



UC3M Working Papers  
Statistics and Econometrics  
16-02  
ISSN 2387-0303  
February 2016

Departamento de Estadística  
Universidad Carlos III de Madrid  
Calle Madrid, 126  
28903 Getafe (Spain)  
Fax (34) 91 624-98-48

## Determining the number of factors after stationary univariate transformations

Francisco Corona<sup>a</sup>, Pilar Poncela<sup>b</sup>, Esther Ruiz<sup>a</sup>

### Abstract

---

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 this paper, we analyze the effects of these changes on the finite sample properties of some popular procedures to determine the number of factors. In particular, we consider the information criteria of Bai and Ng (2002), the edge distribution of Onastki (2010) and the ratios of eigenvalues proposed by Ahn and Horenstein (2013). The performance of these procedures when implemented to differentiated variables depend on both the ratios between variances and dependences 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.

**Keywords:** Dynamic Factor Model, Eigenvalue ratios, Edge distribution, Information criteria, Principal Components factor extraction

---

<sup>a</sup> Department of Statistics, Universidad Carlos III de Madrid.

<sup>b</sup> Departamento de Análisis Económico: Economía Cuantitativa, Universidad Autónoma de Madrid

Acknowledgements: Financial support from the Spanish Ministry of Education and Science projects ECO2012-32401 and ECO2012-32854 is acknowledged by the second and third authors, respectively.

# Determining the number of factors after stationary univariate transformations

Francisco Corona<sup>1</sup>, Pilar Poncela<sup>2</sup> and Esther Ruiz<sup>\*†1</sup>

<sup>1</sup>Departamento de Estadística, Universidad Carlos III de Madrid

<sup>2</sup>Departamento de Análisis Económico: Economía Cuantitativa, Universidad Autónoma de Madrid

February 2016

## Abstract

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 this paper, we analyze the effects of these changes on the finite sample properties of some popular 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.

**Keywords:** Dynamic Factor Models, Principal Components factor extraction, Information Criteria, Edge Distribution, Eigenvalue ratios.

---

\*Corresponding author: [ortega@est-econ.uc3m.es](mailto:ortega@est-econ.uc3m.es). Tel: 34-916249851, Fax: 34-916249849.

†Financial support from the Spanish Ministry of Education and Science projects ECO2012-32401 and ECO2012-32854 is acknowledged by the second and third authors respectively.

# 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 (2002), 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), Buch et al. (2014), Moench et al. (2013), Bräuning and Koopman (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  $I(1)$ <sup>1</sup>. 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

---

<sup>1</sup>Bai (2004) also has asymptotic results for the factors estimated from the original non-stationary data.

has been little discussion in the literature on whether differencing in a univariate fashion 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 with AR and MA parameters 0.5 and two random walk factors, he shows that the number of factor 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 and financial variables. Finally, we illustrate the results determining the number of factors in a system of prices of the euro area.

The rest of this paper is structured as follows. In section 2, we briefly describe the stationary DFM and the factor determination approaches considered. In section 3, we analyze the effects of transforming non-stationary systems by univariate stationary transformations on these procedures. In section 4, we report the results of the Monte Carlo experiments carried out to illustrate their finite sample performance. In section 5, we carry out an empirical application. Finally, we conclude in section 6.

## 2 The stationary Dynamic Factor Model

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

### 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

$$Y_t = PF_t + \varepsilon_t, \tag{1}$$

$$F_t = \Phi F_{t-1} + \eta_t, \tag{2}$$

$$\varepsilon_t = \Gamma \varepsilon_{t-1} + a_t, \tag{3}$$

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,  $\eta_t$  and  $a_t$ , are Gaussian white

noises with positive definite covariance matrices  $\Sigma_\eta$  and  $\Sigma_a$ , 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  $\Gamma$  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  $P$ ,  $\Phi$ ,  $\Sigma_\eta$ ,  $\Gamma$  and  $\Sigma_a$  as fixed parameters. Jungbacker and Koopman (2015) implement the DFM in equations (1) to (3) to the data set of Stock and Watson (2005).

The DFM in equations (1) to (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  $F = (F_1, \dots, F_T)$  is a  $r \times T$  matrix of common factors; see Stock and Watson (2002), 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 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)<sup>2</sup>.

### 2.2.1 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

$$\min_{F_1, \dots, F_T, P} V_r(P, F) \quad (4)$$

subject to  $P'P/N = I_r$  and  $FF'$  being diagonal,

where

$$V_r(P, F) = \frac{1}{NT} \sum_{t=1}^T (Y_t - PF_t)'(Y_t - PF_t) = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \varepsilon_{it}^2 = \frac{1}{NT} \text{tr}(\varepsilon\varepsilon'), \quad (5)$$

where  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_T)$  has dimension  $N \times T$ . The solution to (4) is obtained by setting  $\hat{P}$  equal to  $\sqrt{N}$  times the eigenvectors corresponding to the  $r$  largest eigenvalues of  $YY'$  where

<sup>2</sup>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

$Y = (Y_1, \dots, Y_T)$ . The corresponding PC estimator of  $F$  is given by

$$\hat{F} = \frac{\hat{P}'Y}{N}. \quad (6)$$

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<sup>3</sup>. 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 respect to  $k$ , for  $k = 0, \dots, r_{\max}$ ,

$$IC_1(k) = \ln V_k(\hat{P}, \hat{F}) + k \frac{N+T}{NT} \ln \frac{NT}{N+T}, \quad (7a)$$

$$IC_2(k) = \ln V_k(\hat{P}, \hat{F}) + k \frac{N+T}{NT} \ln m, \quad (7b)$$

$$IC_3(k) = \ln V_k(\hat{P}, \hat{F}) + k \frac{\ln m}{m}, \quad (7c)$$

where  $V_k(\hat{P}, \hat{F})$  is defined as in expression (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  $r \leq r_{\max}$ . The Bai and Ng (2002) criteria in (7) 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  $\frac{1}{NT}YY'$

$$tr(\hat{\varepsilon}\hat{\varepsilon}') = tr(YY') - tr(\hat{P}\hat{F}'\hat{F}\hat{P}') = T \sum_{i=1}^m \hat{\lambda}_i - T \sum_{i=1}^r \hat{\lambda}_i = T \sum_{i=r+1}^m \hat{\lambda}_i,$$

where  $\hat{\lambda}_i$ ,  $i = 1, \dots, m$  are the eigenvalues of  $\hat{\Sigma}_Y = \frac{1}{T}YY'$ , arranged in descending order.

the penalizing power of the function itself and estimating the number of factors using different subsamples. Also, Kapetanios (2010) proposes 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.

<sup>3</sup>Onatski (2012) considers a DFM in which the explanatory power of the factors does not strongly dominate the explanatory power of the idiosyncratic noises.

Therefore,

$$V_r(\widehat{P}, \widehat{F}) = \frac{1}{N} \sum_{i=r+1}^m \widehat{\lambda}_i. \quad (8)$$

Using the expression of  $V_k(\widehat{P}, \widehat{F})$  in (8), the functions in (7) can be written as

$$IC_j(k) = \log\left(\frac{1}{N} \sum_{i=k+1}^m \widehat{\lambda}_i\right) + kg_j(N, T), \quad (9)$$

where  $g_j(N, T)$  is defined accordingly to the criteria in (7) for  $j = 1, 2$  and  $3$ .

### 2.2.2 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  $\widehat{\Sigma}_Y$  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,  $\delta$ , which consistently separates the bounded and diverging eigenvalues of  $\widehat{\Sigma}_Y$ . For any  $j > r$ , the differences  $\widehat{\lambda}_j - \widehat{\lambda}_{j+1}$  converge to 0 while the difference  $\widehat{\lambda}_r - \widehat{\lambda}_{r+1}$  diverges to infinity when both  $N$  and  $T$  tend to infinity. Assuming that  $r_{\max}/N \rightarrow 0$ , Onatski (2010) proposes the following algorithm in order to calibrate  $\delta$  and determine the number of factors:

1. Obtain  $\widehat{\lambda}_i$ ,  $i = 1, \dots, N$  and set  $j = r_{\max} + 1$ .
2. Obtain  $\widehat{\beta}$  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  $\{\widehat{\lambda}_j, \dots, \widehat{\lambda}_{j+4}\}$  and the observations of the regressor variable are  $\{(j-1)^{2/3}, \dots, (j+3)^{2/3}\}$ . Set  $\widehat{\delta} = 2|\widehat{\beta}|$ .
3. Estimate  $\widehat{r} = \max\{k \leq r_{\max} | \widehat{\lambda}_k - \widehat{\lambda}_{k+1} \geq \widehat{\delta}\}$  or  $\widehat{r} = 0$  if  $\widehat{\lambda}_k - \widehat{\lambda}_{k+1} < \widehat{\delta}$ .
4. Set  $j = \widehat{r} + 1$ . Repeat steps 2 and 3 until  $\widehat{r}$  converges.

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

### 2.2.3 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  $\widehat{\Sigma}_Y$  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  $k = 0, \dots, r_{\max}$ , that maximizes the following ratios

$$ER(k) = \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}}, \quad (10)$$

$$GR(k) = \frac{\ln \left[ V_{k-1}(\hat{P}, \hat{F}) / V_k(\hat{P}, \hat{F}) \right]}{\ln \left[ V_k(\hat{P}, \hat{F}) / V_{k+1}(\hat{P}, \hat{F}) \right]} = \frac{\ln(1 + \hat{\lambda}_k^*)}{\ln(1 + \hat{\lambda}_{k+1}^*)}, \quad (11)$$

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  $\hat{\lambda}_0 \rightarrow 0$  and  $m\hat{\lambda}_0 \rightarrow \infty$  as  $m \rightarrow \infty$ .

Note that both the numerator and denominator of  $GR(k)$  are the growth rates 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) 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).

### 2.3 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  $\frac{1}{T}YY'$  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  $\hat{\Sigma}_Y$  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.



### 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.

Consider the DFM given in equations (1) to (3) in which  $\Phi$  and  $\Gamma$  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

$$\Delta Y_t = P \Delta F_t + \Delta \varepsilon_t, \quad (12)$$

$$\Delta F_t = (\Phi - I) F_{t-1} + \eta_t, \quad (13)$$

$$\Delta \varepsilon_t = (\Gamma - I) \varepsilon_{t-1} + a_t. \quad (14)$$

Denote by  $\phi_i$  the  $i$ -th element in the main diagonal of  $\Phi$ . If  $|\phi_i| < 1$ , then the variance of the corresponding differenced factor is given by  $\sigma_{f_i}^2 = 2\sigma_{\eta_i}^2 / (1 + \phi_i)$  where  $\sigma_{\eta_i}^2$  is the variance of  $\eta_i$ . When  $\phi_i = 0.5$ , the difference between the variances of  $F_t$  and  $\Delta F_t$  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  $\Delta F_t$  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  $\varepsilon_t$  and  $\Delta \varepsilon_t$  with respect to  $\gamma_i$ , the  $i$ -th element in the main diagonal of  $\Gamma$ . Note that if  $\varepsilon_t$  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  $\Delta F_t$  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

$$\text{Corr}(\Delta \varepsilon_{it}, \Delta \varepsilon_{it-h}) = 0.5 \gamma_i^{h-1} (\gamma_i - 1). \quad (15)$$

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

$$\text{Corr}(\Delta \varepsilon_{it}, \Delta \varepsilon_{jt}) = \sigma_{\Delta \varepsilon_i}^{-1} \sigma_{\Delta \varepsilon_j}^{-1} (2 - \gamma_i - \gamma_j) \rho \sigma_{\varepsilon_i} \sigma_{\varepsilon_j} = \frac{(2 - \gamma_i - \gamma_j) \rho}{4(1 - \gamma_i)(1 - \gamma_j)},$$

while the pairwise correlations between differenced random walk idiosyncratic noises is given by

$$\text{Corr}(\Delta \varepsilon_{it}, \Delta \varepsilon_{jt}) = \rho.$$

In order to simplify the analysis of the effects of univariate differentiation on the determina-

tion 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

$$\Sigma_{\Delta Y} = P\Sigma_f P' + \sigma_e^2 I, \quad (16)$$

where  $\Sigma_f$  is the covariance matrix of  $\Delta F_t$  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  $i = 1, \dots, N$ , where  $\mu_i$  is the  $i$ -th largest eigenvalue of  $P\Sigma_f P'$ . Furthermore,

$$\text{tr}(P\Sigma_f P') = \text{tr}(P' P\Sigma_f) = \sum_{j=1}^r \sigma_{f_j}^2 \sum_{i=1}^N p_{ij}^2 = \sum_{j=1}^r \mu_j.$$

Therefore, the sum of the  $r$  largest eigenvalues of  $\Sigma_{\Delta Y}$  is given by

$$\sum_{i=1}^r \lambda_i = r\sigma_e^2 + \sum_{j=1}^r \sigma_{f_j}^2 \sum_{i=1}^N p_{ij}^2, \quad (17)$$

while the rest  $N - r$  eigenvalues are given by

$$\lambda_i = \sigma_e^2. \quad (18)$$

Consider the particular case of a unique random walk factor, i.e.  $r = 1$  and  $\phi_1 = 1$ . In this case,

$$\lambda_1 = \sigma_\eta^2 \sum_{i=1}^N p_{i1}^2 + \sigma_e^2, \quad (19a)$$

$$\lambda_i = \sigma_e^2, \quad i = 2, \dots, N. \quad (19b)$$

Consequently, the function to be minimized according to the Bai and Ng (2002) information criteria, is given by

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

The procedure proposed by Onatski (2010) is based on the differences between adjacent eigenvalues. Note that for  $j = 2, \dots, N$ ,  $\lambda_j - \lambda_{j+1} = 0$ . Therefore, the procedure should work as far as the difference between  $\lambda_1$  and  $\lambda_2$  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  $\sigma_e^2$ . Therefore, for given weights and cross-sectional di-

mension, the procedure should work better when  $\sigma_\eta^2$  is large. Also, for a given value of  $\sigma_\eta^2$ , 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  $\lambda_j$  are always  $\sigma_e^2$ .

Consider the  $ER(k)$  criterion of Ahn and Horenstein (2013) given in (10) which looks for a large difference between the ratio of  $\lambda_1$  and  $\lambda_2$  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

$$\lambda_0 = \frac{\sigma_e^2 + N^{-1}\sigma_\eta^2 \sum_{i=1}^N p_{i1}^2}{\log(N)},$$

and, consequently,

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

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

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

Therefore,

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

## 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.

The experiments are based on  $R = 500$  replications generated by the DFM in equations

(1) to (3) with  $T = (100, 500)$  and  $N = (12, 200)$ <sup>4</sup>. 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 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 (1) to (3) with  $r = 1$ ,  $\Phi = 1$  and  $\sigma_\eta^2 = 1$ . The factor loadings are generated by  $p_{i1} \sim U[0, 1]$  with  $\sum_{i=1}^N p_{i1}^2 = 5.59$  and  $65.56$  for  $N = 12$  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  $\Sigma_a = \sigma_a^2 I$  with  $\sigma_a^2 = 1$  so that  $\sigma_e^2 = 10$  and  $1$  for the values of  $\gamma$  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.  $\Gamma$  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  $r_{\max} = 4$  and  $13$  when  $N = 12$  and  $200$ , respectively<sup>5</sup>. 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 1 plots, for  $N = 12$  and  $T = 100$ , the Monte Carlo averages and 95% confidence intervals, for homoscedastic and contemporaneously uncorrelated idiosyncratic noises<sup>6</sup>, of i) the sample ordered eigenvalues (first column); ii) their differences (second column); and iii) their

<sup>4</sup>The time dimension of the multivariate system is generated with  $T^* = T + 100$  observations. The factor extraction is carried out after removing the first 100 observations.

<sup>5</sup>It is important to note that Ahn and Horenstein (2013) recommend 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.

<sup>6</sup>The effect of heteroscedasticity and weak cross-correlation on the estimated eigenvalues is negligible. The results are available upon request.

ratios (third column), together with the corresponding population quantities, when  $\gamma = -0.8$  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). 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  $\Delta Y$  is given by  $\lambda_1 = \sigma_e^2 + \sum_{i=1}^N p_{i1}^2$  while all other eigenvalues are given by  $\sigma_e^2$ . 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 1 shows that, regardless of the value of  $\sigma_e^2$ , the eigenvalues are better estimated when  $\gamma = 1$  than when  $\gamma = -0.8$ , with smaller biases and standard deviations. Obviously, given  $\gamma$ , the eigenvalues are better estimated when  $\sigma_a^2$  is smaller. Therefore, in order to estimate the eigenvalues of the covariance matrix of  $\Delta Y$ , 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 plots the same quantities as in Figure 1 for  $\gamma = -0.8$  and  $\sigma_a^2 = 1$ , when  $N = 12$  and  $T = 100$  (first row),  $N = 12$  and  $T = 500$  (second row),  $N = 200$  and  $T = 100$  (third row) and  $N = 200$  and  $T = 500$  (fourth row). 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 performance 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 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  $\sigma_a^2 = 0.1$  ( $\sigma_e^2 = 1$ ). The first row plots the percentages when  $N = 12$  and  $T = 100$ ; the second row corresponds to  $N = 12$  and  $T = 500$ ; the third row corresponds to  $N = 200$  and  $T = 100$ ; finally, the fourth row plots the results for  $N = 200$  and  $T = 500$ . We consider idiosyncratic noises being homoscedastic and uncorrelated (first column); heteroscedastic and uncorrelated (second column); and homoscedastic and cross-sectionally correlated (third column). 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 4 plots the same quantities as in Figure 3 when  $\gamma = 1$  and  $\sigma_a^2 = 1$ . Note that this case is comparable to that in Figure 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 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 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  $N = 12$  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  $\sigma_e^2$ . The behavior of  $\hat{r}_{ED}$  depends both on  $\gamma$  and  $\sigma_e^2$  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

$$\Sigma_\eta = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}.$$

Finally, the last model considered has a second stationary factor with  $\Sigma_\eta = I$  and

$$\Phi = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}.$$

For each of the three Data Generating Process (DGP) above, Figure 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  $T = 500$ .<sup>7</sup> 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  $\hat{r} = r$ .

---

<sup>7</sup>Monte Carlo results on the estimated eigenvalues are available upon request. They are not included to save space.

## 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,  $P_{it}$ , have been obtained from the OCDE data base<sup>8</sup> and transformed into annual inflation as  $y_{it} = 100 \times \Delta_{12} \log(P_{it})$ . When needed, the inflation rates have been deseasonalized and corrected by outliers using the seasonal adjustment software developed by the United States Census Bureau<sup>9</sup>. Following Stock and Watson (2005), outliers are substituted by the median of the 5 previous observations.

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 data in levels and after differencing. All procedures are implemented with  $r_{\max} = 5$ . Figure 7 displays the number of factors estimated. Note that, when the procedures are implemented using the original inflation rates, the information criteria estimate  $\hat{r} = r_{\max}$  while  $\hat{r}_{ED} = 3$  and  $\hat{r}_{ER} = \hat{r}_{GR} = 1$ . However, after differencing, all procedures detect just one factor. 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 both that  $r = 1$  and  $r = 3$ .

Consider first the results when  $r = 1$ . The extracted factor and its corresponding weights are plotted in Figure 8 for the original (first column) and differenced inflation rates (second column); 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 8, 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 autoregressive parameters of all the idiosyncratic noises are close to one with rather small variances between 0.03 and 0.1. Consequently, the differenced idiosyncratic noises are white noises with small variances.

Finally, Figure 9 plots the factors and weights estimated when  $r = 3$ . The second factor heavily loads (with negative sign) on Ireland while the third one (with positive sign) loads on Greece. This explains why the variances of the idiosyncratic noises associated to these two countries considerably decrease with respect to the one-factor model.

## 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 idiosyn-

---

<sup>8</sup><http://stats.oecd.org/index.aspx?queryid=221>

<sup>9</sup><https://cran.r-project.org/web/packages/seasonal.pdf>

cratic 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.

## References

- Ahn, S. and A. Horenstein (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica* 81(3), 1203–1227.
- Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics and Probability Letters* 80(1), 1806–1813.
- Altissimo, F., B. Mojon, and P. Zaffaroni (2009). Can aggregation explain the persistence of inflation?. *Journal of Monetary Economics* 56, 231–241.
- 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 S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.
- Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica* 72(4), 1127–1177.
- Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecast and inference for factor-augmented regressions. *Econometrica* 74(4), 1133–1150.
- Bai, J. and S. Ng (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics* 3(2), 89–163.
- Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics* 176(1), 18–29.
- Bai, J. and P. Wang (2014). Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics* 178(2), 794–804.
- Barhoumi, K., O. Darné, and L. Ferrara (2013). Testing the Number of Factors: An empirical assessment for forecasting purposes. *Oxford Bulletin of Economics and Statistics* 75(1), 64–79.
- Bräuning, F. and S. J. Koopman (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting* 30(3), 572–584.



- Breitung, J. and I. Choi (2013). Factor models. *Handbook of Research Methods and Applications in Empirical Macroeconomics, Chapter 11*, 249–265.
- Breitung, J. and S. Eickmeier (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics* 163, 71–84.
- Buch, C. M., S. Eickmeier, and E. Prieto (2014). Macroeconomic factors and microlevel bank behavior. *Journal of Money, Credit and Banking* 46(4), 715–751.
- Connor, G. and R. A. Korajczyk (2010). *Encyclopedia of Quantitative Finance: Factor Models of Asset Returns*. Rama Cont, ed., Chicester: Wiley.
- Delle Monache, D., I. Petrella, and F. Venditti (2016). Common faith or parting ways? A time varying parameters factor analysis of euro-area inflation in Dynamic Factor Models. *Advances in Econometrics* 35, 539–565.
- Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent Variables in Socio-Economic Models*, Amsterdam: North-Holland, 365–382.
- Hallin, M. and R. Liska (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* 102, 603–617.
- Jungbacker, B. and S. Koopman (2015). Likelihood-based dynamic factor analysis for measurement and forecasting. *Econometric Journal* 18, 1–21.
- Kapetanios, G. (2010). A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets. *Journal of Business and Economic Statistics* 28(3), 397–409.
- Lahiri, K., G. Monokroussos, and Z. Yongchen (2015). Forecasting Consumption: The Role of Consumer Confidence in Real Time with Many Predictors. *Journal of Applied Econometrics*. (in press).
- Lahiri, K. and X. Sheng (2010). Measuring Forecast Uncertainty by Disagreement: The Missing Link. *Journal of Applied Econometrics* 25, 514–538.
- Lahiri, K. and W. Yao (2004). A dynamic factor model of the coincident indicators for the us transportation sector. *Applied Economic Letters* 11(10), 595–600.
- Moench, E., S. Ng, and S. Potter (2013). Dynamic hierarchical factor models. *Review of Economics and Statistics* 95(5), 1811–1817.
- 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.
- Pinheiro, M., A. Rua, and F. Dias (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 E. Ruiz (2016). Small versus big data factor extraction in dynamic factor models: An empirical assessment in Dynamic Factor Models. *Advances in Econometrics*, 35, 401–434.
- Poncela, P., E. Senra, and L. P. Sierra (2014). Common dynamics of nonenergy commodity prices and their relation to uncertainty. *Applied Economics* 46(30), 3724–3735.
- Reis, R. and M. Watson (2010). Relative Goods’ Prices, Pure Inflation, and The Phillips Correlation. *American Economic Journal: Macroeconomics* 2(3), 128–157.
- Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a priory economic theory. *New Methods in Business Cycle Research Minneapolis: Federal Reserve Bank of Minneapolis*.
- Stock, J. H. and M. W. Watson (2002). Forecasting using Principal Components from a large number of predictors. *Journal of the American Statistical Association* 97(1), 1169–1179.
- Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. *NBER Working Paper 11467*.
- Stock, J. H. and M. W. Watson (2011). Dynamic Factor Models, in Clements, M.P and D.F. Hendry (eds.). *Oxford Handbook of Economic Forecasting*, Oxford: Oxford University Press.
- Stock, J. H. and M. W. Watson (2012a). Disentangling the Channels of the 2007-2009 Recession. *Brookings Papers on Economic Activity*, 81–130.
- Stock, J. H. and M. W. Watson (2012b). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics* 30(4), 481–493.

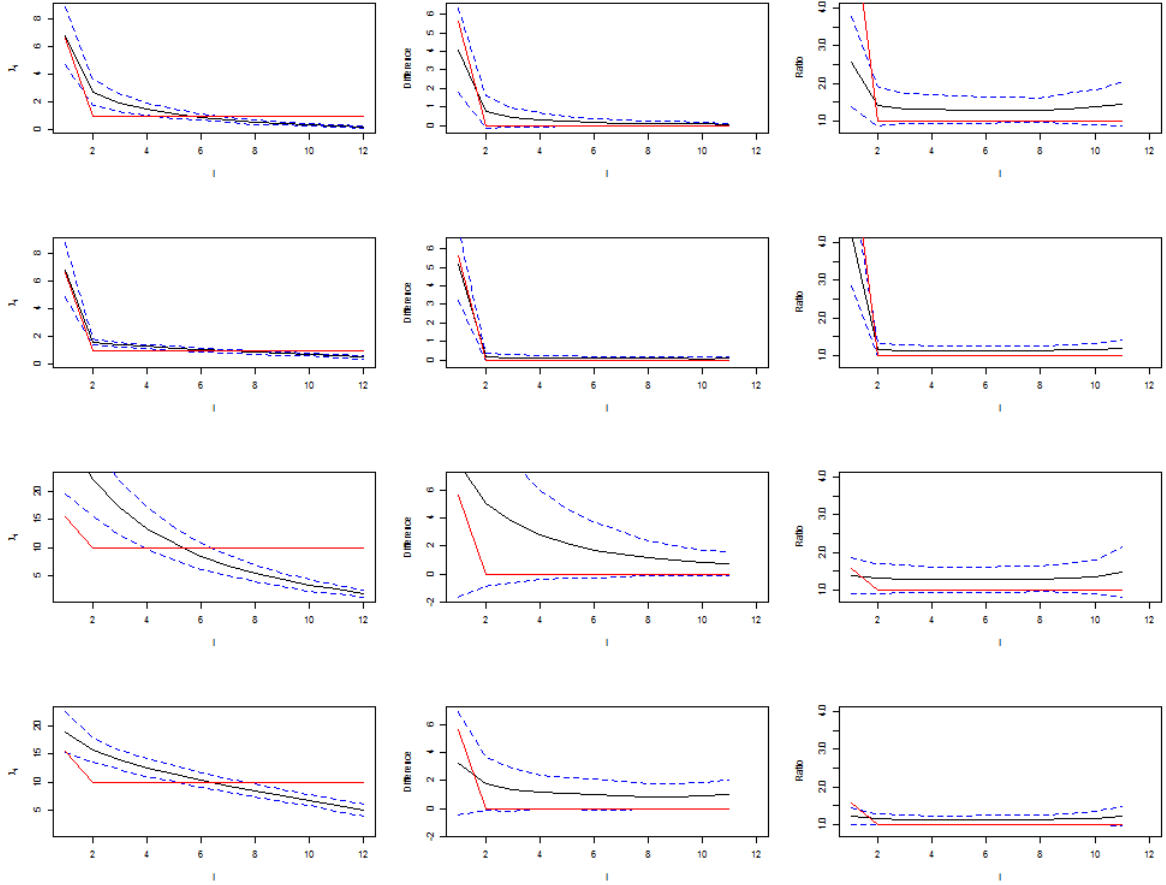


Figure 1: Eigenvalues of DFM with  $N = 12$ ,  $T = 100$ ,  $r = 1$ ,  $\phi = 1$ , with  $\gamma = -0.8$  and  $\sigma_a^2 = 0.1$  (first row),  $\gamma = 1$  and  $\sigma_a^2 = 1$  (second row),  $\gamma = -0.8$  and  $\sigma_a^2 = 1$  (third row) and  $\gamma = 1$  and  $\sigma_a^2 = 10$  (fourth 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.

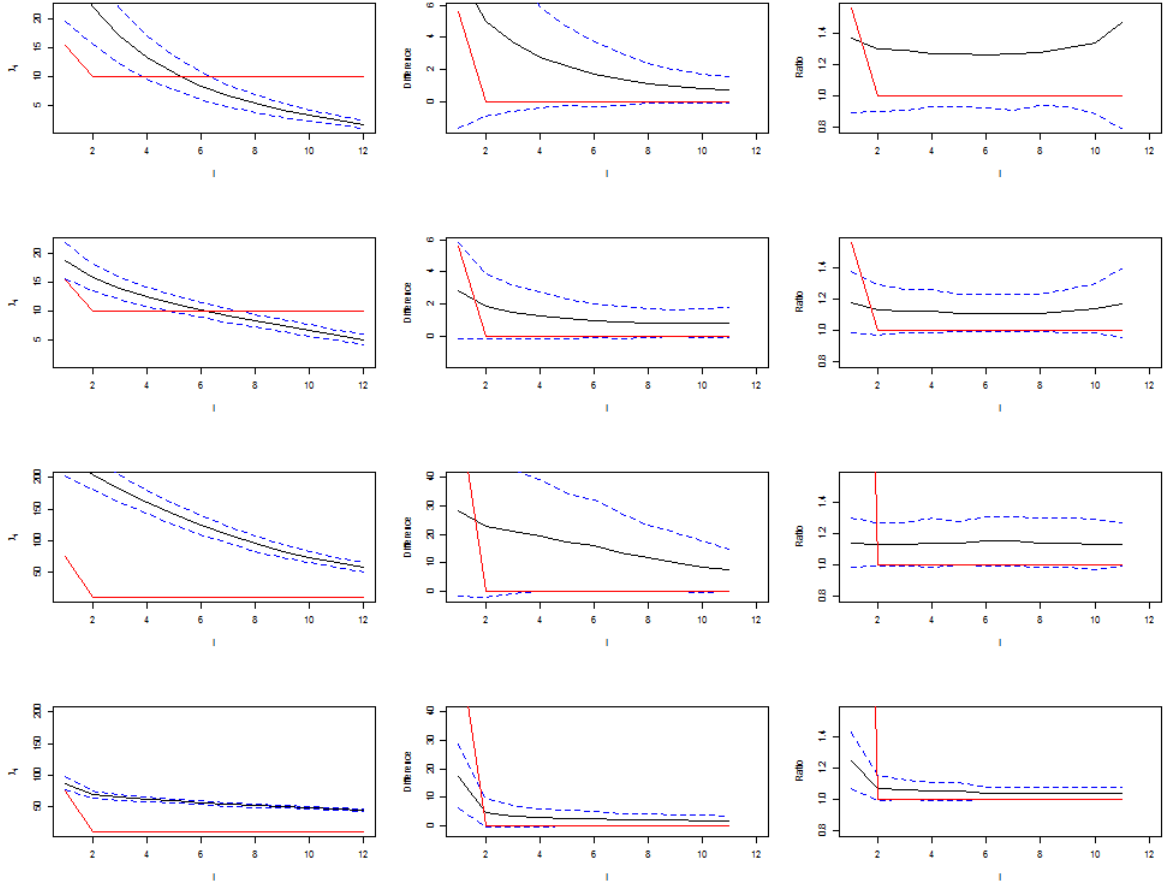


Figure 2: Eigenvalues of DFM with  $r = 1$ ,  $\phi = 1$  and  $\sigma_\eta^2 = 1$  when the idiosyncratic noises are AR(1) process with  $\gamma = -0.8$  and  $\sigma_a^2 = 1$ . The first column plots the eigenvalues while the second and third column plot their differences and ratios respectively. The 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  $N = 200, T = 100$  and fourth row  $N = 200, T = 500$ .

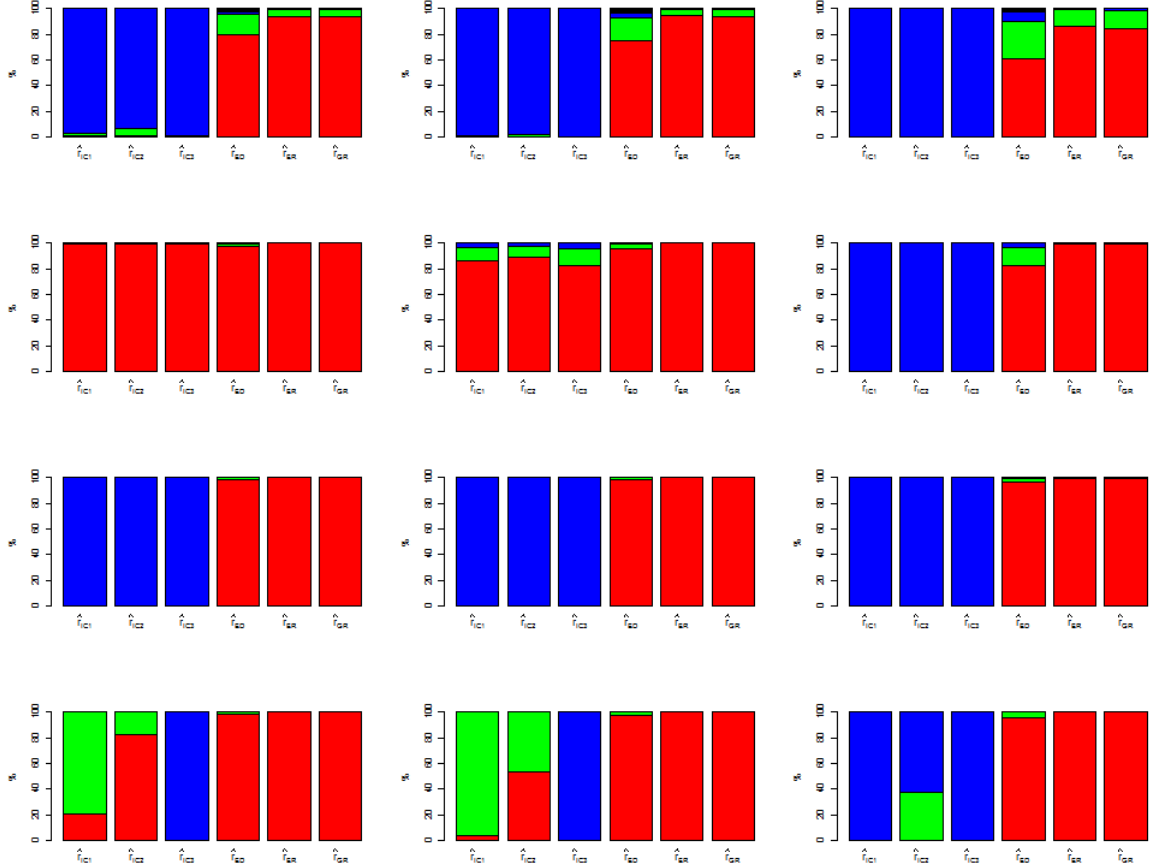


Figure 3: Percentage of  $\hat{r} = r_{\max}$  (blue),  $r_{\max} > \hat{r} > r$  (green),  $\hat{r} = 1$  (red) and  $\hat{r} = 0$  (black) in a DFM with  $r = 1$ ,  $\phi = 1$ ,  $\sigma_{\eta} = 1$ ,  $\gamma = -0.8$  and  $\sigma_a^2 = 0.1$ . First row  $N = 12$ ,  $T = 100$ ; second row  $N = 12$ ,  $T = 500$ ; third row  $N = 200$ ,  $T = 100$ ; fourth 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.

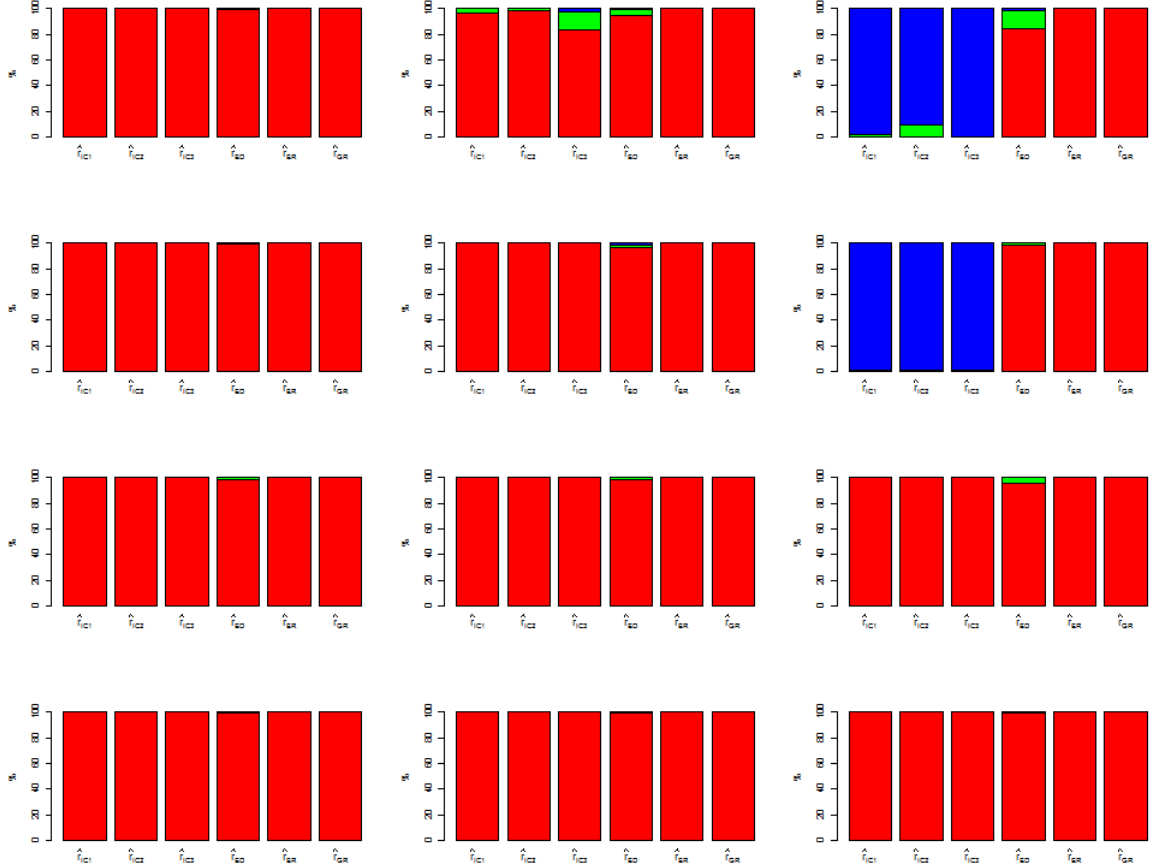


Figure 4: Percentage of  $\hat{r} = r_{\max}$  (blue),  $r_{\max} > \hat{r} > r$  (green),  $\hat{r} = 1$  (red) and  $\hat{r} = 0$  (black) in a DFM with  $r = 1$ ,  $\phi = 1$ ,  $\sigma_{\eta} = 1$ ,  $\gamma = 1$  and  $\sigma_a^2 = 1$ . First row  $N = 12$ ,  $T = 100$ ; second row  $N = 12$ ,  $T = 500$ ; third row  $N = 200$ ,  $T = 100$ ; fourth 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.

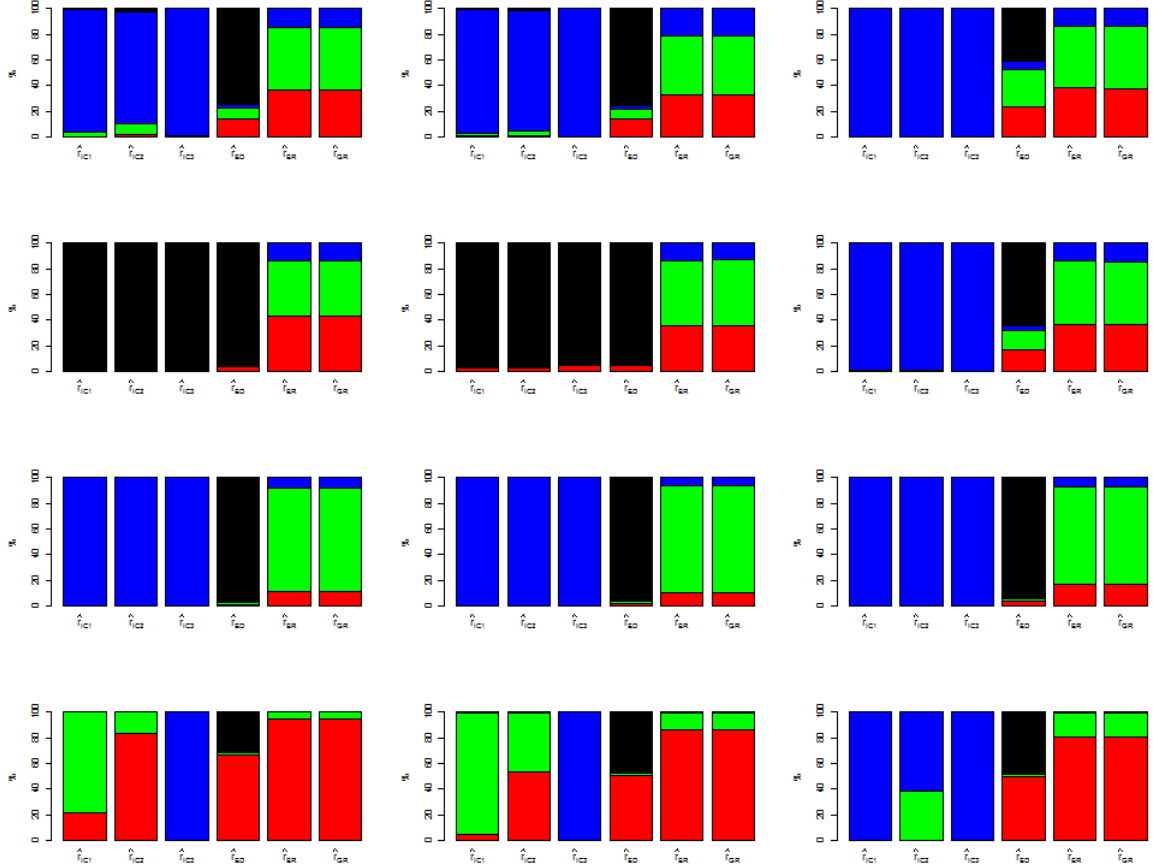


Figure 5: Percentage of  $\hat{r} = r_{\max}$  (blue),  $r_{\max} > \hat{r} > r$  (green),  $\hat{r} = 1$  (red) and  $\hat{r} = 0$  (black) in a DFM with  $r = 1$ ,  $\phi = 1$ ,  $\sigma_{\eta} = 1$ ,  $\gamma = -0.8$  and  $\sigma_a^2 = 1$ . First row  $N = 12$ ,  $T = 100$ ; second row  $N = 12$ ,  $T = 500$ ; third row  $N = 200$ ,  $T = 100$ ; fourth 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.

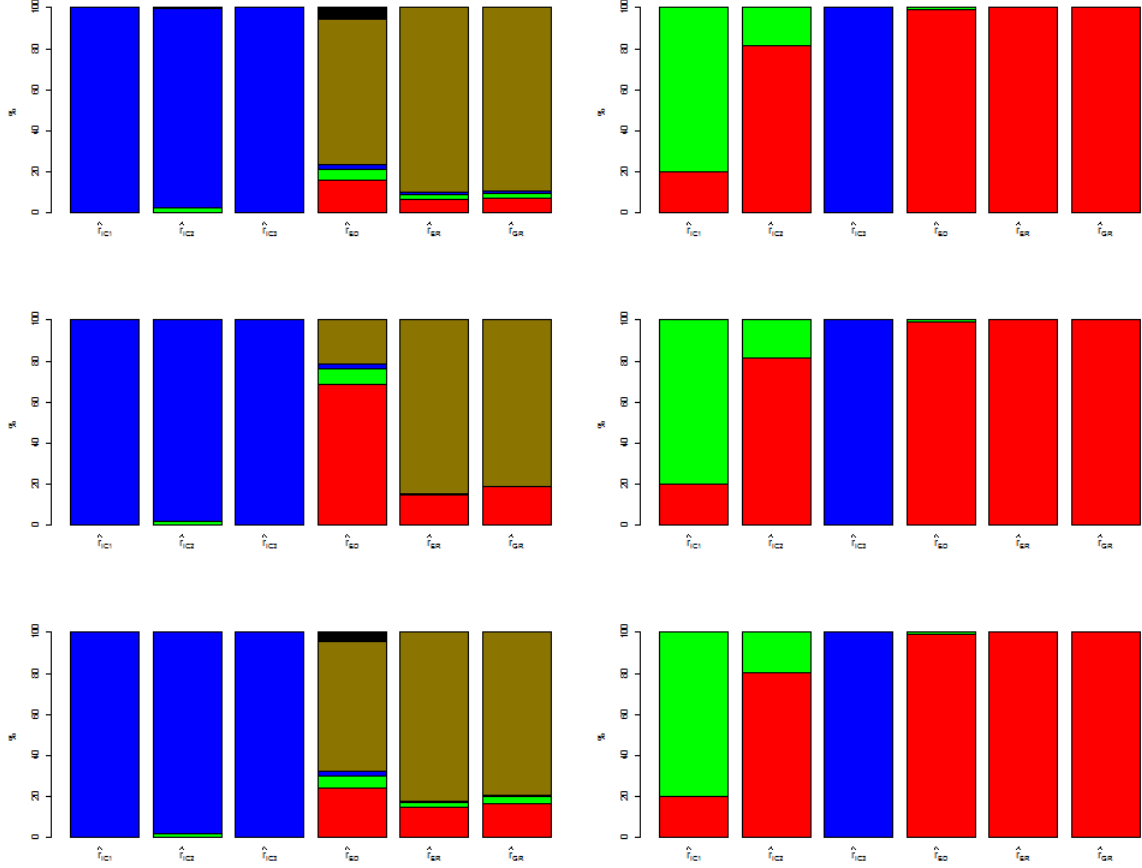


Figure 6: Percentage of  $\hat{r} = r_{\max}$  (blue),  $r_{\max} > \hat{r} > r$  (green),  $\hat{r} = 2$  (red),  $\hat{r} = 1$  (gold) and  $\hat{r} = 0$  (black) in a DFM with  $r = 2$ ,  $\gamma = -0.8$  and  $\sigma_a^2 = 1$ . System dimensions  $N = 12$ ,  $T = 100$  (first column);  $N = 200$ ,  $T = 500$  (second column). The factors are two random walks with variance  $\sigma_\eta^2 = 1$  (first row); two random walks with variances  $\sigma_{\eta_1} = 1$  and  $\sigma_{\eta_2} = 5$  (second row) and a random walk with variance  $\sigma_{\eta_1} = 1$  and a stationary factor with  $\sigma_{\eta_2} = 1$ .



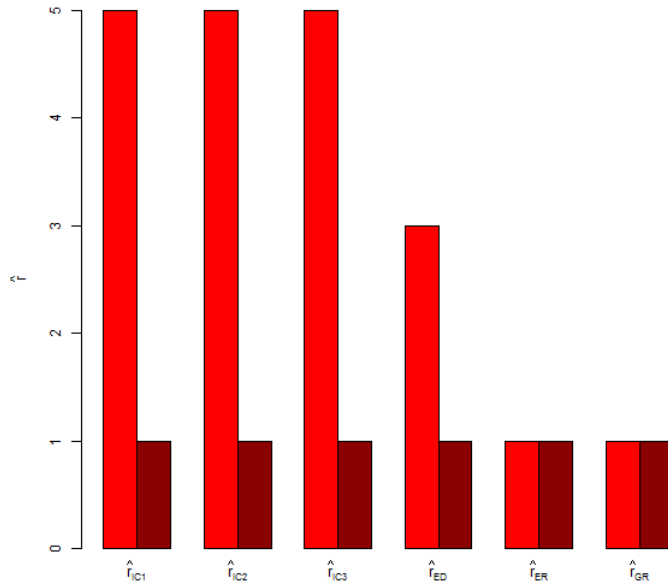


Figure 7: Number of factors detected in the inflation rates system when the procedures are implemented to the levels (red) and to differenced series (dark red).

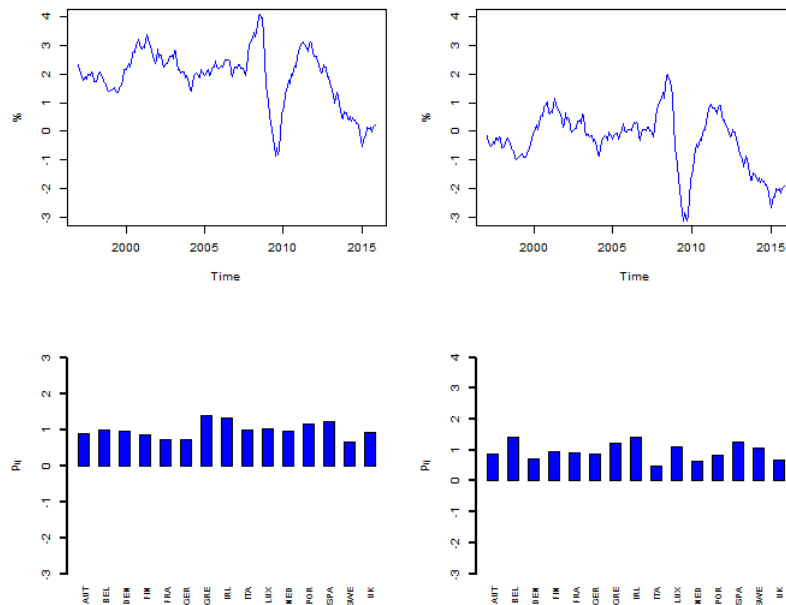


Figure 8: PC estimated factor (first row) and factor weights (second row) obtained assuming  $r = 1$  and using original inflation rates (first column) and differenced rates (second column).

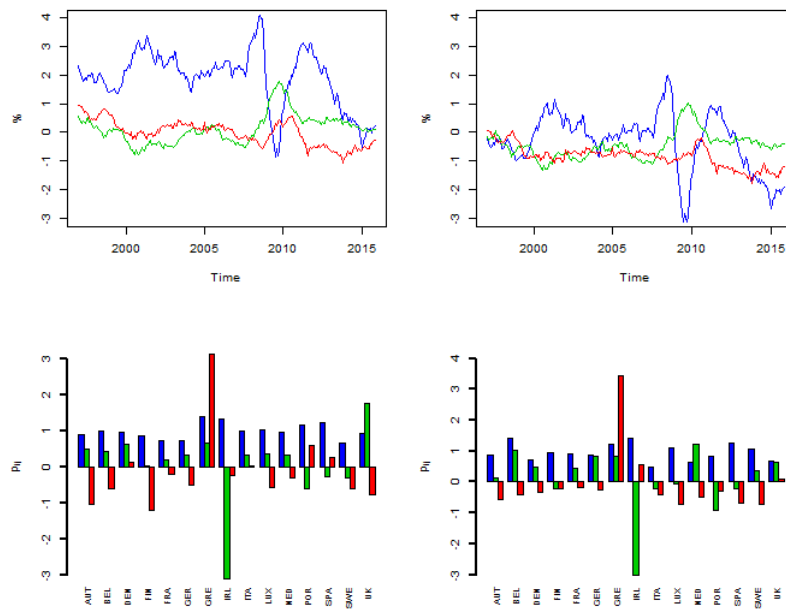


Figure 9: PC estimated factors (first row) and corresponding factor weights (second row) obtained assuming  $r = 3$  and using original inflation rates (first column) and differenced rates (second column).