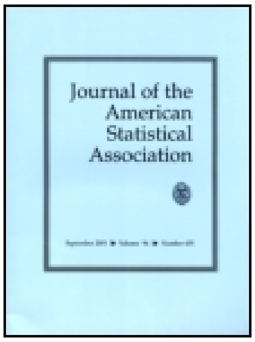

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Generalized Dynamic Principal Components

Daniel Pe a & Victor J. Yohai

Generalized Dynamic Principal Components

Daniel Peña and Victor J. Yohai

Daniel Peña is Professor, Statistics Department, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, España, (E-mail: daniel.peña@uc3m.es). Víctor J. Yohai is Professor Emeritus, Mathematics Department, Faculty of Exact Sciences, Ciudad Universitaria, 1428 Buenos Aires, Argentina (E-mail: victoryohai@gmail.com). This research was partially supported by Grant ECO2012-38442 from MINECOM, Spain, and Grants W276 from Universidad of Buenos Aires, PIP's 112-2008-01-00216 and 112-2011-01- 00339 from CONICET and PICT 2011-0397 from ANPCYT, Argentina. We are grateful to an Associate Editor and two anonymous referees, whose comments lead to important improvements in the paper

ACCEPTED MANUSCRIPT

ABSTRACT

Brillinger defined dynamic principal components (DPC) for time series based on a reconstruction criterion. He gave a very elegant theoretical solution and proposed an estimator which is consistent under stationarity. Here we propose a new enterally empirical approach to DPC. The main differences with the existing methods -mainly Brillinger procedure- are (i) the DPC we propose need not be a linear combination of the observations and (ii) it can be based on a variety of loss functions including robust ones. Unlike Brillinger, we do not establish any consistency results; however, contrary to Brillinger's, which has a very strong stationarity flavor, our concept aims at a better adaptation to possible nonstationary features of the series. We also present a robust version of our procedure that allows to estimate the DPC when the series have outlier contamination. We give iterative algorithms to compute the proposed procedures that can be used with a large number of variables. Our non robust and robust procedures are illustrated with real data sets. Supplemental Material containing mathematical derivations is available on line

Key words: reconstruction of data; vector time series; dimensionality reduction.

1 Introduction

Dimension reduction is very important in vector time series because the number of parameters in a model grows very fast with the dimension m of the vector of time series. Therefore, finding simplifying structures or factors in these models is important to reduce the number of parameters required to apply them to real data. Besides, these factors, as we will see in this paper, may allow to reconstruct with a small error the data sets and therefore to reduce the amount of information to be stored. Dimension reduction is usually achieved by finding linear combinations of the time series variables which have interesting properties. Suppose the time series vector $\mathbf{z}_t = (z_{1,t}, ..., z_{m,t})$, where

$1 \leq t \leq T$, and we assume, for simplicity, that $\bar{\mathbf{z}} = T^{-1}\sum_{t=1}^{T}\mathbf{z}_t$, which will estimate the mean if the process is stationary, is zero and $\mathbf{Z}$ the $T \times m$ matrix whose rows are $\mathbf{z}_1, ..., \mathbf{z}_T$. Let $\mathbf{C} = \mathbf{Z}'\mathbf{Z}/T$, be the sample covariance matrix, $\lambda_1 \geq \lambda_1 \geq \lambda_m$ eigenvalues of C and $\widehat{\mathbf{v}}_i = (\widehat{v}_{1,i}, ..., \widehat{v}_{m,i})'$, $1 \leq i \leq m$, the corresponding eigenvectors satisfying $\mathbf{v}_i'\mathbf{v}_j' = \delta_{ij}$. Then

$$\widehat{\mathbf{p}}_i = (\widehat{p}_{i,t}, ..., \ , \widehat{p}_{i,t})' = \mathbf{Z}\widehat{\mathbf{v}}_i, \ 1 \leq i \leq m \tag{1}$$

is the i-th principal component of $\mathbf{C}$. Let $k < m$, $\widehat{\mathbf{P}}_k$ the $k \times T$ matrix with rows equal $\mathbf{p}_1', ..., \mathbf{p}_k'$ and call $\widehat{\mathbf{V}}_k$ the $m \times k$ matrix whose i-th row is $\mathbf{v}_i'$. Let $\mathbf{P}_k = (p_{i,t})$ be any $k \times T$ matrix and $\mathbf{V}_k = (v_{j,i})$ any $m \times k$ matrix and suppose that we reconstruct all the data $z_{j,t}$ using the k columns of $\mathbf{V}_k$ by $\sum_{i=1}^{k} v_{j,i}p_{i,t}$. Then the mean squared error of the reconstructed series is

$$\mathrm{MSE}(\mathbf{P}_k, \mathbf{V}_k) = (1/mT)\sum_{j=1}^{m}\sum_{t=1}^{T}(z_{j,t} - \sum_{i=1}^{k} v_{j,i}p_{i,t})^2.$$

Okamoto and Kanasawa (1968) showed that

$$(\widehat{P}_k, \widehat{V}_k) = \arg\min_{\mathbf{P}_k \in \mathbb{R}^{k\times m}, \mathbf{V}_k \in \mathbb{R}^{k\times m}} \mathrm{MSE}(P_k, V_k). \tag{2}$$

Note that in (2), the minimization is performed on all possible $k \times m$ matrices $\mathbf{P}_k$, and therefore it is not required that the element $p_{i,t}$ of $\mathbf{P}_k$ are linear combinations of $\mathbf{z}_t$. However, the elements of the optimal matrix $\widehat{P}_k$ given by (1) have this property. One drawback of this reconstruction is that it is static, that is, to reconstruct an observation of period t only the values of the components in that period are used.

Ku, Storer and Georgakis (1995) proposed to apply classical principal components to the augmented observations $\mathbf{z}_t^* = (\mathbf{z}_{t-h}', \mathbf{z}_{t-h+1}', ..., \mathbf{z}_t')'$, $h + 1 \leq t \leq T$, that includes the values of the series up to lag h. These principal components provide linear combinations of the present and past values of the time series with largest variance. However, it is not clear from their definition that these principal components have good reconstruction properties.

An alternative way to find interesting linear combinations was proposed by Box and Tiao (1977) who suggested maximizing the predictability of the linear combinations $c_t = \gamma'\mathbf{z}_t$. Other linear methods for dimension reduction in time series models have been given by the scalar component models, SCM, (Tiao and Tsay, 1989) and the reduced-rank models (Ahn and Reinsel, 1990, Reinsel and Velu, 1998).

Brillinger (1981) addressed the reconstruction problem as follows. Suppose zero mean m dimensional stationary process $\{\mathbf{z}_t\}$, $-\infty < t < \infty$. The dynamic principal components are defined by searching for $m \times 1$ vectors $\mathbf{c}_h$, $-\infty < h < \infty$ and β_j, $-\infty < j < \infty$, so that if we consider as first principal component the linear combination

$$f_t = \sum_{h=-\infty}^{\infty} \mathbf{c}_h' \mathbf{z}_{t-h}, \tag{3}$$

then

$$E\left[(\mathbf{z}_t - \sum_{j=-\infty}^{\infty} \beta_j f_{t+j})'(\mathbf{z}_t - \sum_{j=-\infty}^{\infty} \beta_j f_{t+j})\right]. \tag{4}$$

is minimum. Brillinger elegantly solved this problem by showing that $\mathbf{c}_k$ is the inverse Fourier transform of the principal components of the cross spectral matrices for each frequency, and β_j is the inverse Fourier transform of the conjugates of the same principal components. See Brillinger (1981) and Shumway and Stoffer (2000) for the details of the method. Note that when this procedure is adapted to finite samples the number of lags in (3) and in the reconstruction of the series should be truncated. We can mention two shortcomings of Brillinger's procedure: (i) Brillinger's procedure can be used with nonstationary series, but in this case the mean square error of the reconstructed series may not be close to its possible minimum value (ii) it is not clear how to robustify these principal components using a reconstruction criterion.

A related line of research are factor models for time series. Static factor models assume a contemporaneous relationship between the series and a small number of factors. Some of these models assume stationarity (Peña and Box, 1987, Stock and Watson, 1988, 2002. Bai and Ng, 2002, and Lam and Yao, 2012, among others). All these models use the eigenvalues of the lag covariance matrices of the process and are related to the principal components (PC) of the time series. Some generalizations to the nonstationary case are Peña and Poncela (2006) for integrated processes, Pan and Yao (2008) for general nonstationary processes and Motta, Hafner and von Sachs (2011) and Motta and Ombao (2012) for locally stationary processes.

Dynamic factor models are closely related to dynamic principal components (DPC), because they assume that one important part of the original series can be explained in a dynamic way by a relatively small number of common factors. Forni, Hallin, Lippi, and Reichlin (2000) proposed a very general dynamic factor model allowing for an infinite number of factor lags and low correlation between any two idiosyncratic components. They show that the common component of the series can be consistently estimated by increasing the number of series to infinity. This estimator is obtained by projecting the data in the first q dynamic principal components which include leads and lags. These principal components are obtained as in Brillinger by the Fourier transforms of the eigenvectors of the spectral matrix. This model is applied for prediction in Forni, Hallin, Lippi, and Reichlin (2005), where the authors proposed a one sided method of estimation of a dynamic factor model for forecasting. The forecasts generated with this procedure improve the ones derived by Stock and Watson (2002) using static principal components. Forni, Giannone, Lippi, and Reichlin (2009) proposed a model that can be seen either as a static model with r factors or as a dynamic model with q factor with $q < r$ and develop estimation methods for the factor structure with finite-dimensional factor space. Forni and Lippi (2011) proposed a one sided solution for the general dynamic factor model of Forni at al (2000). Forni, Hallin, Lippi and Zaffaroni (2015) proposed a model with possibly infinite-dimensional factor spaces and obtained a one-sided representation for the dynamic factor model. While some of these models (Stock and Watson, Bai and Ng, 2002, Forni et al., 2009) make assumptions on the relation between the series under study and the way factors are loaded, other ones (Forni et al., 2000, Forni et al., 2015) do not make any assumption of that type. Hallin and Lippi (2013) give a general presentation of the methodological foundations of dynamic factor models.

The procedure we propose is different from these approaches as follows: (1) it is entirely data-analytic and does not assume any given model; (2) it does not assume a fixed number of factors to

ACCEPTED MANUSCRIPT

be identified. Instead, the number of components is chosen to achieve a desired degree of accuracy in the reconstruction of the original series. These differences are the usual ones between the two classical approaches for dimension reduction: principal components and factor models.

In this paper we address the sample reconstruction of a vector of time series from the DPCs by using a finite number of lags. Even if we do not prove consistency, some interesting features of our procedure with respect to previous ones are: (i) the DPC we propose do not need to be a linear or stationary combination of the data and consequently, it may lead to a better adaptation to a possible nonstationary behavior of the series and (ii) it can be easily robustified by changing the loss function to minimize, for example, from the mean square error criterion to a robust scale.

The remaining of this article is organized as follows. In Section 2 we describe the proposed dynamic principal components of a vector time series based on a reconstruction criterion. In Section 3 we study the particular case where the proposed dynamic principal components depend only on one lag. In Section 4 we show the results of a Monte Carlo study that compares the proposed dynamic principal components procedure with the ordinary principal components used in a dynamic way and with the Brillinger's DPC procedure. The comparison is done for both: stationary and non stationary series. We also compare our procedure with the one developed by Forni et al. (2009) for the factor models introduced in this work. We show the good performance of our proposal in all these models, including the case of very large number of series. We also apply our procedure to two nonstationary vector time series real examples. In Section 5 we define robust dynamic principal components using a robust reconstruction criterion and illustrate in one example the good performance of this estimator to drastically reduce the influence of outliers. In Section 6 some final conclusions are presented. Supplemental Material containing mathematical derivations is available on line.

ACCEPTED MANUSCRIPT

2 Finding time series with optimal reconstruction properties

Suppose that we observe $z_{j,t}, 1 \leq j \leq m, 1 \leq t \leq T$, and consider two integer numbers $k_1 \geq 0$ and $k_2 \geq 0$. We can define the first dynamic principal component with k_1 lags and k_2 leads as a vector $\mathbf{f} =(f_t)_{-k_1+1 \leq t \leq T+k_2}$, so that the reconstruction of series $z_{j,t}, 1 \leq j \leq m$, as a linear combination of $\mathbf{f} = (f_{t-k_1}, f_{t-k_1+1}, \ldots . f_t, f_{t+1}, \ldots, f_{t+k_2})$ is optimal with the mean square error (MSE) criterion. More precisely, given a possible factor $\mathbf{f}$, a $m \times (k_1 + k_2)$ matrix of coefficients $\gamma = (\gamma^*_{j,i})_{1 \leq j \leq m, -k_1 \leq i \leq k_2}$,and $\alpha = (\alpha_1, \ldots, \alpha_m)$, the reconstruction of the original series $z_{j,t}$ is defined as

$$\widehat{z}_{j,t} = \sum_{i=-k_1}^{k_2} \gamma_{j,i} f_{t+i} + \alpha_j.$$

Let $k = k_1 + k_2$ and put

$$f^*_t = f_{t-k_1}, 1 \leq t \leq T + k, \ \beta^*_{j,h} = \gamma_{j,h-k_1-1}, 1 \leq h \leq k + 1.$$

and also define

$$f^{**}_t = f^*_{t+k}, 1 - k \leq t \leq T, \beta^{**}_{j.,h} = \beta^*_{j,k+2-h}, 1 \leq h \leq k + 1. \tag{5}$$

Then, the reconstructed series can also be obtained as

$$\widehat{z}_{j,t} = \sum_{i=-k_1}^{k} \beta_{j,i} f_{t+i+k_1} + \alpha_j = \sum_{h=0}^{k} \beta^*_{j,h+1} f^*_{t+h} + \alpha_j = \sum_{h=0}^{k} \beta^{**}_{j,h+1,} f^{**}_{t-h} + \alpha_j.$$

Then, without generality we can use indistinctly k lags or k leads of the principal component to reconstruct the series. Although the reconstruction of the series with k lags is intuitively more appealing , we will derive the optimal solution for the case of k leads. The reason for this is that to derive the optimal solution using lags requires to deal with more cumbersome equations due to the occurrence of negative subscripts. Anyway, once obtained the forward optimal solution we can immediately obtain the backward one using (5).

Let $\mathbf{f} = (f_1, \ldots, f_{T+k})'$, $\beta = (\beta_{j,i})_{1 \leq j \leq m, 1 \leq i \leq k+1}$ and $\alpha = (\alpha_1, \ldots \alpha_m)$, then the MSE loss function when we reconstruct the m series using k leads is given by

$$\mathrm{MSE}(\mathbf{f}, \beta, \alpha) = \frac{1}{Tm} \sum_{j=1}^{m} \sum_{t=1}^{T} (z_{j,t} - \widehat{z}_{j,t}(\mathbf{f}, \beta_j, \alpha_j))^2 = \frac{1}{Tm} \sum_{j=1}^{m} \sum_{t=1}^{T} (z_{j,t} - \sum_{i=0}^{k} \beta_{j,i+1} f_{t+i} - \alpha_j)^2. \tag{6}$$

ACCEPTED MANUSCRIPT

Note that this loss function is well defined and makes sense even in the case of nonstationary vector time series. The optimal choices of $\mathbf{f} = (f_1, ..., f_{T+k})'$ and $\beta = (\beta_{j,i})_{1\leq j\leq m, 1\leq i\leq k+1}$, $\alpha = (\alpha_1, ...\alpha_m)$ are defined by

$$(\widehat{\mathbf{f}}, \widehat{\beta}, \widehat{\alpha}) = \arg \min_{\mathbf{f}\in\mathbb{R}^{T+k}, \beta\in\mathbb{R}^{m\times(k+1)}, \alpha\in\mathbb{R}^m} \mathrm{MSE}(\mathbf{f}, \beta, \alpha). \tag{7}$$

Clearly if $\mathbf{f}$ is optimal, $\gamma\mathbf{f}+\delta$ is optimal too. Thus, we can choose $\mathbf{f}$ so that $\sum_{t=1}^{T+k} f_t = 0$ and $(1/(T+k))\sum_{t=1}^{T+k} f_t^2 = 1$. We call $\widehat{\mathbf{f}}$ the first DPC of order k of the observed series $\mathbf{z}_1, ..., \mathbf{z}_t$. Note that the first DPC of order 0 corresponds to the first regular principal component of the data.

Let $\mathbf{C}_j(\alpha_j) = (c_{j,t,q}(\alpha_j))_{1\leq t\leq T+k, 1\leq q\leq k+1}$ be the $(T+k)\times(k+1)$ matrix defined by

$$c_{j,t,q}(\alpha_j) = \begin{cases} (z_{j,t-q+1} - \alpha_j) & \text{if} \quad 1\vee(t-T+1) \leq q \leq (k+1)\wedge t \\ 0 & \text{if} \quad \text{otherwise} \end{cases}, \tag{8}$$

where $a\vee b = \max(a,b)$ and $a\wedge b = \min(a,b)$. Let $\mathbf{D}_j(\beta_j) = (d_{j,t,q}(\beta_j))$ be the $(T+k)\times(T+k)$ given by

$$d_{j,t,q}(\beta_j) = \begin{cases} \sum_{v=(t-k)\vee 1}^{t\wedge T} \beta_{j,q-v+1}\beta_{j,t-v+1} & \text{if} \quad (t-k)\vee 1 \leq q \leq (t+k)\wedge(T+k) \\ 0 & \text{if} \quad \text{otherwise} \end{cases}$$

and

$$\mathbf{D}(\beta) = \sum_{j=1}^{m} \mathbf{D}_j(\beta_j). \tag{9}$$

Differentiating (6) with respect to f_t in Section 1 of the Supplemental Material we derive the following equation

$$\mathbf{f} = \mathbf{D}(\beta)^{-1}\sum_{j=1}^{m} \mathbf{C}_j(\alpha)\beta_j. \tag{10}$$

Obviously, the coefficients β_j and α_j, $1\leq j\leq m$, can be obtained using the least squares estimator, that is

$$\begin{pmatrix} \beta_j \\ \alpha_j \end{pmatrix} = (\mathbf{F}(\mathbf{f})'\mathbf{F}(\mathbf{f}))^{-1}\,\mathbf{F}(\mathbf{f})'\,\mathbf{z}^{(j)}, \tag{11}$$

where $\mathbf{z}^{(j)} = (z_{j,1}, ..., z_{j,T})'$ and $\mathbf{F}(\mathbf{f})$ is the $T \times (k+2)$ matrix with t-th row $(f_t, f_{t+1}, ..., f_{t+k}, 1)$. Then the first DPC is determined by equations (10) and (11). The second DPC is defined as the first DPC of the residuals $r_{j,t}(\mathbf{f}, \beta)$. Higher order DPC are defined in a similar manner.

2.1 Computational algorithm for the DPC

To define an iterative algorithm to compute $(\widehat{\mathbf{f}}, \widehat{\beta}, \widehat{\alpha})$ is enough to give $\mathbf{f}^{(0)}$ and a rule describing how to compute $\beta^{(h)}, \alpha^{(h)}$, $\mathbf{f}^{(h+1)}$ once $\mathbf{f}^{(h)}$ is known. The following two steps based on (10) and (11) describe a natural rule to perform this recursion.

step 1 Based on (11), define $\beta_j^{(h)}$ and $\alpha_j^{(h)}$, for $1 \leq j \leq m$, by

$$\begin{pmatrix} \beta_j^{(h)} \\ \alpha_j^{(h)} \end{pmatrix} = \left(\mathbf{F}(\mathbf{f}^{(h)})'\mathbf{F}(\mathbf{f}^{(h)})\right)^{-1} \mathbf{F}(\mathbf{f}^{(h)})'\mathbf{z}^{(j)}.$$

step 2 Based on (10), define $\mathbf{f}^{(h+1)}$ by

$$\mathbf{f}^{*} = \mathbf{D}(\beta^{(h)})^{-1}\mathbf{C}(\alpha^{(h)})\beta^{(h)}$$

and

$$\mathbf{f}^{(h+1)} = (\mathbf{T}+k)^{1/2}(\mathbf{f}^{*} - \bar{\mathbf{f}}^{*})/\|\|\mathbf{f}^{*} - \bar{\mathbf{f}}^{*}\|.$$

The initial value $\mathbf{f}^{(0)}$ can be chosen equal to the standard (non dynamic) first principal component, completed with k zeros. We stop the iterations when

$$\frac{\mathrm{MSE}(\mathbf{f}^{(h)}, \beta^{(h)}, \alpha^{(h)}) - \mathrm{MSE}(\mathbf{f}^{(h+1)}, \beta^{(h+1)}, \alpha^{(h+1)})}{\mathrm{MSE}(\mathbf{f}^{(h)}, \beta^{(h)}, \alpha^{(h)})} < \varepsilon$$

for some value ε.

Remark 1. Note that the dimension of the matrices to be inverted to compute $\mathbf{f}^{(h)}, \beta^{(h)}, \alpha^{(h)}$ are independent of the number of time series and therefore we can deal with large number of variables.

ACCEPTED MANUSCRIPT

Remark 2. Note also that there are no restrictions on the values of $\mathbf{f}$ and in particular we do not assume, as in Brillinger, that their components must be linear combinations of the series. In this way the values of $\mathbf{f}$ can be adapted to the nonstationarity character of the time series.

Assume we are considering p dynamic principal components of order k and let $\beta_{j,i,s}$ $1 \leq j \leq m, 1 \leq i \leq k+1$, be the coefficient $\beta_{j.i}$ corresponding to the $s-$th component, $1 \leq s \leq p$. Then, the number of values required to reconstruct the original series are the $(T+k)p$ values of the p factors plus $(k+1)mp$ values for the coefficients $\beta_{j,i,s}$ plus the m intercepts α_j. Thus the proportion of the original information required to reconstruct the series is $((T+k)p+(k+1)mp+m)/mT$ and when T is large compared to k and m, this ratio is close to p/m. In practice, the number of lags to reconstruct the series, k, and the number of principal components, p, need to be chosen. Of course the accuracy of the reconstruction improves when any of these two numbers is enlarged, but also the size of the information required will also increase. For large T increasing the number of components introduces more values to store than increasing the number of lags. However, we should also take into account the reduction in MSE due to enlarging each of these components. In general, increasing the number of lags after some point will have a negligible effect on the MSE. Then, if the level of the MSE is larger than desired, a new component should be added. Thus one possible strategy is to start with one principal component and increase the number of lags until the reduction of further lags is smaller than some value ϵ. Then a new principal component is introduced and the same procedure is applied. The process stops when the MSE reaches some satisfactory value. Note that this rule is similar to what is generally used for determining the number p in ordinary principal components.

ACCEPTED MANUSCRIPT

3 Dynamic Principal Components when $k = 1$

To illustrate the computation of the first DPC, let us consider the simplest case of $k = 1$. Then, we search for $\widehat{\beta}$=$(\widehat{\beta}_{ji})_{1\leq j\leq m,1\leq i\leq 2}$ and $\widehat{\mathbf{f}}$= $(\widehat{f}_1, ..., \widehat{f}_{T+1})'$ such that

$$(\widehat{\mathbf{f}},\widehat{\beta}) = \arg\min_{1} \sum_{t=1}^{T}\sum_{j=1}^{m}(z_{j,t} - \beta_{j,1} f_t - \beta_{j,2} f_{t+1})^2. \tag{12}$$

Suppose now that $\mathbf{z}_t$ is stationary, then in Section 2 of the Supplemental Material is shown that, except in both ends, $\widehat{f}_t$, can be approximated by

$$\widehat{f}_t^* = \frac{1}{\alpha}\left[\sum_{j=1}^{m}\widehat{\beta}_{j,1}\sum_{q=-\infty}^{\infty}c^{|t-q|}z_{j,q} + \sum_{j=1}^{m}\widehat{\beta}_{j,2}\sum_{q=-\infty}^{\infty}c^{|t-q|}z_{j,q-1}\right], \tag{13}$$

where $|c| < 1$. Therefore the DPC is approximated by linear combinations of the stationary series $z_{j,t} + \sum_{i=1}^{\infty}c^i(z_{j,t+i} + z_{j,t-i})$, and $z_{j,t-1} + \sum_{i=1}^{\infty}c^i(z_{j,t-1+i} + z_{j,t-1-i})$, $1 \leq j \leq m$. These series give the largest weight to the periods t and $t-1$ respectively and the weights decrease geometrically when we move away of these values. We conjecture that in the case of the first DPC of order k, a similar approximation outside both ends of $\widehat{f}_t$ by an stationary process can be obtained.

4 Monte Carlo simulation and examples

4.1 Simulation results for the stationary case

We perform a Monte Carlo study using as vector series $\mathbf{z}_t = (z_{1,t}, z_{2,t}, ..., z_{m,t})'$, $1 \leq t \leq T$ generated as follows:

$$z_{i,t} = 10\sin(2pi(i/m))f_t + 10\cos(2pi(i/m))f_{t-1} + 10(i/m)f_{t-2} + u_{i,t}, 1 \leq i \leq m, 1 \leq t \leq T, \tag{14}$$

where $f_t, -2 \leq t \leq T$ and $u_{i,t}, \leq t \leq T$, $1 \leq i \leq m$ are i.i.d. random variables with distribution N(0, 1). We compute three different principal components: (i) The ordinary principal component used in a dynamic way with k lags to reconstruct the original series (OPC_k) (ii) the dynamic principal component (DPC_k) proposed here, (iii) Brillinger dynamic principal components (BDPC_k) adapted for finite samples as follows:

$$f_t = \sum_{i=-k}^{k} \mathbf{c}_i' \mathbf{z}_{t-i}, \; k+1 \leq t \leq T-k, \tag{15}$$

where $\mathbf{c}_k$ are the coefficients defined below (4) in Section 1. In this simulation we used OPC_2, DPC_2 and BDPC_{10}. To reconstruct the original series with the three procedures we used least squares. The BDPC_k components were computed using the R code developed for Hörmann, Kidziński and Hallin (2014) that was kindly provided by the authors. However, we were not able to run this program for dimension $m = 1000$ because of lack of enough memory (our computer has 8 GB of installed memory). We performed for each case 1000 Monte Carlo replications and in Table 1 we show the MSE of the reconstructed series.

We observe that the performances of DPC_2 and BDPC_{10} are quite similar and that the MSEs are close to one, that is, equal to the variance of the error terms $u_{i,t}$s. We also observe that the MSE of the reconstructed series with the OPC_2 procedure is much larger.

In Figure 1 we plot the estimated factor loadings for a data set satisfying (14) with $T = 200$ and $m = 1000$.

We observe that the estimated loadings follow a pattern quite close to the one satisfied by the true values: sin, cosine and a linear trend.

4.2 Simulation results in the nonstationary case: VARI$(1, 1)$ model.

In this case we consider a VARI$(1, 1)$ $m-$dimensional vector series $\mathbf{z}_t$ generated as follows. Consider an stationary VAR(1) model $\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{u}_t$, $1 \leq t \leq T$, where the $\mathbf{u}_t$s are i.i.d. m-dimensional vectors with distribution $\mathrm{N}_m(\mathbf{0}, \mathbf{I})$ and let $\mathbf{z}_t = \mathbf{z}_{t-1} + \mathbf{x}_t$. We consider 1000 replications and in each replication we generate a new matrix A of the form $A = V\Lambda V'$, where V is an orthogonal matrix generated at random with uniform distribution and Λ is a diagonal matrix, where the diagonal elements are independent with uniform distribution in the interval $[0, 0.9]$. We took $m = 20, 100$ and 200 and $T = 400$. We obtained the first principal component using the OPC_{10}, DPC_{10} and BDPC_{10} procedures. In Table 2 we show the percentage of explained variance for each procedure. We consider that, since in this case we are not dealing with a factor model, this measure of performance is easier to interpret than the MSE of the reconstructed series.

We observe that the best performance is achieved by the DPC_{10}. We also tried to use the different procedures with larger number of lags obtaining essentially the same results.

4.3 Simulation results using the factor model in Forni et al. (2009)

In this subsection we compare by means of a Monte Carlo simulation the dynamic principal components procedures with the one developed by Forni et. al (2009) (FGLR) for a special class of factor models that may be seen either as a static model with r factors or as a dynamic model with q factor with $q < r$. In our Monte Carlo study we consider a vector series $\mathbf{z}_t = (z_{1,t}, z_{2,t}, ..., z_{m,t})'$, $1 \leq t \leq T$, where

$$z_{i,t} = \sin(2\pi i/m) f_{1t} + \cos(2i/m) f_{2t} + u_{i,t}, 1 \leq i \leq m, 1 \leq t \leq T,$$

where all the $u_{i,t}$ are independent with distribution $N(0, 1)$. The vector of static factors $\mathbf{f}_t = (f_{1t}, f_{2t})$ satisfies the autoregressive model $\mathbf{f}_t = \mathbf{A}\mathbf{f}_{t-1} + v_t\mathbf{b}$,$1 \leq t \leq T$, where all v_t are independent N(0, 1), A is a 2×2 diagonal matrix with diagonal equal to $(-0.8, 0.7)$ and $\mathbf{b} = (1, 1)'$. We took $m = 5$, 100 and 1000, $T = 200$ and 400 and the number of replications was 1000. Note that in this case $r = 2$ and $q = 1$. In Table 3 we show the MSEs of the reconstructed series. The code for the FGLR procedure was kindly provided by the authors of Forni et al. (2009). As was explained in Subsection 4.1, we were not able to run the BDPC_{10} procedure with $m = 1000$.

We observe that for small m the MSEs of the reconstructed series with the principal components procedures are smaller than those reconstructed with the FGLR. Instead, for large m, both reconstruction errors are close to the variance of the idiosyncratic component $u_{i,t}$. The reason for this is that the goal of factor analysis models is accounting for cross-correlations (the instantaneous ones, in the static case, the instantaneous and lagged ones in the dynamic case) while the goal of principal components is dimension reduction, that is, to obtain an approximate reconstruction of the original data based on an small number of unobserved variables, However, it can be proved that when m increases the MSEs of the reconstructed series using dynamic factor analysis or dynamic principal components converge to the variance of the idiosyncratic component. Therefore the comparison between both principal component methods and the FGLR factor analysis method is only relevant when m is large. The results presented in Table 3 indicates that both DPC and FGLR give very good results in this case.

4.4 Example 1

We illustrate the DPC with a small data sets with six series corresponding to the Industrial Production Index (IPI) of France, Germany, Italy, United Kingdom, USA and Japan. We use monthly data from January 1991 to December 2012 and the data are taken from Eurostat. The six series are plotted in Figure 2.

In Table 4 we show the percentage of the variance explained by the OPC$_k$ and DPC$_k$ procedures for $k = 1, 5, 10, 12$ and for the BDPC$_{12}$ procedure. The reason to take only $k = 12$ for the BPC is to be close to the original Brillinger definition.

We note that the reconstruction of the series using the DPC is notably better that the one obtained by means of the OPC and BDPC procedures with the same lags. Increasing the number of lags obviously improves the reconstruction obtained by both components, although the improvement is larger with the DPC.

Figure 3 shows the boxplots of the absolute value of the reconstructed series errors for the OPC$_{12}$ and DPC$_{12}$ procedures. Note that the reconstruction errors are significantly smaller when

ACCEPTED MANUSCRIPT

using the DPC_{12} procedure.

4.5 Example 2.

In this example the data set is composed of 30 daily stock prices in the stock market in Madrid corresponding to the 251 trading days of the year 2004. The source of the data is the Ministry of Economy, Spain. In Table 5 we show the percentage of the variance explained by the different procedures using the OPC_k,DPC_k BDPC_k procedures.

We observe that the best performance corresponds to the DPC_k procedure. As shown in Table 5 including lags in the OPC does not make much difference in the results, but it has an important effect on the DPC. In Figure 4 we show the percentage of the variance explained be the DPC_5 against the one explained by the $\mathrm{DPC}_{5.}$ for the 30 stock prices. We note the for most of the variables the best performance correspond to the DPC_5 procedure.

Figure 5 presents the first OPC_5 and DPC_5. The first DPC_5, which is much smoother than the first OPC_5, seems to be very useful to represent the general trend of the set of time series.

5 Robust Generalized DPCs

As most procedures defined by minimizing the mean square error, the DPC given by (7) is not robust. In fact a very small fraction of outliers may have an unbounded influence on $(\mathbf{f}, \alpha, \beta)$. The procedure proposed by Brillinger seems to be very difficult to robustify. At first sight, it may seems that it may be robustified by using a robust estimator of the spectral matrix. For example, Spangl and Dutter (2005) and Li (2012) proposed robust estimators for the spectral matrix. However this is not enough to obtain robust DPCs. In fact, a robust estimator of the spectral matrix only guarantees the robustness of the coefficients in the linear combinations defining the principal components. However, the result of applying these coefficients to the original series may be largely affected by outlying observations. Instead, the DPC procedure defined by (7) is easier to robustify. An standard way to obtain robust estimates for many statistical models is to replace the minimization of the mean square scale for the minimization of a robust M-scale. This strategy was used for many statistical models, including among others linear regression (Rousseeuw and Yohai, 1984), the estimation of a scatter matrix and multivariate location for multivariate data (Davis, 1987) and to estimate the ordinary principal components (Maronna, 2005). The estimators defined by means of a robust M-scale are called S-estimators. In this section we introduce S-estimators for DPC.

Special care is required for time series with strong seasonality. The reason is that when the values corresponding to a particular season are very different to the others they will be considered as outliers by a robust procedure, and therefore they will be downweighted. As a consequence, the reconstruction of this season may be affected by large errors. Thus, the procedure we present here assumes that the series have been adjusted by seasonality to avoid this problem.

5.1 Generalized S -DPCs

Let ρ_0 be a symmetric, non-decreasing function for $x \geq 0$ and $\rho_0(0) = 0$.Given a sample $\mathbf{x} = (x_1, ..., x_n)$, the M-scale estimator $S(\mathbf{x})$ is defined as the value s solution of

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{x_i}{s}\right) = b. \tag{16}$$

If ρ_0 is bounded, then the breakdown point to ∞ of $S(\mathbf{x})$, that is, the minimum fraction of outliers than can take $S(\mathbf{x})$ to ∞ is $b/\max\rho_0$. Moreover, the breakdown point to 0, that is, the minimum fraction of inliers that can take $S(\mathbf{x})$ to 0, is $1-(b/\max\rho_0)$. Note that if $b/\max\rho_0 = 0.5$ both breakdown points are 0.5 (see section 3.2.2. in Maronna, Martin and Yohai, 2006). In what follows we assume without loss of generality that $\max\rho_0 = 1$. Moreover ρ_0 is chosen so that $E_\phi(\rho_0(\mathbf{x})) = b$,where ϕ is the standard normal distribution. This condition guarantees that for normal samples $S(\mathbf{x})$ is a consistent estimator of the standard deviation. One very popular family of ρ

functions is the Tukey biweight family defined by

$$\rho_c^T(x) = \begin{cases} 1 - \left(1 - (x/c)^2\right)^3 & \text{if} \quad |x| \leq c \\ 1 & \text{if} \quad |x| > c \end{cases}.$$

Given $\mathbf{f} = (f_1, ..., f_{T+k}), \beta_j = (\beta_{j,0}, \beta_{j,1}, ..., \beta_{j,k})$ and α_j, $1 \leq j \leq m$, let $r_{j,t}(\mathbf{f}, \beta_j, \alpha_j) = z_{j,t} - \sum_{i=0}^{k} \beta_{j,i} f_{t+i} - \alpha_j$, $\mathbf{r}_j(\mathbf{f}, \beta_j, \alpha_j) = (r_{j,1}(\mathbf{f}, \beta_j, \alpha_j), ..., r_{j,T}(\mathbf{f}, \beta_j, \alpha_j))$. Let β the $m \times (k+1)$ matrix whose j-th row is β_j, $\alpha = (\alpha_1, ...\alpha_m)$ and

$$\text{SRS}(\mathbf{f}, \beta, \alpha) = \frac{1}{m} \sum_{j=1}^{m} S^2(\mathbf{r}_j(\mathbf{f}, \beta_j, \alpha_j)). \tag{17}$$

We define the first S-DPC $\widehat{\mathbf{f}}$ by

$$(\widehat{\mathbf{f}}, \widehat{\beta}, \widehat{\alpha}) = \arg \min_{\mathbf{f} \in \mathbb{R}^{T+k}, \beta \in \mathbb{R}^{m \times (k+1)}, \alpha \in \mathbb{R}^m} \text{SRS}(\mathbf{f}, \beta, \alpha). \tag{18}$$

Note that $\widehat{\beta}$ and $\widehat{\alpha}$ are the coefficients that should be used to reconstruct the $z_{j,t}$'s from $\widehat{\mathbf{f}}$.

We can observe that the only difference with the definition given in (7) is that instead of minimizing the MSE of the residuals, we minimize the sum of squares of the robust M-scales applied to the residuals of the m series. Put $\psi = \rho'$, $w(u) = \psi(u)/u$,

$$s_j = s_j(\mathbf{f}, \beta_j, \alpha_j) = S(\mathbf{r}(\mathbf{f}, \beta_j, \alpha_j)). \tag{19}$$

Then s_j satisfies

$$\frac{1}{T} \sum_{t=1}^{T} \rho \left(\frac{z_{v-} \sum_{i=0}^{k} \beta_{j,i+1} f_{v+i} - \alpha_j}{s_j} \right) = b. \tag{20}$$

Define the weights

$$w_{j,t} = w_{j,t}(\mathbf{f}, \beta_j, \alpha_j) = w\left(\frac{r_{j,t}(\mathbf{f}, \beta_j)}{s_j}\right), \ 1 \leq j \leq m, \ 1 \leq t \leq T \tag{21}$$

and

$$W_{j,t,v} = W_{j,t,v}(\mathbf{f}, \beta, \alpha, s) = \frac{s_j^2 w_{j,v}(\mathbf{f}, \beta_j, \alpha_j, s_j)}{\sum_{h=(t-k)\vee 1}^{t \wedge T} w_{j,h}(\mathbf{f}, \beta_j, \alpha_j, s_j) r_{j,h}^2}, \tag{22}$$

where $\mathbf{s} = (s_1, ...s_m)$. Let $\mathbf{C}_j(\mathbf{f}, \beta_j, s) = (c_{j,t,q}(\mathbf{f}, \beta_j, s))_{1\leq t\leq T+k, 0\leq q\leq k}$ be the $(T+k)\times(k+1)$ matrix defined by

$$c_{j,t,q}(\mathbf{f},\beta,\alpha,s) = \begin{cases} W_{j,t,t-q+1}(\mathbf{f},\beta,\alpha,\mathbf{s})(z_{j,t-q+1}-\alpha_j) & \text{if} \quad 1\vee(t-T+1)\leq q\leq(k+1)\wedge t \\ 0 & \text{if} \quad \text{otherwise} \end{cases}, \tag{23}$$

$\mathbf{D}_j(\mathbf{f},\beta,\alpha,s) = (d_{j,t,q}(\mathbf{f},\beta,\alpha,s))$ the $(T+k)\times(T+k)$ matrix with elements

$$d_{j,t,q}(\mathbf{f},\beta,\alpha,s) = \begin{cases} \sum_{v=(t-k)\vee 1}^{t\wedge T} W_{j,t,v}\beta_{j,q-v+1}\beta_{j,t-v+1} & \text{if} \quad (t-k)\vee 1\leq q\leq(t+k)\wedge(T+k) \\ 0 & \text{if} \quad \text{otherwise} \end{cases}$$

and

$$\mathbf{D}(\mathbf{f},\beta,\alpha,s) = \sum_{j=1}^{m}\mathbf{D}_j(\mathbf{f},\beta,\alpha,s). \tag{24}$$

Differentiating (20) with respect to f_t we derive in Section 3 of the Supplemental Material the following equation

$$\mathbf{f} = \mathbf{D}(\mathbf{f},\beta,\alpha,s)^{-1}\sum_{j=1}^{m}\mathbf{C}_j(\mathbf{f},\beta,\alpha,s)\beta_j. \tag{25}$$

Let $\mathbf{F}(\mathbf{f})$ be the $T\times(k+2)$ matrix with t-th row $(f_t, f_{t+1}, ..., f_{t+k}, 1)$ and $W_j(\mathbf{f},\beta,s)$ be the diagonal matrix with diagonal equal to $w_{j,1}((\mathbf{f},\beta_j,s), ..., w_{j,T}(\mathbf{f},\beta_j,s)$. Then differentiating (20) with respect to $\beta_{j,i}$ and $\alpha_j$we derive in Section 3 of the Supplemental Material

$$\begin{pmatrix}\beta_j \\ \alpha_j\end{pmatrix} = \left(\mathbf{F}(\mathbf{f})'W_j(\mathbf{f},\beta_j,s)\mathbf{F}(\mathbf{f})\right)^{-1}\mathbf{F}(\mathbf{f})\mathbf{W}_j(\mathbf{f},\beta_j,s)'\mathbf{z}^{(j)}. \tag{26}$$

Then the first S-PDC is determined by equation (19),(25)and (26). Note that the estimator defined by (7) is an S-estimate corresponding to $\rho_0^2(u) = u^2$ and $b = 1$. Then for this case (25) and (26) become (10) and (11) respectively. The second S-DPC is defined as the first S-DPC of the residuals $r_{j,t}(\mathbf{f},\beta)$. Higher order S-DPC are defined in a similar manner.

One important point is the choice of b. At first sight, $b = 0.5$ may seem a good choice, since in this case we are protected against up to 50 % of large outliers. However, the following argument shows that this choice may not be convenient. The reason is that with this choice, the procedure has the so called 50% exact fitting property. This means that when 50 % of the $r_{j,t}(\mathbf{f},\beta_j,\alpha_j)$**s** are zero the scale $S(\mathbf{r}_j(\mathbf{f},\beta_j,\alpha_j))$ is 0 no matter the value of the remaining values. Moreover, if 50 % of the $|r_{j,t}(\mathbf{f},\beta_j,\alpha_j)|$ are small the scale $S(\mathbf{r}_j(\mathbf{f},\beta_j,\alpha_j))$ is small too. Then when $b = 0.5$, the procedure may choose $\mathbf{f},\beta$ and α so to reconstruct the values corresponding to 50% of the periods even if the dataset do not contain outliers.. For this reason it is convenient to choose a smaller value as b, as for example $b = .10$. In that case to obtain $S(\mathbf{r}_j(\mathbf{f},\beta_j,\alpha_j)) = 0$, it is required that 90% of the $r_{j,t}(\mathbf{f},\beta_j,\alpha_j)$**s** be 0. One may wonder why for regression is common to use $b = 0.5$ and the 50% exact fitting property does not cause the problems mentioned above. The reason is that in this case, if there are no outliers, the regression hyperplane fitting 50% of the observations also fits the remaining 50%. This does not occur in the case of the dynamic principal components.

5.2 Computational algorithm for the S-DPC

An iterative algorithm similar to the one described for the DPC in Section 2 can be used to compute the S-DPC. The only difference is that it should be based on (25)and (26) instead on (10) and (11).

The initial value $\mathbf{f}^{(0)}$ can be chosen equal to a regular (non dynamic) robust principal component, for example the one proposed in Maronna (2005). Once $\mathbf{f}^{(0)}$ is computed we can use this value to compute a matrix $\mathbf{F}^{(0)} = \mathbf{F}$ with i-th row $(f_i^{(0)}, f_{i+1}^{(0)}, ..., f_{i+k}^{(0)}, 1)$. The j-th row of $\beta^{(0)}$ and $\alpha_j^{(0)}$can be obtained using a regression S-estimate taking $\mathbf{z}^{(j)}$ as response and $\mathbf{F}^{(0)}$ as design matrix. Finally $s_j^{(0)} = S(\mathbf{r}_j(\mathbf{f}^{(0)},\beta^{(0)})$.

A procedure similar to the one described at the end of Section 2 can be used to determine a convenient number of lags and components replacing the MSE by the SRS.

5.3 Example 3

We will use the data of example 2 to illustrate the performance of the robust DPC. This dataset was modified as follows: each of the 7530 values composing the dataset was modified with probability 0.05 adding 20 to the real value. In Table 6 we show the MSEs of the series reconstructed with the DPC. Since the DPC is very sensitive to the presence of outliers, we also compute the S-DPC. Since the MSE is also very sensitive to outliers, we evaluate the performance of the dynamic principal components procedures using the SRS criterion. We take as ρ the bisquare function with $c = 5.13$ and $b = 0.1$. These values make the M-scale consistent to the standard deviation in the Gaussian case. Table 6 gives the MSEs and the SRSs for the DPC$_k$ and S-DPC$_k$ procedures for $k = 1, 5$ and 10.

We observe that the robust measure of performance SRS corresponding to the S-DPC is much smaller than the one corresponding to the DPC. In Figure 6 we show the boxplots for the first 16 stock prices with the DPC$_k$ and with the S-DPC$_k$. For a better visualization, we have eliminated the outliers larger than 10. These boxplots shows that the S-DPC$_k$ is much less affected by the outliers than the DPC$_k$.The boxplots of the fourteen remaining stocks are similar, but are not shown by shortness sake.

6 Conclusions

We have proposed two dynamic principal components procedures for multivariate time series: the first one using a minimum squared error criterion to evaluate the reconstruction of the original time series and the second one using a robust scale criterion. Both criteria make sense even if in the case of nonstationary series. A Monte Carlo study shows that the proposed dynamic principal component based on the MSE criterion can improve considerably the reconstruction obtained using ordinary principal components in a dynamic way. In the case of stationary series the performance of the proposed procedure is comparable with a finite sample version of Brillinger's approach and in the case of nonstationary series our procedure seems to behave better. We have also shown in an example that the robust procedure based on a robust scale is not much affected by the presence of outliers.

A simple heuristic rule to determine a convenient value for the number of components, p, and the number of lags, k, is suggested. However, further research may lead to better methods to choose these parameters trading off accuracy in the series reconstruction and economy in the number of values stored for that purpose.

Although the proposed DPC seems to be very powerful for data reconstruction they have some limitations for forecasting, because they use information from leads and lags to reconstruct the series, an this is not convenient for forecasting. However, they may be useful to find the dimension of the factors space in factor models and this will be the subject of further research.

7 Supplemental Material

The Supplemental Material available on line contains the proofs of (10), (13), (25) and (26).

References

Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced-rank autoregressive models for multiple time series, *Journal of the American Statistical Association*, **83**, 849–856.

Ahn, S. K. and Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association*, **85**, 813–823.

Bai,J. and Ng, S.(2002). Determining the number of factors in approximate factor models, *Econometrica*, **70**, 191–221,

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

Brillinger, D. R. (1981). *Time Series Data Analysis and Theory*, Expanded edition, Holden-Day, San Francisco.

Davies, P.L. (1987). Asymptotic Behavior of S-Estimators of Multivariate Location Parameters and Dispersion Matrices, *The Annals of Statistics*, 15, 1269-1292.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation, *The Review of Economic and Statistics*, **82**, 540–554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005). The generalized dynamic factor model: one sided estimation and forecasting, *Journal of the American Statistical Association*, **100**, 830–840.

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

Forni, M. and Lippi, M. (2011). The general dynamic factor model: One-sided representation results, *Journal of Econometrics*, **163**, 23– 28.

Forni, M., Hallin, M., Lippi, M. and Zaffaroni, P. (2015). Dynamic Factor Models with Infinite-Dimensional Factor Spaces: One-Sided Representations. *Journal of Econometrics*, **185**, 359-371.

Hallin, M. and Lippi, M.(2013). Factor models in high-dimensional time series —A time-domain approach, *Stochastic Processes and their Applications*, **123**, 2678–2695.

Hörmann, S. ,Kidziński, L. and Hallin, M. (2014). Dynamic functional principal components. *Journal of the Royal Statistical Society, Series B.* To appear.

Ku, W., R.H. Storer, and C. Georgakis (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30, 179-196.

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.

Li, T. (2012). On robust spectral analysis by least absolute deviations, *Journal of Time Series Analysis*, **33,** 298-303.

Maronna, R.A. (2005). Principal Components and Orthogonal Regression Based on Robust Scales, *Technometrics*, 47, 264-273.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust Statistics*, Wiley, Chichester.

Motta, G. Hafner, C.M. and von Sachs, R. (2011) Locally stationary factor models:identification an nonparametric estimation, *Econometric Theory*, **27**, 2011, 1279–1319.

Motta, G. an Ombao, H. (2012). Evolutionary factor analysis of replicated time series, *Biometrics*, **68**, 825-836.

Okamoto, M. and Kanazawa,M. (1968). Minimization of eigenvalues of a matrix and optimality of principal components, *Annals of Mathematical Statistics*, **39, 859-863.**

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

Peña, D. and Box, G.E.P. (1987). Identifying a simplifying structure in time series, *Journal of the American Statistical Association*, **82**, 836–843.

Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis, *Journal of Statistical Planning and Inference*, 136,4, 1237-1256.

Reinsel, G. C. and Velu, R. P. (1998). *Multivariate Reduced-Rank Regression*, Springer, New York.

Rousseeuw, P.J. and Yohai, V. (1984). “Robust Regression by Means of S estimators”, in *Robust and Nonlinear Time Series Analysis*, edited by J. Franke, W. Härdle, and R.D. Martin, Lecture Notes in Statistics 26, Springer Verlag, New York, 256-274.

Shumway, R.H. & Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*. New York: Springer.

Spangl, B. and Dutter, R. (2005). On robust estimation of power spectra, *Austrian Journal of Statistics*, **34**, 199–210.

Stock, J.H. and Watson, M.W. (1988). Testing for Common Trends. *Journal of the American Statistical Association*, **83**, 1097-1107.

Stock, J.H. and Watson, M.W (2002). Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association*, **97**, 1167–1179.

Tiao, G. C. and Tsay, R.S. (1989). Model specification in multivariate time series. *Journal of* *the Royal Statistical Society, Series B*, **51**, 157-195.

ACCEPTED MANUSCRIPT

m	T	**OPC**$_2$	**DPC**$_2$	BDPC$_{10}$
20	100	52.53	0.91	0.94
	200	55.86	0.92	0.95
100	100	54.89	0.95	0.99
	200	57.65	0.97	0.99
500	100	53.56	0.96	1.00
	200	57.14	0.98	1.00
1000	100	54.88	0.96	-
	200	59.09	1.00	-

Table 1: MSE of the Reconstructed Series for the Stationary Model with one Factor

ACCEPTED MANUSCRIPT

m	**OPC**$_{10}$	**DPC**$_{10}$	BDPC$_{10}$
20	67	83	55
100	67	86	62
200	69	86	62

Table 2: Percentage of Explained Variance in the VARI(1,1) Model

ACCEPTED MANUSCRIPT

m	T	**DPC**$_5$	BDPC$_{10}$	FGLR
5	200	0.66	0.71	0.91
	400	0.71	0.74	0.89
100	200	0.93	0.97	1.02
	400	0.95	0.97	1.01
1000	200	0.96	-	1.02
	400	0.98	-	1.01

Table 3: MSE of the Reconstructed Series for a Factor Model with r=2 and q=1

ACCEPTED MANUSCRIPT

k	OPC_k	DPC_k	$BDPC_k$
1	66	82	-
5	77	90	-
10	78	95	-
12	80	97	89

Table 4: Percentage of Explained Variance of the IPI Series Using the OPC, DPC and BDPC Procedures

ACCEPTED MANUSCRIPT

k	OPC_k	DPC_k	$BDPC_k$
0	60	60	-
1	60	82	-
5	61	87	-
10	62	88	60

Table 5: Explained variability of the OPC and DPC for the stock prices series with different number of lags

k	MSE of the DPC$_k$	SRS of the DPC$_k$	SRS of the S-DPC$_k$
1	18.81	6.05	0.84
5	17.75	6.64	0.50
10	16.90	7.63	0.48

Table 6: MSE and SRS of the DPC$_k$ and S DPC$_k$ for the contaminated stock prices series

ACCEPTED MANUSCRIPT

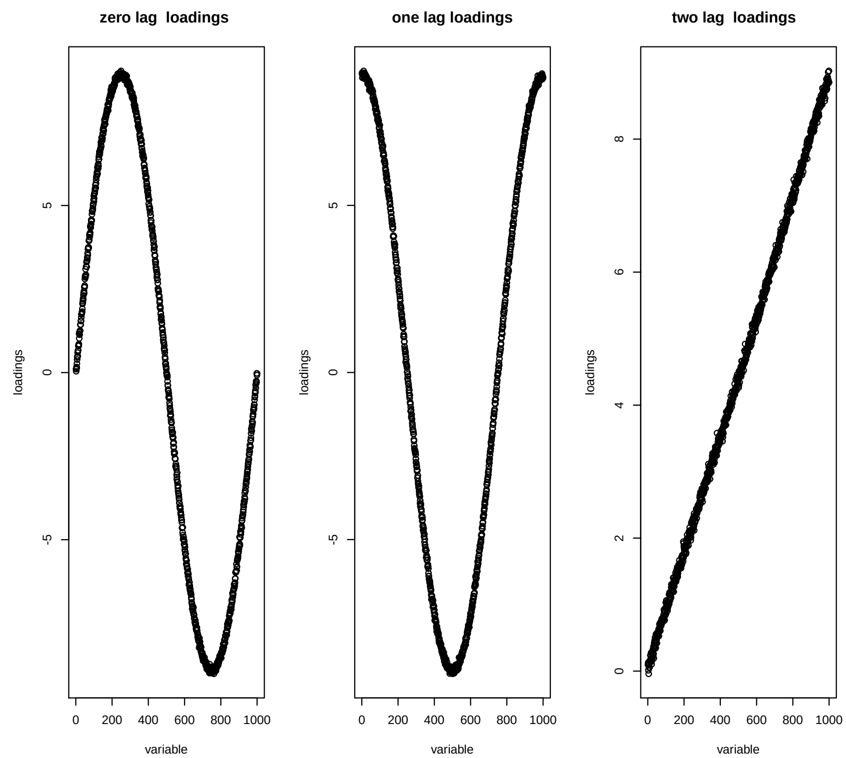

Figure 1: Loadings for one Replication of the Stationary Model with T=200 and m=1000

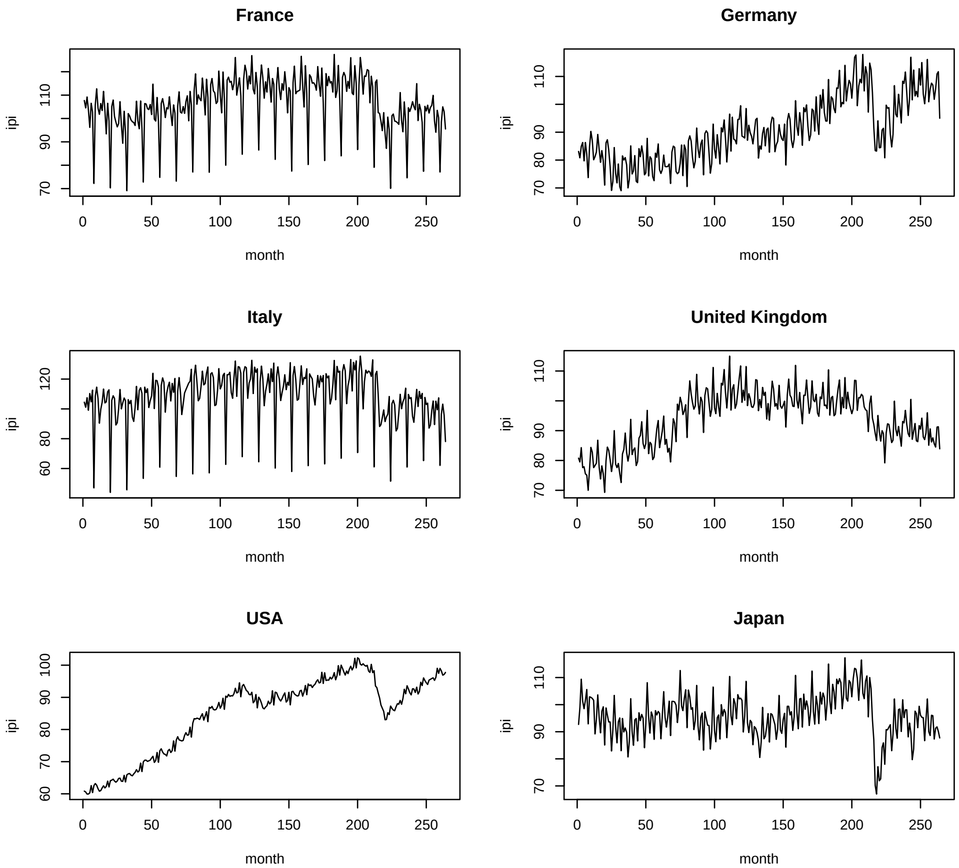

Figure 2: Industrial Production Index of Six Countries 1991-2012

ACCEPTED MANUSCRIPT

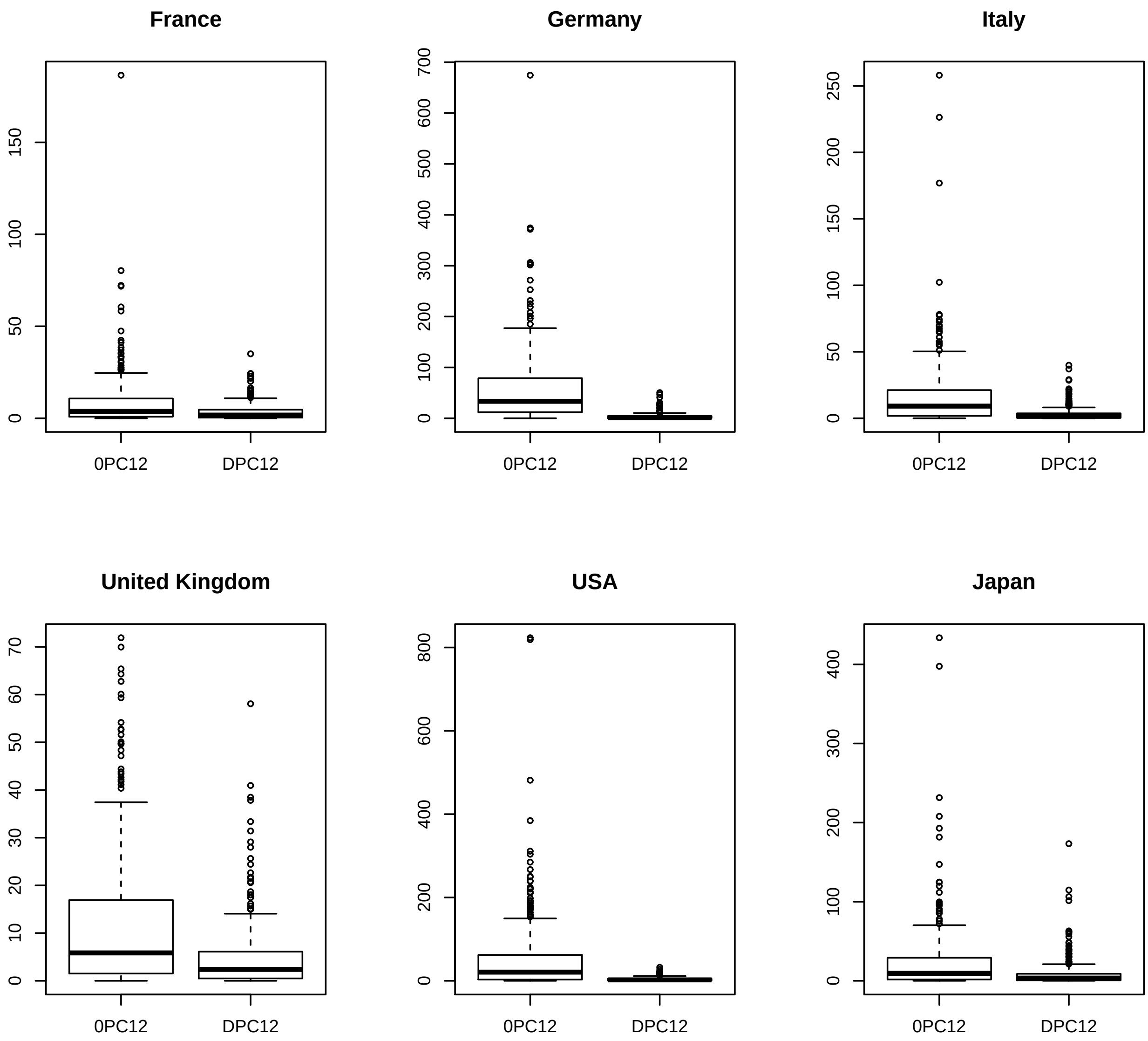

Figure 3: Boxplots of the Absolute Values of the Errors of the Reconstructed IPI Series

ACCEPTED MANUSCRIPT

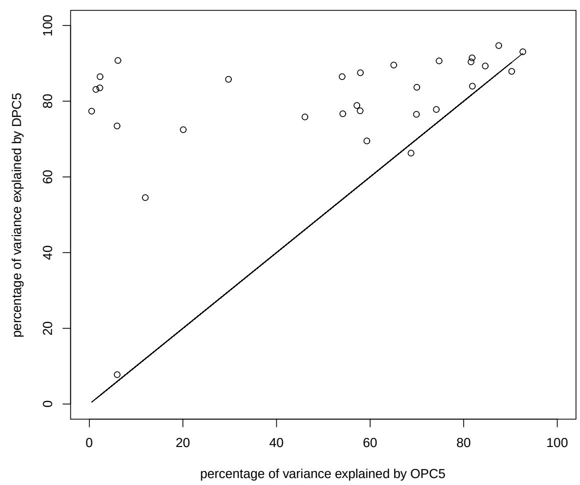

Figure 4: Percentage of the Variance Explained by the DPC_5 Against the Percentage Explained by the OPC_5 Procedure.

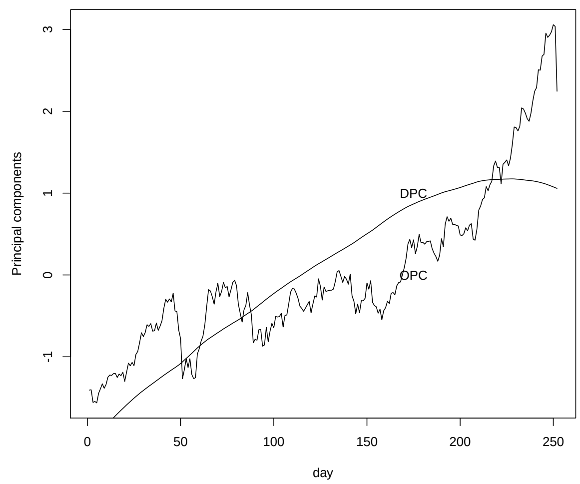

Figure 5: First OPC_5 and DPC_5 for the Stock Prices Series

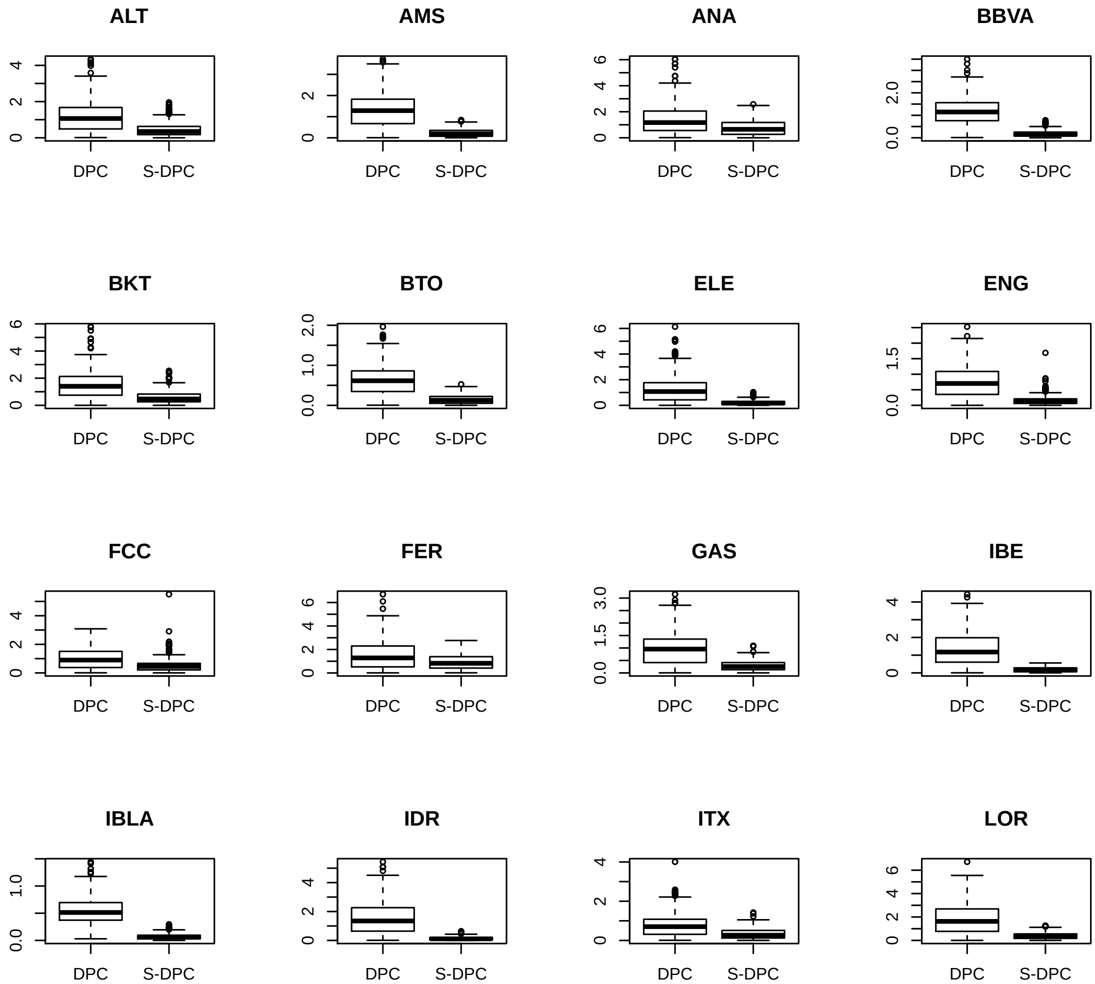

Figure 6: Boxplots of the Residual Absolute Values of the Stock Prices Obtained with the DPC_5 and $S\text{-}DPCC_5$ Procedures

ACCEPTED MANUSCRIPT