and temporal dimensions of the dataset as arguments of the function penalizing overparametrization. Alessi et al. (2010) propose a refinement of these criteria based on multiplying the penalty function by a constant that tunes the penalizing power of the function itself. Furthermore, they suggest estimating the number of factors using different subsamples. These criteria are linked to the eigenvalues of the sample covariance matrix from the variables in the system. In particular, the number of factors is selected as the number of eigenvalues larger than a threshold specified by a penalty function. Alternative criteria based on random matrix theory and the behavior of the eigenvalues of the sample data covariance matrix have been proposed by Kapetanios (2010) and Onatski (2010). In this last paper he proposes an alternative estimator based on using differenced adjacent eigenvalues arranged in descending order. More recently, Ahn and Horenstein (2013) propose two alternative estimators based on ratios of adjacent eigenvalues. Other procedures to detect the number of common factors are given by Harding (2013), who proposes a consistent procedure by imposing restrictions on the time and spatial correlation patterns of the error terms with improved finite sample properties when it is compared with Bai and Ng (2002) and Onatski (2010) in the presence of weak factors. Also Caner and Han (2014) propose a procedure based on a group bridge estimator while Han and Caner (2016) put forward a modification of the penalty function of Bai and Ng (2002) which is data dependent.

These procedures tend to perform very well under the traditional assumptions of PC factor extraction such as: pervasive factor loadings, strong common factors, weak dependences in the idiosyncratic errors, large sample sizes, etc. Consequently, the finite sample performance depends on the data generating process (DGP). Differentiating a non-stationary DFM may affect the original relation between the variances of the common component and the idiosyncratic terms. When differencing in a univariate fashion and increase the variance of the idiosyncratic component with respect to the variance of the common component, distortions may be introduced in the determination of the number of factors.

Focusing on non-stationary DFMs, there are some studies that analyze the determination of the number of non-stationary common factors. Bai (2004) proposes a new criteria for data in levels and studies the performance of Bai and Ng (2002) information criteria for first differenced data. The author implements a Monte Carlo analysis with contemporaneously uncorrelated idiosyncratic noises following an ARMA model and two and four random walk factors. The study concludes that the static common factors are correctly determined for both criteria. Furthermore, Bai and Ng (2004) disentangle the non-stationarity of the common factors, once determined by Bai and Ng (2002) information criteria allowing non-stationary idiosyncratic errors. However, as far as we know, there is not a complete study over whether differencing in a univariate fashion affects the correct determination of the number of factors.

In this dissertation, we contribute by studying the effects of univariate stationary transformations of non-stationary systems when determining the number of factors using the procedures proposed by Bai and Ng (2002); Onatski (2010) and Ahn and Horenstein (2013), which are frequently used in empirical applications. We focus on the case known in the literature as static factors given that, the statistical properties are very well known and this approach is the most popular in empirical applications. When $r = 1$ and the idiosyncratic noises are mutually uncorrelated and homoscedastic, we analytically derive the eigenvalues of the covariance matrix examining how the procedures are affected by univariate differentiation. We carry out Monte Carlo experiments considering several designs selected to represent different situations that can be potentially encountered when dealing with the empirical analysis of real macroeconomic variables. Finally, we apply the methodology implemented in the Monte Carlo analysis for a system of prices from the euro area.

We conclude that when differencing in a univariate fashion a cointegrated DFM, a trade-off may be introduced between the variance of the idiosyncratic component with respect to the variance of the common component, affecting the determination of the number of factors. Furthermore, it is important to consider the variance and the dependence structure of the differenced idiosyncratic noises. When $r = 1$, the ratios of adjacent eigenvalues of Ahn and Horenstein (2013) tend to perform well under several specifications. However, the performance of all procedures tend to decrease when $r = 2$. In this case, Onatski (2010) performs better. The Bai and Ng (2002) information criteria only perform well under traditional assumptions of PC factor extraction. In the empirical application, we determine between 1 and 3 common factors. The first common factor is the common inflation in the euro area while the second and third common factors can be attributed to Ireland and Greece respectively.

1.3. Estimating non-stationary common factors: Implications for risk sharing

In the context of large DFMs, the estimation of their components is mainly determined by applying PC. This fact can be attributed to the fact that PC extraction allows a consistent estimation of the factor space without assuming any particular error distribution or specifications of the factors and idiosyncratic noises. The traditional assumptions are that the variability of common component are not small and the idiosyncratic component has weak serial and cross-correlation. Furthermore, the latter allows for heteroscedasticity. PC factor extraction separates the common component from the idiosyncratic noises through cross-sectional averaging of the observations. In large sample sizes, only the effects of the common component are pervasive over the observations, such that the weighted averages of the idiosyncratic terms converge to zero. Under stationarity, Bai (2003) derives the rate of convergence and the limiting distributions of the estimated factors, factor loadings and the common component when N and T tend to infinity.

Under non-stationarity in large DMFs, Bai (2004) proposes a PC factor extraction procedure using data in levels, deriving the rates of convergence and the limiting distributions for the estimated common trends, the estimated loading weights and the common component. To obtain consistent estimations, we require stationarity in idiosyncratic noises. Furthermore, Bai and Ng (2004) put forward PC to first differenced data, using the "differencing and recumulating" method to obtain a consistent estimation for the non-stationary common factors. Their results are consistent even if the idiosyncratic components are $I(1)$. Additionally, Choi (2016) extends the Generalized PC Estimator (GPCE) under similar assumptions from Bai (2004). In his work, the author derives the corresponding asymptotic theory, showing that the GPCE is more efficient than the PC estimator. Also, Barigozzi et al. (2016) propose to project the original observations in the factor loadings estimated using first-differenced data. Furthermore, they study the cointegrated DFM and develop the asymptotic theory for non-stationary DFMs, proving that its estimator is consistent even if the idiosyncratic errors are non-stationary.

In this dissertation, we contribute in analyzing the finite sample performance of these procedures under several specifications in the DGP, which summarize situations presented in empirical applications. We consider different sample sizes, serial and cross-sectional correlations, and heteroscedasticity in idiosyncratic noises. Moreover, we take into account different variance sizes in the disturbances of the idiosyncratic term and systems with one and two non-stationary factors. We also consider the situation where one factor is stationary and the other is non-stationary.

As it is well known in stationary DFMs, Doz et al. (2011, 2012) prove that incorporating the factor and idiosyncratic dynamics in the model by combining PC and Kalman Smoothing (2SKS), we obtain consistent estimates of the common factors. Additionally, we achieve an outperformance in the precision of the factor estimates with respect to PC factor extraction, see the Monte Carlo analysis given by Poncela and Ruiz (2016). In this dissertation we extend the 2SKS procedure by allowing non-stationary common factors. Furthermore, we apply this approach to first-differenced data, using the "differencing and recumulating" estimator. For these two procedures, we also analyze the finite sample performance.

Another contribution is the empirical application. As far as we know, risk sharing has been analyzed as a short-run issue. In this study, we apply the procedures to extract non-stationary common factors and disentangling whether risk sharing is a short or long-run issue.

We conclude that the procedures that extract non-stationary common factors using data in levels perform better when the idiosyncratic noises are stationary, even when this component has large variance and/or the serial correlation is drastically negative. In any case, these last procedures are more robust to the presence of heterocedasticity, cross-sectional correlation and different sample sizes with respect to the approaches that use first differenced data. On the other hand, when the idiosyncratic component is non-stationary, the procedures which estimate the non-stationary common factors using first differenced data, perform better. Regardless of the approach, when the system has one non-stationary common factor and the other is stationary, the first common factor is better extracted. In the empirical application we obtain that at least four common factors are non-stationary. Applying Panel Analysis of Non-stationarity in Idiosyncratic and Common components (PANIC) to the idiosyncratic component extracted by PC to first-differenced data, we conclude that the idiosyncratic noises are $I(1)$. The non-stationary factor model point outs the lack of risk sharing both in the short and long-run.

1.4. Organization

The rest of this dissertation is organized as follows. In Chapter 2 we analyze the performance of three of the most popular criteria to determine the number of factors under the univariate stationary transformations frequently implemented in empirical applications. Second, we analytically derive the eigenvalues of the covariance matrix to investigate the effects of transforming non-stationary systems by univariate stationary transformations on these procedures. Third, we carry out the Monte Carlo analysis to study the finite sample performance of the approaches when determining the number of factors under several DGP. Fourth, we implement an empirical application to a system of prices from euro area countries. Finally, we summarize the main results and conclusions. In Chapter 3 we describe the procedures to extract non-stationary common factors in large DFMs and extend the hybrid procedure based on PC and Kalman Smoothing to non-stationary systems. The finite sample performance of these procedures is studied in the Monte Carlo analysis and we carry out an empirical application to determine whether international risk sharing is a short or long-run phenomenon. Finally, in Chapter 4 we conclude and provide directions for future research.

Chapter 2

Determining the number of factors after stationary univariate transformations

2.1. Introduction

In recent years, due to the availability of data on a vast number of macroeconomic and financial variables, there has been an increasing interest in modeling large systems of economic time series. In order to reduce the dimensionality and extract the underlying factors, one can use Dynamic Factor Models (DFMs), originally introduced in economics by Geweke (1977) and Sargent and Sims (1977). The aim of DFMs is to represent the dynamics of the system through a small number of hidden common factors which are mainly used for forecasting and macroeconomic policy-making; see Stock and Watson (2011) and Breitung and Choi (2013), for recent reviews of the existing literature. Kajal Lahiri has contributed to the DFMs literature with several empirical works. For example, Lahiri and Yao (2004) implement a DFM to analyze the business cycle features of the transportation sector and Lahiri and Sheng (2010) to measure the forecast uncertainty by disagreement. Lahiri et al. (2015) also implement a DFM to a real-time jagged-edge data set of over 160 explanatory variables to re-examine the role of consumer confidence surveys in forecasting personal consumption expenditure. The properties of many popular factor extraction procedures rely on the number of factors in the system being known. However, in practice, the number of factors is unknown and needs to be determined. Among the most popular procedures proposed with this purpose are the criteria proposed by Bai and Ng (2002), which are now standard in the literature. These criteria are based on modifications of the Akaike (AIC) and Bayesian (BIC) information criteria taking into account the cross-sectional and temporal dimensions of the dataset as arguments of the function penalizing overparametrization. Alternatively, Onatski (2010) proposes an estimator of the number of factors based on using differences between adjacent eigenvalues of the sample covariance matrix of the variables contained in the system, arranged in descending order while Ahn and Horenstein (2013) propose two alternative estimators based on ratios of adjacent eigenvalues.

It is well known that macroeconomic time series are frequently non-stationary and possibly cointegrated. Within the context of Principal Components (PC) factor extraction, and following Stock and Watson (2002a), the most popular way of dealing with large systems of non-stationary macroeconomic variables is by differencing the variables in a univariate fashion; see, for example, Breitung and Eickmeier (2011), Stock and Watson (2012a,b), Barhoumi et al. (2013), Moench et al. (2013), Bräuning and Koopman (2014), Buch et al. (2014), Poncela et al. (2014) and Jungbacker and Koopman (2015) for recent references. The theoretical justification of this extended practice is analyzed in Bai and Ng (2004) who show that applying PC to first-differenced data and recovering the original factors by "recumulating" is consistent regardless of whether the factors and/or idiosyncratic errors are $I(0)$ or $I(1)$.[1] However, their theory proceeds assuming that the number of common factors in the system is known. On the other hand, as mentioned above, macroeconomic variables are not only non-stationary but can also be cointegrated. Differencing a cointegrated system may distort the determination of the number of factors due to the introduction of non-invertible moving average (MA) components and/or the trade-off introduced between the variances of the common and idiosyncratic components. Surprisingly, there has been little discussion in the literature on whether differencing in a univariate fashion

[1] Bai (2004) also has asymptotic results for the factors estimated from the original non-stationary data.

affects the correct determination of the number of factors. As far as we know, only Bai (2004) analyzes the performance of the information criteria proposed by Bai and Ng (2002) when implemented to differenced data. In his Monte Carlo experiments, carried out for a unique DFM with contemporaneously uncorrelated idiosyncratic noises following an ARMA model and two random walk factors, he shows that the number of factors is correctly determined.

The main objective of this paper is to fill this gap by analyzing the effects of univariate stationary transformations of cointegrated systems when determining the number of factors using the approaches proposed by Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013). In the context of a DFM with mutually uncorrelated and homoscedastic idiosyncratic noises, we first derive analytically the eigenvalues of the covariance matrix and show how they are affected by univariate differentiation. We also carry out Monte Carlo experiments considering several designs selected to represent different situations that can be potentially encountered when dealing with the empirical analysis of real macroeconomic variables. Finally, we illustrate the results determining the number of factors in a system of prices of the euro area. It is important to note that the procedures for determining the number of factors considered in this paper are designed for what is known in the literature as static factors. Alternatively, several factor determination procedures have been proposed in the context of dynamic factors; see, for example Amengual and Watson (2007), Bai and Ng (2007), Hallin and Liska (2007), Jacobs and Otter (2008) and Breitung and Pigorsch (2013). The difference between static and dynamic factors is described by, for example, Bai and Ng (2008). They argue that, although dynamic factors can be useful to establishing the number of primitive shocks in the economy, the properties of estimated static factors are better understood from a theoretical point of view. Furthermore, we focus the analysis on procedures to detect the number of static factors as they are more popular in empirical economics.

The rest of this paper is structured as follows. In section 2.2, we briefly describe the stationary DFM and the factor determination approaches considered. In section 2.3, we analyze the effects of transforming non-stationary systems by univariate stationary transformations on these procedures. In section 2.4, we report the results of the Monte Carlo experiments carried out to illustrate their finite sample performance. In section 2.5, we carry out an empirical application. Finally, we conclude in section 2.6.

2.2. The stationary Dynamic Factor Model

In this section, we introduce notation and the stationary DFM and describe the factor determination procedures considered.

2.2.1. The model

We consider a DFM with cross-sectional dimension N, where the unobserved $r < N$ common factors, $F_t = (F_{1t}, \dots, F_{rt})'$, and the idiosyncratic noises, $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$, follow VAR(1) processes. The factors explain the common evolution of a vector of time series, $Y_t = (y_{1t}, \dots, y_{Nt})'$ observed from $t = 1, \dots, T$. The basic DFM considered is given by

$$Y_t = PF_t + \varepsilon_t, \tag{2.1}$$

$$F_t = \Phi F_{t-1} + \eta_t, \tag{2.2}$$

$$\varepsilon_t = \Gamma \varepsilon_{t-1} + a_t, \tag{2.3}$$

where the factor disturbances, $\eta_t = (\eta_{1t}, \dots, \eta_{rt})'$, are $r \times 1$ vectors, distributed independently from the idiosyncratic noises for all leads and lags. Furthermore, η_t and a_t, are Gaussian white noises with positive definite covariance matrices Σ_η and Σ_a, respectively, and $P = (p_1^{'}, \dots, p_N^{'})'$, is the $N \times r$ matrix of factor loadings, where, $p_i = (p_{i1}, \dots, p_{ir})$. Finally, $\Phi = \text{diag}(\phi_1, \dots, \phi_r)$ and Γ are $r \times r$ and $N \times N$ matrices containing the autoregressive parameters of the factors and the idiosyncratic components, respectively. These autoregressive matrices satisfy the usual stationarity assumptions. Furthermore, we assume that the structure of the idiosyncratic noises is such that they are weakly correlated. Following Bai and Ng (2002), Onatski (2012, 2015) and

Ahn and Horenstein (2013), we consider the entries in P, Φ, Σ_η, Γ and Σ_a as fixed parameters. Jungbacker and Koopman (2015) and Alvarez et al. (2016) implement the DFM in equations (2.1) to (2.3) to the data set of Stock and Watson (2005).

The DFM in equations (2.1) to (2.3) is not identified because, for any $r \times r$ nonsingular matrix H, the system can be expressed in terms of a new loading matrix and a new set of common factors. A normalization is necessary to solve this identification problem and uniquely define the factors. In the context of PC factor extraction, it is common to impose the restriction $P^{'}P/N = I_r$ and FF' being diagonal, where $F = (F_1, \ldots, F_T)$ is a $r \times T$ matrix of common factors; see Stock and Watson (2002a), Bai and Ng (2002, 2008, 2013), Connor and Korajczyk (2010) and Bai and Wang (2014) for papers dealing with identification issues. Note that these are normalization restrictions, and they may not have an economic interpretation.

2.2.2. Determining the number of factors

The DFM described above assumes that the number of factors, r, is known. However, in practice, it needs to be estimated. Obtaining the correct value of r is crucial for an adequate estimation of the space spanned by the factors. There are several alternative procedures designed to determine r in DFMs. In this paper, we consider the information criteria proposed by Bai and Ng (2002) and the estimators proposed by Onatski (2010) and Ahn and Horenstein (2013).[2]

The Bai and Ng (2002) information criteria

The most popular information criteria to select the number of factors in DFMs, proposed by Bai and Ng (2002), are based on a consistent PC estimator of P and F_t which is given by the solution to the following least squares problem

$$\min_{F_1,\ldots,F_T,P} V_r(P,F) \tag{2.4}$$

[2] Alternatively, based on the estimator proposed by Hallin and Liska (2007), Alessi et al. (2010) propose a refinement of Bai and Ng (2002) criteria based on multiplying the penalty function by a constant that tunes the penalizing power of the function itself and estimating the number of factors using different subsamples. Also, Kapetanios (2010) proposes

subject to $P^{'}P/N = I_r$ and $FF^{'}$ being diagonal, where

$$V_r(P,F) = \frac{1}{NT}\sum_{t=1}^{T}(Y_t - PF_t)^{'}(Y_t - PF_t) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{it}^2 = \frac{1}{NT}tr(\varepsilon\varepsilon'), \tag{2.5}$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)$ has dimension $N \times T$. The solution to (2.4) is obtained by setting $\hat{P}$ equal to $\sqrt{N}$ times the eigenvectors corresponding to the r largest eigenvalues of $YY^{'}$ where $Y = (Y_1, \ldots, Y_T)$. The corresponding PC estimator of F is given by $\hat{F} = N^{-1}\hat{P}^{'}Y$.

PC factor extraction separates the common component, PF_t, from the idiosyncratic noises by averaging cross-sectionally the variables within Y_t such that when N and T tend simultaneously to infinity, the weighted averages of the idiosyncratic noises converge to zero, remaining only the linear combinations of the factors. Therefore, it requires that the cumulative effects of the common component increase proportionally with N, while the eigenvalues of $\Sigma_\varepsilon = E(\varepsilon_t\varepsilon_t^{'})$ remain bounded; see the review of Breitung and Choi (2013) for a description of these conditions.[3] Bai (2003) proves that the PC estimators of factors, factor loadings and common components are asymptotically equivalent to the maximum likelihood estimators and, consequently, consistent. Also, he derives the rate of convergence and their corresponding limiting distributions when N and T tend simultaneously to infinity.

In order to determine r, Bai and Ng (2002) propose minimizing the following functions with

determining the number of factors using resampling to choose the normalizing constants to be used in order to have an asymptotic distribution for the eigenvalues of the sample covariance matrix of Y. Given that these procedures are very intensive computationally, we do not consider them further in this paper. Recently, Harding (2013) proposes a consistent procedure with improved finite sample properties when compared with Bai and Ng (2002) and Onatski (2010) in the presence of weak factors. Also Caner and Han (2014) propose a procedure based on a group bridge estimator while Han and Caner (2016) propose a modification of the penalty function of Bai and Ng (2002) which is data dependent.

[3] Onatski (2012) considers a DFM in which the explanatory power of the factors does not strongly dominate the explanatory power of the idiosyncratic noises.

respect to k, for $k = 0, \ldots, r_{\max}$,

$$IC_1(k) = \ln V_k(\hat{P}, \hat{F}) + k\frac{N+T}{NT}\ln\frac{NT}{N+T}, \tag{2.6a}$$

$$IC_2(k) = \ln V_k(\hat{P}, \hat{F}) + k\frac{N+T}{NT}\ln m, \tag{2.6b}$$

$$IC_3(k) = \ln V_k(\hat{P}, \hat{F}) + k\frac{\ln m}{m}, \tag{2.6c}$$

where $V_k(\hat{P}, \hat{F})$ is defined as in expression (2.5) with P and F_t substituted by their respective PC estimates, $m = \min\langle N, T\rangle$ and $r_{\max}$ is a bounded integer such that $r \le r_{\max}$. The criteria in (2.6) are quite sensitive to the choice of $r_{\max}$; see the Monte Carlo results in Ahn and Horenstein (2013). Bai and Ng (2002) use $r_{\max} = 8$ in their Monte Carlo experiments. On the other hand, in the context of first-differenced data, Bai and Ng (2004) use $IC_1(k)$, with $r_{\max} = 6$. Under appropriate assumptions, Bai and Ng (2002) prove the consistency of the information criteria above to determine the number of common factors.

If $\hat{\varepsilon}_t = Y_t - \hat{P}\hat{F}_t$ are the residuals of the regression of the variables in Y on the r first principal components of $\frac{1}{NT}YY'$, then $tr(\hat{\varepsilon}\hat{\varepsilon}') = tr(YY') - tr(\hat{P}\hat{F}\hat{F}'\hat{P}') = T\sum_{i=1}^{m}\hat{\lambda}_i - T\sum_{i=1}^{r}\hat{\lambda}_i = T\sum_{i=r+1}^{m}\hat{\lambda}_i$, where $\hat{\lambda}_i$, $i = 1, \ldots, m$ are the eigenvalues of $\hat{\Sigma}_Y = \frac{1}{T}YY'$, arranged in descending order. Therefore,

$$V_r(\hat{P}, \hat{F}) = \frac{1}{N}\sum_{i=r+1}^{m}\hat{\lambda}_i. \tag{2.7}$$

Using the expression of $V_k(\hat{P}, \hat{F})$ in (2.7), the functions in (2.6) can be written as

$$IC_j(k) = \ln(\frac{1}{N}\sum_{i=k+1}^{m}\hat{\lambda}_i) + kg_j(N, T), \tag{2.8}$$

where $g_j(N, T)$ is defined accordingly to the criteria in (2.6) for $j = 1, 2$ and 3.

Differenced eigenvalues

Onatski (2010) proposes an alternative procedure to select r, called edge distribution (ED), and shows that it outperforms the criteria proposed by Bai and Ng (2002) when the proportion of the variance attributed to the factors is small relative to the variance due to the idiosyncratic noises or when these are substantially correlated. Furthermore, computationally, the procedure proposed by Onatski (2010) allows the determination of the number of factors without previous estimation of the common component. Finally, it relaxes the standard assumption of PC factor extraction about the r eigenvalues of $\hat{\Sigma}_Y$ growing proportionally to N. Instead of requiring that the cumulative effect of factors grow as fast as N, Onatski (2010) imposes a structure on the idiosyncratic noises. Under the assumption of Normality, both cross-sectional and temporal dependence are allowed. This procedure is based on determining a sharp threshold, δ, which consistently separates the bounded and diverging eigenvalues of $\hat{\Sigma}_Y$. For any $j > r$, the differences $\hat{\lambda}_j - \hat{\lambda}_{j+1}$ converge to 0 while the difference $\hat{\lambda}_r - \hat{\lambda}_{r+1}$ diverges to infinity when both N and T tend to infinity. Assuming that $r_{\max}/N \to 0$, Onatski (2010) proposes the following algorithm in order to calibrate δ and determine the number of factors:

1. Obtain $\hat{\lambda}_i$, $i = 1, ..., N$ and set $j = r_{\max} + 1$.

2. Obtain $\hat{\beta}$ as the ordinary least squares (OLS) estimator of the slope of a simple linear regression with constant, where the observations of the dependent variable are $\left\{\hat{\lambda}_j, \ldots, \hat{\lambda}_{j+4}\right\}$ and the observations of the regressor variable are $\{(j-1)^{2/3}, \ldots (j+3)^{2/3}\}$. Set $\hat{\delta}= 2|\hat{\beta}|$.

3. Estimate $\hat{r} = \max\{k \leq r_{\max} | \hat{\lambda}_k - \hat{\lambda}_{k+1} \geq \hat{\delta}\}$ or $\hat{r} = 0$ if $\hat{\lambda}_k - \hat{\lambda}_{k+1} < \hat{\delta}$.

4. Set $j = \hat{r} + 1$. Repeat steps 2 and 3 until $\hat{r}$ converges.

Under suitable conditions, Onatski (2010) proves the consistency of $\hat{r}$ for any fixed $\delta > 0$. He sets the number of iterations to four although the convergence of the above algorithm is often achieved at the second iteration. Additionally, he sets $r_{\max} = 8$ when $r = 1, 2, 5$ and $r_{\max} = 20$ when $r = 15$.

Ratios of eigenvalues

Recently, Ahn and Horenstein (2013) propose two further estimators of the number of factors based on the fact that the r largest eigenvalues of $\hat{\Sigma}_Y$ grow unbounded as N increases, while the other eigenvalues remain bounded. They show that these estimators are less sensitive to the choice of $r_{\max}$ than those based on the Bai and Ng (2002) information criteria. The two new estimators are defined as the value of k, for $k = 0, \ldots, r_{\max}$, that maximizes the following ratios

$$ER(k) = \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}}, \tag{2.9}$$

$$GR(k) = \frac{\ln\left[V_{k-1}(\hat{P}, \hat{F})/V_k(\hat{P}, \hat{F})\right]}{\ln\left[V_k(\hat{P}, \hat{F})/V_{k+1}(\hat{P}, \hat{F})\right]} = \frac{\ln(1+\hat{\lambda}_k^*)}{\ln(1+\hat{\lambda}_{k+1}^*)}, \tag{2.10}$$

where $\hat{\lambda}_0 = \frac{1}{m}\sum_{k=1}^{m}\hat{\lambda}_k/\ln(m)$ and $\hat{\lambda}_k^* = \hat{\lambda}_k/\sum_{j=k+1}^{m}\hat{\lambda}_j$. The value of $\hat{\lambda}_0$ has been chosen following the definition of Ahn and Horenstein (2013) according to which $\hat{\lambda}_0 \to 0$ and $m\hat{\lambda}_0 \to \infty$ as $m \to \infty$.[4]

Note that both the numerator and denominator of $GR(k)$ are the growth rates of sums of residual variances computed with j and $j+1$ factors. Ahn and Horenstein (2013) show that, contrary to the estimator proposed by Bai and Ng (2002), their estimators are not dependent on $r_{\max}$ and suggest to chose it as $\min(r_{\max}^*, 0.1m)$ where $r_{\max}^* = \#\left\{k \mid N^{-1}\hat{\lambda}_k \geq V_0/m, k \geq 1\right\}$. Under the same assumptions of Bai and Ng (2006) and Onatski (2010), and allowing for some variables in Y to be perfectly multicollinear or with zero idiosyncratic variances, they establish consistency of the $ER(k)$ and $GR(k)$ estimators. The results obtained in their Monte Carlo analysis show that the two estimators outperform the Bai and Ng (2002) information criteria and Onatski (2010) estimator mainly when the idiosyncratic components are simultaneously cross-sectionally and serially correlated. However, the estimator proposed by Onatski (2010)

[4]Ideas similar to the ER estimator have also been considered by Luo et al. (2009) and Wang (2012). Furthermore, Lam and Yao (2012) study the properties of ratio of eigenvalues of $M = \sum_{i=1}^{i_0}\hat{\Sigma}_{Y(i)}\hat{\Sigma}_{Y(i)}'$ where $\hat{\Sigma}_{Y(i)}$ is the sample covariance matrix of Y_{t-i}. The estimator of the number of factors is given by $\hat{r} = \min\{k \leq r_{\max} \parallel \hat{\mu}_{k+1}/\hat{\mu}_k\}$, where $\hat{\mu}_k$ is the k-th eigenvalue of M. They show that its estimator performs well even if when the factors are weak but N is large.

outperforms the $ER(k)$ and $GR(k)$ ratios when the variance of the idiosyncratic component is larger than that of the common component (weak factors).

2.2.3. A note on the convergence of eigenvalues

The procedures to determine the number of common factors described above are based on the eigenvalues of the sample covariance matrix, $\hat{\Sigma}_Y$. One of the main contributions of Bai and Ng (2002) is to show that the convergence of the eigenvalues of $\frac{1}{TN}YY'$ depends on m. Later, Kapetanios (2010) reviews the available literature about the topic pointing out that the distribution of the largest eigenvalue depends in complicated ways on the parameters of the model. It seems that serial correlation affects both the parameters of the asymptotic limits and their functional form. Furthermore, he shows that the first r eigenvalues of Σ_Y increase at rate N which follows from the fact that the r largest eigenvalues of $F'F$ will grow at rate N as long as the loading matrix P is not sparse and suggests that it is reasonable to expect a similar behavior from the eigenvalues of the sample covariance matrix.

More recently, Onatski (2012, 2015) develops new asymptotics for the eigenvalues of the sample covariance matrix by considering that both the weights and the factors are fixed parameters.

2.3. Determining the number of factors after differencing

As mentioned in the Introduction, macroeconomic systems are often non-stationary. In this section, we analyze the effects on the performance of the number of factors determination procedures described above of transforming the data in a univariate fashion in order to achieve stationarity. Note that differencing affects the ratio between the variances of the factors and idiosyncratic components, the temporal dependence structure and the cross-correlations among the idiosyncratic noises.

Consider the DFM given in equations (2.1) to (2.3) in which Φ and Γ are diagonal matrices which may have 1's in the main diagonal. Consequently, both the factors and the idiosyncratic noises can be either stationary or non-stationary random walks. Under this specification, the system of first-differenced data satisfies all conditions of Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013). After differencing the data in a univariate fashion, the DFM takes the following form

$$\Delta Y_t = P\Delta F_t + \Delta\varepsilon_t, \tag{2.11}$$

$$\Delta F_t = (\Phi - I)F_{t-1} + \eta_t, \tag{2.12}$$

$$\Delta\varepsilon_t = (\Gamma - I)\varepsilon_{t-1} + a_t. \tag{2.13}$$

Denote by ϕ_i the i-th element in the main diagonal of Φ. If $|\phi_i| < 1$, then the variance of the corresponding differenced factor is given by $\sigma^2_{f_i} = 2\sigma^2_{\eta_i}/(1+\phi_i)$ where $\sigma^2_{\eta_i}$ is the variance of η_i. When $\phi_i = 0.5$, the difference between the variances of F_t and ΔF_t is zero. Therefore, in this case, the variance of the factor is not changed after differencing the data. However, if $\phi_i < 0.5$, the variance of ΔF_t is larger than that of F_t while if $\phi_i > 0.5$, it is smaller. The same relation can be established for the variances of the elements in ε_t and $\Delta\varepsilon_t$ with respect to γ_i, the i-th element in the main diagonal of Γ. Note that if ε_t is stationary, with autoregressive parameters smaller than 0.5 while F_t is non-stationary, then overdifferencing the idiosyncratic components may introduce distortions on the determination of the number of factors given that the relation between the variances of the common and idiosyncratic components is modified with the variances of ΔF_t being smaller and the variances of $\Delta\varepsilon_t$ being larger. The dynamic dependence of the idiosyncratic noises of the differenced model are given by

$$Corr(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-h}) = 0.5\gamma_i^{h-1}(\gamma_i - 1).$$

Finally, note that differencing also affects the cross-correlations of the idiosyncratic noises. Consider, for example, that the correlation between ε_{it} and ε_{jt} is given by ρ. If the idiosyncratic noises are stationary, then

$$Corr(\Delta\varepsilon_{it}, \Delta\varepsilon_{jt}) = \sigma_{\Delta\varepsilon_i}^{-1}\sigma_{\Delta\varepsilon_j}^{-1}\left(2-\gamma_i-\gamma_j\right)\rho\sigma_{\varepsilon_i}\sigma_{\varepsilon_j} = \frac{0.5\left(2-\gamma_i-\gamma_j\right)\rho}{\sqrt{(1-\gamma_i)(1-\gamma_j)}}.$$

In order to simplify the analysis of the effects of univariate differentiation on the determination of r, we consider $\Gamma = \gamma I$ and $\Sigma_a = \sigma_a^2 I$, so that the idiosyncratic noises are homoscedastic and mutually uncorrelated and all of them are governed by the same autoregressive parameter. Given that there is no correlation between the factors and the idiosyncratic components, the covariance matrix of the first-differenced data is given by $\Sigma_{\Delta Y} = P\Sigma_f P' + \sigma_e^2 I$, where Σ_f is the covariance matrix of ΔF_t and $\sigma_e^2 = 2\sigma_a^2/(1+\gamma)$ is the variance of each element in $\Delta\varepsilon_t$. The ordered eigenvalues of $\Sigma_{\Delta Y}$ are equal to $\sigma_e^2 + \mu_i$ for $i = 1, \ldots, N$, where μ_i is the i-th largest eigenvalue of $P\Sigma_f P'$. Furthermore, $tr\left(P\Sigma_f P'\right) = tr\left(P' P\Sigma_f\right) = \sum_{j=1}^{r}\sigma_{f_j}^2\sum_{i=1}^{N}p_{ij}^2 = \sum_{j=1}^{r}\mu_j$. Therefore, the sum of the r largest eigenvalues of $\Sigma_{\Delta Y}$ is given by $\sum_{i=1}^{r}\lambda_i = r\sigma_e^2 + \sum_{j=1}^{r}\sigma_{f_j}^2\sum_{i=1}^{N}p_{ij}^2$, while the rest $N-r$ eigenvalues are given by $\lambda_i = \sigma_e^2$.

Consider the particular case of a unique random walk factor, i.e. $r = 1$ and $\phi_1 = 1$. In this case, $\lambda_1 = \sigma_\eta^2\sum_{i=1}^{N}p_{i1}^2 + \sigma_e^2$ and $\lambda_i = \sigma_e^2$, for $i = 2, \ldots, N$. Consequently, the function to be minimized according to the Bai and Ng (2002) information criteria, is given by

$$IC(k) = \begin{cases} \ln\left(N^{-1}\sigma_\eta^2\sum_{i=1}^{N}p_{i1}^2 + \sigma_e^2\right), & k = 0 \\ \ln(N-k) - \ln(N) + \ln(\sigma_e^2) + kg(N,T), & k \geq 1. \end{cases}$$

The procedure proposed by Onatski (2010) is based on the differences between adjacent eigenvalues. Note that for $j = 2, \ldots, N$, $\lambda_j - \lambda_{j+1} = 0$. Therefore, the procedure should work as far as the difference between λ_1 and λ_2 is large. This difference is given by $\lambda_1 - \lambda_2 = \sigma_\eta^2\sum_{i=1}^{N}p_{i1}^2$ and does not depend on the value of σ_e^2. Therefore, for given weights and cross-sectional dimension, the procedure should work better when σ_η^2 is large. Also, for a given value of σ_η^2, the procedure should work better as N increases. Note that in the first step of the algorithm proposed by Onatski (2010), $\hat{\delta} = 0$ because for $j = r_{\max} + 1$ eigenvalues λ_j are always σ_e^2.

Consider the $ER(k)$ criterion of Ahn and Horenstein (2013) given in (2.9) which looks for a large difference between the ratio of λ_1 and λ_2 with respect to the ratios between other adjacent eigenvalues. Note that, in the particular case we are considering, if $N < T$, the mock eigenvalue is given by $\lambda_0 = \ln(N)^{-1}\left(\sigma_e^2 + N^{-1}\sigma_\eta^2 \sum_{i=1}^N p_{i1}^2\right)$, and, consequently,

$$ER(k) = \begin{cases} \frac{1+N^{-1}q\sum_{i=1}^N p_{i1}^2}{\ln(N)\left(1+q\sum_{i=1}^N p_{i1}^2\right)}, & k=0 \\ 1+q\sum_{i=1}^N p_{i1}^2, & k=1 \\ 1, & k \geq 2, \end{cases}$$

where $q = \frac{\sigma_\eta^2(1+\gamma)}{2\sigma_a^2}$. Note that if N is large enough, $ER(0)$ should be close to 0. Therefore, for given weights, the criteria should work better when q is larger.

Finally, consider the $GR(k)$ criterion of Ahn and Horenstein (2013). In this case, note that

$$\lambda_i^* = \begin{cases} (N\ln(N))^{-1}, & i=0 \\ (N-1)^{-1}(q\sum_{i=1}^N p_{i1}^2 + 1), & i=1 \\ (N-i)^{-1}, & i \geq 2. \end{cases}$$

Therefore,

$$\frac{\ln(1+\lambda_k^*)}{\ln(1+\lambda_{k+1}^*)} = \begin{cases} \frac{\ln(N\ln N+1)-\ln(N\ln N)}{\ln(N+\sum_{i=1}^N p_{i1}^2)-\ln(N-1)}, & k=0 \\ \frac{\ln(N+\sum_{i=1}^N p_{i1}^2)-\ln(N-1)}{\ln(N-1)-\ln(N-2)}, & k=1 \\ \frac{\ln(N+1-k)-\ln(N-k)}{\ln(N-k)-\ln(N-k-1)}, & k \geq 2. \end{cases}$$

2.4. Finite sample performance

The results in the previous section are based on population covariance matrices and their corresponding eigenvalues. However, in practice, when determining the number of common factors in empirical applications, one should estimate the covariance matrix by its sample version and obtain the corresponding estimated eigenvalues. As mentioned above, the asymptotic distribution of estimated eigenvalues is complicated and not always known. The finite sample properties of the estimated eigenvalues depend on the temporal sample size used for their estimation, T, the cross-sectional dimension, N, the ratio between the variances of the common and idiosyncratic components and the structure of the temporal and cross-sectional dependencies of the idiosyncratic noises. In this section, we carry out Monte Carlo experiments in order to analyze how the determination of the number of factors is affected by univariate differentiation of non-stationary data when implemented in finite samples. We should note that the procedures considered have been developed for N and T going to infinity. However, when the procedures are implemented in practice, both N and T are finite. Our interest in this paper is to study the performance of the criteria under different combinations of N and T similar to those often encountered when dealing with systems of macroeconomic and financial variables. Furthermore, we want to investigate how small N and T can be for the procedures to be reliable under different structures of the factors and idiosyncratic noises. In this way, our results can be of interest for practitioners in empirical applications.

The experiments are based on $R = 500$ replications generated by the DFM in equations (2.1) to (2.3) with $N = (12, 50, 100, 200)$ and $T = (100, 500)$.[5] Our simulations are categorized into two parts. The first part is designed to investigate how the alternative estimators considered behave when detecting a unique random walk factor under different temporal and cross-sectional structures of the idiosyncratic noises. The second part is designed to analyze models with more than one factor.

Consider first a DFM defined as in equations (2.1) to (2.3) with $r = 1$, $\Phi = 1$ and $\sigma^2_\eta = 1$. The factor loadings are generated by $p_{i1} \sim U[0, 1]$ with $\sum_{i=1}^{N} p^2_{i1} = 5.59$, 18.70, 34.63 and 65.56 for $N = 12$, 50, 100 and 200, respectively; Bai and Ng (2006) and Poncela and Ruiz (2016) also generate the factor loadings by the same distribution. We consider several structures for the idiosyncratic noises. First, the idiosyncratic noises are mutually uncorrelated and homoscedastic. In particular, the autoregressive coefficient matrix of the idiosyncratic components is diagonal, $\Gamma = \gamma I$, with $\gamma = (-0.8, 1)$ and $\Sigma_a = \sigma^2_a I$ with $\sigma^2_a = 1$ so that $\sigma^2_e = 10$ and 1 for the values of

[5] The time dimension of the multivariate system is generated with $T^* = T + 100$ observations. The factor extraction

γ considered. Note that, differently from simulations carried out in related works, we consider both positive and negative values for the autoregressive parameter of the idiosyncratic noises; see, Pinheiro et al. (2013) who estimate correlations for $\Delta\varepsilon_t$ between -0.6 and 0.9 when dealing with the U.S. monthly macroeconomic data set of Stock and Watson (2005). In order to separate the effects of the temporal dependence and the variance of the differenced idiosyncratic noises on the results, we also consider the combinations $\gamma = -0.8$ and $\sigma_a^2 = 0.1$ ($\sigma_e^2 = 1$) and $\gamma = 1$ and $\sigma_a^2 = 10$ ($\sigma_e^2 = 10$). We introduce contemporaneous correlations among the idiosyncratic noises. Σ_a is generated with $\sigma_a^2 = 0.1,\ 1$ and 10 in the main diagonal and, following Onatski (2012), a Toeplitz structure with parameter $b = 0.5$. Finally, we consider models with heteroscedastic idiosyncratic noises. The variances are generated by $\sigma_{a_i}^2 \sim U\,[0.5, 1.5]\,,\ \sigma_{a_i}^2 \sim U\,[0.05, 0.15]$ and $\sigma_{a_i}^2 \sim U\,[5, 15]$; see Bai and Ng (2006) and Breitung and Eickmeier (2011) for the same design to simulate heteroscedastic idiosyncratic noises. In these two latter cases, we consider $\gamma = -0.8$ and 1.

For each replica, we generate observations Y_t and differentiate the data in a univariate fashion. Then, the eigenvalues of the sample covariance matrix of $\frac{1}{T-1}(\Delta Y)(\Delta Y)'$ are computed and r is determined using each of the procedures described above with $r_{\max} = 4$, 7 and 13 when $N = 12$, 50 and 200, respectively.[6] The number of factors determined using the three criteria proposed by Bai and Ng (2002) are denoted by $\hat{r}_{IC_1}$, $\hat{r}_{IC_2}$, $\hat{r}_{IC_3}$, while the number of factors determined implementing the procedure due to Onatski (2010) is denoted by $\hat{r}_{ED}$. Finally, the number of factors estimated using the two ratios proposed by Ahn and Horenstein (2013) are denoted by $\hat{r}_{ER}$ and $\hat{r}_{GR}$.

Figure 2.1 plots, for $N = 12$ and $T = 100$, the Monte Carlo averages and 95% confidence intervals, for homoscedastic and contemporaneously uncorrelated idiosyncratic noises,[7] of i)

is carried out after removing the first 100 observations.

[6]It is important to note that Ahn and Horenstein (2013) recomend double demeaned the data for their estimators to have a better behaviour. However, in our Monte Carlo experiments, we observe a deterioration of the performance of all criteria to determine the number of factors. Consequently, we compute the covariance matrix using the original differenced observations.

[7]The effect of heteroscedasticity and weak cross-correlation on the estimated eigenvalues is negligible. The results are available upon request.

the sample ordered eigenvalues; ii) their differences; and iii) their ratios, together with the corresponding population quantities, when $\gamma = -0.8$ and $\sigma_a^2 = 0.1$, $\gamma = 1$ and $\sigma_a^2 = 1$, $\gamma = -0.8$ and $\sigma_a^2 = 1$ and $\gamma = 1$ and $\sigma_a^2 = 10$. When the idiosyncratic noises are homoscedastic and white noise, according to the results in previous section, the largest eigenvalue of the population covariance matrix of ΔY is given by $\lambda_1 = \sigma_e^2 + \sum_{i=1}^{N} p_{i1}^2$ while all other eigenvalues are given by σ_e^2. Note that in the first two cases, $\sigma_e^2 = 1$ and the population eigenvalues are equal. In the two latter cases, $\sigma_e^2 = 10$. Figure 2.1 shows that, regardless of the value of σ_e^2, the eigenvalues are better estimated when $\gamma = 1$ than when $\gamma = -0.8$, with smaller biases and standard deviations. Obviously, given γ, the eigenvalues are better estimated when σ_a^2 is smaller. Therefore, in order to estimate the eigenvalues of the covariance matrix of ΔY, it is important not only the relative variance of the differenced idiosyncratic noises but also their temporal dependence.

In order to analyze the separate effect of the cross-sectional and temporal dimensions of the system on the estimation of the eigenvalues, Figure 2.2 plots the same quantities as in Figure 2.1 for $\gamma = -0.8$ and $\sigma_a^2 = 1$, when $N = 12, 50$ and 200, and $T = 100$ and 500. Note that when N increases, the first eigenvalue of the population covariance matrix is different and is estimated with larger biases and standard deviations. All other eigenvalues are also estimated with larger biases and standard deviations. Therefore, given T, increasing N could lead to an even worse estimation of the sample eigenvalues. However, as expected, given N, an increase in T leads to smaller biases and standard deviations of the estimated eigenvalues.

The finite sample properties of the estimated eigenvalues have effects on the properties of the procedures to detect the number of factors. Figure 2.3 plots, for each of the procedures considered, the percentage of replicates in which the estimated number of common factors is: i) $\hat{r} = 0$; ii) $\hat{r} = r$; iii) $\hat{r} = r_{\max}$; and iv) $\hat{r} > r$, when $\gamma = -0.8$ and $\sigma_a^2 = 0.1$ ($\sigma_e^2 = 1$), when $N = 12, 50$ and 200 and $T = 100$ and 500. We consider idiosyncratic noises being homoscedastic and uncorrelated; heteroscedastic and uncorrelated; and homoscedastic and cross-sectionally correlated. We can observe that, regardless of the structure of the idiosyncratic noises and the cross-sectional dimension, when $T = 100$, the three information estimators tend to overestimate

r and in most of the replicates $\hat{r}_{IC} = r_{\max}$. However, when $T = 500$, the percentage of $\hat{r}_{IC} = r$ is close to 100% if the idiosyncratic errors are homoscedastic and cross-sectionally uncorrelated even if $N = 12$. However, if there is cross-sectional correlation $\hat{r}_{IC} = r_{\max}$. On the other hand, increasing N leads to a larger percentage of $\hat{r}_{IC} > r$. The performance of the two estimators based on ratios of eigenvalues, $\hat{r}_{ER}$ and $\hat{r}_{GR}$, is very similar and always better than that of the estimator based on differenced eigenvalues, $\hat{r}_{ED}$. The percentages of correct estimation of r when implementing the $\hat{r}_{ER}$ and $\hat{r}_{GR}$ estimators are close to 90% when $N = 12$ and $T = 100$ and increase to 100% when increasing either N or T. The results for heteroscedastic and cross-correlated idiosyncratic noises are very similar.

Figure 2.4 plots the same quantities as in Figure 2.3 when $\gamma = 1$ and $\sigma_a^2 = 1$. Note that this case is comparable to that in Figure 2.3 in the sense that the variance of the differenced idiosyncratic noises is the same, $\sigma_e^2 = 1$, but the differentiated idiosyncratic noises are cross-sectionally uncorrelated white noises. We can observe that the performance of the alternative procedures to estimate r is rather different to that in Figure 2.3. All procedures have correct estimations close to 100% except the information criteria when $N = 12$ and the idiosyncratic errors are cross-correlated. In this latter case, $\hat{r}_{IC} = r_{\max}$. Consequently, not only the variance of the differenced idiosyncratic noises but also its dependence structure have effects on the procedures to detect the number of factors. Only the $\hat{r}_{ER}$ and $\hat{r}_{GR}$ estimators seem to be robust to them.

Finally, Figure 2.5 considers the case when $\gamma = -0.8$ and $\sigma_a^2 = 1$ with $\sigma_e^2 = 10$. In this case, the information criteria behave very similarly than when $\sigma_a^2 = 0.1$ and $T = 100$ with $\hat{r}_{IC} = r_{\max}$. However, when $N = (12, 50)$ and $T = 500$, the information criteria procedures estimate $\hat{r}_{IC} = 0$. Therefore, it seems that they are more affected by the temporal dependence of the differenced idiosyncratic noises than by their variance. On the other hand, when looking at the performance of $\hat{r}_{ER}$ and $\hat{r}_{GR}$, we can observe that it clearly deteriorates when $\sigma_e^2 = 10$. Therefore, their performance clearly depends on σ_e^2. The behavior of $\hat{r}_{ED}$ depends both on γ and σ_e^2 with a rather large percentage of cases in which $\hat{r}_{IC} = 0$.

In the second part of the Monte Carlo experiments, we consider models in which $r = 2$.

First, we consider a second non-stationary common factor, i.e. $\Phi = I$ and $\Sigma_\eta = I$. Second, the covariance matrix of the factor disturbances is given by $\Sigma_\eta = \text{diag}(1, 5)$. Finally, the last model considered has a second stationary factor with $\Sigma_\eta = I$ and $\Phi = \text{diag}(1, 0.5)$.

For each of the three Data Generating Process (DGP) above, Figure 2.6 plots the percentages of i) $\hat{r} = 0$; ii) $\hat{r} = 1$; iii) $\hat{r} = r$; iv) $\hat{r} = r_{\max}$; and v) $\hat{r} > r$, when $\gamma = -0.8$ and $\sigma_a^2 = 0.1$ ($\sigma_e^2 = 1$) and for $N = 12$ with $T = 100$ and $N = 200$ with $T = 500$.[8] First of all, observe that when $N = 12$ and $T = 100$, the information criteria chose $\hat{r} = r_{\max}$ in all cases. Increasing the dimensions of the system helps for $\hat{r}_{IC1}$ and $\hat{r}_{IC2}$ but not for $\hat{r}_{IC3}$. When looking at the ED, ER and GR criteria, we can observe that, regardless of the structure of the two factors, when $N = 200$ and $T = 500$, all of them have percentages of determination of the true number of factors close to 100%. However, when $N = 12$ and $T = 100$, there is a large percentage of replicates in which $\hat{r} = 1$. In this case, the ED procedure is better than the two procedures based on ratios. When the two common random walks in the original data have different variances, the ED procedure has an acceptable proportion of cases in which $\hat{r} = r$.

2.5. Empirical analysis

In this section, we implement the procedures considered in this paper to determine the number of common factors in a system of inflation rates in 15 euro area countries, namely, Austria (AUT), Belgium (BEL), Denmark (DEN), Finland (FIN), France (FRA), Germany (GER), Greece (GRE), Ireland (IRL), Italy (ITA), Luxemburg (LUX), Netherlands (NED), Portugal (POR), Spain (SPA), Sweden (SWE) and United Kingdom (UK). Prices, observed monthly from January 1996 to November 2015,[9] P_{it}, have been obtained from the OCDE data base[10] and transformed into annual inflation as $y_{it} = 100 \times \Delta_{12} \log(P_{it})$. When needed, the inflation rates have been corrected

[8] Monte Carlo results on the estimated eigenvalues are available upon request. They are not included to save space.

[9] Note that the sample period includes the global financial crisis of 2008-2009. Stock and Watson (2011) point out that the factor space can be consistently estimated by PC even with certain types of breaks or time variation in the factor loadings. Intuitively, if under weak assumptions $\hat{F}$ consistently estimates a rotation of F, then, the factors can break or evolve in some limited fashion and the PC estimator will remain consistent.

[10] http://stats.oecd.org/index.aspx?queryid=221

by outliers using the software developed by the United States Census Bureau.[11] Following Stock and Watson (2005), outliers are substituted by the median of the 5 previous observations. Furthermore, the inflation series have been deseasonalized when appropriate.[12]

Then, as in Reis and Watson (2010) and Altissimo et al. (2009), we carry out the determination of the number of factors using both the inflation data in levels and after differencing. All procedures are implemented with $r_{\max} = 5$. Regardless of whether the procedures are implemented using the original or differenced inflation rates, the information criteria estimate $\hat{r} = 5$ and $\hat{r}_{ER} = \hat{r}_{GR} = 1$. However, after differencing, the ED procedure detects just one factor while $\hat{r}_{ED} = 3$ in the original inflation series. According to our Monte Carlo experiments, if the number of true factors is $r \geq 2$, then the ED, ER and GR procedures tend to detect $\hat{r} < r$ when implemented to differentiated data. Therefore, we could expect the true number of factors to be larger than one. Consequently, we extract the factors assuming that $r = 3$ both from the original and differenced inflation series. In the latter case, the extracted factors are reaccumulated as proposed by Bai and Ng (2004). The extracted factors and their corresponding weights are plotted in Figure 2.7; compare with the factor extracted by Delle Monache et al. (2016) using quarterly inflation for a panel of 12 inflation rates from a sample of EMU countries. In Figure 2.7, there are not significant differences between the factors estimated using the original and differenced inflation rates but for the centering of the latter. This result could be expected since the variances of all the idiosyncratic noises are rather small with values between 0.03 and 0.1. Consequently, the differenced idiosyncratic noises are white noises with small variances.

Finally, we should point out that the main difference between extracting factors assuming that $r = 1$ or $r = 3$ is the interpretability. Recall that PC consistently estimates the space spanned by the factors. Therefore, assuming that $r = 3$ we can obtain rotations that are not allowed when assuming that $r = 1$.

[11] https://cran.r-project.org/web/packages/seasonal.pdf

[12] Camacho et al. (2015) show that the performance of deseasonalized data is comparable to using non-seasonally adjusted data in the context of estimating factors with forecasting purposes. Previous deseasonalizing apparently provides the best of two worlds: not working with incorrect assumptions about common seasonality while keeping a limited number of parameters to be estimated.

2.6. Conclusions

Differencing non-stationary cointegrated systems have effects on the properties of factor determination procedures. We show that both the variance and the dependence structure of the differenced idiosyncratic noises are important when measuring these effects. If $r = 1$, the ER and GR procedures work well even in relatively small sizes under all the structures of the idiosyncratic noises considered in this paper. Only when the variance of the differenced idiosyncratic noises is very large with respect to the variance of the differenced factor, the performance is worse although better than the alternatives. However, the performance of all procedures deteriorates when $r = 2$. In this case, the ED procedure seems to work better.

Figures

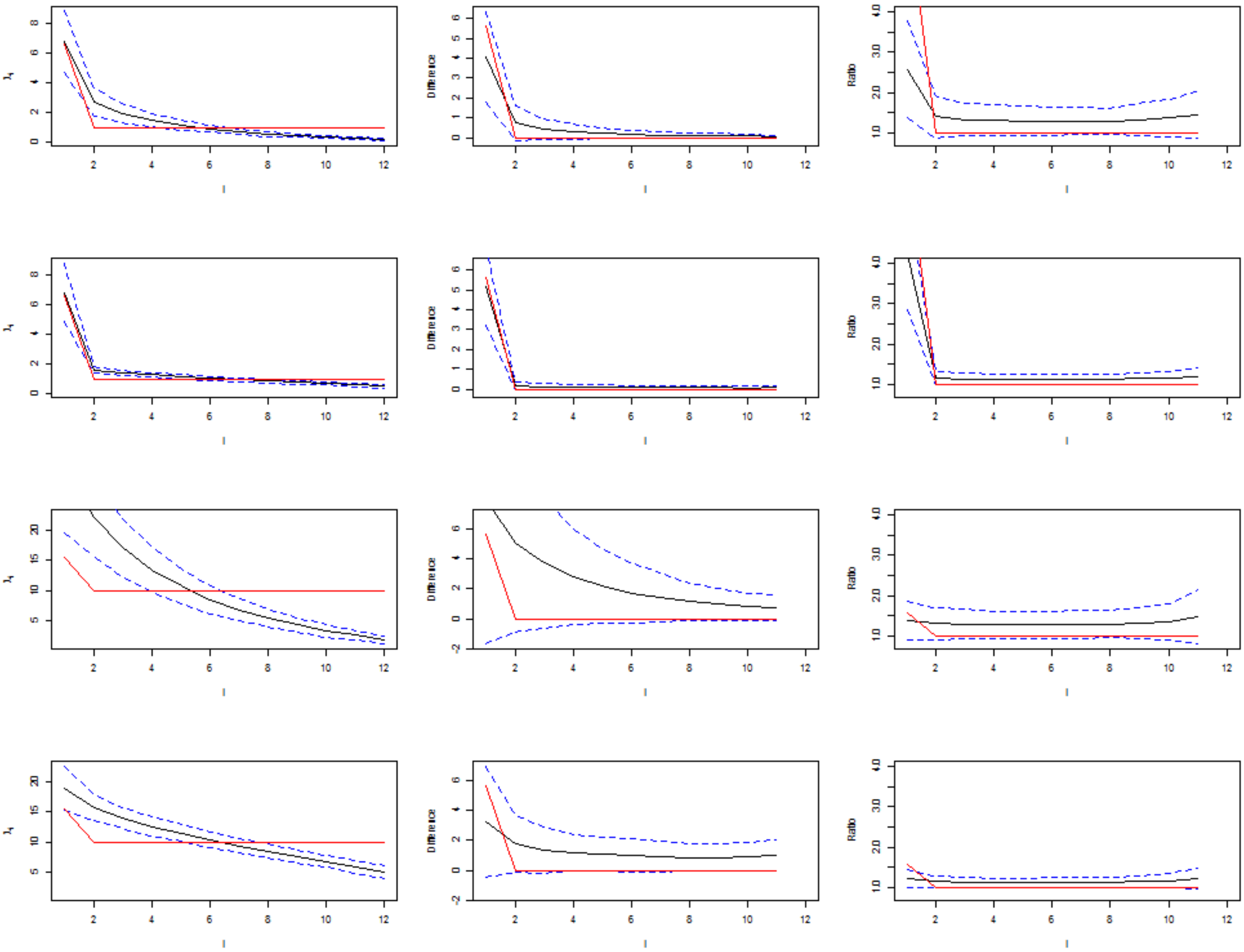

Figure 2.1: Eigenvalues of DFM with $N = 12$, $T = 100$, $r = 1$, $\phi = 1$, with $\gamma = -0.8$ and $\sigma_a^2 = 0.1$ (first row), $\gamma = 1$ and $\sigma_a^2 = 1$ (second row), $\gamma = -0.8$ and $\sigma_a^2 = 1$ (third row) and $\gamma = 1$ and $\sigma_a^2 = 10$ (fourth row). The first column plots the eigenvalues while the second and third columns plot their differences and ratios respectively. The population eigenvalues are plotted in red, the Monte Carlo averages in black and the corresponding 95% intervals in blue.

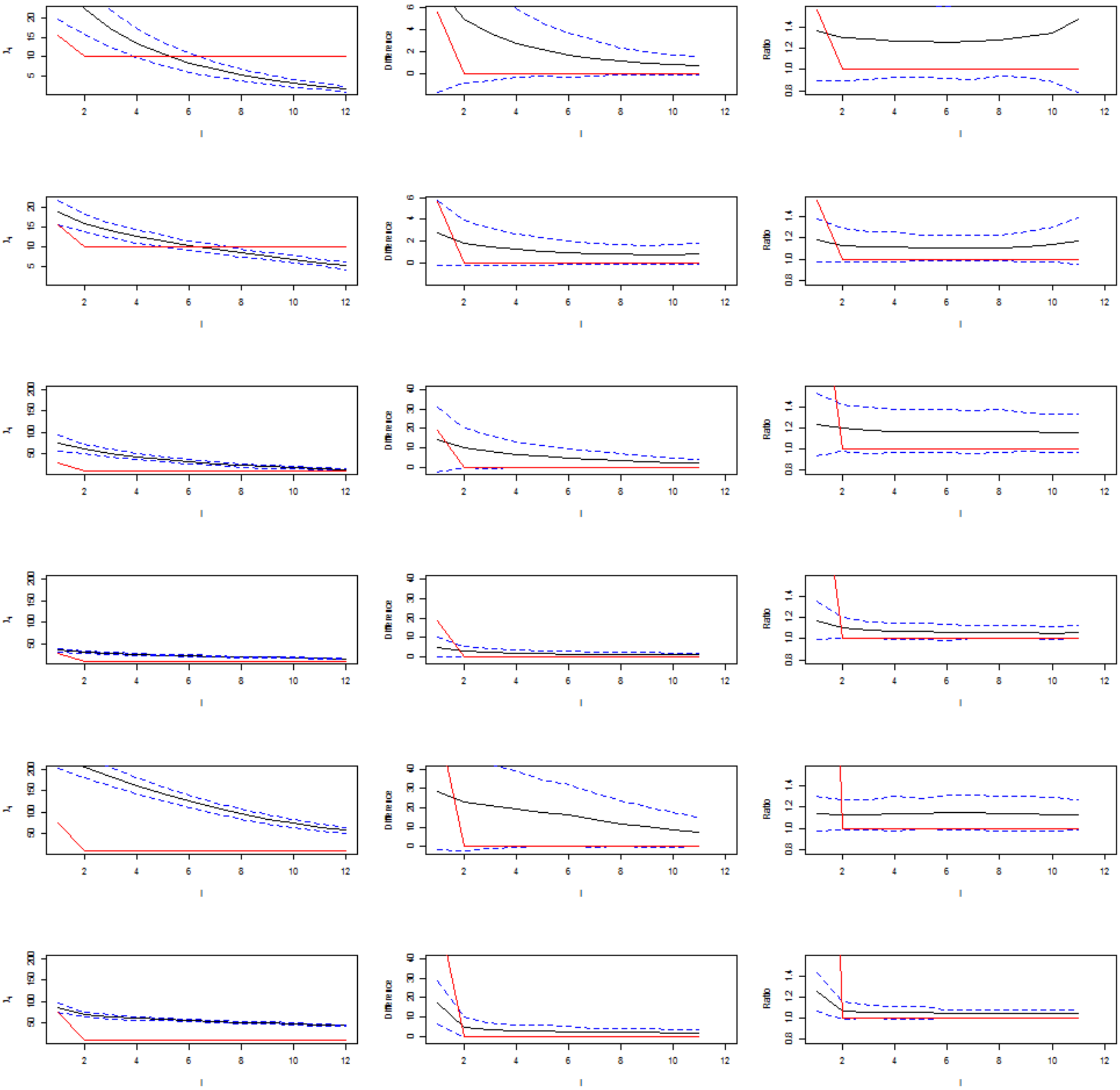

Figure 2.2: Eigenvalues of DFM with $r = 1$, $\phi = 1$ and $\sigma_\eta^2 = 1$ when the idiosyncratic noises are AR(1) process with $\gamma = -0.8$ and $\sigma_a^2 = 1$. The first column plots the eigenvalues while the second and third column plot their differences and ratios respectively. The population eigenvalues are plotted in red, the Monte Carlo averages in black and the corresponding 95% intervals in blue. First row $N = 12, T = 100$; second row $N = 12, T = 500$; third row $N = 50, T = 100$; fourth row $N = 50, T = 500$; fifth row $N = 200, T = 100$ and sixth row $N = 200, T = 500$.

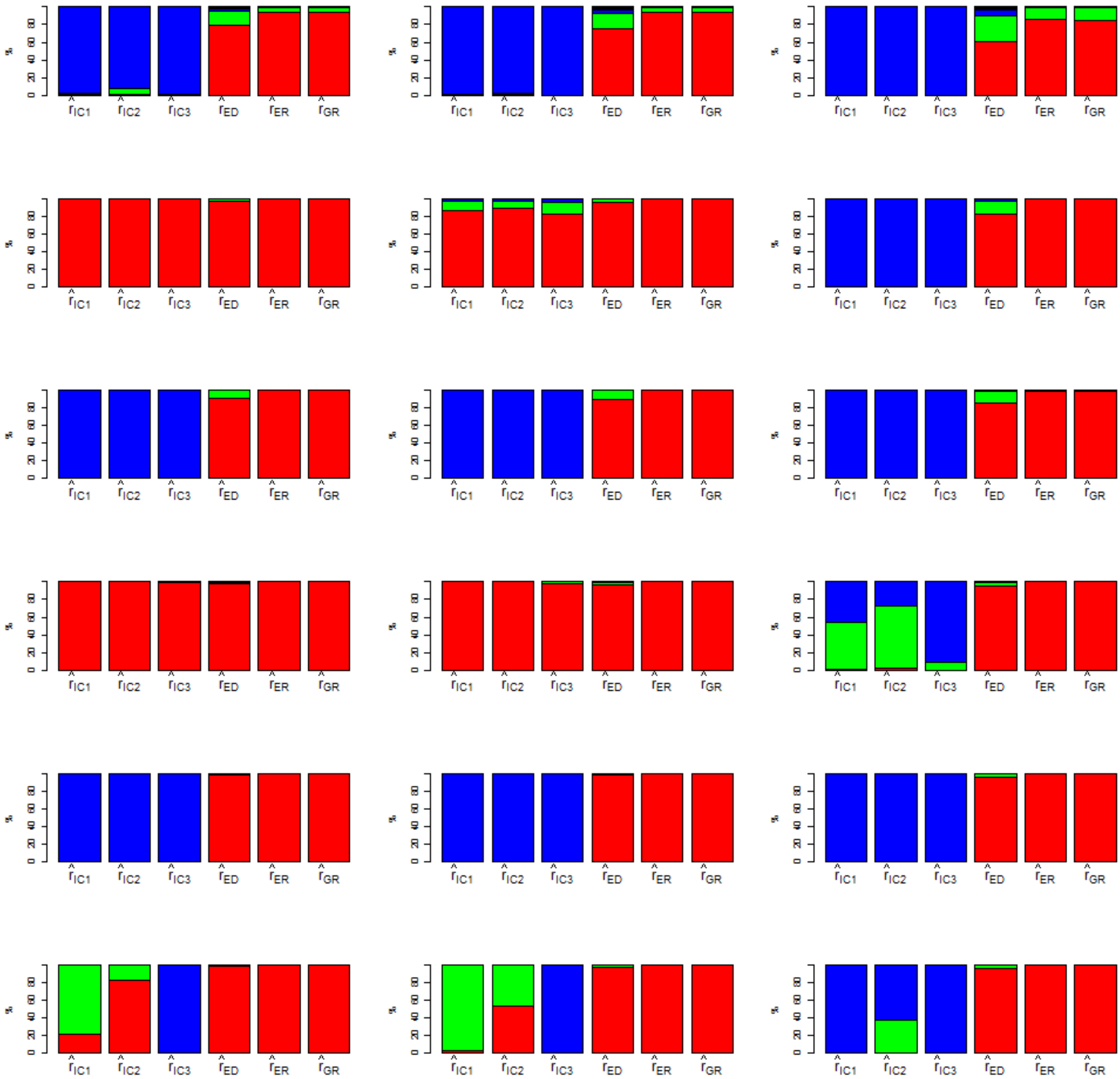

Figure 2.3: Percentage of $\hat{r} = r_{\max}$ (blue), $r_{\max} > \hat{r} > r$ (green), $\hat{r} = 1$ (red) and $\hat{r} = 0$ (black) in a DFM with $r = 1$, $\phi = 1$, $\sigma_\eta^2 = 1$, $\gamma = -0.8$ and $\sigma_a^2 = 0.1$. First row $N = 12, T = 100$; second row $N = 12, T = 500$; third row $N = 50, T = 100$; fourth row $N = 50, T = 500$; fifth row $N = 200, T = 100$ and sixth row $N = 200, T = 500$. The first column has homoscedastic and uncorrelated idiosyncratic noise; the second column the noises are heteroscedastic while in the third column they are cross-sectionally correlated.

Figure 2.4: Percentage of $\hat{r} = r_{\max}$ (blue), $r_{\max} > \hat{r} > r$ (green), $\hat{r} = 1$ (red) and $\hat{r} = 0$ (black) in a DFM with $r = 1$, $\phi = 1$, $\sigma_\eta^2 = 1$, $\gamma = 1$ and $\sigma_a^2 = 1$. First row $N = 12, T = 100$; second row $N = 12, T = 500$; third row $N = 50, T = 100$; fourth row $N = 50, T = 500$; fifth row $N = 200, T = 100$ and sixth row $N = 200, T = 500$. The first column has homoscedastic and uncorrelated idiosyncratic noise; the second column the noises are heteroscedastic while in the third column they are cross-sectionally correlated.

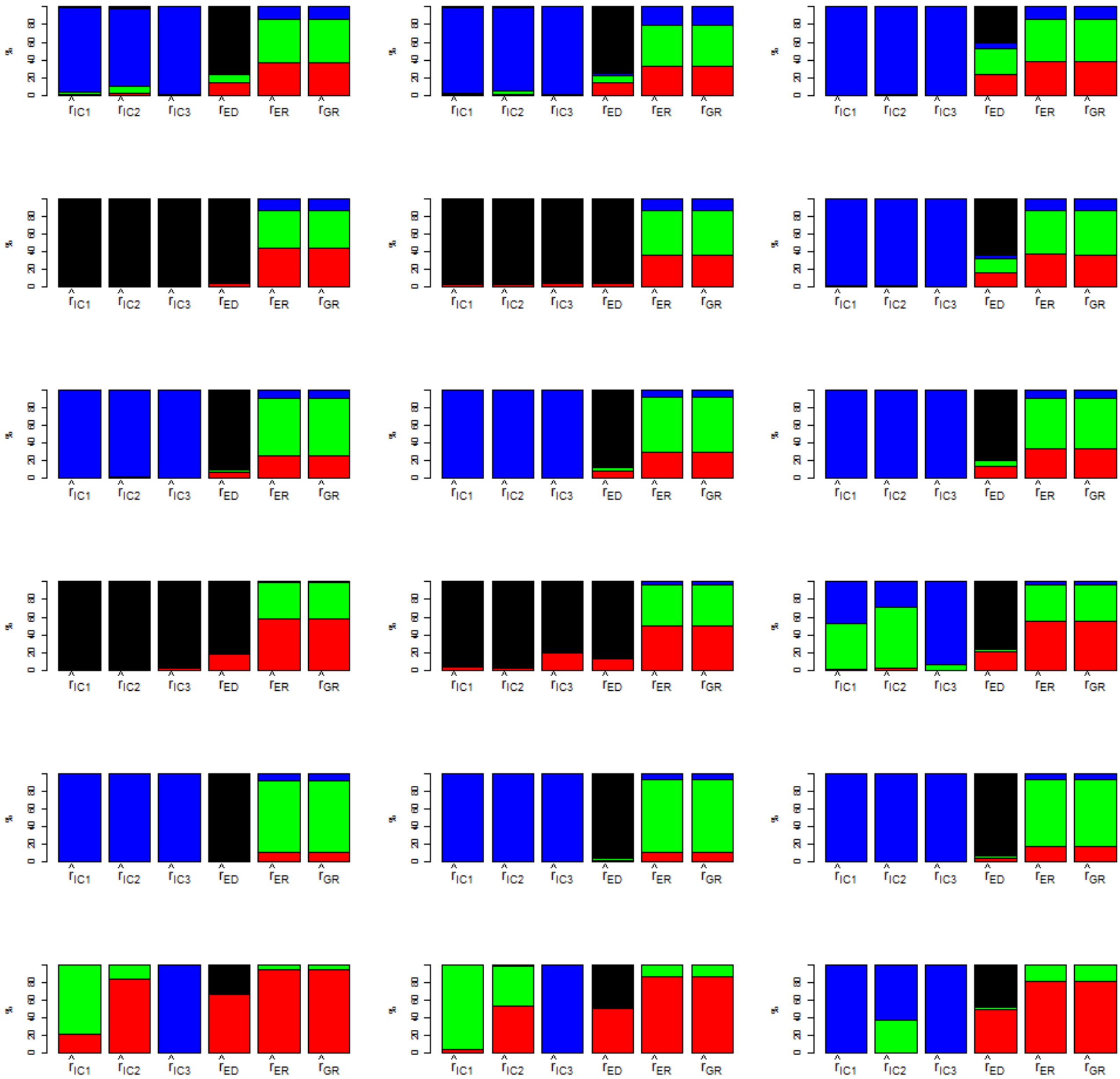

Figure 2.5: Percentage of $\hat{r} = r_{\max}$ (blue), $r_{\max} > \hat{r} > r$ (green), $\hat{r} = 1$ (red) and $\hat{r} = 0$ (black) in a DFM with $r = 1$, $\phi = 1$, $\sigma^2_\eta = 1$, $\gamma = -0.8$ and $\sigma^2_a = 1$. First row $N = 12, T = 100$; second row $N = 12, T = 500$; third row $N = 50, T = 100$; fourth row $N = 50, T = 500$; fifth row $N = 200, T = 100$ and sixth row $N = 200, T = 500$. The first column has homoscedastic and uncorrelated idiosyncratic noise; the second column the noises are heteroscedastic while in the third column they are cross-sectionally correlated.

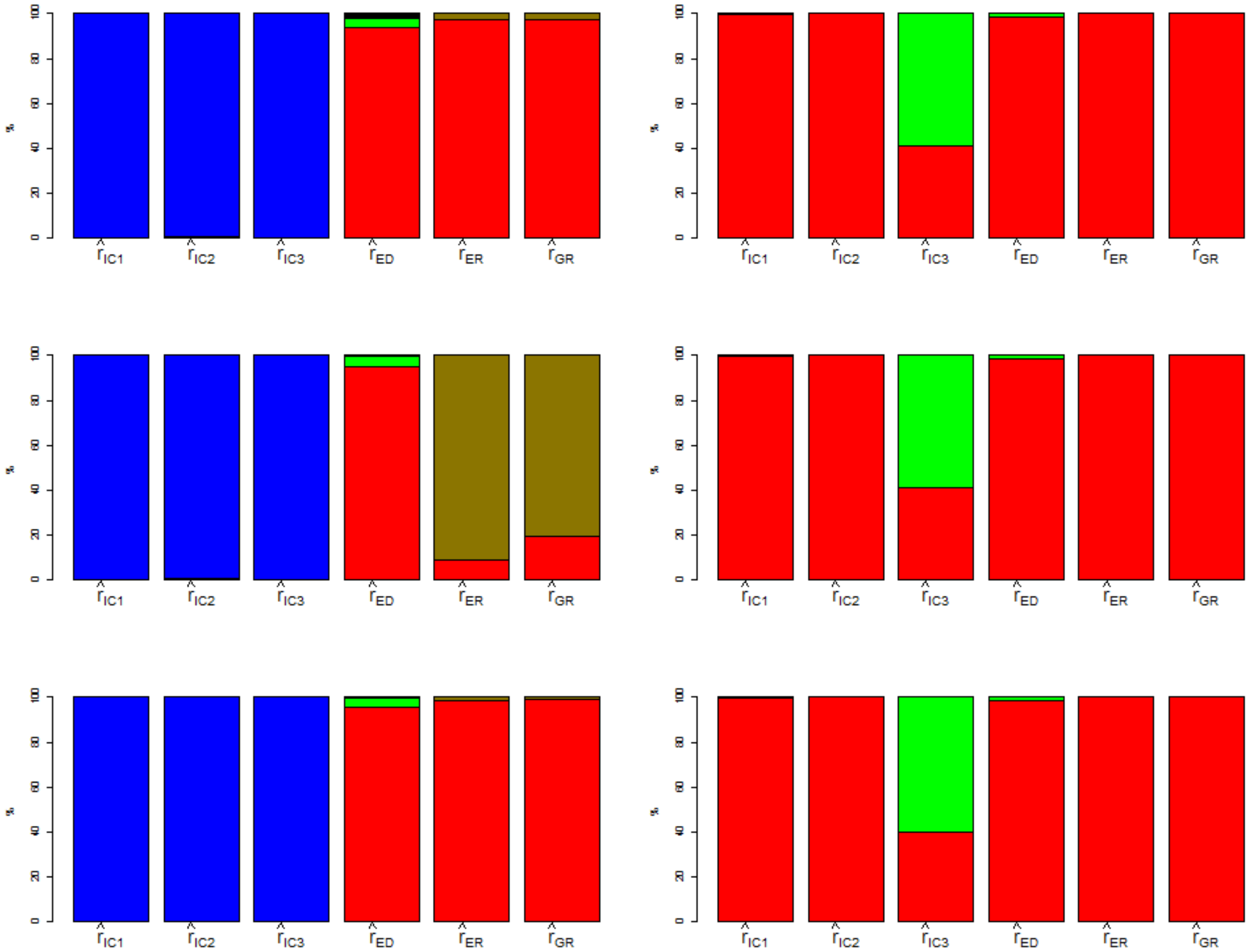

Figure 2.6: Percentage of $\hat{r} = r_{\max}$ (blue), $r_{\max} > \hat{r} > r$ (green), $\hat{r} = 2$ (red), $\hat{r} = 1$ (gold) and $\hat{r} = 0$ (black) in a DFM with $r = 2$, $\gamma = -0.8$ and $\sigma_a^2 = 0.1$. System dimensions $N = 12$, $T = 100$ (first column); $N = 200$, $T = 500$ (second column). The factors are two random walks with variance $\sigma_\eta^2 = 1$ (first row); two random walks with variances $\sigma_{\eta_1}^2 = 1$ and $\sigma_{\eta_2}^2 = 5$ (second row) and a random walk with variance $\sigma_{\eta_1}^2 = 1$ and a stationary factor with $\sigma_{\eta_2}^2 = 1$.

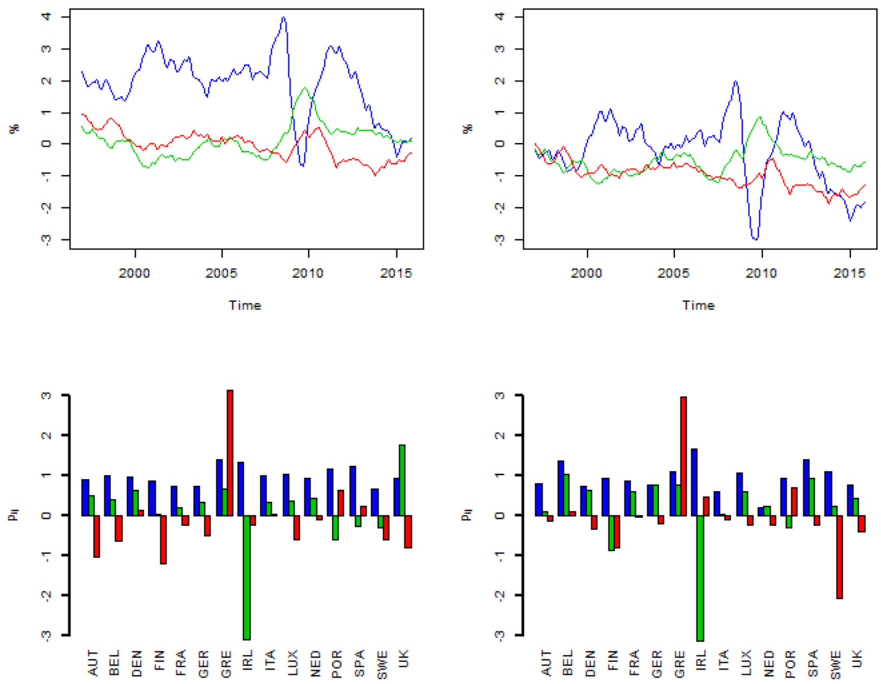

Figure 2.7: PC estimated factors (first row) and corresponding factor weights (second row) obtained assuming $r = 3$ and using original inflation rates (first column) and differenced rates (second column).

Chapter 3

Estimating non-stationary common factors: Implications for risk sharing

3.1. Introduction

Dynamic Factor Models (DFMs) were first introduced in economics by Geweke (1977) and Sargent and Sims (1977) with the aim of extracting the underlying common factors in a system of time series. In macroeconomics, these common factors are useful for building indicators and to predict key variables of the economy, among many other applications. Recently, econometricians have to deal with data sets consisting of hundreds of series, making the use of large dimensional DFMs very attractive in practice; see Breitung and Eickmeier (2006), Bai and Ng (2008), Stock and Watson (2011), Breitung and Choi (2013) and Bai and Wang (2016) for reviews of the existing literature.

It is well known that macroeconomic time series are frequently non-stationary and cointegrated. The connection between cointegration and common factors is analyzed by Stock and Watson (1988), Johansen (1991), Vahid and Engle (1993), Escribano and Peña (1994), Gonzalo and Granger (1995), Bai (2004), Bai and Ng (2004), Moon and Perron (2004), Banerjee et al. (2014a,b) and Barigozzi et al. (2016, 2017), among others. Although differencing has advantages in univariate time series to deal with non-stationarity, it should be made with great care when dealing with multivariate systems; see Box and Tiao (1977). It is well known that when differencing a cointegrated system, the long-run information, crucial to understand co-movements between the variables, is lost. Canova (1998) qualifies the detrending issue as "delicate and controversial" and compares the properties of the cyclical components of a system of seven real macroeconomic series obtained using seven univariate and three multivariate techniques. He concludes that the properties of the extracted business cycles vary widely across detrending methods. Sims (2012) claims that "when cointegration may be present, simply getting rid of the non-stationarity by differencing individual series so that they are all stationary throws away vast amounts of information and may distort inference". Consequently, the number of works dealing with non-stationary and possibly cointegrated DFMs is increasing. In the context of non-stationary systems, Bai (2004) proposes factor extraction implementing Principal Components (PC) to data in levels and derives the rates of convergence and limiting distributions of the estimated common trends, loading weights and the common component when the idiosyncratic components are stationary; see Engel et al. (2015) for an application to exchange rates. However, Barigozzi et al. (2016, 2017) point out that stationarity of the idiosyncratic components would produce an amount of cointegration for the observed system that it is not observed in the systems that are standard in the DFMs literature as, for example, those of Stock and Watson (2012a) and Forni et al. (2009). The idiosyncratic component in those datasets is likely to be non-stationary and, consequently, an estimation strategy robust to the assumption that some of the idiosyncratic components are non-stationary should be preferred. Alternatively, PC can be implemented to first difference data. Then, the estimated factors can be either obtained by integration of their estimated first differences as proposed by Bai and Ng (2004) or by projecting the original system onto the space spanned by the estimated loading as proposed by Barigozzi et al. (2016).[1] Bai and Ng (2004) prove the consistency of PC factor estimates when they are obtained from first differenced data using the "differencing and recumulating"

[1] In this paper we focus on DFMs without deterministic components. In this case, both approaches are equivalent.

method; see Greenway-McGrevy et al. (2016) who obtain recumulated factors in the context of exchange rates. Additionally, in their Monte Carlo analysis, they evaluate and compare the finite sample properties of both PC procedures and show that the non-stationary common factors can be properly recovered by both approaches when the idiosyncratic components are stationary. However, when the idiosyncratic components are non-stationary, PC cannot be directly implemented to the original data as proposed by Bai (2004) and it is convenient to use the "differencing and recumulating" method proposed by Bai and Ng (2004). Finally, Choi (2016) extends the Generalized PC estimator (GPCE) to the case of unit roots in the common factors, deriving the asymptotic distribution of the common factors and factor loadings. He shows that the GPCE is more efficient than the traditional PC estimator. Although consistent, PC based approaches have a major limitation in that they are not exploiting in any way the dynamic nature of the factors, nor the serial and cross-sectional dependence, or the heterocedasticity of the idiosyncratic components. Consequently, they are not efficient.[2]

Instead of implementing PC procedures, factor extraction can be carried out using two-step Kalman Smoothing (2SKS) techniques based on combining PC factor extraction and a Kalman Smoother. The main advantage of the 2SKS comes from the flexibility of the Kalman filter to explicitly model the factor and idiosyncratic dynamics. In the stationary case, Doz et al. (2011, 2012) show that 2SKS outperforms PC in terms of the precision of the factor estimates and derive its asymptotic properties; see also Poncela and Ruiz (2016). 2SKS has been implemented to non-stationary systems by Seong et al. (2013) in a low-dimensional setting and in Quah and Sargent (1993) in a large but finite cross-sectional dimension case with orthogonal idiosyncratic components.

The contributions of this paper are twofold. First, we extend the analysis of Bai and Ng (2004) comparing the factors extracted using PC implemented to the original non-stationary system with those obtained by "differencing and recumulating". In the case of a single factor, we consider a

[2]Other authors dealing with non-stationary DFMs are Eickmeier (2009), who analyzes the comovements and heterogeneity in the euro area by fitting a non-stationary DFM similar to Bai and Ng (2004), augmented with a structural factor setup from Forni and Reichlin (1998). Also, Bai and Ng (2010) extend the results of Bai and Ng (2004),

wide range of structures of the idiosyncratic noises, including heteroscedasticity and temporal and/or cross-sectional dependences. We also consider systems with two factors with the factors being either both non-stationary or one stationary and another non-stationary. With respect to the idiosyncratic components, we consider cases in which all of them are either stationary or non-stationary and cases in which some of them are stationary and others are not. We also include in the comparison the GPCE proposed by Choi (2016). Finally, we compare PC and 2SKS factor extraction. We analyze the performance of the 2SKS procedure when extracting factors using the first differenced data and estimating the original factors by recumulating. Furthermore, we propose a new 2SKS procedure which can be implemented to the original non-stationary system.[3]

The second contribution of this paper is an empirical application in which all factor extraction procedures are implemented to a non-stationary system of aggregate output and consumption variables of 21 OECD industrialized countries. International risk sharing focus on cross-border mechanisms to smooth consumption when a country is hit by a negative output shock. The goal is to check international risk sharing is a short or long-run issue. This is helpful to check if GDP fluctuations are directly passed to consumption on the contrary, can be at least partially cross-border smoothed (and therefore not totally passed to consumption). The use of possible non-stationary DFMs allows us to distinguish between long-run and short-run issues in consumption smoothing through international risk sharing. As far as we know, this is the first time that non-stationary DFMs are used in this context.

The rest of this paper is structured as follows. Section 3.2 describes the DFM and the factor extraction procedures considered. Section 3.3 presents the results of Monte Carlo experiments. Section 3.4 contains the empirical application to measure risk sharing. Finally, Section 3.5 concludes.

and Forni et al. (2014) who evaluate the role of news shocks in generating the business cycle. In this paper, we focus on non-stationary DFMs based on time domain. For non-stationary DFMs based on frequency domain, see Eichler et al. (2011).

[3] In independent work, Barigozzi and Luciani (2017) also propose a generalization of Doz et al. (2011, 2012) to the non-stationary case. They show empirically that the 2SKS extraction is more efficient than integrating the PC estimator of the first differences of the factors. However, they do not consider the comparison with recumulating the 2SKS.

3.2. Factor extraction algorithms

In this section we introduce notation and describe the DFM considered. Furthermore, the PC and 2SKS factor extraction procedures are described.

3.2.1. Dynamic Factor Model

We consider the following static DFM where the unobserved common factors, F_t, and the idiosyncratic noises, ε_t, follow potentially non-stationary VAR(1) processes:

$$Y_t = PF_t + \varepsilon_t, \tag{3.1}$$

$$F_t = \Phi F_{t-1} + \eta_t, \tag{3.2}$$

$$\varepsilon_t = \Gamma \varepsilon_{t-1} + a_t, \tag{3.3}$$

where $Y_t = (y_{1t}, \dots, y_{Nt})'$ and $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$ are $N \times 1$ vectors of the variables observed at time t and idiosyncratic noises respectively. The common factors, $F_t = (F_{1t}, \dots, F_{rt})'$ and the factor disturbances, $\eta_t = (\eta_{1t}, \dots, \eta_{rt})'$, are $r \times 1$ vectors, with r $(r < N)$ being the number of common factors which is assumed to be known. The $N \times 1$ vector of idiosyncratic disturbances, a_t, is distributed independently from the factor disturbances, η_t, for all leads and lags. Furthermore, η_t and a_t, are assumed to be Gaussian white noises with positive definite covariance matrices $\Sigma_\eta = \text{diag}(\sigma^2_{\eta_1}, ..., \sigma^2_{\eta_r})$ and Σ_a, respectively. $P = (p_1, \dots, p_N)'$, is the $N \times r$ matrix of factor loadings, where, $p_i = (p_{i1}, \dots, p_{ir})'$ is an $r \times 1$ vector. For identification, we assume that $P'P/N = I_r$. Finally, $\Phi = \text{diag}(\phi_1, \dots, \phi_r)$ and $\Gamma = \text{diag}(\gamma_1, ..., \gamma_N)$ are $r \times r$ and $N \times N$ matrices containing the autoregressive parameters of the factors and idiosyncratic components, respectively, which can be equal to one; see, for example, Stock and Watson (1989) and Barigozzi and Luciani (2017) for static DFMs for non-stationary data. Note that according to economic theory, there is full aggrement that some factors (related with, for example, technology shocks) have permanent effects while others (such as monetary policy shocks) have only transitory effects. Furthermore, there is also arguments to assume non-stationary idiosyncratic components. Barigozzi et al. (2016, 2017) point out that stationarity of the idiosyncratic components would produce an amount of cointegration for the observed system that it is not consistent with that observed in the systems that are standard in the DFMs literature as, for example, those of Stock and Watson (2002a) and Forni et al. (2009). The idiosyncratic component in those datasets is likely to be non-stationary. The implausibility of a stationary idiosyncratic component is also confirmed empirically by Barigozzi et al. (2016) in a large macroeconomic system of quarterly series describing the US economy with about half of the estimated idiosyncratic components found to be non-stationary according to the test proposed by Bai and Ng (2004).

The DFM in equations (3.1) to (3.3) is not identified. To solve the identification problem and uniquely define the factors, a normalization is necessary. In the context of PC factor extraction, it is common to impose the restriction $P'P/N = I_r$ and $F'F$ being diagonal, where $F = (F_1, \ldots, F_T)$ is the $r \times T$ matrix of common factors; see, for example, Bai and Wang (2014) and Barigozzi et al. (2016).

3.2.2. PC factor extraction

The most popular procedures for factor extraction in large datasets are based on the PC procedure. The distinctive feature of PC is that it allows a consistent factor extraction without assuming any particular error distribution and specifications of the factors and idiosyncratic noises further than the cross-correlation of the latter being weak and the variability of the common factors being not too small.[4] Furthermore, PC is computationally simple which explains its wide implementation among practitioners when dealing with very large systems of economic variables.

PC factor extraction separates the common component, PF_t, from the idiosyncratic component, ε_t, through cross-sectional averages of Y_t in such a way that when N and T tend

[4] Onatski (2012) considers a DFM in which the explanatory power of the factors does not strongly dominate the

to infinity, the effect of the idiosyncratic component converges to zero remaining only the effects associated to the common factors. The PC estimators of P and F_t, are obtained as the solution to the following least squares problem

$$\min_{F_1,\ldots,F_T,P} V_r(P,F) \tag{3.4}$$

subject to $P'P/N = I_r$ and $F'F$ being diagonal where $V_r(P,F) = \frac{1}{NT}\sum_{t=1}^{T}(Y_t - PF_t)'(Y_t - PF_t)$. The solution to (3.4) is obtained by setting $\hat{P}^{PCL}$ equal to $\sqrt{N}$ times the eigenvectors corresponding to the r largest eigenvalues of YY' where $Y = (Y_1, \ldots, Y_T)$ is a $N \times T$ matrix of observable. The corresponding PC estimator of F using data in levels is given by

$$\hat{F}^{PCL} = N^{-1}\hat{P}^{PCL'}Y. \tag{3.5}$$

Alternatively, when the common factors are $I(0)$, Bai and Ng (2002) give the restriction $FF'/T = I_r$ with $P'P$ being diagonal, such that, the estimator of the matrix of common factors, $\hat{F}^{PCL}$, is the $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix $Y'Y$, with estimated factor loadings, $\hat{P}^{PCL} = Y\hat{F}^{PCL'}/T$. When the common factors are $I(1)$, Bai (2004) proposes to use the restriction $FF'/T^2 = I_r$ with $P'P$ being diagonal. In this case, $\hat{F}^{PCL}$, is the T times the eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix $Y'Y$ and $\hat{P}^{PCL} = Y\hat{F}^{PCL'}/T^2$. The difference is only computational, these latest restrictions are less costly when $N > T$, while that $P'P/N = I_r$ with FF' being diagonal are less costly when $N < T$.

In the context of stationary systems, if the common factors are pervasive and the serial and cross-sectional correlation of the idiosyncratic components is weak, Bai (2003) proves the consistency of $\hat{F}^{PCL}$, $\hat{P}^{PCL}$ and the common component, deriving their asymptotic distributions when N and T tend simultaneously to infinity, allowing for heteroscedasticity in both the temporal and cross-sectional dimensions; see also Bai and Ng (2002) and Stock and Watson

explanatory power of the idiosyncratic noises.

(2002a). Additionally, Bai (2004) extends the asymptotic results when F_t is $I(1)$ and ε_t is $I(0)$. When the idiosyncratic components are $I(1)$, Bai and Ng (2008) show that PC factor extraction implemented to data in levels yields inconsistent estimates of the common factors.

In order to obtain more efficient estimates of F_t and P relative to the PC factor extraction, Choi (2016) proposes a GPCE implemented to the original non-stationary system. Using the standardization $FF'/T^2 = I_r$, the feasible estimator of the factor space spanned by F_t, denoted by $\hat{F}_t^{GPCE}$, is T times the eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix $Y'\hat{\Sigma}_\varepsilon^{-1}Y$ where $\hat{\Sigma}_\varepsilon^{-1/2} = \text{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_N)$ with $\hat{\sigma}_i^2 = \sum_{t=1}^T \hat{\varepsilon}_{it}^2/T$, and $\hat{\varepsilon}_{it}^2$ are obtained after implementing the PC estimator of F_t proposed by Bai (2004) as in equation (3.5). The corresponding weights are given by $\hat{P}^{GPCE} = T^{-2}Y\hat{F}^{GPCE'}$; see Choi and Hwang (2012) for an application to forecasting the Korean inflation. Choi (2016) shows that the GLS version of the PC estimator is asymptotically equivalent to the original PC estimator.

Alternatively, instead of extracting the factors implementing PC to the original data, Bai and Ng (2004) propose differencing the data in a univariate fashion and extract the factors from the following differenced model

$$\Delta Y_t = P\Delta F_t + \Delta\varepsilon_t, \tag{3.6}$$

$$\Delta F_t = \Phi\Delta F_{t-1} + \Delta\eta_t, \tag{3.7}$$

$$\Delta\varepsilon_t = (\Gamma - I)\varepsilon_{t-1} + a_t, \tag{3.8}$$

where $\Delta = (1 - L)$ with L being the lag operator such that $LY_t = Y_{t-1}$. The weights are estimated as $\sqrt{N}$ times the first r normalized eigenvectors of the $N \times N$ sample covariance matrix of ΔY_t and denoted by $\hat{P}^{PCD}$. The corresponding estimated factors are given by

$$\hat{f}_t = N^{-1}\hat{P}^{PCD'}\Delta Y_t, \quad t = 2, \ldots, T. \tag{3.9}$$

Once the factors are extracted from the first differenced variables, the estimated factors can be obtained either by integration of their estimated first differences as proposed by Bai and Ng (2004) or by projecting the original system onto the space spanned by the estimated loading as proposed by Barigozzi et al. (2016). The "differencing and recumulating" estimated factor is given by

$$\hat{F}_t^{PCD} = \sum_{s=2}^{t} \hat{f}_s, \quad t = 2, \ldots, T. \tag{3.10}$$

Note that assuming $Y_0 = 0$, the estimated differenced factor at time 1 is given by $\hat{f}_1 = N^{-1}\hat{P}^{PCD'}Y_1$ and, consequently, the estimated recumulated factor coincides with the projected factor which is given by

$$\hat{F}_t^{BLL} = N^{-1}\hat{P}^{PCD'}Y_t \quad t = 1, \ldots, T. \tag{3.11}$$

Bai and Ng (2004) and Barigozzi et al. (2016) show that, respectively, $\hat{F}_t^{PCD}$ and $\hat{F}_t^{BLL}$ are a consistent estimator for a rotation of F_t up to a level shift regardless of whether the idiosyncratic component, ε_t, is $I(0)$ or $I(1)$. Note that the factor estimators proposed by Bai and Ng (2004) and Barigozzi et al. (2016) are asymptotically equivalent with some finite sample differences when there are deterministic trends in the DFMs.

3.2.3. Two-step Kalman Smoother

The 2SKS procedure was proposed by Doz et al. (2011) for stationary DFMs. Therefore, 2SKS can be implemented to ΔY_t. The 2SKS factor extraction procedure is based on combining PC and Kalman Smoother techniques. First, the common factors and factor loadings are estimated using PC obtaining $\hat{P}^{PCD}$ and $\hat{f}_t$ and the corresponding idiosyncratic and factor residuals, $\Delta\hat{\varepsilon} = \Delta Y - \hat{P}^{PCD}\hat{f}$ and $u_t = \hat{f}_t - \hat{\Phi}\hat{f}_{t-1}$ where $\hat{\Phi}$ is the ordinary least squares (OLS) estimator of the regression of $\hat{f}_t$ on $\hat{f}_{t-1}$. These residuals are used to estimate the covariance matrices $\hat{\Psi} = \text{diag}\left(\hat{\Sigma}_{\Delta\varepsilon}\right)$ where $\hat{\Sigma}_{\Delta\varepsilon} = \Delta\hat{\varepsilon}\Delta\hat{\varepsilon}'/(T-1)$ with $\Delta\hat{\varepsilon} = (\Delta\hat{\varepsilon}_2, \ldots, \Delta\hat{\varepsilon}_T)$ is an $N \times T - 1$ matrix and $\hat{\Sigma}_\eta = uu'/(T-1)$ where $u = (u_2, \ldots, u_T)$ is an $r \times T - 1$ matrix. Assuming that $f_0 \sim N(0, \Sigma_f)$, the unconditional covariance of the factors can be estimated as $\text{vec}\left(\hat{\Sigma}_f\right) = \left(I_{r^2} - \hat{\Phi} \otimes \hat{\Phi}\right)^{-1} \text{vec}\left(\hat{\Sigma}_\eta\right)$. After writting the DFM in equations (3.6) to (3.8) in state-space form, with the system matrices substituted by $\hat{P}^{PCD}$, $\hat{\Psi}$, $\hat{\Phi}$, $\hat{\Sigma}_\eta$ and $\hat{\Sigma}_f$, the Kalman smoother is run to obtain an updated estimation of the factors denoted by $\hat{f}_t^{KS}$. Finally, estimates of the common factors, $\hat{F}_t^{KSD}$, are obtained by recumulating analogously to equation (3.10).

Doz et al. (2011) prove the consistency of $\hat{f}_t^{KS}$ when N and T are large considering assumptions slightly different than those in Bai and Ng (2002), Stock and Watson (2002a) and Bai (2003) but with a similar role. The 2SKS works well in finite samples obtaining more accurate factor estimates of $f_t = \Delta F_t$ even in the presence of cross-sectional heteroscedasticity in the idiosyncratic noises, see Doz et al. (2011). Finally, Doz et al. (2012) propose iterating the 2SKS procedure until convergence is achieved in terms of two consecutive log-likelihood values.

Considering the possibility of non-stationary common factors, we propose to extend the 2SKS algorithm as follows[5]

1. Obtain PC estimates of P and F_t with data in levels given by expression (3.5). Compute the idiosyncratic residuals $\hat{\varepsilon} = Y - \hat{P}^{PCL}\hat{F}^{PCL}$ and the covariance matrix of the idiosyncratic residuals, $\hat{\Psi} = \text{diag}\left(\hat{\Sigma}_\varepsilon\right)$.

2. For each estimated factor, $\hat{F}_{jt}^{PCL}$, $j = 1, \dots, r$, carry out the Augmented Dickey Fuller (ADF) test.

   a) If the null hypothesis of a unit root is rejected, obtain the OLS estimate of the autoregressive coefficient, $\hat{\phi}_j$, the residuals $u_{jt} = \hat{F}_{jt}^{PCL} - \hat{\phi}\hat{F}_{jt-1}^{PCL}$ and the sample variance of the factor disturbance, $\hat{\sigma}^2_{\eta_j} = \sum_{t=1}^T u_{jt}^2/T$. The initial state of the factor is assumed to have zero mean and variance estimated by $\hat{\sigma}^2_{F_j} = \hat{\sigma}^2_{\eta_j}/(1-\hat{\phi}_j^2)$.

   b) If the null hypothesis is not rejected, then $\hat{\phi}_j = 1$ and the residuals are computed as $u_{jt} = \Delta\hat{F}_{jt}^{PCL}$. Calculate the variance of the factor residuals, $\hat{\sigma}^2_{\eta j} = \sum_{t=2}^T \Delta\hat{F}_{jt}^{PCL^2}/(T-$

[5] Barigozzi and Luciani (2017) propose an alternative extension in which, in order to isolate common trends and stationary factors, they use a nonparametric approach which identifies the common trends as those linear combinations of the factors obtained by the leading eigenvectors of a transformation of the long-run covariance matrix as proposed by Peña and Poncela (2006), Pan and Yao (2008), Lam et al. (2011) and Zhang et al. (2016).

1). Assume a diffuse prior for the initial factor with mean zero and variance $\hat{\sigma}^2_{F_j} = \kappa$, where κ is a large constant that empirically performs well (for instance $\kappa = 10^7$); see Harvey and Phillips (1979), Burridge and Wallis (1985) and Harvey (1989).[6]

3. Obtain $\hat{\Phi} = \text{diag}(\hat{\phi}_1, \ldots, \hat{\phi}_r)$, $\hat{\Sigma}_\eta = \text{diag}(\hat{\sigma}^2_{\eta_1}, \ldots, \hat{\sigma}^2_{\eta_r})$, $\hat{\Sigma}_F = \text{diag}(\hat{\sigma}^2_{F_1}, \ldots, \hat{\sigma}^2_{F_r})$ and use them together with $\hat{P}^{PCL}$ and $\hat{\Psi}$ in the KS to obtain the estimated common factors $\hat{F}^{KSL}$.

3.3. Finite sample performance

In this section, we carry out Monte Carlo experiments in order to study the performance of the factor extraction procedures described in the previous section. The experiments are based on $R = 500$ replicas generated by the DFM in equations (3.1)-(3.3) with sample sizes $T = (100, 500)$ and $N = (12, 50, 200)$. The factor loadings are generated once as $P \sim U[0, 1]$ and the autoregressive matrix of the idiosyncratic components is diagonal, $\Gamma = \gamma I$, with $\gamma = (-0.8, 0, 0.7, 1)$.[7] We consider three specifications of dependence of the idiosyncratic noises: a) homoscedastic and cross-sectionally uncorrelated, with $\Sigma_a = \sigma^2_a I$ where $\sigma^2_a = (0.1, 1, 10)$; b) heteroscedastic and cross-sectionally uncorrelated with the variances generated by $\sigma^2_{a_i} \sim U[0.05, 0.15]$, $\sigma^2_{a_i} \sim U[0.5, 1.5]$ and $\sigma^2_{a_i} \sim U[5, 15]$; c) homoscedastic and cross-sectionally correlated with weak cross-correlation generated following Kapetanios (2010) as $\Sigma^{1/2}\varepsilon_t$ where $\Sigma = [\sigma_{i,j}], \sigma_{i,j} = \sigma_{j,i} \sim U(-0.1, 0.1)$ for $|i - j| \leq 5$ for $i, j = 1, \ldots N$. Finally, with respect to the unobserved factors, we consider four different data generating processes (DGPs). The first DGP, denoted as model 1 (M1), has $r = 1$, $\Phi = 1$ and $\sigma^2_\eta = 1$ so that the factor is given by a random walk. The second and third models (M2 and M3) introduce a second random walk with $r = 2$ and $\Phi = I$ while $\Sigma_\eta = I$ (M2) and $\Sigma_\eta = \text{diag}(1, 5)$ (M3). Finally, the fourth model considered

[6] Koopman (1997) gives an exact solution for the initialization of the Kalman filter and smoothing for state space models with diffuse initial conditions.

[7] Alternatively, we generate artificial systems by model M1 where the temporal dependence of the idiosyncratic errors is $\gamma = \text{diag}(-0.8 I_{N/2}, 1 I_{N/2})$ and $\gamma = \text{diag}(0 I_{N/2}, 0.7 I_{N/2})$. The results are very similar to those when all idiosyncratic errors have the same dependence with $\gamma = -0.8$ and $\gamma = 0$, respectively. It seems that the results are driven by the smallest temporal dependence among the idiosyncratic noises. These results are available upon request.

(M4) also has two factors but one is stationary while the other is not. In particular, in model M4, $\Sigma_\eta = I$ and $\Phi = \text{diag}(1, 0.5)$.

For each DGP considered, the common factors are estimated using the procedures described in Section 3.2 obtaining $\hat{F}_t^{PCD}$ and $\hat{F}_t^{KSD}$, based on "differencing and recumulating", and $\hat{F}_t^{PCL}$, $\hat{F}_t^{GPCE}$ and $\hat{F}_t^{KSL}$, based on data in levels.[8] Following Bai (2004), the performance of the factor extraction procedures is evaluated by computing the sample correlation between the true factor, F_t, and a rotation of the estimated factors, $\hat{\delta}_j' \widehat{F}_t^{(j)}$, estimated by the following regression

$$F_{jt} = \hat{\delta}_j' \widehat{F}_t^{(j)} + \widehat{\nu}_t.$$

Figure 3.1 plots the Box-plots of the sample correlations between the true and rotated estimated factors obtained through the Monte Carlo replicates when the systems are generated by the M1 model with homoscedastic idiosyncratic errors with $\sigma_a^2 = 10$ when the temporal and cross-sectional dimensions are $(N, T) = (12, 50), (12, 100), (50, 100), (200, 100)$ and $(200, 500)$. Several conclusions can be obtained from Figure 3.1. First, all procedures based on differencing and recumulating are similar among them. The same can be said about the procedures based on extracting factors directly from the data in levels. Second, regardless of N and T, the correlations of the "differencing and recumulating" PC procedure can be rather low when the temporal dependence of the idiosyncratic component is negative. Furthermore, using the "differencing and recumulating" estimator implemented with the 2SKS procedure generates even smaller correlations, mainly when $\gamma = -0.8$. Note that, when the serial dependence of the idiosyncratic components is such that $\gamma < 0.5$, the variance of the differenced idiosyncratic component, $\sigma_{\Delta\varepsilon}^2$, is larger than the corresponding variance of the original component, σ_ε^2; see, for example, Corona et al. (2016). Consequently, the performance of the procedures using data in first differences deteriorates in this case. However, if $\gamma \geq 0.5$, then $\sigma_{\Delta\varepsilon}^2 < \sigma_\varepsilon^2$ and, consequently, the procedures based on "differencing and recumulating" may have an advantage. Third, if the idiosyncratic

[8]Note that, in the context of the DFM considered in this paper, the Monte Carlo results for the procedure proposed by Barigozzi et al. (2016) (BLL) are almost identical to those obtained by the procedure proposed by Bai and Ng (2004).

noises are white noise, the 2SKS procedures implemented to raw data generate correlations which are always close to 1. Note that the two-step procedure proposed in this paper does a remarkably good job. Only when the cross-sectional and temporal dimensions are very large, the procedures based on first differences estimate factors with correlations close to one. Fourth, if the dependence of the idiosyncratic noises is positive, differencing or extracting the factors using the original non-stationary system yields similar correlations. Only when N and T are relatively small, differencing performs worse. Finally, when the idiosyncratic errors are non-stationary, i.e. $\gamma = 1$, extracting the factors using differenced or original data yields similar moderate correlations. Only when N is very large, we observe the result established by the asymptotic theory with the procedures based on "differencing and recumulating" having correlations close to one while the non-consistent procedures based on original non-stationary data having smaller correlations.

The Box-plots plotted in Figure 3.1, help to understand the role of the dynamic dependence of the idiosyncratic noises on the performance of the alternative factor extraction procedures considered. In order to evaluate the effect of the variance of the disturbance of the idiosyncratic noises, Figure 3.2 plots the Box-plots of the correlations of the common factor estimates and the simulated ones for model M1 with $\gamma = -0.8$ and the same dimensions considered above and $\sigma_a^2 = 0.1$, 1 and 10. Note that if σ_a^2 is small, then all procedures have correlations close to 1 regardless of the cross-sectional and temporal dimensions and whether they are based on first differences or original data. The deterioration of the procedures based on "differencing and recumulating" is already observed for $\sigma_a^2 = 1$ with the exception of very large N and T. Finally, in Figure 3.3, we study the role of the variance of the idiosyncratic noises when $\gamma = 1$. In this case, it is clearly better to take first differences to the original series. The performance of the procedures based on extracting factors from the original data is only reasonable when $\sigma_a^2 = 0.1$.

To evaluate the precision of the factor estimates and summarizing the results, we carry out a response surface analysis by regressing the sample correlation averages on the cross-sectional and temporal dimensions, N and T, and the temporal dependence and variance of the idiosyncratic noises, γ and σ_a^2, for model M1 with homoscedastic, heteroscedastic and cross-correlated idiosyncratic noises. In the case of heteroscedastic idiosyncratic errors, the value of σ_a^2 considered as regressor is the expected value of the variances for each idiosyncratic noise. The regression parameter estimates together with the corresponding standard errors and adjusted R^2 are reported in Table 3.1. First, we can observe that the average correlation of the procedures based on "differencing and recumulating" is clearly smaller than that of the procedures implemented to original data. As above, we also observe that the correlations are similar among methods based on first differences and among methods based on original systems. Second, it is also clear that the correlations between the true factors and the rotated estimates obtained using procedures based on differenced data increase with γ, the temporal dependence of the idiosyncratic noise. This result could be expected given that, as explained above, when $\gamma < 0.5$, the variance of the differenced idiosyncratic component, $\sigma_{\Delta\varepsilon}^2$, is larger than the corresponding original variance, σ_ε^2, and, consequently, the recovery of the common factors is less precise. Furthermore, note that the increase in the correlations between true and rotated extracted factors is larger for KSL than for the PCL procedure, as expected given the flexibility of the Kalman filter to explicitly model the idiosyncratic dynamics. However, the correlations decrease with γ when the factor extraction procedures are implemented to original data. Third, increasing σ_a^2 negatively affects factor extraction for all procedures. However, for the same reasons explained above, the effect of σ_a^2 is less important if the factors are extracted using original non-stationary observations than when they are extracted using first-differenced data. Finally, Table 3.1 shows that the results are almost the same regardless of the particular specifications of the idiosyncratic components. It is remarkable that, for the particular specifications of the heteroscedasticity considered in this paper, the correlations between the true and rotated estimated factors obtained when the PCL and GPCE procedures are implemented are very similar.

Finally, we consider the three models with two factors. Figure 3.4 plots the Box-plots of the correlations across the Monte Carlo experiments between the true and rotated estimated common factors through the Monte Carlo experiments for models M2, M3 and M4 (by rows) with $\sigma_a^2 = 10$

and $\gamma = -0.8$. In each case, we consider homoscedastic, heteroscedastic and cross-correlated idiosyncratic errors (by columns) The cross-sectional and temporal dimensions are $N = 50$ and $T = 100$. First of all, as far as the two factors are non-stationary, models M2 and M3, we can observe the same patterns as those described for the case of one single factor. However, when one factor is a random walk and the second factor is stationary, model M4, none of the procedures estimate this factor adequately. The results are drastically deteriorated when extracting the stationary common factor.[9] Finally, Figure 3.5 plots the Box-plots of the correlations across Monte Carlo replicates when the idiosyncratic noise is $I(1)$ and $\sigma_a^2 = 1$. As expected, we can observe that the common factors are better extracted when we use first-differenced data.

In the context of determination of the number of factors, Corona et al. (2016) conclude that if ε_t is stationary, with autoregressive parameters smaller than 0.5 while F_t is non-stationary, then overdifferencing the idiosyncratic components may introduce distortions on the determination of the number of factors given that the relation between the variances of the common and idiosyncratic components is modified with the variances of ΔF_t decreasing and the variances of $\Delta\varepsilon_t$ increasing in relation to the variance of F_t and ε_t, respectively. Recall as well, that some procedures do not yield consistent estimates when the idiosyncratic noises are $I(1)$.

3.4. Empirical analysis

International or cross-border risk sharing focuses on the smoothing of consumption when a country is hit by a negative output shock. In an ideal world of perfect risk sharing, consumption should be insured. However, in practice, risk sharing is far from being full or complete and a percentage of GDP shocks are passed into consumption and are not smoothed. In a time series context, risk sharing has been traditionally addressed in the literature as a short-run issue and, consequently, analyzed within the context of stationary models. Nevertheless, more recently, some authors question this view and bring in the long-run perspective to the problem, although

[9] The results are similar even if the cross-sectional and temporal dimensions are increased.

the results are not conclusive. For instance, Becker and Hoffmann (2006) and Pierucci and Ventura (2010) analyse risk sharing within a cointegration context. Artis and Hoffmann (2008, 2012) argue that risk sharing has increased at lower frequencies and relate their results to the permanent income hypothesis. On the contrary, Leibrecht and Scharler (2008) using cointegration techniques and vector error correction models found that while consumption risk sharing in the short-run was around 30%, only accounts for a 10% in the long-run. As regards factor models, Del Negro (2002) implement a stationary DFM to disentangle movements in US state output and consumption due to national, regional or state-specific factors. Very recently, for capital flows, Byrne and Fiess (2016) apply non-stationary factor models to analyze the common and idiosyncratic elements in emerging markets' capital inflows.

The economic interpretation of the common factor analysis in our model should be as follows. If there is full risk sharing, idiosyncratic consumption and output cannot share a common factor since these two variables should be orthogonal in an ideal case of complete risk sharing where, under certain assumptions, domestic consumption should be a constant fraction of the aggregate world output. Hence, lack of complete full risk sharing should be detected through commonalities between domestic output and consumption. If we can find non-stationary common factors among the series of output and GDP we could conclude that there is no risk sharing in the long-run.

Our sample covers the following 21 industrialized OECD countries: Australia (AUS), Austria (AUT), Belgium (BEL), Canada (CAN), Denmark (DEN), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Ireland (IRL), Italy (ITA), Japan (JPN), Netherlands (NLD), New Zealand (NZL), Norway (NOR), Portugal (PRT), Spain (ESP), Sweden (SWE), Switzerland (CHE), United Kingdom (GBR) and United States (USA). The data are annual observations of Gross Domestic Product (GDP) and Consumption (C) from National Accounts and cover the time span 1960-2014 with $N = 42$ and $T = 55$. The main source of data is AMECO, the annual macro-economic database of the European Commission's Directorate General for Economic and Financial Affairs (DG ECFIN), which provides harmonized statistics on all of the variables required to perform the analysis. The nominal GDP and C have been transformed in Purchasing Power Standard (PPS) units by dividing the nominal aggregates by the appropriate PPS exchange rate reported by AMECO. To compute per-capita variables, the real aggregates expressed in PPS are divided by the population taken from the OECD Statistics. We build the aggregate GDP and C for the set of countries included in the analysis. To build the aggregates, we use weighted averages in order to reflect the importance of each country in the group of economies. Then, starting from the real indicators computed for each country in PPS, we followed the weighting procedure described in Beyer et al. (2001), where the aggregation is performed directly on growth rates (first difference of logs) but using time-varying weights of countries that are given by their relative share in real GDP, in levels. The aggregate GDP and consumption growth rates are integrated to get the log of the aggregate variables. As initial condition for the aggregated GDP (consumption), we aggregate the levels of real GDP (consumption) and take logs. To define the idiosyncratic variables or gaps in log levels we subtract the log of the aggregate from the log level of a specific country. The resulting gap could be interpret as the log of the percentage of a particular country GDP (consumption) over the aggregate variable (see Giannone and Reichlin, 2006, for the same interpretation).

Unit root tests are performed for the GDP and consumption gaps for all countries and, overall, we can consider that the series are $I(1)$. In order to determine the number of common factors, we implement the procedure proposed by Onatski (2010) and choose $r = 5$ regardless whether it is implemented to data in levels or first differences; see Corona et al. (2016) for a comparison on alternative procedures to determine the number of common factors in non-stationary DFMs.

Since, we do not know if the idiosyncratic errors are stationary or not, we differentiate the data and extract 5 common factors using PCD. Then, we recumulate the extracted common factors and the specific components. We use PANIC to check if the idiosyncratic errors are non-stationary. We performed individual tests for each idiosyncratic error and the pooled test proposed by Bai and Ng (2004) where the pooled statistic of the log of the pvalues (p_i) of the individual tests follows a standard normal distribution

$$P = \frac{-2\sum_{i=1}^{N} \log p_i - 2N}{\sqrt{4N}}. \tag{3.12}$$

Pooled tests could not be used in the original data because of strong cross correlation due to the common factors but they can be used in the specific components since this strong cross correlation has been removed after extracting the common factors. Both the individual tests over the idiosyncratic components as well as the pooled test (the p statistic was 0.19) indicate the idiosyncratic components are non-stationary. In this case, we have to choose any of the methods to extract the common factors that work with the data in first differences, since if the errors are non-stationary the procedures that work with the data in levels do not yield to consistent estimates. This was reflected in our simulations by the low correlations between the generated common factors and the estimated ones.

The rationale for finding that the idiosyncratic errors are non-stationary should be as follows. A large part of the commonality has been removed when generating the data as the variables that enter into the model are already deviations from the aggregate. This aggregate might proxy world comovements. Nevertheless, there are still strong correlations in the data that we remove through the common factors. If what it is left is non-stationary, as it might seem the case, it means that there are persistent movements that are generated internally and not shared among countries or due to interactions with third countries, as it might happen with the U.S. and Mexico. Another way of looking at this result is as follows: if after removing r_1 non-stationary common factors, what is left is stationary, it means that we should find $2N - r_1$ cointegrating relations among the data. This is not the case and, therefore, we conclude that in our model after removing r common factors (r_1 being non-stationary), what is left is non-stationary as well.

We proceed using PCD to recover the common factors and the factor loadings.[10] As mentioned before, we applied the "differencing and recumulating" method suggested by Bai and Ng (2004), although any method that works with the data in first differences could be used as well. We test how many of the common factors are non-stationary. The extracted sample factors in first differences are orthogonal as this condition is imposed for identifcation purposes, however the recumulated common factors do not need to be orthogonal. Therefore, we test how many of the

[10]The factor extraction results using the other procedures are available upon request.

common factors are non-stationary using the variant of the test for common trends of Stock and Watson (1988) proposed by Bai and Ng (2004). Basically, the test consist of deciding how many of the eigenvalues of the first order autoregressive matrix, after correcting for serial correlation in the residuals are close enough to 1. The estimated eigenvalues are 0.66, 0.83, 0.90, 0.91 and 1.02. We cannot reject the null hypothesis of 5 common trends, even though the fifth eigenvalue is only 0.66. Since T is not so large, we can conclude that there are 5 common factors in the data and, at least, 4 of them are non-stationary factors.

The next step is to decide if the factor loadings are different from zero and if we find a loadings different from zero associated to GDP and consumption for the same country. Since the factor loading matrix is the same for the model in first differences than for the model in levels, and in the model in levels the idiosyncratic errors are I(1), we perform inference about the factor loadings using the factor model in first differences (the asymptotic distribution of the loadings is given in Bai, 2003).

We analyze the factor loadings for the first common factor (see Figure 3.6) related to GDP series. The factor loadings could be considered different from zero for all countries but Australia, Canada, Denmark, UK and Switzerland. It gives positive weight to the Anglo-Saxon countries (USA, CAN, GBR, NZL and AUS) although it can be only considered different from zero for US and New Zealand while the weights have the opposite sign for the rest of European countries (other than the United Kingdom) and Japan. Within the last set, the highest, in absolute value, are given to Greece, Portugal, Spain followed by Japan. Curious enough, Greece, Portugal, Spain (jointly with Ireland and Italy that also have significant factor loadings of the same sign) constitute the so called PIIGS group, peripheral European countries where risk sharing has collapsed during the last recession and subsequent sovereign debt crisis faced. Kalemli-Ozcan et al. (2014) point out that the governments of these countries did not save during the expansionary phases of the business cycle and were not able to borrow on the international markets during the crisis due to the high levels of outstanding public debt. Ireland is also included in this set although its case is slightly different, with government deficits related to banking failures (see Kalemli-Ozcan et al., 2014). This might be the reason why Ireland is included in this group instead of within the Anglo-Saxon countries. Japan has experienced a long lasting recession and sluggish output growth since the early 1990s. We check the results through other estimation methods. No matter the estimation method, the factor loadings in domestic or idiosyncratic consumption seem to follow very closely those of idiosyncratic output, indicating lack of risk sharing. This interpretation should be in accordance with Becker and Hoffmann (2006) and Pierucci and Ventura (2010).

The second common factor gives the highest positive weight to New Zealand. On the negative side appears Japan. The next 2 common factors are devoted to separate Greece from other countries. Basically, the 3rd common factor separates Greece from Portugal and the 4th one to separates Greece from Ireland and Norway. The fifth common factor loads on several countries and has a difficult interpretation.

There are $21 \times 5 = 105$ loadings associated to each country for GDP and the same quantity associated to consumption. Only in 27 out of the 105 possible cases, factor loadings were significant for one of the variables (GDP or consumption) and not for the other (which could be an indication of risk sharing). However, we find that when a loading is significant for GDP for one country, it is usually significant and of the same sign for consumption for the same country.

3.5. Conclusions

In this study we examine the finite sample performance of alternative factor extraction procedures to estimate non-stationary common factors in the context of large DMFs. Furthermore, we extend the hybrid method from Doz et al. (2011) based on combining PC and Kalman smoothing, applying the technique to original non-stationary observations. We show that, when the idiosyncratic errors are non-stationary, the approaches based on estimating the common factors using non-stationary time series in levels do not perform well and that the procedures based on first differences should be used. This fact was pointed out by Bai and Ng (2008) for the basic PC estimator and we have checked that the same holds for the remaining methods that use data in levels. The empirical application shows that for a non-stationary system of 21 OECD industrialized economies, at least four common factors are non-stationary, such that, consumption and GDP share common trends. Furthermore, we apply PANIC to the estimated idiosyncratic errors, concluding that this component is non-stationary. Hence, these facts suggest the lack of full risk sharing both in the short and long-run.

Tables and Figures

Table 3.1: Response surface analysis by regressing sample correlations averages on the sample size, serial correlation and the variance of the idiosyncratic disturbance. Standard errors between parenthesis.

Dependent variable: Sample correlation averages						
M1 with homoscedastic idiosyncratic errors						
Regressor	PCD	KSD	PCL	BLL	GPCE	KSL
Constant	0.8517	0.7901	0.9548	0.8523	0.9437	0.9611
	(0.0445)	(0.0591)	(0.0316)	(0.0444)	(0.0328)	(0.0315)
N	0.0002	0.0004	0.0001	0.0002	0.0001	0.0001
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
T	0.0003	0.0003	0.0000	0.0003	0.0001	0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
γ	0.1331	0.2145	-0.1009	0.1329	-0.1093	-0.1062
	(0.0281)	(0.0374)	(0.0200)	(0.0281)	(0.0207)	(0.0199)
σ_a^2	-0.0296	-0.0353	-0.0100	-0.0102	-0.0296	-0.0090
	(0.0044)	(0.0058)	(0.0031)	(0.0044)	(0.0032)	(0.0031)
$\bar{R}^2$	0.5035	0.5026	0.3175	0.5031	0.3266	0.3215
M1 with heteresocedastic idiosyncratic errors						
	PCD	KSD	PCL	BLL	GPCE	KSL
Constant	0.8454	0.7879	0.9542	0.8459	0.9372	0.9618
	(0.0440)	(0.0580)	(0.0317)	(0.0440)	(0.0333)	(0.0314)
N	0.0003	0.0004	0.0001	0.0003	0.0001	0.0001
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
T	0.0003	0.0003	0.0000	0.0003	0.0001	0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
γ	0.1309	0.2150	-0.0993	0.1306	-0.1086	-0.1043
	(0.0278)	(0.0367)	(0.0200)	(0.0278)	(0.0211)	(0.0198)
σ_a^2	-0.0322	-0.0367	-0.0108	-0.0103	-0.0321	-0.0091
	(0.0043)	(0.0057)	(0.0031)	(0.0043)	(0.0033)	(0.0031)
$\bar{R}^2$	0.5362	0.5236	0.3139	0.5358	0.3291	0.3162
M1 with cross-correlated idiosyncratic errors						
	PCD	KSD	PCL	BLL	GPCE	KSL
Constant	0.8537	0.7929	0.9538	0.8543	0.9453	0.9599
	(0.0439)	(0.0583)	(0.0316)	(0.0439)	(0.0327)	(0.0313)
N	0.0002	0.0003	0.0001	0.0002	0.0001	0.0001
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
T	0.0003	0.0003	0.0001	0.0003	0.0001	0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
γ	0.1296	0.2116	-0.0986	0.1293	-0.1075	-0.1040
	(0.0278)	(0.0369)	(0.0199)	(0.0277)	(0.0206)	(0.0198)
σ_a^2	-0.0299	-0.0358	-0.0105	-0.0299	-0.0102	-0.0093
	(0.0043)	(0.0057)	(0.0031)	(0.0043)	(0.0032)	(0.0031)
$\bar{R}^2$	0.5097	0.5098	0.3166	0.5094	0.3260	0.3205

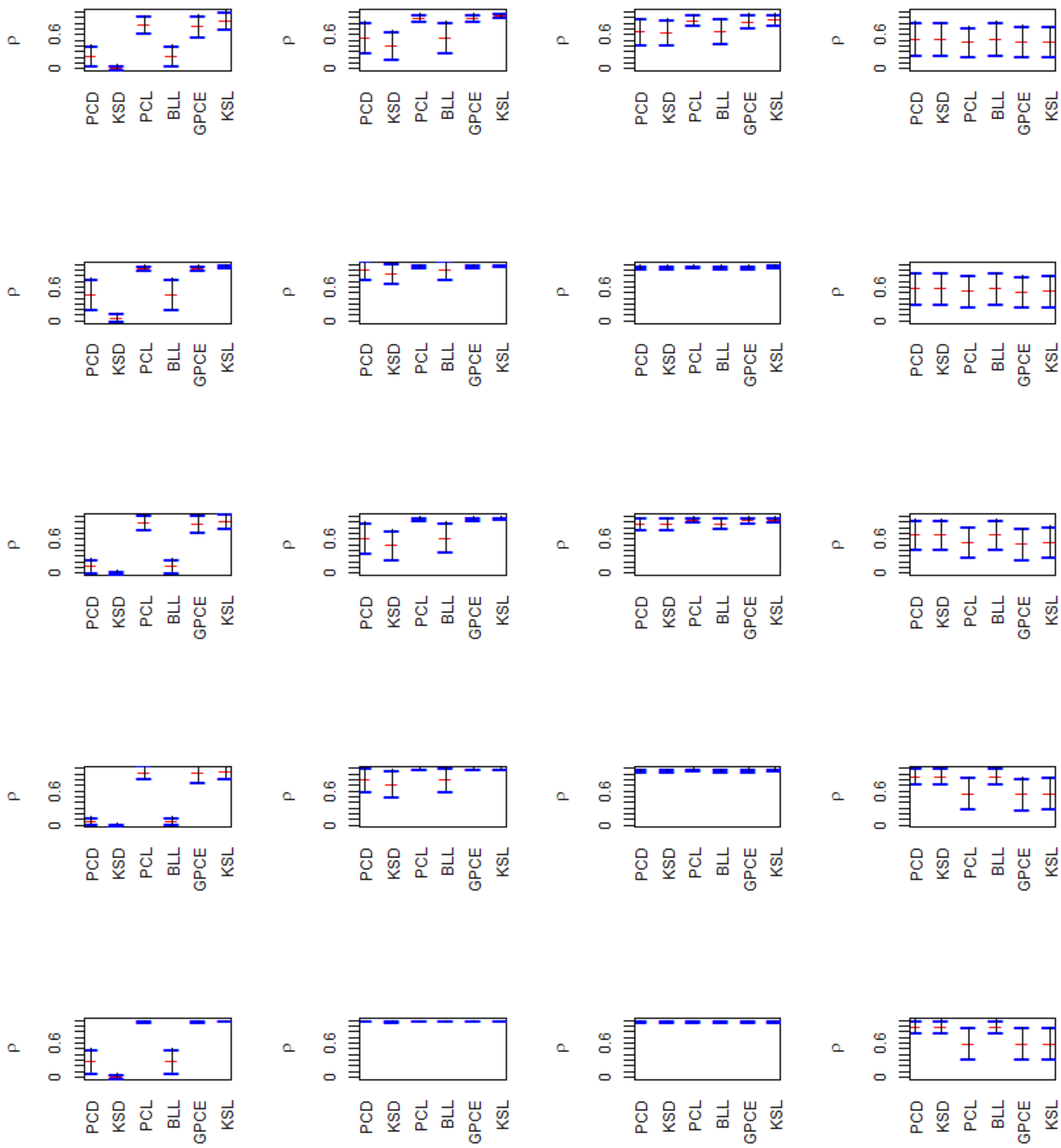

Figure 3.1: Box-plots of the sample correlations between $\{\hat{\delta}_j'\hat{F}_t^{PCD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{KSD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{PCL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{BLL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{GPCE}\}$ and $\{\hat{\delta}_j'\hat{F}_t^{KSL}\}$ with $\{F_t\}$. We consider the M1 model with homoscedasticity in idiosyncratic errors with $\sigma_a^2 = 10$. First row indicates $N = 12$ and $T = 50$; second row $N = 12$ and $T = 100$; third row $N = 50$ and $T = 100$; fourth row $N = 200$ and $T = 100$ and fifth row $N = 200$ and $T = 500$. The first column plots $\gamma = -0.8$, second column $\gamma = 0$, third column $\gamma = 0.7$ and fourth column $\gamma = 1$.

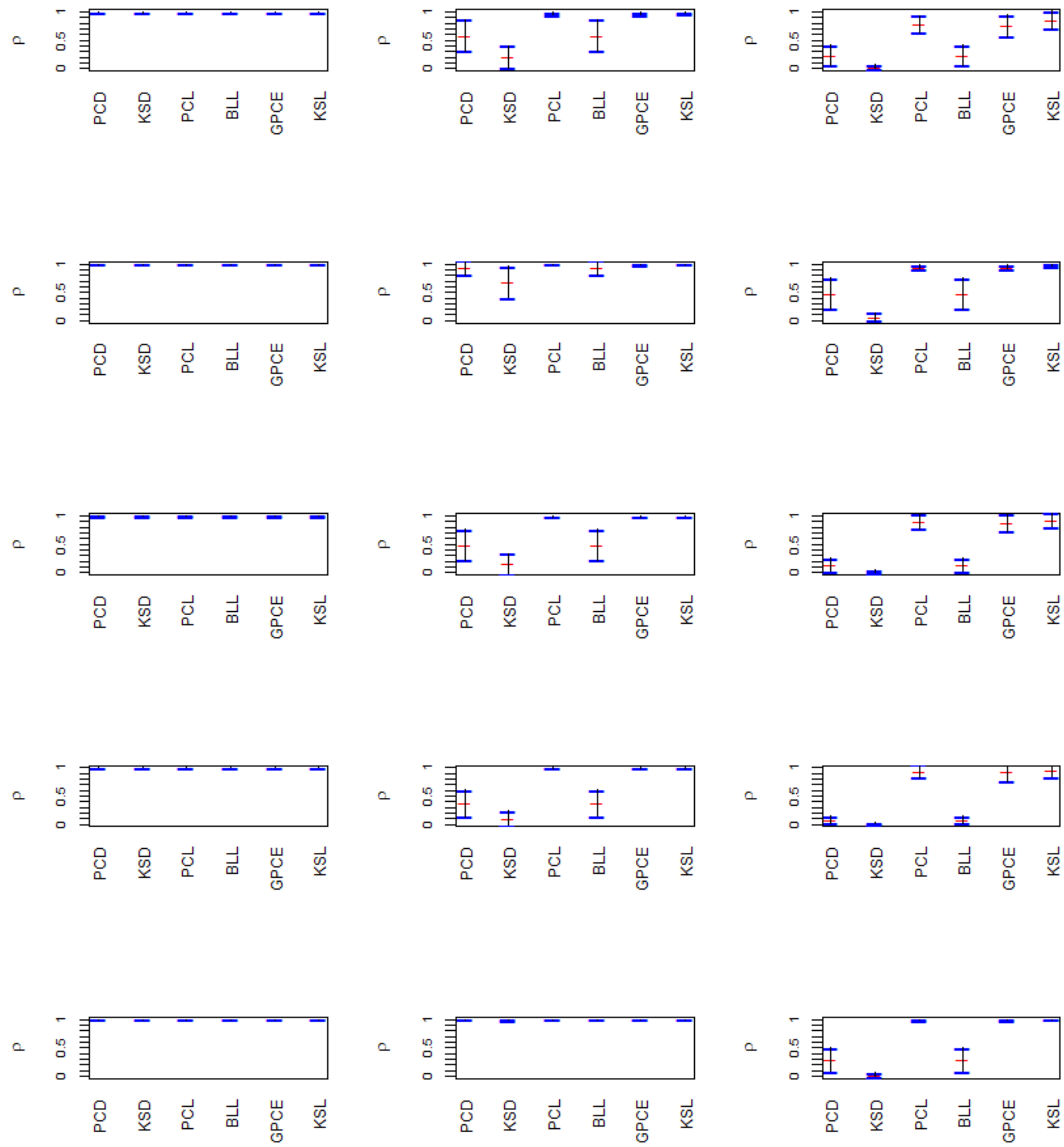

Figure 3.2: Box-plots of the sample correlations between $\{\hat{\delta}_j' \hat{F}_t^{PCD}\}$, $\{\hat{\delta}_j' \hat{F}_t^{KSD}\}$, $\{\hat{\delta}_j' \hat{F}_t^{PCL}\}$, $\{\hat{\delta}_j' \hat{F}_t^{BLL}\}$, $\{\hat{\delta}_j' \hat{F}_t^{GPCE}\}$ and $\{\hat{\delta}_j' \hat{F}_t^{KSL}\}$ with $\{F_t\}$. We consider the M1 model with homoscedasticity in idiosyncratic errors with $\gamma = -0.8$. First row indicates $N = 12$ and $T = 50$; second row $N = 12$ and $T = 100$; third row $N = 50$ and $T = 100$; fourth row $N = 200$ and $T = 100$ and fifth row $N = 200$ and $T = 500$. The first column plots $\sigma_a^2 = 0.1$, second column $\sigma_a^2 = 1$, and third column $\sigma_a^2 = 10$.

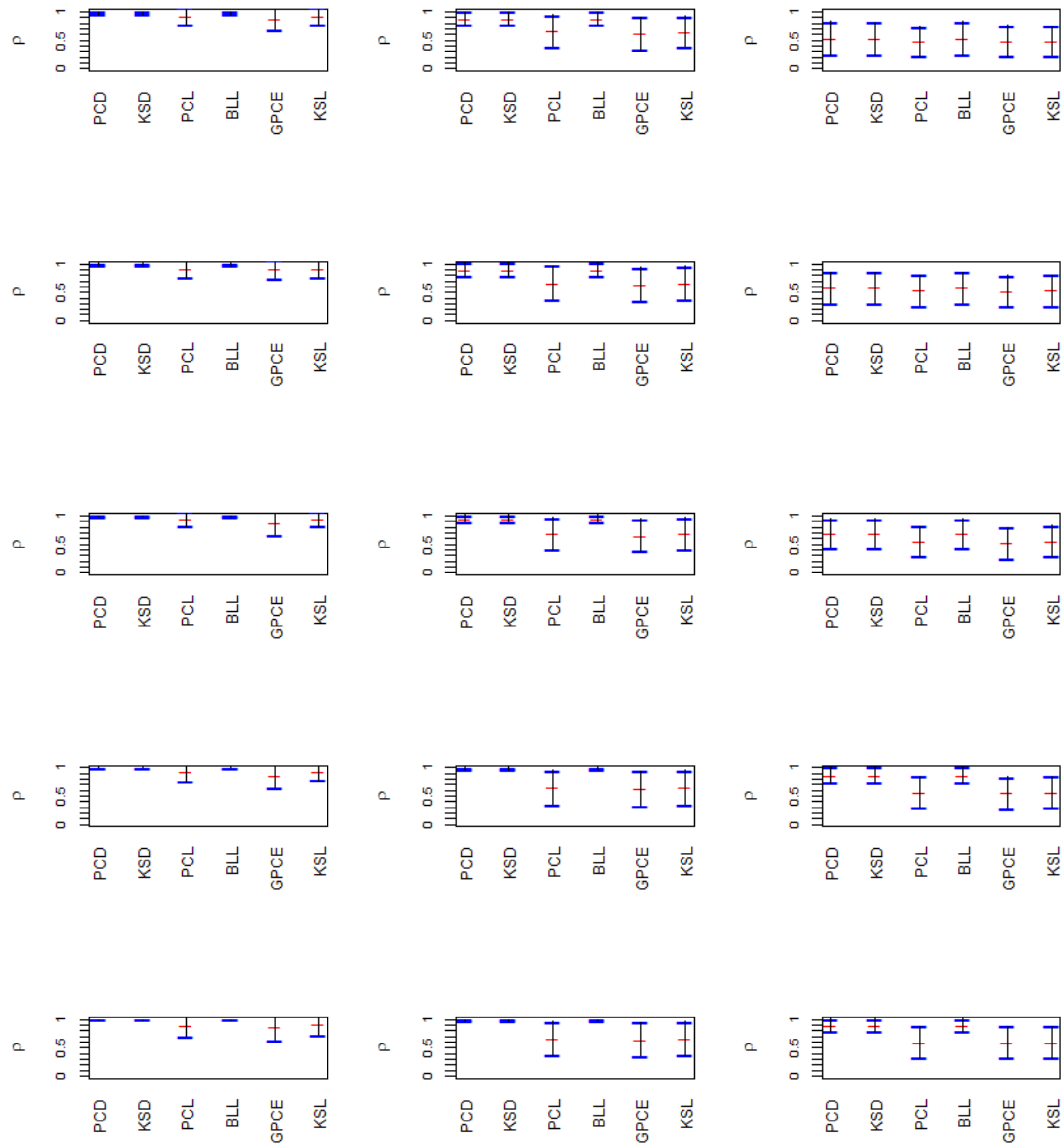

Figure 3.3: Box-plots of the sample correlations between $\{\hat{\delta}_j'\hat{F}_t^{PCD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{KSD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{PCL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{BLL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{GPCE}\}$ and $\{\hat{\delta}_j'\hat{F}_t^{KSL}\}$ with $\{F_t\}$. We consider the M1 model with homoscedasticity in idiosyncratic errors with $\gamma = 1$. First row indicates $N = 12$ and $T = 50$; second row $N = 12$ and $T = 100$; third row $N = 50$ and $T = 100$; fourth row $N = 200$ and $T = 100$ and fifth row $N = 200$ and $T = 500$. First column plots $\sigma_a^2 = 0.1$, second column $\sigma_a^2 = 1$ and third column $\sigma_a^2 = 10$.

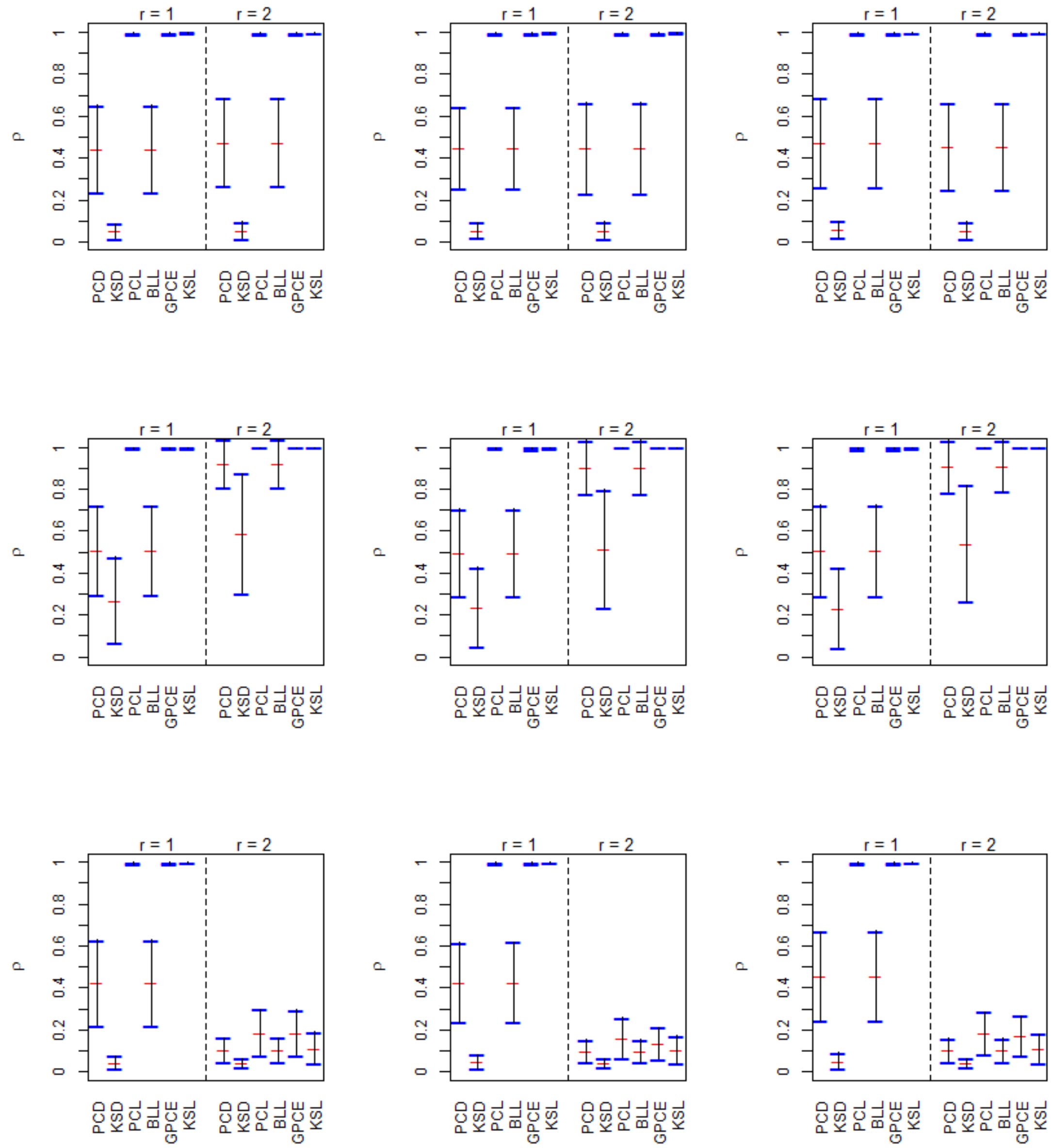

Figure 3.4: Box-plots of the sample correlations between $\{\hat{\delta}_j'\hat{F}_t^{PCD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{KSD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{PCL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{BLL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{GPCE}\}$ and $\{\hat{\delta}_j'\hat{F}_t^{KSL}\}$ with $\{F_t\}$. We consider the $N = 50$ and $T = 100$ with $\sigma_a^2 = 10$ and $\gamma = -0.8$. First row plots M2 model, second row M3 model and third row M4 model. First column indicates the homoscedasticity, second column heteroscedasticity and third column cross-sectionally correlated idiosyncratic errors.

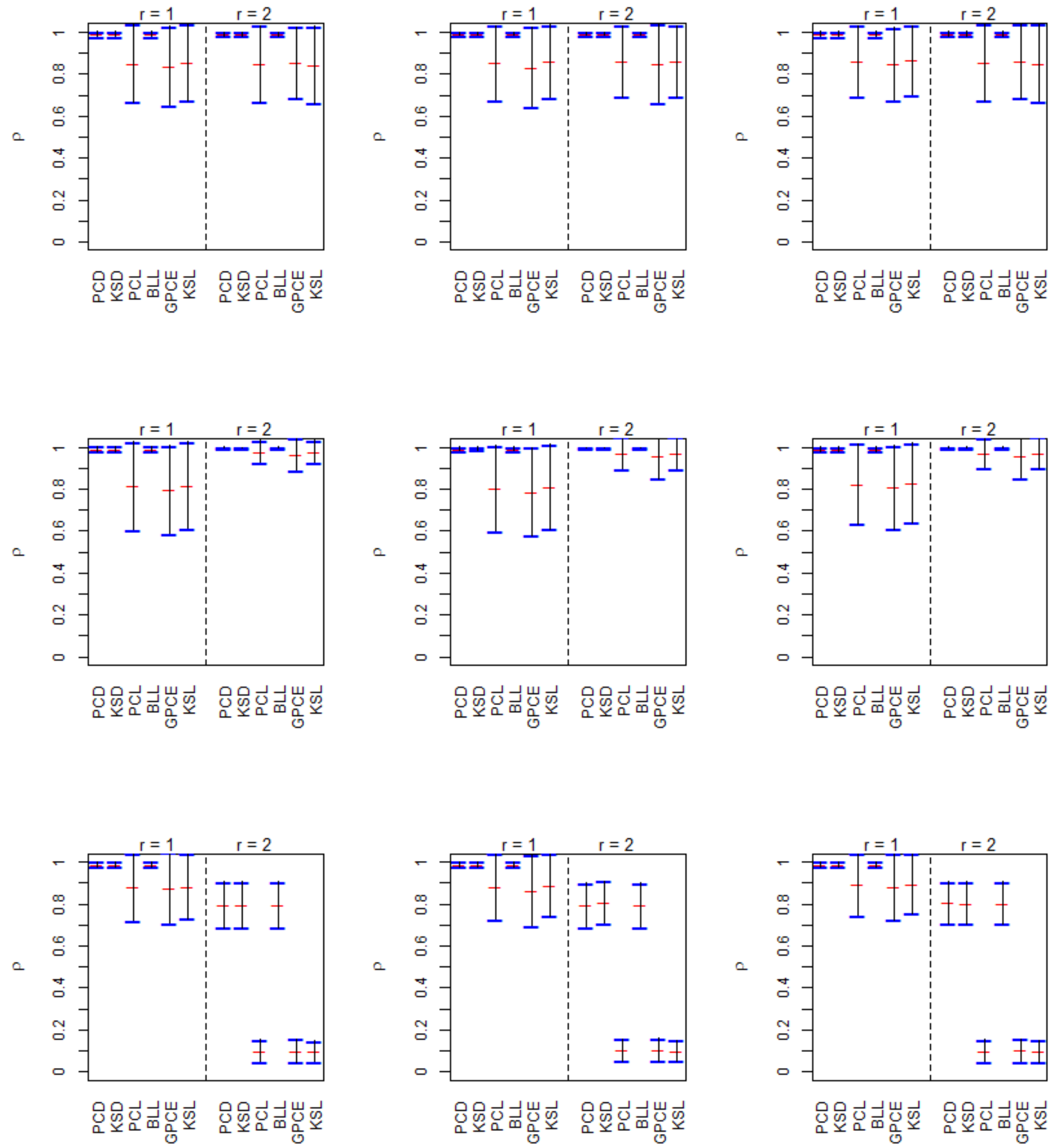

Figure 3.5: Box-plots of the sample correlations between $\{\hat{\delta}_j'\hat{F}_t^{PCD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{KSD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{PCL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{BLL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{GPCE}\}$ and $\{\hat{\delta}_j'\hat{F}_t^{KSL}\}$ with $\{F_t\}$. We consider the $N = 50$ and $T = 100$ with $\sigma_a^2 = 1$ and $\gamma = 1$. First row plots M2 model, second row M3 model and third row M4 model. First column indicates the homoscedasticity, second column heteroscedasticity and third column cross-sectionally correlated idiosyncratic errors.

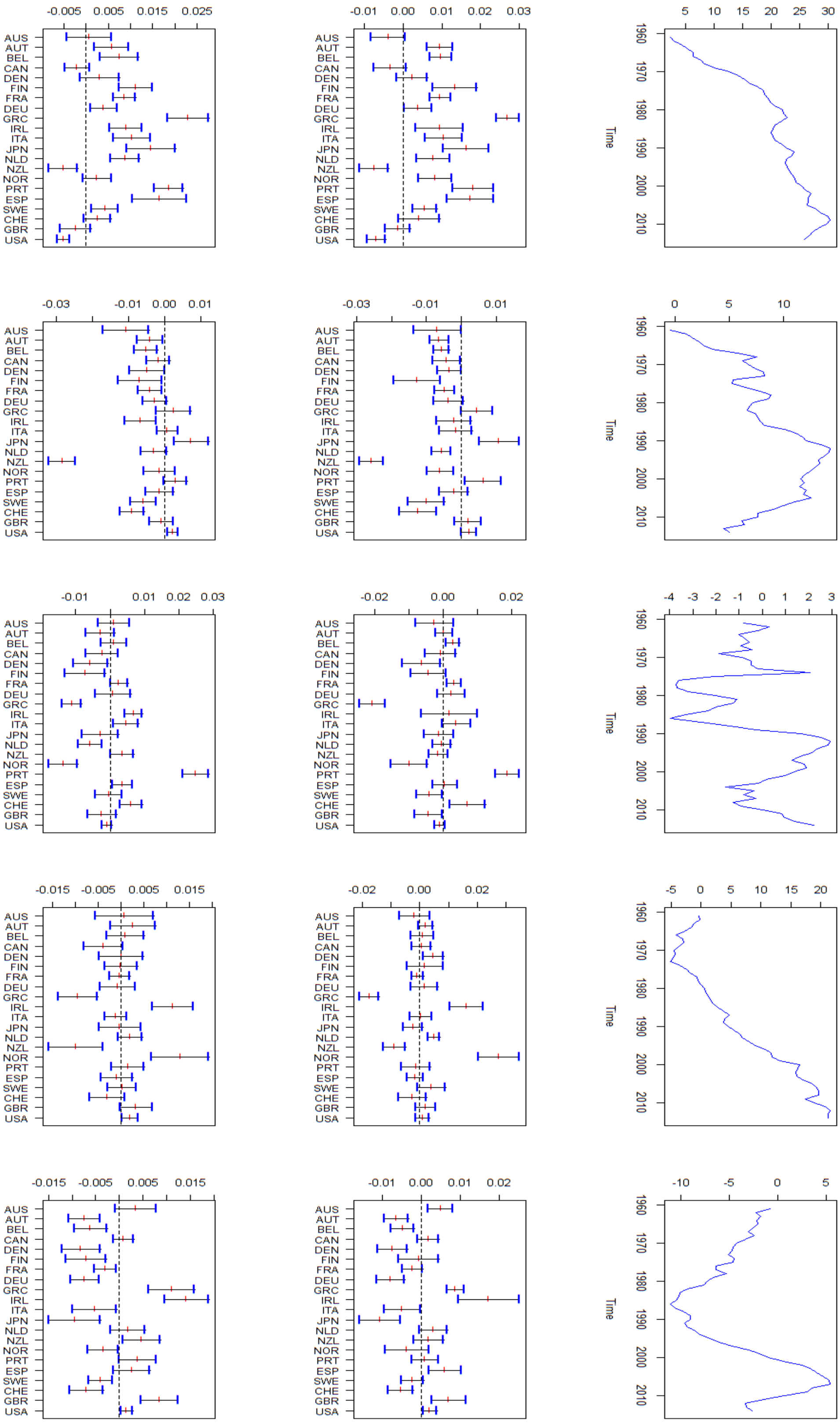

Figure 3.6: Top panel $\hat{F}_{jt}$, middle panel $\hat{p}_{ij}$ for GDP ($i = 1, \ldots 21$) and bottom panel $\hat{p}_{ij}$ for C ($i = 22, \ldots 42$) for $j = 1, \ldots, 5$.. We plot the corresponding 95% confidence intervals for $\hat{P}$ (middle and bottom panels). All estimations are obtained using PCD.

Chapter 4

Summary and Future Research

This dissertation analyzes topics about large Non-stationary Dynamic Factor Models (DFMs). In Chapter 2, we focus on studying the effects of stationary univariate transformations when determining the number of factors. We consider the Bai and Ng (2002) information criteria, the edge distribution given by Onatski (2010) and the ratios of adjacent eigenvalues of the sample covariance matrix proposed by Ahn and Horenstein (2013). These are three of the most popular procedures to determine the number of factors frequently implemented in empirical analysis. Bai and Ng (2002) propose minimizing a penalized likelihood or log sum of squares, where the penalty functions increase linearly in the number of factors. These criteria is the most popular in the context of PC factor extraction and are based on a consistent estimation of the factor space. Onatski (2010) proposes an alternative procedure to select the number of factors and shows that it outperforms the criteria proposed by Bai and Ng (2002) when the proportion of the variance due to the idiosyncratic component is large relative to the variance atributted to the common component. Furthermore, this procedure performs well when the noises are substantially correlated. Ahn and Horenstein (2013) carry out two ratios of adjacent eigenvalues of the sample covariance matrix under the assumption that the eigenvalues r largest eigenvalues of $\hat{\Sigma}_Y$ grow unbounded as N increases, while the other eigenvalues remain bounded. For a DFM with $r = 1$ and uncorrelated and homocesdastic idiosyncratic errors, we first analytically derive the eigenvalues of the covariance matrix and show how the procedures considered are affected when the system is differenced. We implement a Monte Carlo analysis considering DGP which can often occur in empirical applications. We conclude that that the Ahn and Horenstein (2013) procedure performs well when $r = 1$ even if when the sample size is small and under several dynamics and structure dependence in the idiosyncratic noises. When $r = 2$, the performance from these approaches deteriorates, being the Onatski (2010) procedure the one that outperforms. The Bai and Ng (2002) information criteria perform well when the common component variance is large with respect to the variance of the idiosyncratic term. In the empirical application, when we use first-differenced data, all approaches detect one common factor. When differencing the system, Bai and Ng (2002) information criteria tend to $r_{\max}$, Onatski (2010) procedure identifies 3 common factors and the ratios of adjacent eigenvalues of Ahn and Horenstein (2013) remain determining one common factor. Considering three factors, the first is associated to the common inflation of the euro area, while the second and the third are directly determined by Ireland and Greece respectively. The main conclusion is that, if the idiosyncratic error is stationary (cointegrated DFM), with autoregressive parameters smaller than 0.5 while the common factor is non-stationary, then overdifferencing the idiosyncratic components may introduce distortions on the determination of the number of factors. This occurs given that the relation between the variances of the common and idiosyncratic components is modified with the variances from the differenced common factor being smaller and the variances from the differenced idiosyncratic term being larger.

In Chapter 3, we analyze the finite sample performance of the approaches existing in literature to extract consistently non-stationary common factors. Additionally, we extend the Doz et al. (2011) procedure to extract common factors to the non-stationary case, which are based on PC and Kalman smoothing. Furthermore, we estimate factor loadings and non-stationary common factors for a set of OECD countries, with the goal of disentangling whether risk sharing is a short or long-run phenomenon. We take into account the Bai and Ng (2004) procedure which extracts the non-stationary common factors applying PC to first-differeced data, using the "differencing and recumulating" estimator regardless of whether the idiosyncratic noises are non-stationary. In a similar way, we consider the hybrid procedure of Doz et al. (2011). Furthermore, we analyze procedures which extract non-statioanary common factors using data in levels. In this context, Bai (2004) proposes to use PC in data in levels when the idiosyncratic errors are stationary. Barigozzi et al. (2016) project the original observations in the factor loadings estimated as in Bai and Ng (2004) allowing non-stationary idiosyncratic errors. Choi (2016) uses GPCE under similar assumptions to Bai (2004). Finally, we consider the hybrid technique for data in levels, allowing non-stationary common factors in the DFM. In the Monte Carlo analysis, we study the finite sample performance of the approaches considered. The main goal is to determine whether we can obtain close estimations of the simulated common factors considering several DGP that can be encountered in empirical applications. The main conclusion is that, the procedures that perform with data in levels, tend to recover better the simulated common factors when the idiosyncratic component is stationary. On the other hand, the procedures that use first-differenced data and the Barigozzi et al. (2016) approach have an outperformance. In the empirical analysis, we apply the Onatski (2010) procedure to determine the number of common factors. We obtain $\hat{r} = 5$ and extract the common factors using the procedure given by Bai and Ng (2004) to extract non-stationary common factors. To disentangle the non-stationarity in the system, we apply PANIC. We conclude that the idiosyncratic errors are non-stationary and at least four common factors are non-stationary. The fact that the idiosyncratic errors are non-stationary can be attributed to persistent movements that are generated internal and not shared among countries or due to interactions with third countries, as it might happen with the U.S. and Mexico. The non-stationary factor model points out the lack of risk sharing both in the short and long-run.

Further research should focus on extending the finite sample performance of the procedures to determine the number of factors in the context of dynamic factors, see for example Amengual and Watson (2007), Bai and Ng (2007), Hallin and Liska (2007), Jacobs and Otter (2008) and Breitung and Pigorsch (2013). It is well known that in presence of break points the null hypothesis tends to be rejected by the usual ADF and PP tests. Hence, we can implement, for example, the Busetti and Harvey (2001) unit root test in order to improve the performance of the 2SKS procedure to extract stationary or non-stationary common factors. Also, it is interesting to study the statistical properties of the hybrid procedure when the common factors are non-stationary, which is proposed in this dissertation, initially considering stationary idiosyncratic errors. Furthermore, we can extend the algorithm allowing non-stationary idiosyncratic errors following the ideas given by Barigozzi and Luciani (2017).

Bibliography

Ahn, S. and Horenstein, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics and Probability Letters*, 80(1):1806–1813.

Alonso, A. M., Bastos, G., and García-Martos, C. (2016). Electricity price forecasting by averaging dynamic factor models. *Energies*, 9(8):600;doi:10.3390/en9080600.

Alonso, A. M., García-Martos, C., and Sánchez, M. J. (2011). Seasonal dynamic factor analysis and bootstrap inference: Application to electricity market forecasting. *Technometrics*, 53(2):137–151.

Altissimo, F., Mojon, B., and Zaffaroni, P. (2009). Can aggregation explain the persistence of inflation?. *Journal of Monetary Economics*, 56:231–241.

Alvarez, R., Camacho, M., and Perez-Quiros, G. (2016). Aggregate versus disaggregate information in dynamic factor models. *International Journal of Forecasting*, 32:680–694.

Amengual, D. and Watson, M. W. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics*, 25(1):91–96.

Andersen, T. G., Fusari, N., and Todorov, V. (2015). The risk premia embedded in index options. *Journal of Financial Economics*, 117(3):558–584.

Artis, M., Marcelino, M., and Proietti, T. (2004). Dating business cycle: a methodological contribution with an application to the euro area. *Oxford Bulletin of Economics and Statistics*, 66(4):537–565.

Artis, M. J. and Hoffmann, M. (2008). Financial globalization, international business cycles and consumption risk sharing. *Scandinavian Journal of Economics*, 110(3):447–471.

Artis, M. J. and Hoffmann, M. (2012). The home bias, capital income flows and improved long-term consumption risk sharing between industrialized countries. *International Finance*, 14(3):481–505.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, 122(1):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). Confidence intervals for diffusion index forecast and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25(1):52–60.

Bai, J. and Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3(2):89–163.

Bai, J. and Ng, S. (2010). Panel unit root tests with cross-section dependence: a further investigation. *Econometric Theory*, 26:1088–1114.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29.

Bai, J. and Wang, P. (2014). Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics*, 178(2):794–804.

Bai, J. and Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8:53–80.

Banerjee, A., Marcellino, M., and Masten, I. (2014a). Forecasting with factor-augmented error correction models. *International Journal of Forecasting*, 30(3):589–612.

Banerjee, A., Marcellino, M., and Masten, I. (2014b). Structural factor error correction models: cointegration in large-scale structural FAVAR models. *Working Paper*.

Barhoumi, K., Darné, O., and Ferrara, L. (2013). Testing the number of factors: an empirical assessment for forecasting purposes. *Oxford Bulleting of Economics and Statistics*, 75(1):64–79.

Barigozzi, M., Lippi, M., and Luciani, M. (2016). Non-Stationary Dynamic Factor Models for Large Datasets. *Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C.*, 024.

Barigozzi, M., Lippi, M., and Luciani, M. (2017). Dynamic factor models, cointegration, and error correction mechanisms. *Working Paper, arXiv:1510.02399v3*.

Barigozzi, M. and Luciani, M. (2017). Common factors, trends, and cycles in large datasets, manuscript.

Becker, S. and Hoffmann, M. (2006). Intra- and international risk-sharing in the short run and the long run. *European Economic Review*, 50(6):777–806.

Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(1):526–546.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120(1):381–422.

Beyer, A., Doornik, J., and Hendry, D. (2001). Constructing historical euro-zone data. *The Economic Journal*, 111(469):102–121.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Box, G. and Tiao, G. (1977). A canonical analysis of multiple time series. *Biometrika*, 64:355–365.

Bräuning, F. and Koopman, S. J. (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting*, 30(3):572–584.

Breitung, J. and Choi, I. (2013). Factor models, in Hashimzade, N. and Thorthon, M.A. (eds.). *Handbook of Research Methods and Applications in Empirical Macroeconomics*, United Kingdom: Edward Elgar Publishing.

Breitung, J. and Eickmeier, S. (2006). Dynamic factor models, in Hübler, O. and J. Frohn (eds.). *Modern Econometric Analysis*, Berlin: Springer.

Breitung, J. and Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163:71–84.

Breitung, J. and Pigorsch, U. (2013). A canonical correlation approach for selecting the number of dynamic factors. *Oxford Bulletin of Economics and Statistics*, 75(1):23–36.

Buch, C. M., Eickmeier, S., and Prieto, E. (2014). Macroeconomic factors and microlevel bank behavior. *Journal of Money, Credit and Banking*, 46(4):715–751.

Burridge, P. and Wallis, K. F. (1985). Calculationg the variance of seasonally time series models. *Journal of American Statistical Association*, 80:541–552.

Busetti, F. and Harvey, A. (2001). Testing for the presence of a random walk in series with structural breaks. *Journal of Multivariate Time Series*, 22(2):127–150.

Byrne, J. and Fiess, N. (2016). International capital flows to emerging markets: National and global determinants. *Journal of International Money and Finance*, 61:82–100.

Camacho, M., Lovcha, Y., and Perez-Quiros, G. (2015). Can we use seasonally adjusted variables in dynamic factor models? *Studies in Nonlinear Dynamics and Econometrics*, 19(3):377–391.

Caner, M. and Han, X. (2014). Selecting the correct number of factors in approximated factor models: the large panel case with group bridge estimators. *Journal of Business & Economic Statistics*, 32(3):359–374.

Canova, F. (1998). Detrending and business cycle facts. *Journal of Monteray Economics*, 41:475–512.

Choi, I. (2016). Efficient estimation of nonstationary factor models. *Journal of Statistical Planning and Inference*, (forthcoming).

Choi, I. and Hwang, S. (2012). Forecasting Korean inflation. *Applied Korean Economics*, 14(3):133–169.

Connor, G. and Korajczyk, R. A. (2010). *Encyclopedia of Quantitative Finance: Factor Models of Asset Returns*. Wiley, Rama Cont, ed., Chicester.

Corona, F., Poncela, P., and Ruiz, E. (2016). Determining the number of factors after stationary univariate transformations. *Empirical Economics*, (forthcoming).

Costantini, M. (2013). Forecasting the industrial production using alternative factor models and business survey data. *Journal of Applied Statistics*, 40(10):2275–2289.

Del Negro, M. (2002). Assymetric shocks among U.S. states. *Journal of International Economics*, 56:273–297.

Delle Monache, D., Petrella, I., and Venditti, F. (2016). Common faith or parting ways? A time varying parameters factor analysis of euro-area inflation in dynamic factor models, in Hillebrand, E. and Koopman, S.J. (eds.). *Advances in Econometrics*, 35:539–565.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Doz, C., Giannone, D., and Reichlin, L. (2012). A quasi maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics*, 94(4):1014–1024.

Eichler, M., Motta, G., and von Sachs, R. (2011). Fitting dynamic factor models to non-stationary time series. *Journal of Econometrics*, 163(1):51–70.

Eickmeier, S. (2009). Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model. *Journal of Applied Econometrics*, 24(6):933–959.

Engel, C., Mark, N., and West, K. (2015). Factor model forecasts of exchange rates. *Econometric Reviews*, 34(1-2):32–55.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(1):251–276.

Escribano, A. and Peña, D. (1994). Cointegration and common factors. *Journal of Time Series Analysis*, 15(6):577–586.

Everaert, G., Heylen, F., and Schoonackers, R. (2015). Fiscal policy and TFP in the OECD: measuring direct and indirect effects. *Empirical Economics*, 49:605–640.

Favero, C., Marcellino, M., and Neglia, F. (2005). Principal components at work: the empirical analysis of monetary policy with large datasets. *Journal of Applied Econometrics*, 20(1):603–620.

Forni, M., Gambetti, L., and Sala, L. (2014). No news in business cycles. *Economic Journal*, 124:1168–1191.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models versus structural VARs. *Econometric Theory*, 25:1319–1347.

Forni, M. and Reichlin, L. (1998). Let's get real: a factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies*, 65(3):453–473.

García-Martos, C., Rodríguez, J., and Sánchez, M. J. (2011). Forecasting electricity prices and their volatilities using Unobserved Components. *Energy Economics*, 33(6):1227–1239.

Geweke, J. (1977). The dynamic factor analysis of economic time series, in Aigner, D.J. and Goldberger, A.S. (eds.). *Latent Variables in Socio-Economic Models*, Amsterdam: North-Holland.

Giannone, D. and Reichlin, L. (2006). Does information help recovering structural shocks from past observations? *Journal of the European Economic Association*, 4(2-3):455–465.

Gonzalo, J. and Granger, C. W. J. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*, 13(1):27–35.

Greenway-McGrevy, R., Mark, N., Sul, D., and Wu, J.-L. (2016). Identifying exchange rate common factors. *Manuscript*.

Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Associaton*, 102:603–617.

Han, X. and Caner, M. (2016). Determining the number of factors with potentially strong within-block correlation in error terms. *Manuscript*.

Harding, M. (2013). Estimating the number of factors in large dimensional factor models. *Manuscript*.

Harvey, A. (1989). *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Harvey, A. and Phillips, G. (1979). Maximum Likelihood Estimation of Regression Models With Autoregressive-Moving Averages Disturbances. *Biometrika*, 152:49–58.

Jacobs, J. P. A. M. and Otter, P. W. (2008). Determining the number of factors and lag order in dynamic factor models: a minimum entropy approach. *Econometric Reviews*, 27(4-6):385–397.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(1):231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegrating vectors in gaussian vector autoregressive models. *Econometrica*, 59(1):1551–1580.

Jungbacker, B. and Koopman, S. J. (2015). Likelihood-based dynamic factor analysis for measurement and forecasting. *Econometric Journal*, 18:1–21.

Kalemli-Ozcan, S., Luttini, E., and Sensen, B. (2014). Debt crises and risk-sharing: The role of markets versus sovereigns. *Scandinavian Journal of Economics*, 116(1):253–276.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics*, 28(3):397–409.

Koopman, S. (1997). Exact Initial Kalman Filtering and smoothing for nonstationary time series models. *Journal of American Statistical Association*, 92(440):1630–1638.

Kunst, R. and Neusser, K. (1997). Cointegration in a macroeconomic system. *Journal of Applied Econometrics*, 5(1):351–365.

Lahiri, K., Monokroussos, G., and Yongchen, Z. (2015). Forecasting consumption: the role of consumer confidence in real time with many predictors. *Journal of Applied Econometrics*. (in press).

Lahiri, K. and Sheng, X. (2010). Measuring Forecast Uncertainty by Disagreement: The Missing Link. *Journal of Applied Econometrics*, 25:514–538.

Lahiri, K. and Yao, W. (2004). A dynamic factor model of the coincident indicators for the US transportation sector. *Applied Economic Letters*, 11(10):595–600.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors using high-dimensional time series. *Biometrika*, 98(4):901–918.

Leibrecht, M. and Scharler, J. (2008). Reconsidering consumption risk sharing among oecd countries: Some evidence based on panel cointegration. *Open Economies Review*, 19(4):493–505.

Luo, R., Wang, H., and Tsai, C. L. (2009). Contour projected dimension reduction. *The Annals of Statistics*, 37:3743–3778.

Marcellino, M., Stock, J. H., and Watson, M. W. (2003). Macroeconomic forecasting in the euro area: country specific versus euro wide information. *European Economic Review*, 47(1):1–18.

Mestekemper, T., Kauermann, G., and Smith, M. S. (2013). A comparison of periodic autoregressive and dynamic factor models in intraday energy demand forecasting. *International Journal of Forecasting*, 29(1):1–12.

Moench, E., Ng, S., and Potter, S. (2013). Dynamic hierarchical factor models. *Review of Economics and Statistics*, 95(5):1811–1817.

Moon, H. R. and Perron, B. (2004). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics*, 122(1):81–126.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168:244–258.

Onatski, A. (2015). Asymptotic analysis of the squared estimation error in misspecified factor models. *Journal of Econometrics*, 186(2):388–406.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95(1):365–379.

Panopoulou, E. and Vrontos, S. (2015). Hedge fund return predictability: to combine forecast or combine information? *Journal of Banking and Finance*, 56:103–122.

Peña, D. and Poncela, P. (2006). Non-stationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, 136(1):237–257.

Pierucci, E. and Ventura, L. (2010). Risk sharing: a long run issue? *Open Economies Review*, 21(5):705–730.

Pinheiro, M., Rua, A., and Dias, F. (2013). Dynamic factor models with jagged edge panel data: taking on board the dynamics of the idiosyncratic components. *Oxford Bulletin of Economics and Statistics*, 75(1):80–102.

Poncela, P. and Ruiz, E. (2016). Small versus big data factor extraction in dynamic factor models: An empirical assessment in dynamic factor models, in Hillebrand, E. and Koopman, S.J. (eds.). *Advances in Econometrics*, 35:401–434.

Poncela, P., Senra, E., and Sierra, L. P. (2014). Common dynamics of nonenergy commodity prices and their relation to uncertainty. *Applied Economics*, 46(30):3724–3735.

Quah, D. and Sargent, T. (1993). A dynamic index model for large cross sections, in. *Business Cycles, Indicators and Forecasting*, University of Chicago Press.

Reis, R. and Watson, M. W. (2010). Relative goods' prices, pure inflation, and the Phillips correlation. *American Economic Journal: Macroeconomics*, 2(3):128–157.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priory economic theory, in Sims, C.A. (ed.). *New Methods in Business Cycle Research*, Minneapolis: Federal Reserve Bank of Minneapolis.

Schumacher, C. (2005). Forecasting german GDP using alternative factor models based on large datasets. *Bundesbank*, pages Discussion Paper 24–2005.

Seong, B., Ahn, A., and Zadrozny, P. (2013). Estimation of vector error correction models with mixed-frequency data. *Journal of Time Series Analysis*, 34:194–205.

Sims, C. (2012). Comments and discussion: disentangling the channels of the 2007-2009 recession. *Brookings Papers on Economic Activity*, pages 141–148.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, 83(1):1097–1107.

Stock, J. H. and Watson, M. W. (1989). New indexes of coincident and leading economic indicators, in blanchard, o.j. and s- fischer (eds.). *NBER Macroeconomics Annual 1989*, MIT Press.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(1):1169–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(1):147–163.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for VAR analysis. *NBER*, Working Paper 11467.

Stock, J. H. and Watson, M. W. (2011). Dynamic factor models, in Clements, M.P and Hendry, D.F. (eds.). *Oxford Handbook of Economic Forecasting*, Oxford: Oxford University Press.

Stock, J. H. and Watson, M. W. (2012a). Disentangling the channels of the 2007-2009 recession. *Brookings Papers on Economic Activity*, Spring 2012:81–130.

Stock, J. H. and Watson, M. W. (2012b). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, 30(4):481–493.

Vahid, F. and Engle, R. F. (1993). Common trends and common cycles. *Journal of Applied Econometrics*, 8(1):341–360.

Wang, H. (2012). Factor profiled sure independence screening. *Biometrika*, 99:15–28.

Zhang, R., Robinson, P., and Yao, Q. (2016). Identifying cointegration by eigen analysis. *Biometrika*.

Appendix A

Appendix to Chapter 2

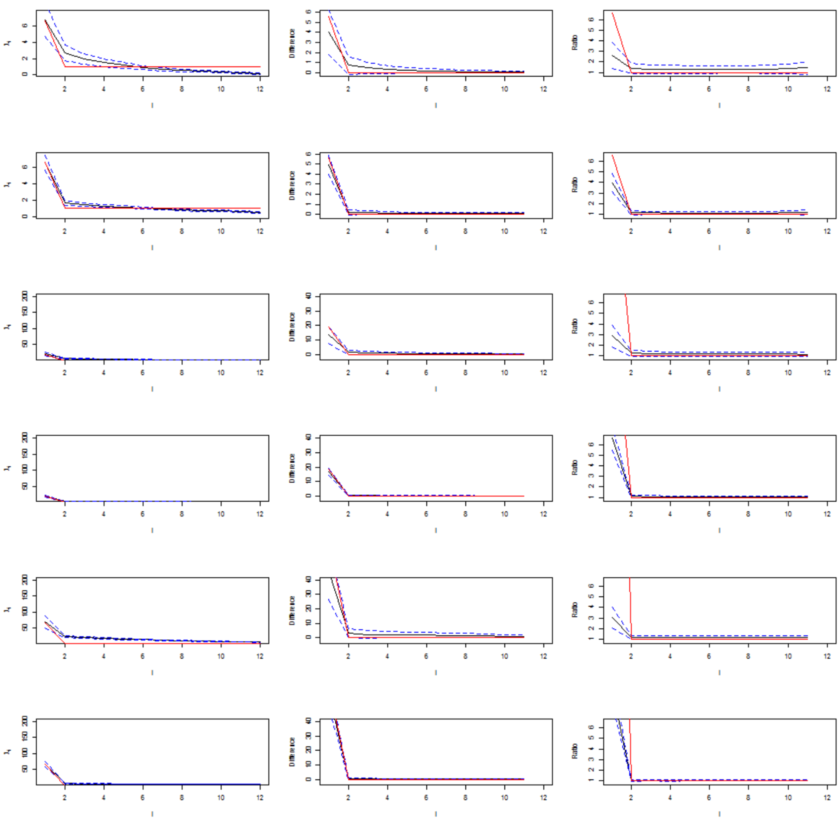

Figure A.1: Eigenvalues of DFM with $r = 1$, $\phi = 1$ and $\sigma_\eta^2 = 1$ when the idiosyncratic noises are AR(1) process with $\gamma = -0.8$ and $\sigma_a^2 = 0.1$. The first column plots the eigenvalues while the second and third column plot their differences and ratios respectively. The population eigenvalues are plotted in red, the Monte Carlo averages in black and the corresponding 95% intervals in blue. First row $N = 12, T = 100$; second row $N = 12, T = 500$; third row $N = 50, T = 100$; fourth row $N = 50, T = 500$; fifth row $N = 200, T = 100$ and sixth row $N = 200, T = 500$.

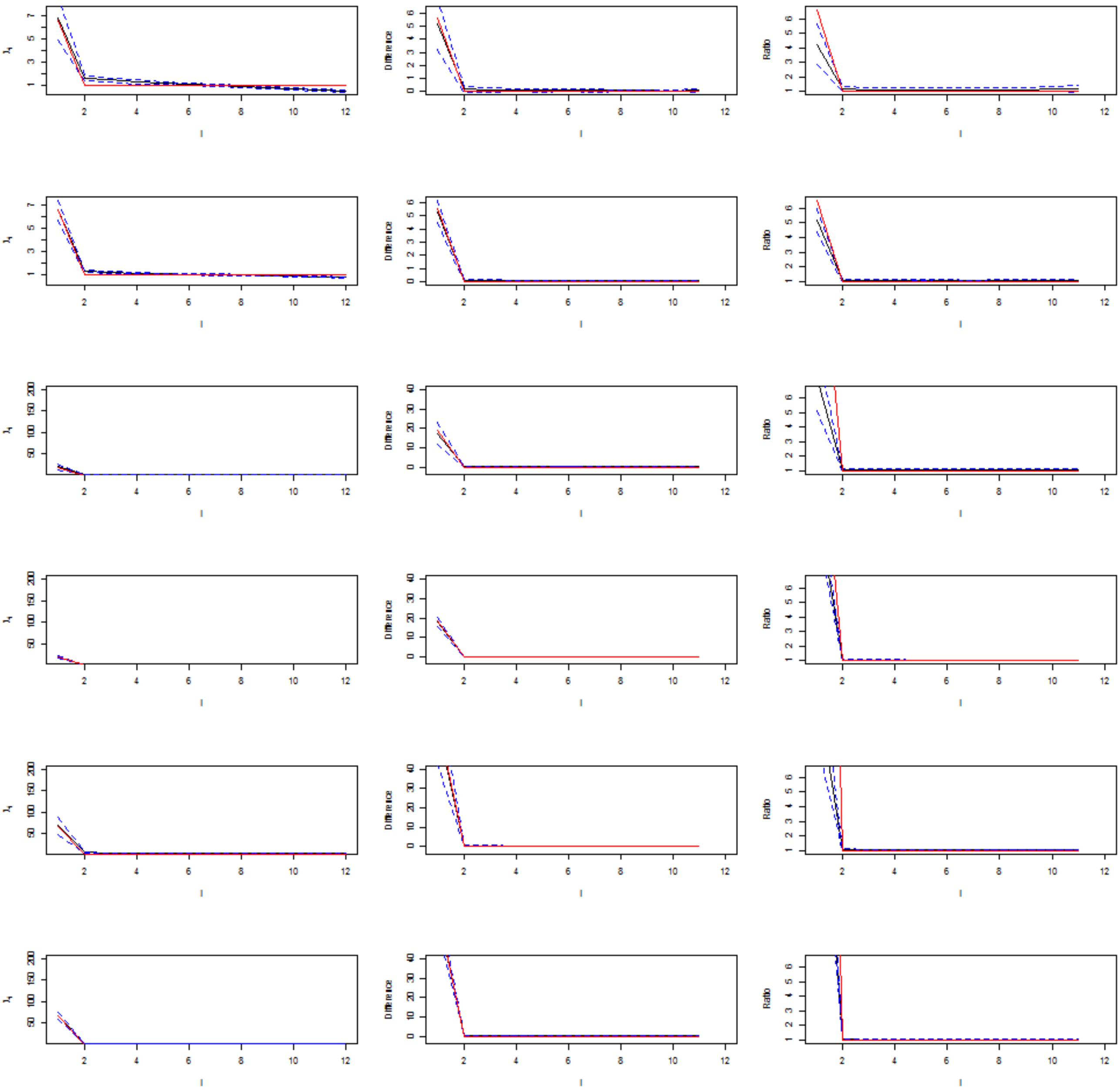

Figure A.2: Eigenvalues of DFM with $r = 1$, $\phi = 1$ and $\sigma_\eta^2 = 1$ when the idiosyncratic noises are AR(1) process with $\gamma = 1$ and $\sigma_a^2 = 1$. The first column plots the eigenvalues while the second and third column plot their differences and ratios respectively. The population eigenvalues are plotted in red, the Monte Carlo averages in black and the corresponding 95% intervals in blue. First row $N = 12, T = 100$; second row $N = 12, T = 500$; third row $N = 50, T = 100$; fourth row $N = 50, T = 500$; fifth row $N = 200, T = 100$ and sixth row $N = 200, T = 500$.

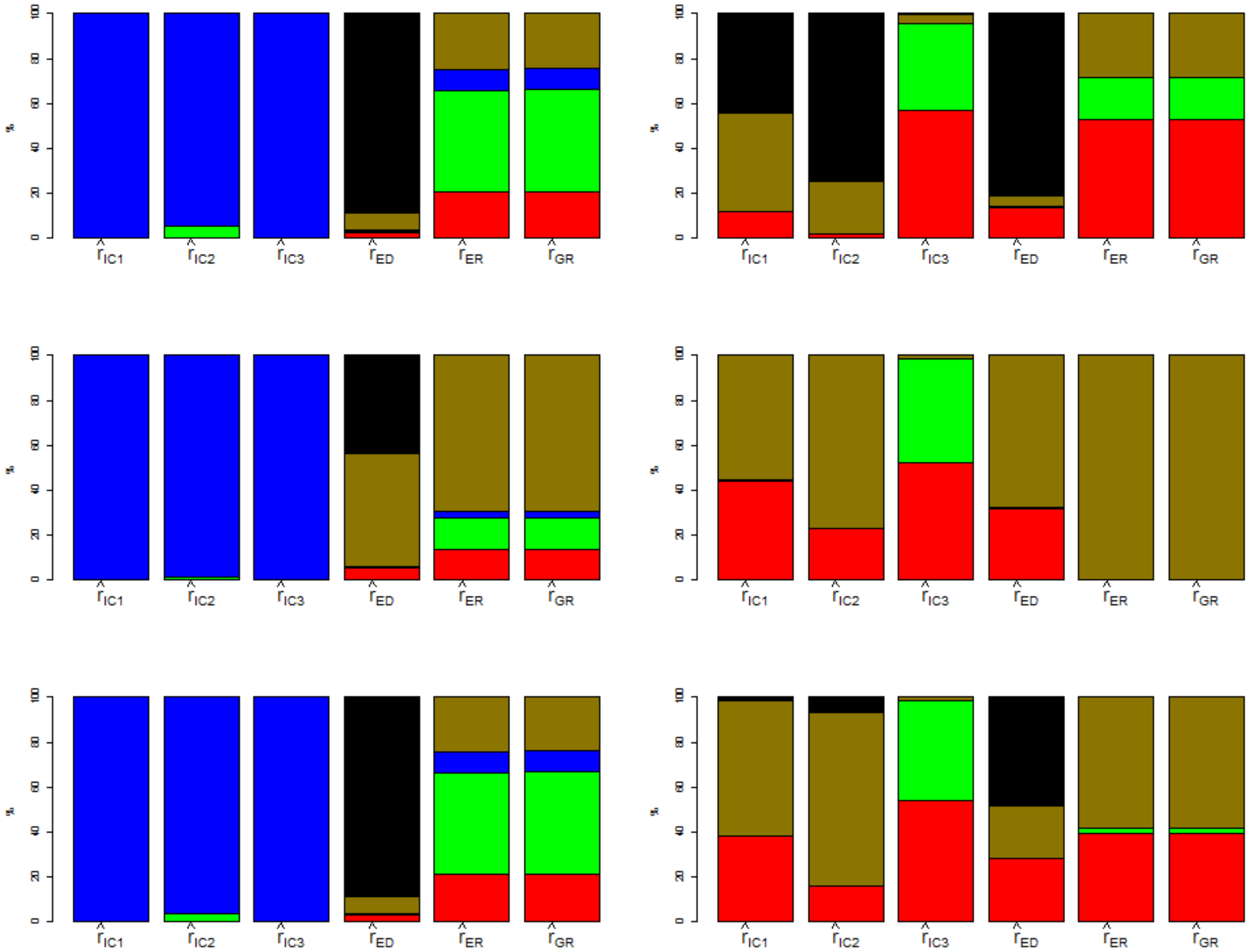

Figure A.3: Percentage of $\hat{r} = r_{\max}$ (blue), $r_{\max} > \hat{r} > r$ (green), $\hat{r} = 2$ (red), $\hat{r} = 1$ (gold) and $\hat{r} = 0$ (black) in a DFM with $r = 2$, $\gamma = -0.8$ and $\sigma_a^2 = 1$. System dimensions $N = 12$, $T = 100$ (first column); $N = 200$, $T = 500$ (second column). The factors are two random walks with variance $\sigma_\eta^2 = 1$ (first row); two random walks with variances $\sigma_{\eta_1}^2 = 1$ and $\sigma_{\eta_2}^2 = 5$ (second row) and a random walk with variance $\sigma_{\eta_1}^2 = 1$ and a stationary factor with $\sigma_{\eta_2}^2 = 1$.

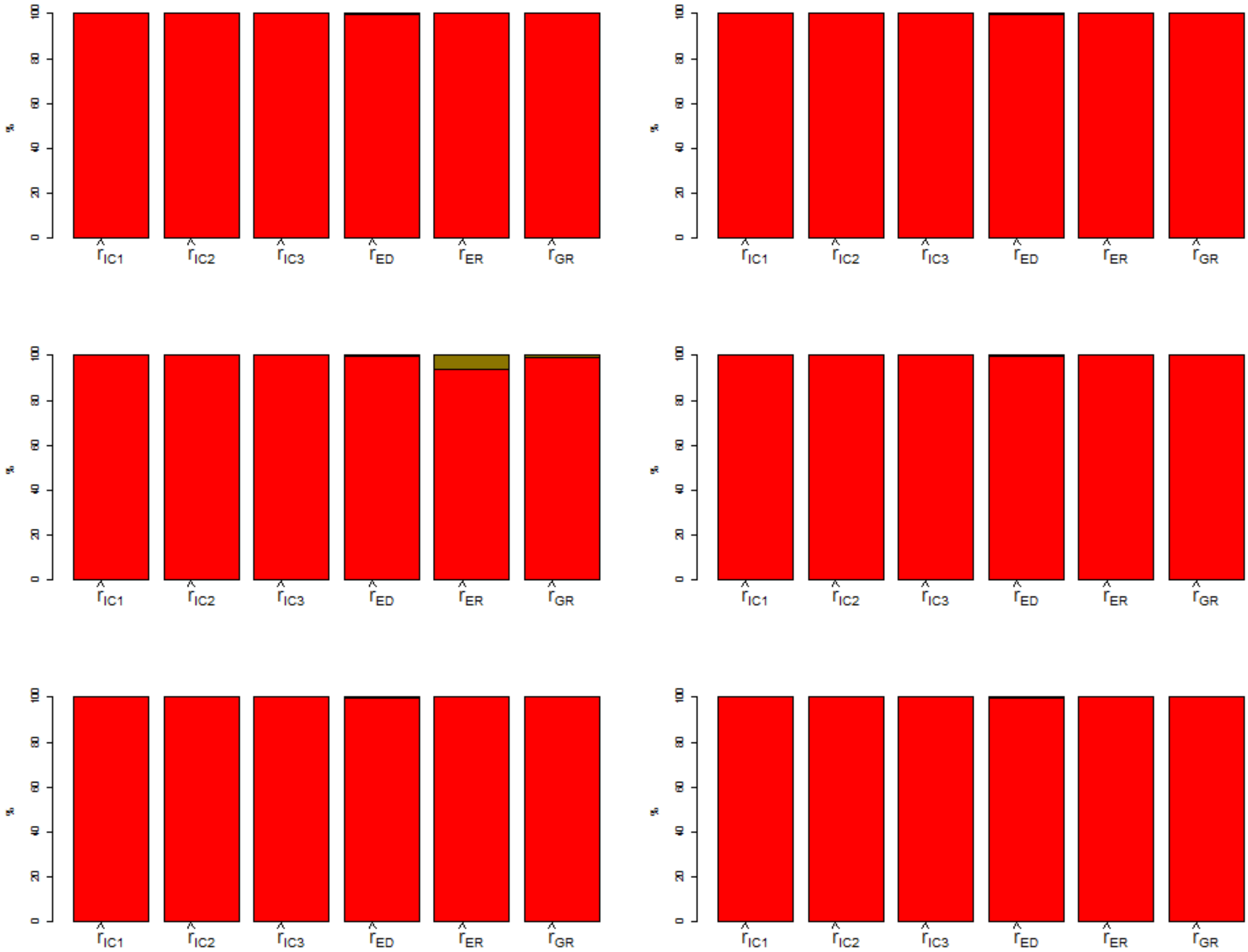

Figure A.4: Percentage of $\hat{r} = r_{\max}$ (blue), $r_{\max} > \hat{r} > r$ (green), $\hat{r} = 2$ (red), $\hat{r} = 1$ (gold) and $\hat{r} = 0$ (black) in a DFM with $r = 2$, $\gamma = 1$ and $\sigma_a^2 = 1$. System dimensions $N = 12$, $T = 100$ (first column); $N = 200$, $T = 500$ (second column). The factors are two random walks with variance $\sigma_\eta^2 = 1$ (first row); two random walks with variances $\sigma_{\eta_1}^2 = 1$ and $\sigma_{\eta_2}^2 = 5$ (second row) and a random walk with variance $\sigma_{\eta_1}^2 = 1$ and a stationary factor with $\sigma_{\eta_2}^2 = 1$.

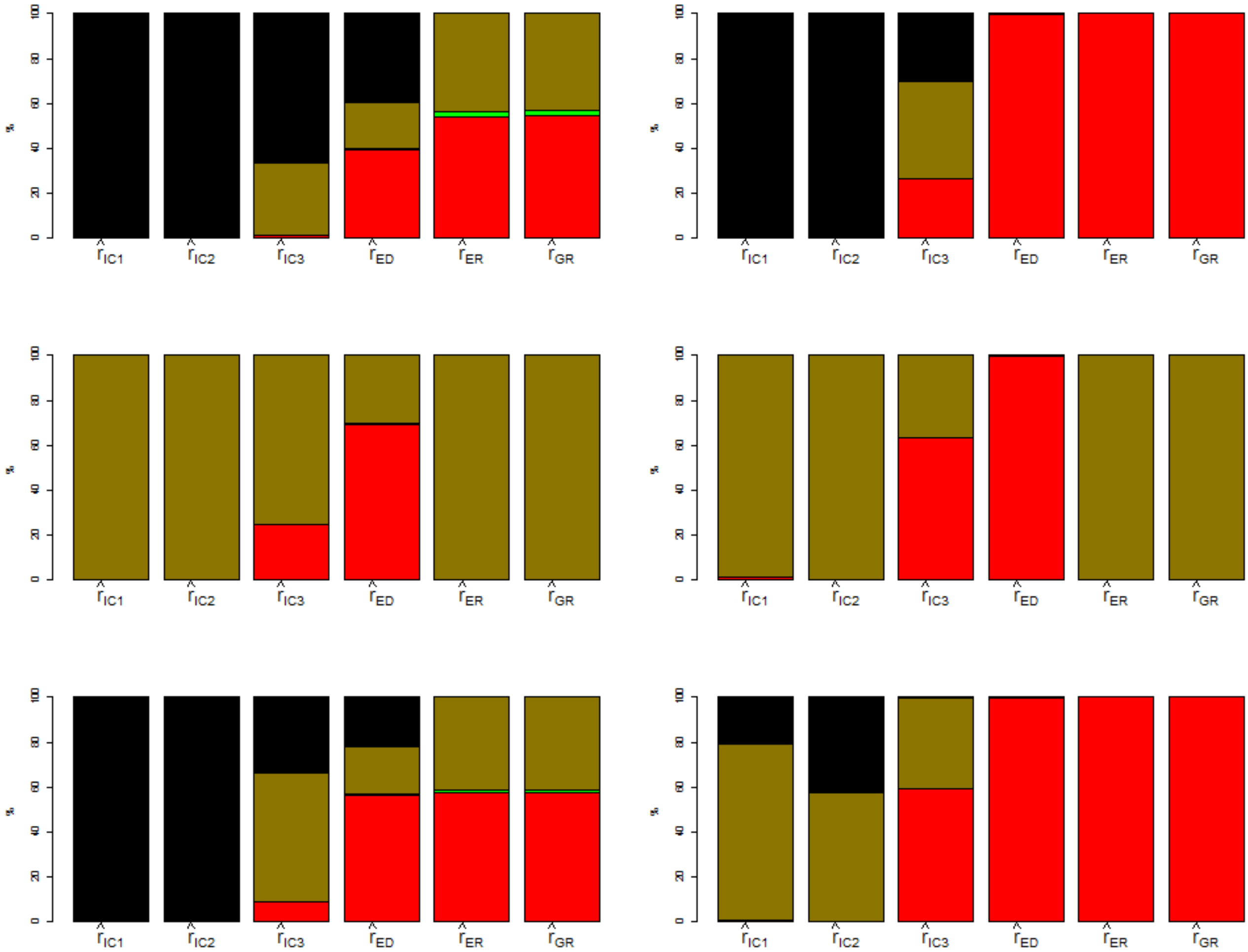

Figure A.5: Percentage of $\hat{r} = r_{\max}$ (blue), $r_{\max} > \hat{r} > r$ (green), $\hat{r} = 2$ (red), $\hat{r} = 1$ (gold) and $\hat{r} = 0$ (black) in a DFM with $r = 2$, $\gamma = 1$ and $\sigma_a^2 = 10$. System dimensions $N = 12$, $T = 100$ (first column); $N = 200$, $T = 500$ (second column). The factors are two random walks with variance $\sigma_\eta^2 = 1$ (first row); two random walks with variances $\sigma_{\eta_1}^2 = 1$ and $\sigma_{\eta_2}^2 = 5$ (second row) and a random walk with variance $\sigma_{\eta_1}^2 = 1$ and a stationary factor with $\sigma_{\eta_2}^2 = 1$.

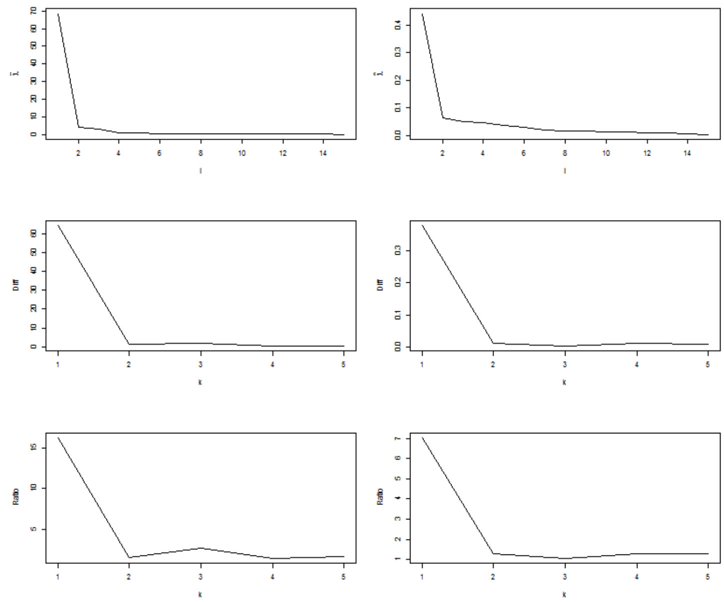

Figure A.6: Eigenvalues (first row), difference of eigenvalues (second row) and ratio of eigenvalues (third row) from $\hat{\Sigma}_Y$ (left column) and $\hat{\Sigma}_{\Delta Y}$ (right column) in empirical application for $i = 1, \ldots, r_{\max}$.

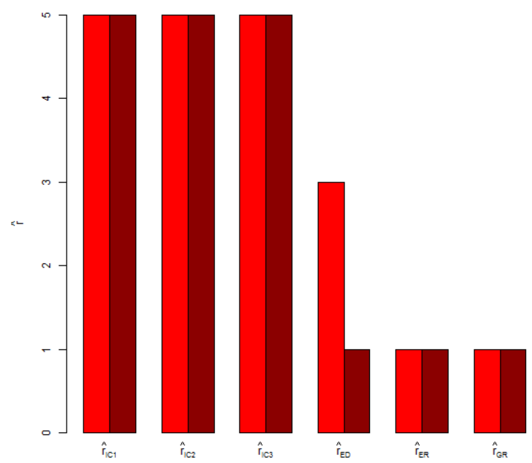

Figure A.7: Determination of number of factor for each considered procedure. Red color is using data in levels and wine color using first-differenced data.

Appendix B

Appendix to Chapter 3

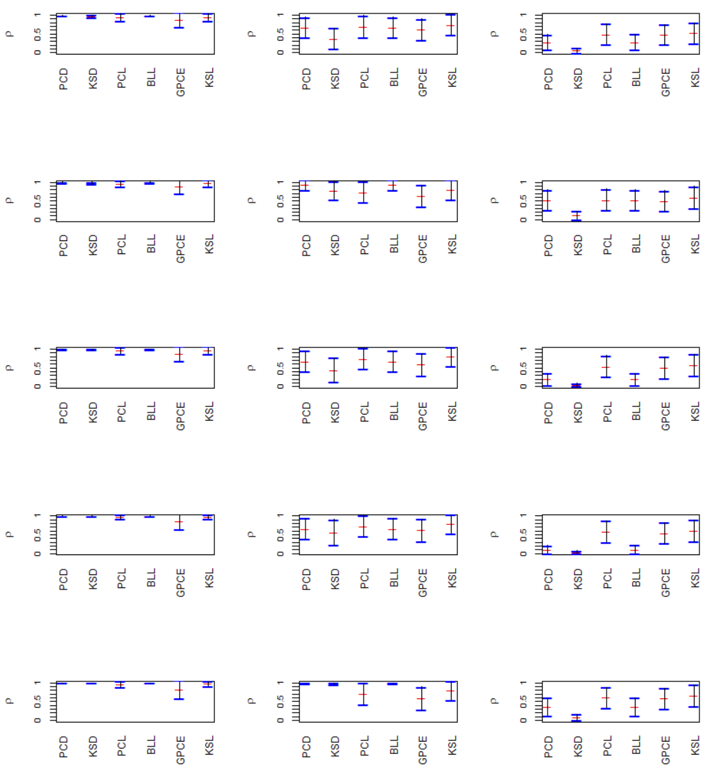

Figure B.1: Box-plots of the sample correlations between $\{\hat{\delta}_j'\hat{F}_t^{PCD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{KSD}\}$, $\{\hat{\delta}_j'\hat{F}_t^{PCL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{BLL}\}$, $\{\hat{\delta}_j'\hat{F}_t^{GPCE}\}$ and $\{\hat{\delta}_j'\hat{F}_t^{KSL}\}$ with $\{F_t\}$. We consider the M1 model with homoscedasticity in idiosyncratic errors with $\gamma = \text{diag}(-0.8I_{N/2}, 1I_{N/2})$. First row indicates $N = 12$ and $T = 50$; second row $N = 12$ and $T = 100$; third row $N = 50$ and $T = 100$; fourth row $N = 200$ and $T = 100$ and fifth row $N = 200$ and $T = 500$. The first column plots $\sigma_a^2 = 0.1$, second column $\sigma_a^2 = 1$, and third column $\sigma_a^2 = 10$.

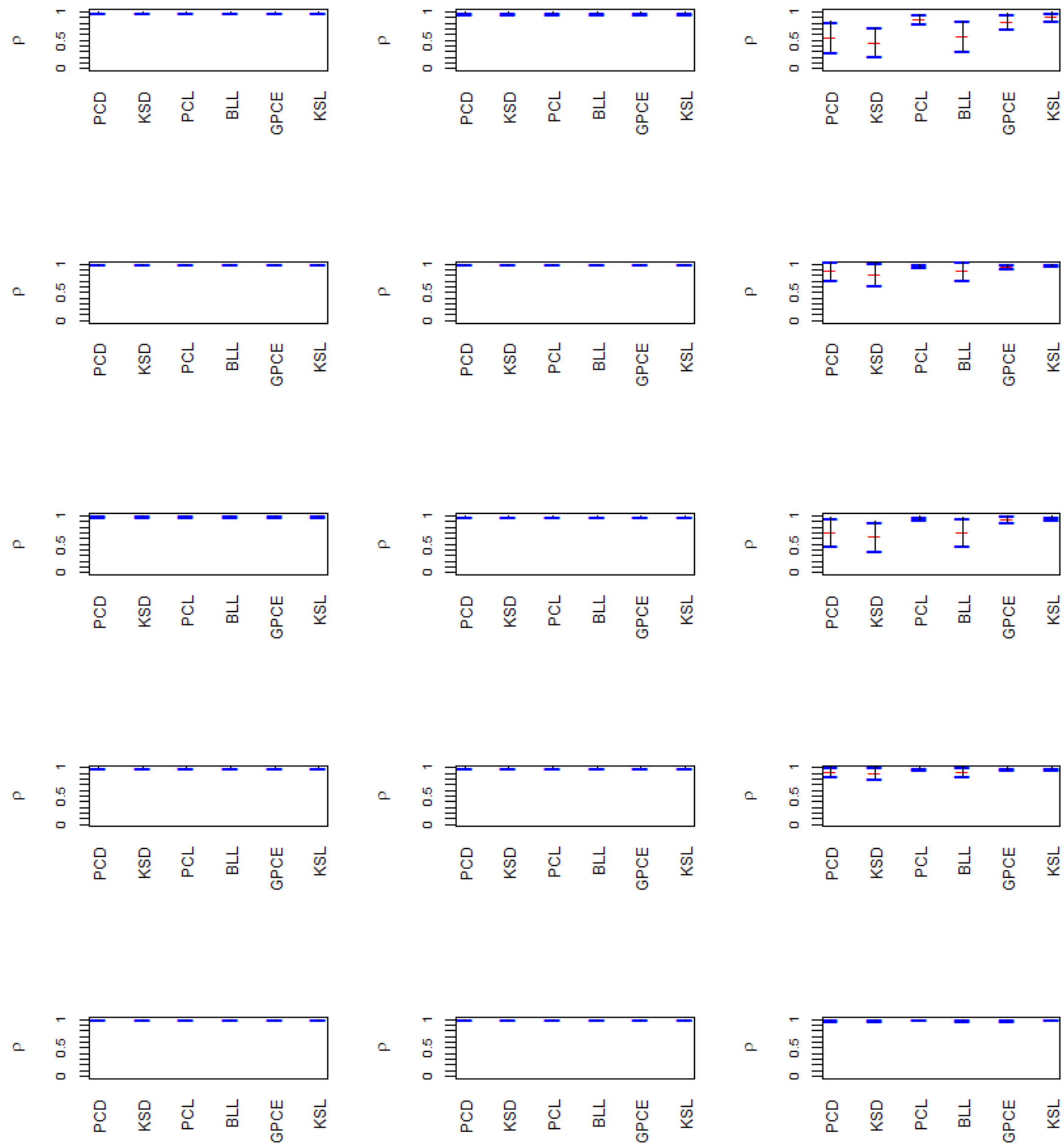

Figure B.2: Box-plots of the sample correlations between $\{\hat{\delta}_j' \hat{F}_t^{PCD}\}$, $\{\hat{\delta}_j' \hat{F}_t^{KSD}\}$, $\{\hat{\delta}_j' \hat{F}_t^{PCL}\}$, $\{\hat{\delta}_j' \hat{F}_t^{BLL}\}$, $\{\hat{\delta}_j' \hat{F}_t^{GPCE}\}$ and $\{\hat{\delta}_j' \hat{F}_t^{KSL}\}$ with $\{F_t\}$. We consider the M1 model with homoscedasticity in idiosyncratic errors with $\gamma = \text{diag}(0I_{N/2}, 0.7I_{N/2})$. First row indicates $N = 12$ and $T = 50$; second row $N = 12$ and $T = 100$; third row $N = 50$ and $T = 100$; fourth row $N = 200$ and $T = 100$ and fifth row $N = 200$ and $T = 500$. The first column plots $\sigma_a^2 = 0.1$, second column $\sigma_a^2 = 1$, and third column $\sigma_a^2 = 10$.

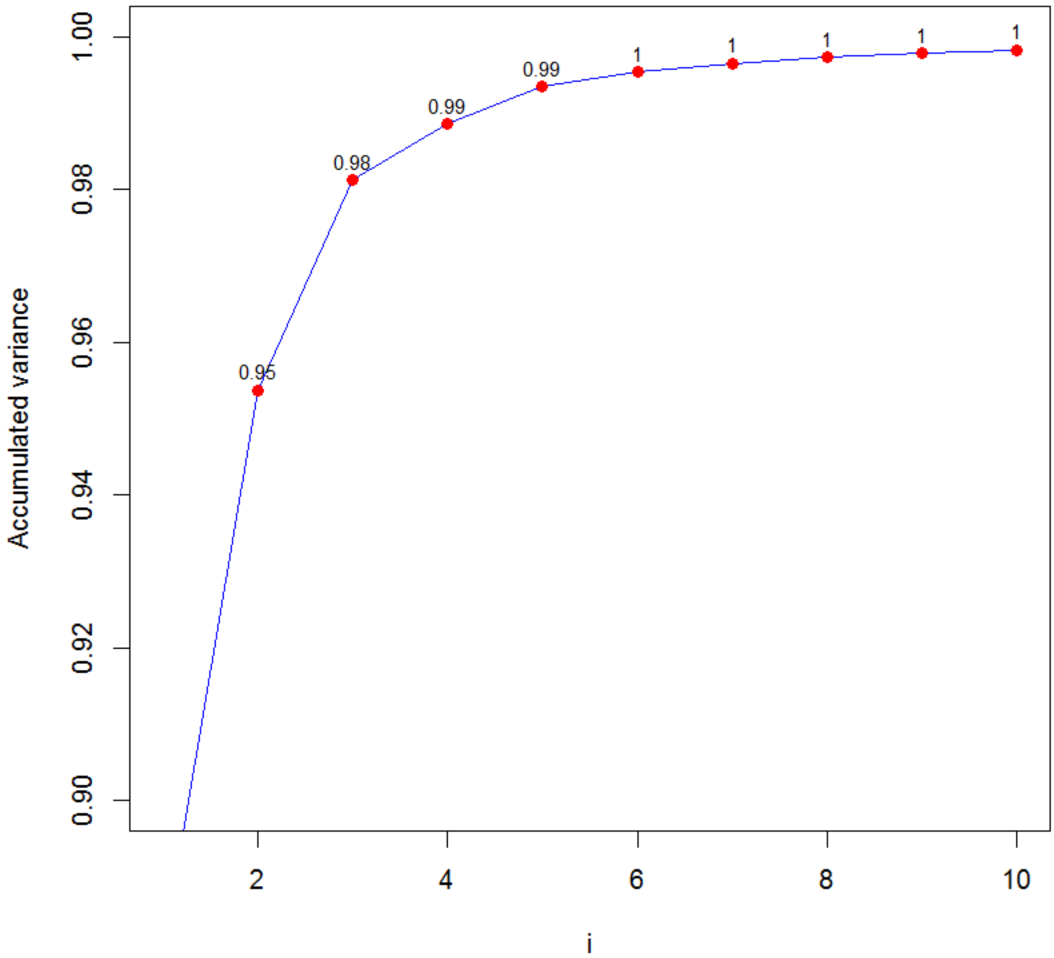

Figure B.3: $\sum_{s=1}^{i} \hat{\lambda}_s$ from $\hat{\Sigma}_Y$ for $i = 1, \ldots, r_{\max}$.

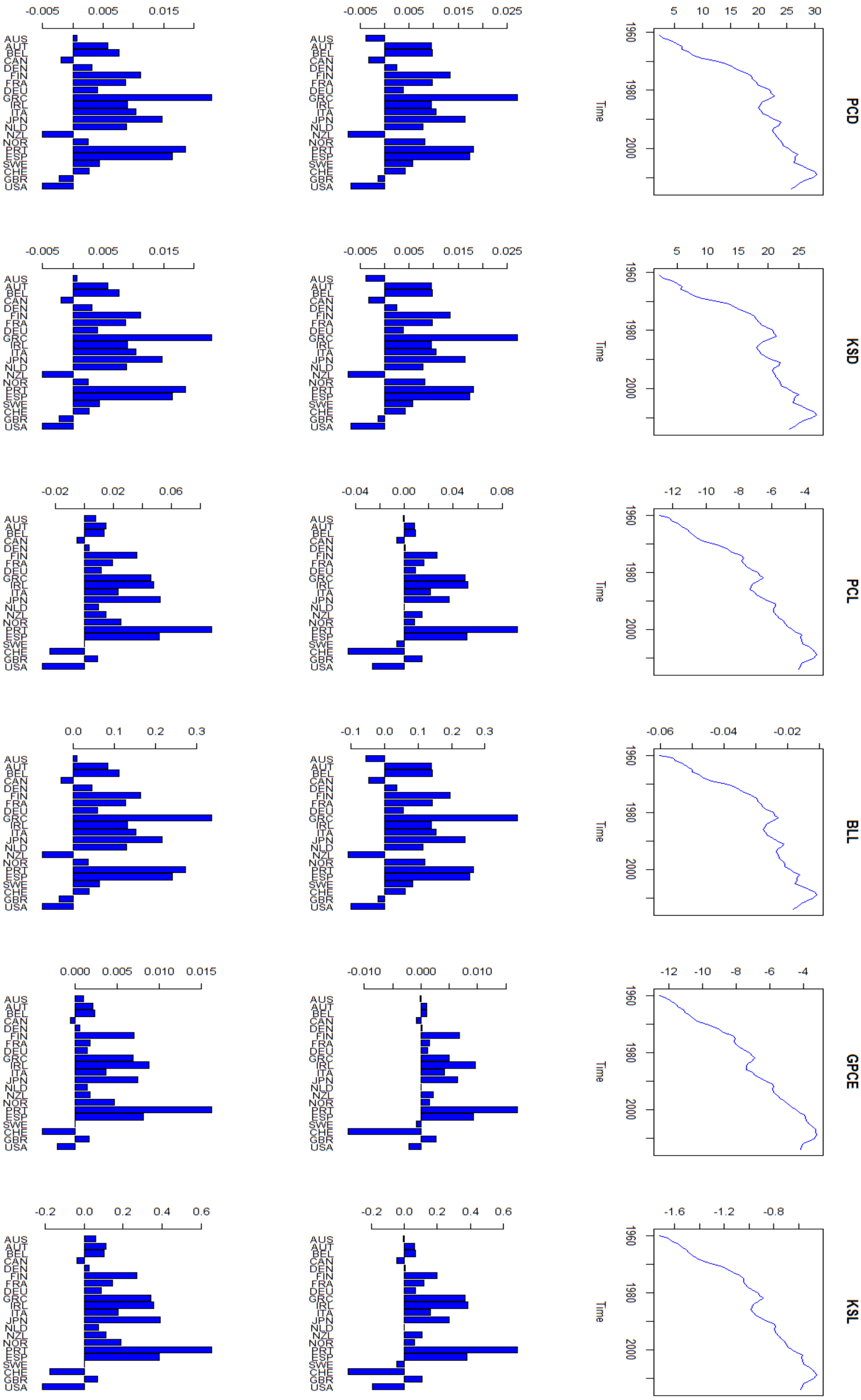

Figure B.4: Factor extraction for each considered procedure (first factor). Top panel $\hat{F}_{1t}$, middle panel $\hat{p}_{i1}$ for GDP ($i = 1, \dots 21$) and bottom panel $\hat{p}_{i1}$ for C ($i = 22, \dots 42$).

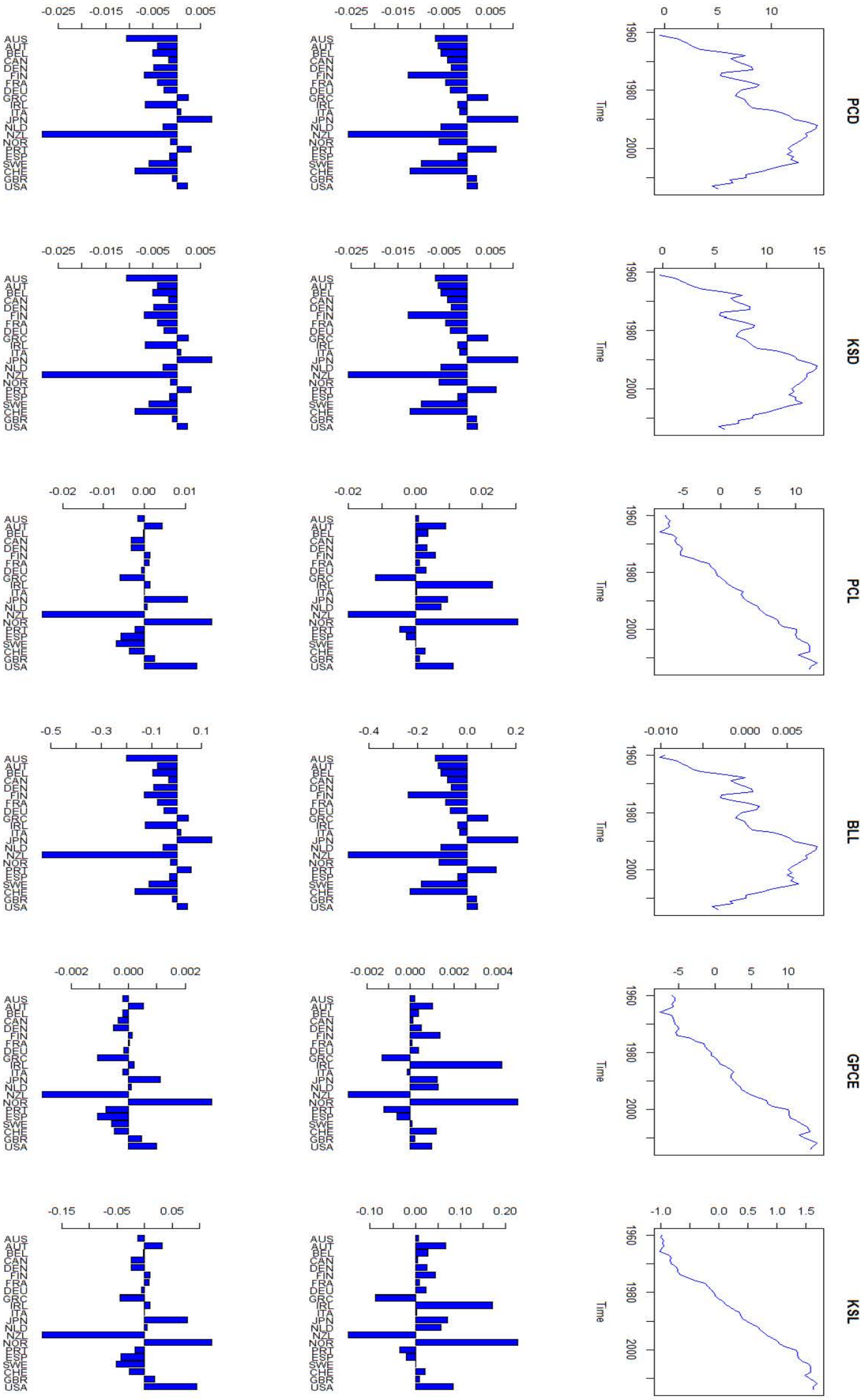

Figure B.5: Factor extraction for each considered procedure (second factor). Top panel $\hat{F}_{2t}$, middle panel $\hat{p}_{i2}$ for GDP ($i = 1, \ldots 21$) and bottom panel $\hat{p}_{i2}$ for C ($i = 22, \ldots 42$).

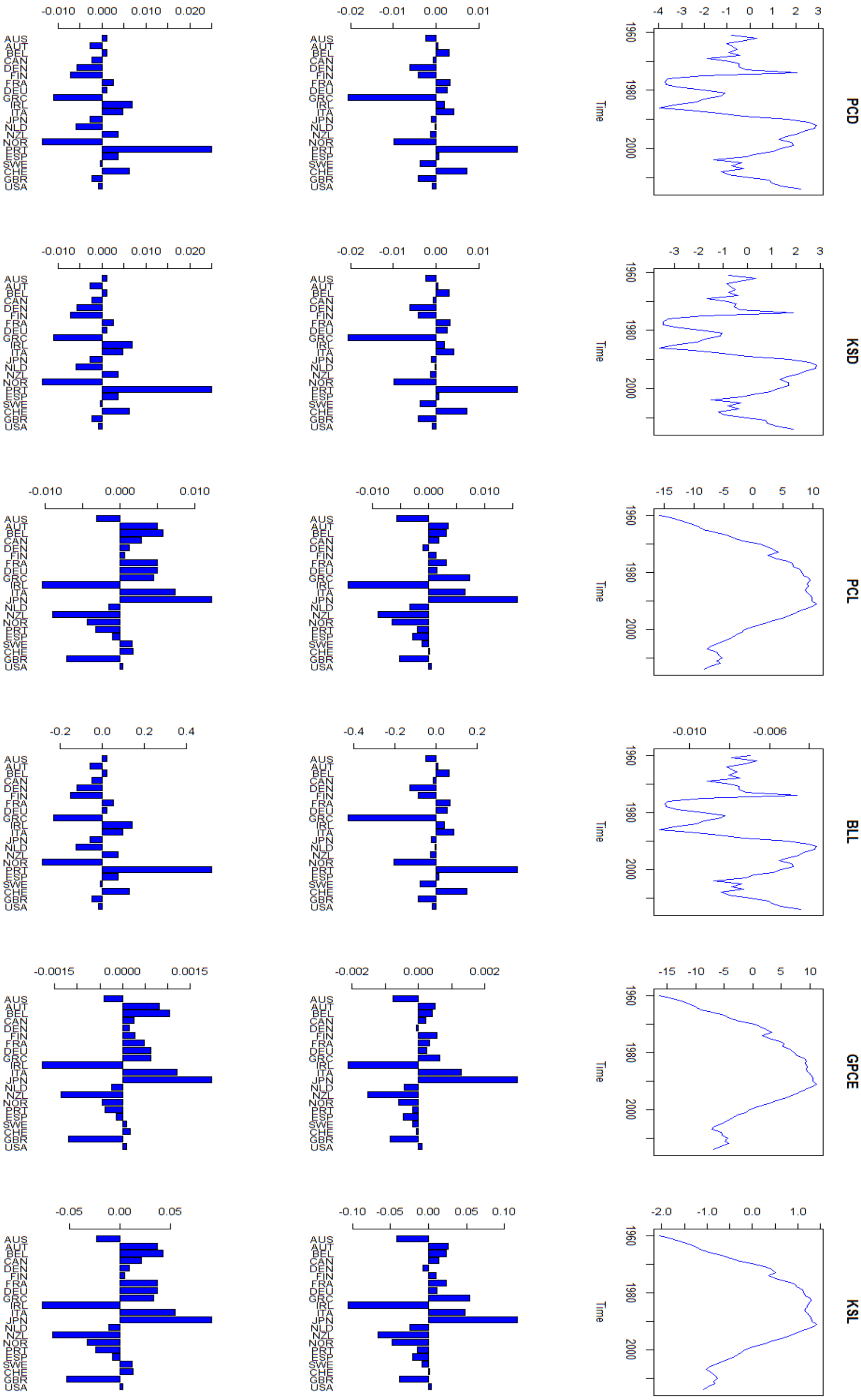

Figure B.6: Factor extraction for each considered procedure (third factor). Top panel $\hat{F}_{3t}$, middle panel $\hat{p}_{i3}$ for GDP ($i = 1, \ldots 21$) and bottom panel $\hat{p}_{i3}$ for C ($i = 22, \ldots 42$).

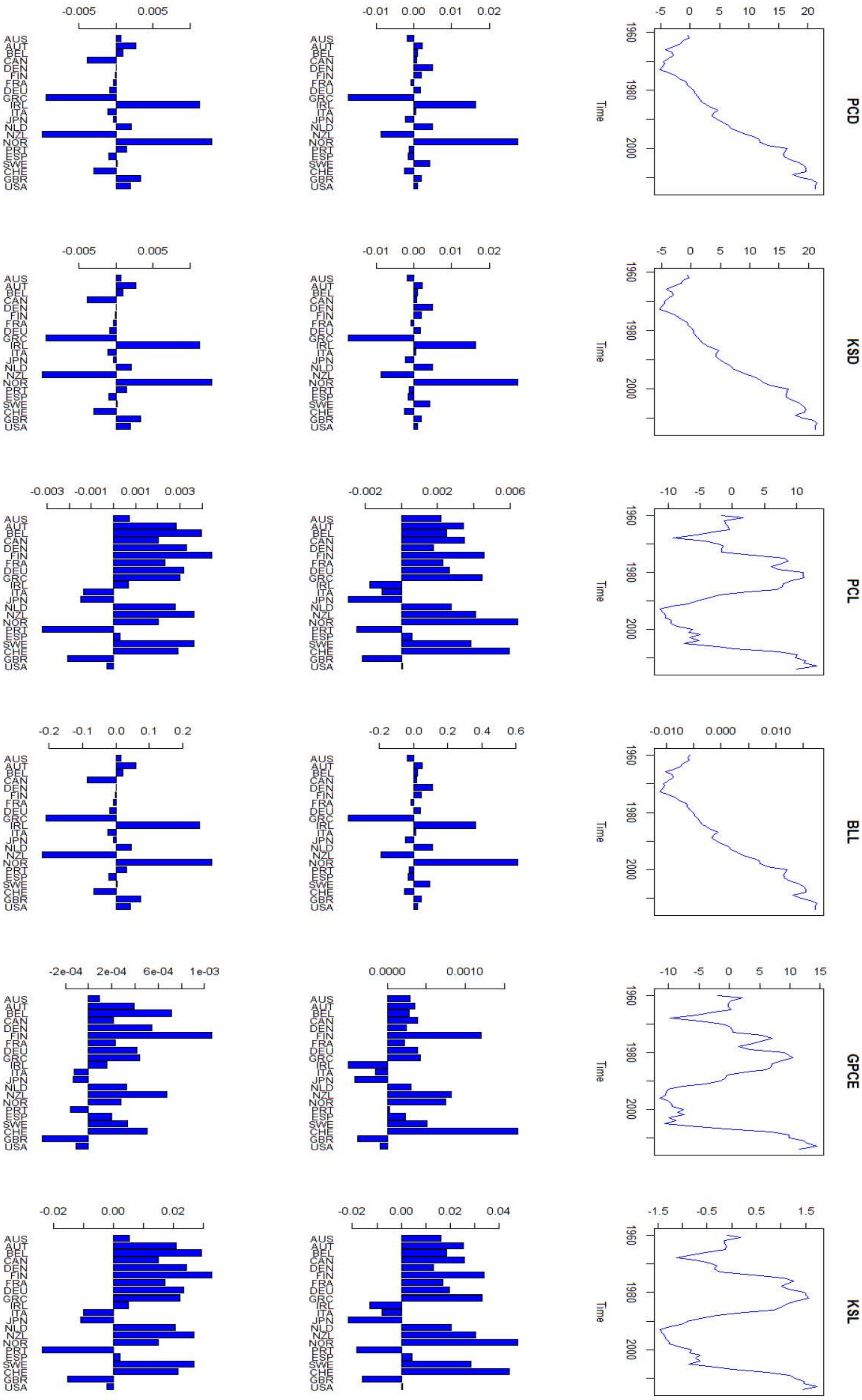

Figure B.7: Factor extraction for each considered procedure (fourth factor). Top panel $\hat{F}_{4t}$, middle panel $\hat{p}_{i4}$ for GDP ($i = 1, \dots 21$) and bottom panel $\hat{p}_{i4}$ for C ($i = 22, \dots 42$).

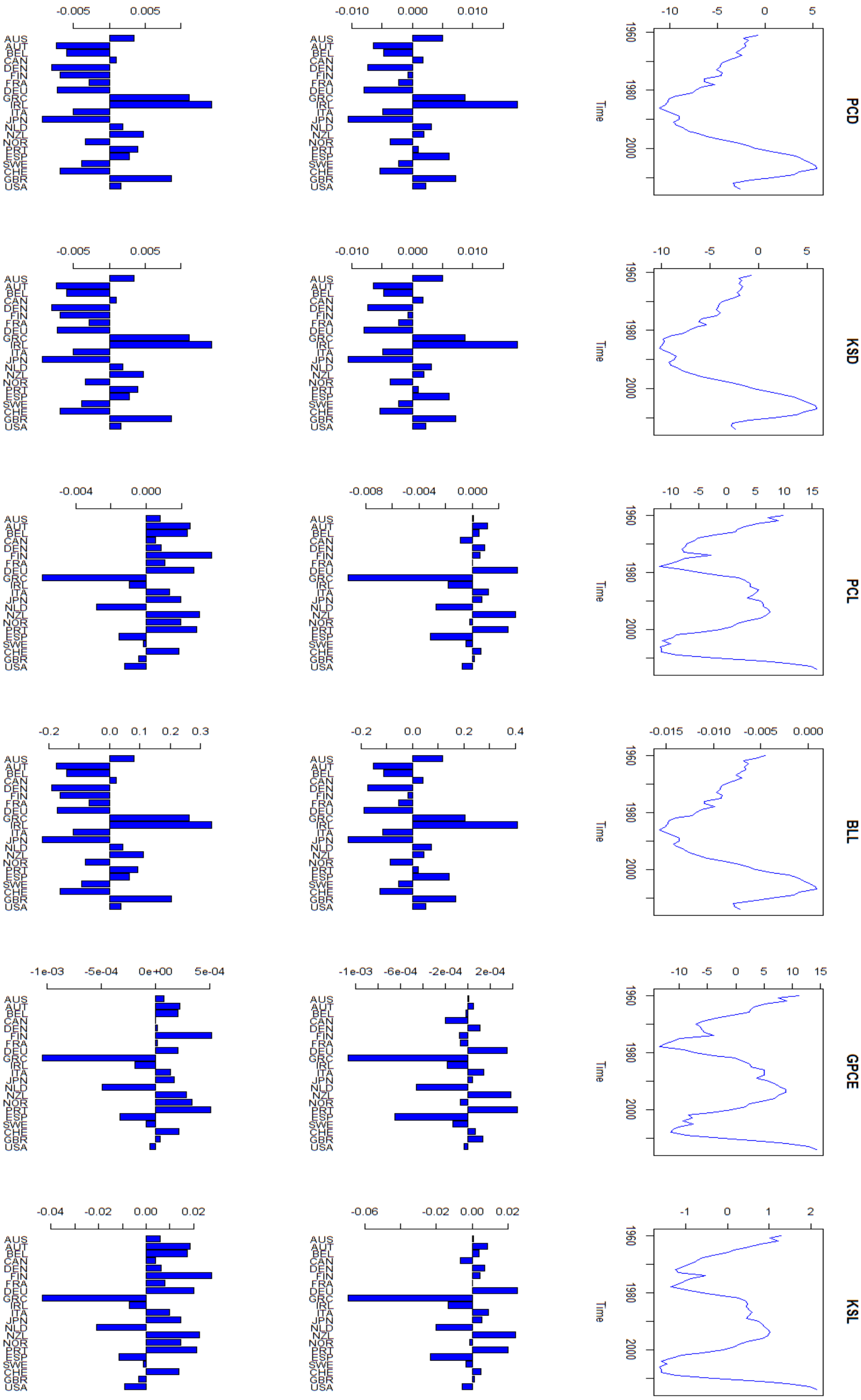

Figure B.8: Factor extraction for each considered procedure (fifth factor). Top panel $\hat{F}_{5t}$, middle panel $\hat{p}_{i5}$ for GDP ($i = 1, \dots 21$) and bottom panel $\hat{p}_{i5}$ for C ($i = 22, \dots 42$).