Dynamic Principal Components in the Time Domain

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.
ABSTRACT

We propose a time domain approach to define dynamic principal components (DPC) using a reconstruction of the original series criterion. This approach to define DPC was introduced by Brillinger, who gave a very elegant theoretical solution in the stationary case using the cross spectrum. Our procedure can be applied under more general conditions including the case of non stationary series and relatively short series. We also present a robust version of our procedure that allows to estimate the DPC when the series have outlier contamination. Our non robust and robust procedures are illustrated with real datasets.

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 set of data and therefore reducing the amount of information to be stored. In this article, we will consider linear time series models and we will concentrate in the time domain approach. 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 $\mathbf{z} = T^{-1} \sum_{t=1}^{T} \mathbf{z}_t$, which will estimate the mean if the
process is stationary, is zero. It is well known that the first principal component, $p_{1,t}, 1 \leq t \leq T$, minimizes the mean squared prediction error of the reconstruction of the vector time series, given by $\sum_{j=1}^{m} \sum_{t=1}^{T} (z_{j,t} - \alpha_j p_{1,t})^2$ and, in general, the first $k$ principal components, $k \leq m$, $p_{1,t}, \ldots, p_{kt}, 1 \leq t \leq T$, minimize the mean squared prediction error $\sum_{j=1}^{m} \sum_{t=1}^{T} (z_{j,t} - \sum_{i=1}^{k} \alpha_{j,i} p_{i,t})^2$ to reconstruct the vector of time series. Let $C = \sum_{t=1}^{T} z_t z_t^\prime / T$, be the sample covariance matrix and let $\lambda_1 \geq \lambda_1 \geq \lambda_m$ be the eigenvalues of $C$. Then $\alpha_i = (\alpha_{1,i}, \ldots, \alpha_{m,i})^\prime$, $1 \leq i \leq m$, is the eigenvectors of $C$ corresponding to the eigenvalue $\lambda_i$.

Ku, Storer and Georgakis (1995) propose to apply principal components to the augmented observations $z_t^* = (z_{t-h}^\prime, z_{t-h+1}^\prime, \ldots, z_t^\prime)^\prime$, $h + 1 \leq t \leq T$, 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, and using the well know properties of standard principal components we conclude that the first component obtained from this approach is a solution to the following reconstruction problem

$$M_1 = \sum_{j=1}^{m} \left[ \sum_{t=h+1}^{T} (z_{j,t} - \alpha_j p_{t-h})^2 + \sum_{t=h}^{T-1} (z_{j,t} - \alpha_j p_{t-h+1})^2 + \ldots + \sum_{t=1}^{T-h} (z_{j,t} - \alpha_j p_t)^2 \right],$$

which implies that, apart from the end effect, we minimize for each observation $z_{j,t}$, for $h + 1 \leq t \leq T - h$, the sum $\sum_{j=1}^{m} \sum_{t=0}^{h} (z_{j,t} - \alpha_j p_{t-l})^2$. Thus, this approach does not optimize a useful reconstruction criterion.

An alternative way to find interesting linear combinations was proposed by Box and Tiao (1977) who suggested maximizing the predictability of the linear combinations $c_t = \gamma^\prime z_t$. Other linear methods for dimension reduction in time series models have been given by the scalar component models, SCM, (Tiao and Tsay, 1989), the reduced-rank models (Ahn and Reinsel, 1990, Reinsel and Velu, 1998), and dynamic factor models (Peña and Box, 1987, Stock and Watson, 1988,
Forni el al. 2000, Peña and Poncela 2006 and Lam and Yao 2012), among others. None of the previous mentioned methods has as a goal to reconstruct the original series by using the principal components as in the classical case.

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

$$f_t = \sum_{h=-\infty}^{\infty} c'_h z_{t-h},$$

then

$$E \left[ (z_t - \sum_{j=-\infty}^{\infty} \beta_j f_{t+j})'(z_t - \sum_{j=-\infty}^{\infty} \beta_j f_{t+j}) \right].$$

is minimum. Brillinger elegantly solved this problem by showing that $c_h$ is the inverse Fourier transform of the principal components of the cross spectral matrices for each frequency, and $\beta_j$ 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. Although this result solves the theoretical problem it has the following shortcomings: (i) It can be applied only to stationary series; (ii) The optimal solution requires the unrealistic assumption that infinite series are observed, and it is not clear how to modify it when the observed series are finite; (iii) It is not clear how to robustify these principal components using a reconstruction criterion. The second shortcoming seems specially serious. In fact in Section 4 we show by means of a Monte Carlo simulation that what seems a natural modification for finite series of the Brillinger’s procedure does not work well.
In this paper we address the sample reconstruction of a vector of time series avoiding the drawbacks of Brillinger method. Our procedure provides an optimal reconstruction of the vector of time series from a finite number of lags. Some of the advantages of our procedure are: (i) it does not require stationarity and (ii) it can be easily made robust by changing the minimization of the mean squared error criterion by the minimization of a robust scale. The rest of this article is organized as follows. In Section 2 we describe the proposed dynamic principal components based on the 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, with the ordinary principal components and those proposed by Brillinger and we show the performances of these three types of principal components in two 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 eliminate the influence of outliers. In Section 6 some final conclusions are presented. Section 7 is an Appendix containing mathematical derivations.

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 $k$ lags (first DPC$_k$) as a vector $f = (f_t)_{-k_1+1 \leq t \leq T+k_2}$, so that the reconstruction of series $z_{j,t}, 1 \leq j \leq m$, as a linear combination of $f_{t-k_1}, f_{t-k_1+1}, ..., f_t, f_{t+1}, ..., f_{t+k_2}$ is optimal with the mean squared error (MSE)
criterion. More precisely, suppose that given a possible factor $f$, the $m \times (k_1 + k_2)$ matrix of coefficients $\beta = (\beta_{j,i})_{1 \leq j \leq m, -k_1 + 1 \leq i \leq k_2}$, and $\alpha = (\alpha_1, ..., \alpha_m)$ are used to reconstruct the values $z_{j,t}$ as

$$\hat{z}_{j,t}(f, \beta, \alpha) = \sum_{i=-k_1}^{k_2} \beta_{j,i} f_{t+i} + \alpha_j,$$

where $\beta_j$ is the $j$-th row of $\beta$. Let \( k = k_1 + k_2 \) and put $f_t^* = f_{t-k_1}$, $1 \leq t \leq T + k$ and $\beta_{j,h}^* = \beta_{j,h-k_1}$, $0 \leq h \leq k$, then, the reconstructed series are obtained as

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

Therefore we can always assume that $k_1 = 0$ and we will use $k$ to denote the number of forward lags.

Consider the MSE loss function

$$\text{MSE}(f, \beta, \alpha) = \sum_{j=1}^{m} \frac{1}{T} \sum_{t=1}^{T} (z_{j,t} - \hat{z}_{j,t}(f, \beta, \alpha))^2 = \sum_{j=1}^{m} \sum_{t=1}^{T} (z_{j,t} - \sum_{i=0}^{k} \beta_{j,i+1} f_{t+i} - \alpha_j)^2.$$

(3)

The optimal choices of $f = (f_1, ..., f_{T+k})'$ and $\beta = (\beta_{j,i})_{1 \leq j \leq m, 1 \leq i \leq k+1}$, $\alpha = (\alpha_1, ... \alpha_m)$ are given by

$$\hat{(f, \beta)} = \arg \min_{f, \beta, \alpha} \text{MSE}(f, \beta, \alpha).$$

(4)

Clearly if $f$ is optimal, $\gamma f + \delta$ is optimal too. Thus, we can choose $f$ so that $\sum_{t=1}^{T+k} f_t^2/(T+k) = 1$, and $\sum_{t=1}^{T+k} f_t/(T+k) = 0$. We call $\hat{f}$ the first DPC of order $k$ of the observed series $z_1, ..., z_T$. Note that the first DPC of order 0 corresponds to the first regular principal component of the data. Moreover, the matrix $\hat{\beta}$ contains the coefficients to be used to reconstruct the $m$ series from $\hat{f}$ in an optimal way.

Let $C_j(\alpha_j) = (c_{j,t,q}(\alpha_j))_{1 \leq t \leq T+k, 1 \leq q \leq k+1}$ be the $(T+k) \times (k+1)$ matrix defined
by
\[ c_{j,t,q}(\alpha_j) = \begin{cases} 
(z_j,t-q+1 - \alpha_j) & \text{if } 1 \lor (t - T + 1) \leq q \leq (k + 1) \land t \\
0 & \text{otherwise}
\end{cases}. \tag{5} \]

where \( a \lor b = \max(a, b) \) and \( a \land b = \min(a, b) \). Let \( D_j(f, \beta_j) = (d_{j,t,q}(f, \beta_j)) \) be the \((T + k) \times (T + k)\) given by
\[ d_{j,t,q}(f, \beta_j) = \begin{cases} 
\sum_{v=(t-k)\lor 1}^{t\land T} \beta_{j,q-v+1} \beta_{j,t-v+1} & \text{if } (t-k) \lor 1 \leq q \leq (t+k) \land (T+k) \\
0 & \text{otherwise}
\end{cases} \]

and
\[ D(f, \beta) = \sum_{j=1}^{m} D_j(f, \beta_j). \tag{6} \]

Differentiating (3) with respect to \( f_t \) in Subsection 7.1 we get the following equation
\[ f = D(f, \beta)^{-1} \sum_{j=1}^{m} C_j(\alpha) \beta_j. \tag{7} \]

Obviously, the coefficients \( \beta_j \) and \( \alpha_j \), \( 1 \leq j \leq m \), can be obtained using the least squares estimator, that is
\[ \begin{pmatrix} \beta_j \\
\alpha_j \end{pmatrix} = \left( F(f)' F(f) \right)^{-1} F(f)' z^{(j)}, \tag{8} \]

where \( z^{(j)} = (z_{j,1}, ..., z_{j,T})' \) and \( F(f) \) 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 (7) and (8). The second DPC is defined as the first DPC of the residuals \( r_{j,t}(f, \beta) \). Higher order DPC are defined in a similar manner. We will call \( p \) the selected number of components.

To define an iterative algorithm to compute \((\hat{f}, \hat{\beta}, \hat{\alpha})\) is enough to give \( f^{(0)} \) and to describe how to compute \( \beta^{(h)}, \alpha^{(h)}, f^{(h+1)} \) once \( f^{(h)} \) is known. According to (7) and (8) a natural such a rule is given by the following two steps:
step 1 Based on (8), define $\beta_j^{(h)}$ and $\alpha_j^{(h)}$, for $1 \leq j \leq m$, by

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

step 2 Based on (7), define $f^{(h+1)}$ by

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

and

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

The initial value $f^{(0)}$ can be chosen equal to the standard (non dynamic) first principal component, completed with $k$ zeros. The iterative procedure is stopped when

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

for some value $\varepsilon$.

Note that we start with $m$ series of size $T$. Assuming that we consider $p$ dynamic principal components let $\beta_{j,i,s}$, $1 \leq j \leq m$, $1 \leq i \leq k + 1$, the coefficients $\beta_{j,i}$ 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 $\alpha_j$. 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 $m$ is close to $p/m$. In applications 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. Is clear that increasing the number of lags after some point will have a negligible effect on the reduction in MSE. Then, if the level of the MSE is larger than desired, adding an additional component is call for. Thus one possible strategy will be start with one factor and increase the number of lags until the reduction of further lags is smaller than $\epsilon$. Then a new factor 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.

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 $\hat{\beta}=(\beta_{ji})_{1\leq j \leq m,1 \leq i \leq 2}$ and $\hat{f}=(\hat{f}_1, ..., \hat{f}_{T+1})'$ such that

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

(9)

Put $a_1 = \sum_{j=1}^m \beta_{j,1}^2$, $a_2 = \sum_{j=1}^m \beta_{j,2}^2$ and $b = \sum_{j=1}^m \beta_{j,1}\beta_{j,2}$, then the matrix $D = \sum_{j=1}^m D_j$ defined in (10) can be written as

$$D = a_2 \begin{pmatrix} a_1/a_2 & b/a_2 & 0 & 0 & ... & ... \\ b/a_2 & 1 + a_1/a_2 & b/a_2 & 0 & ... & ... \\ 0 & b/a_2 & 1 + a_1/a_2 & b/a_2 & ... & ... \\ ... & ... & ... & ... & ... & ... \\ 0 & ... & ... & b/a_2 & 1 + a_1/a_2 & b/a_2 \\ 0 & ... & ... & 0 & b/a_2 & 1 \end{pmatrix}.$$ 

9
Let $\hat{\beta}^{(i)} = (\hat{\beta}_{i,1}, ..., \hat{\beta}_{i,m}), i = 1, 2$. It is shown in the appendix that if $\hat{\beta}^{(1)} \neq \lambda \hat{\beta}^{(2)}$ there exists $|c| < 1, \alpha, w_1$ and $w_2$ so that

$$D = \alpha \begin{pmatrix}
w_1 & -c & 0 & 0 & ... & ... \\
-c & 1 + c^2 & -c & 0 & ... & ... \\
0 & -c & 1 + c^2 & -c & ... & ... \\
... & ... & ... & ... & ... & ... \\
0 & ... & ... & -c & 1 + c^2 & -c \\
0 & ... & ... & 0 & -c & w_2
\end{pmatrix}$$

Note that $\hat{\beta}^{(1)} = \lambda \hat{\beta}^{(2)}$ implies that putting $\hat{f}^* = (\hat{f}_t^*)_{1 \leq t \leq T}$ where $\hat{f}_t^* = \hat{f}_t + \lambda \hat{f}_{t+1}$ we have

$$\sum_{t=1}^{T} \sum_{j=1}^{m} (z_{j,t} - \hat{\beta}_{j,1} \hat{f}_t - \hat{\beta}_{j,2} \hat{f}_{t+1})^2 = \sum_{t=1}^{T} \sum_{j=1}^{m} (z_{j,t} - \hat{\beta}_j \hat{f}_t^*)^2,$$

and therefore, in this case the first DPC is as good for reconstructing the series as the first classical PC.

Let $A_0$ be defined by

$$A_0 = \begin{pmatrix}
1 & -c & 0 & 0 & ... & ... \\
-c & 1 + c^2 & -c & 0 & ... & ... \\
0 & -c & 1 + c^2 & -c & ... & ... \\
... & ... & ... & ... & ... & ... \\
0 & ... & ... & -c & 1 + c^2 & -c \\
0 & ... & ... & 0 & -c & 1
\end{pmatrix},$$

put $m_1 = w_1 - 1, m_2 = w_2 - 1$ and let $G = (G_1, G_2)$ be the $(T + 1) \times 2$ dimensional matrix where $G_1' = (m_1^{1/2}, 0, ..., 0)$ and $G_2' = (0, ..., 0, m_1^{1/2})$. We can write $D = \alpha(A_0 + GG')$, and then according to the Proposal A.3.3 of Seber (1984)
we have

\[
D^{-1} = \frac{1}{\alpha} \left( A_0^{-1} - A_0^{-1}G(I + G^tA_0^{-1}G)^{-1}G^tA_0^{-1} \right) \tag{11}
\]

\[
= \frac{1}{\alpha} (A_0^{-1} - A_0^{-1}GHG^tA_0^{-1}),
\]

where \( H = (I + G^tA_0^{-1}G)^{-1} = (h_{i,h}) \) is a \( 2 \times 2 \) matrix. We also have that \( A_0^{-1} \) is of the form

\[
(A_0^{-1})_{i,h} = \frac{1}{1 - c^2} c^{[i-h]}.
\]

and then we get

\[
A_0^{-1}GH = \frac{1}{1 - c^2} \begin{pmatrix}
    m_1^{1/2}h_{11} + m_2^{1/2}c^T h_{21} & m_1^{1/2}h_1 + m_2^{1/2}c^T h_{22} \\
    m_1^{1/2}h_{11}c^{i-1} + m_2h_{21}c^{T-i+1} & m_1^{1/2}h_{12}c^{i-1} + m_2h_{22}c^{T-i+1} \\
    m_1^{1/2}h_{11}c^T + m_2h_{21} & m_1^{1/2}h_{12}c^T + m_2h_{22}
\end{pmatrix}
\]

and

\[
(A_0^{-1}GHG^tA_0^{-1})_{ih} = 1/(1 - c^2)^2 \left[ (m_1^{1/2}h_{11}c^{i-1} + m_2^{1/2}h_{21}c^{T-i+1})m_1^{1/2}c^{h-1} \\
+ (m_1^{1/2}h_{12}c^{i-1} + m_2h_{22}c^{T-i+1})m_2^{1/2}c^{T-h+1} \right] = A_1c^{i+h-2} + A_2c^{T-i+h} + A_3c^{2T-i-h+2}. \tag{13}
\]

By (7) we have \( \hat{f} = D^{-1} \sum_j C_j \hat{\beta}_j \), where \( \hat{\beta}_j \) is given by (5) and \( C_j = (Z_1, Z_2) \) where \( Z_1' = (z_{j,1}, ..., z_{j,T}, 0) \) and \( Z_2' = (0, z_{j,1}, ..., z_{j,T}) \). Therefore, by (11), (12) and (13) we obtain

\[
\hat{f}_t = \frac{1}{\alpha} \left[ \sum_{j=1}^m \hat{\beta}_{j,1} \sum_{q=1}^T c^{[t-q]} z_{j,q} + \sum_{j=1}^m \hat{\beta}_{j,2} \sum_{q=2}^{T+1} c^{[t-q]} z_{j,q-1} \right] + R_t
\]
where \( R_t \to 0 \) except for \( t \) close to 1 or to \( T \).

Suppose now that \( z_t \) is stationary, then except in both ends \( \hat{f}_t \) can be approximated by the stationary process

\[
\hat{f}_t^* = \frac{1}{\alpha} \left[ \sum_{j=1}^{m} \beta_{j,1} \sum_{q=-\infty}^{\infty} c^{t-q} z_{j,q} + \sum_{j=1}^{m} \beta_{j,2} \sum_{q=-\infty}^{\infty} c^{t-q} z_{j,q} - 1 \right],
\]

and the DPC is approximated as linear combinations of the geometrically and symmetrically filtered series

\[
z_{j,t} + \sum_{i=1}^{\infty} c^{t} (z_{j,t+i} + z_{j,t-i}), \quad z_{j,t-1} + \sum_{i=1}^{\infty} c^{t} (z_{j,t-1+i} + z_{j,t-1-i}), \quad 1 \leq j \leq m.
\]

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 \( \hat{f}_t \) by an stationary process can be obtained.

### 4 Monte Carlo simulation and two real examples

We perform a Monte Carlo study using as vector series

\[
z_t = (z_{1,t}, z_{2,t}, z_{3,t})', \quad 1 \leq t \leq T
\]

generated as follows: let \( v_t, 1 \leq t \leq T + 2, w_{i,t}, 1 \leq i \leq 3, 1 \leq t \leq T \), i.i.d random variables with distribution \( N(0, 1) \), then

\[
z_{i,t} = v_{t+i-1} + 0.1 w_{i,t}, \quad 1 \leq i \leq 3, \quad 1 \leq t \leq T.
\]

We compute three different principal components: (i) The ordinary principal component (OPC), (ii) the dynamic principal component (DPC\(_k\)) proposed here with \( k; 1, 5 \) and 10, (iii) Brillinger dynamic principal components (BDPC\(_M\)) adapted for finite samples as follows:

\[
f_t = \sum_{k=(-M)\lor(t-T)}^{M\land(t-1)} c_k^t z_{t-k}, \quad (14)
\]

where \( c_k \) are the coefficients defined below \([2]\) in Section 1. The values of \( M \) where taken 10, 20 and 50. To reconstruct the original series with the OPC we
used $k = 1, 5, 10$ lags and the corresponding coefficients were obtained using least squares. To reconstruct the series with $\text{DPC}_k$ we proceed as described in Section 2. Finally, the original series $z_t$ were reconstructed using the $\text{BDPC}_M$ by

$$\hat{z}_{i,t} = \sum_{j=\lceil (M-t) \rceil}^{\lceil (M-t+1) \rceil} \beta_{i,j} f_{t+j}$$

where the $\beta_{i,j}$ are described below (2) in Section 1. The cross spectrum matrix was computed using the function \texttt{mvspec} in the ASTSA package with the R software. We took two values of $T$: 100 and 500 and we make 500 replications. Table 1 shows the MSE of the prediction residuals obtained with $\text{OPC}$, $\text{DPC}_k$ and $\text{BDPC}_M$. We observe that the procedure $\text{DPC}_k$ proposed here produces a much better reconstruction of the original series than the $\text{OPC}_k$ and the $\text{BDPC}_M$.

| $T$ | $\text{OPC}_k$ | $\text{DPC}_k$ | $\text{BDPC}_M$ |
|-----|----------------|----------------|-----------------|
|     | $k$ | $k$ | $M$          |
| 100 | 1.31| 0.78| 0.67 0.89    |
|     | 0.016| 2.05| 2.08 2.17  |
| 500 | 1.42| 0.79| 0.66 0.97    |
|     | 0.034| 2.03| 2.03 2.03  |

Table 1: Mean Square Errors obtained in the Monte Carlo study.

4.1 Example 1

We use six series corresponding to the Industrial Production Index (IPI) of France, Germany, Italy, United Kingdom, USA and Japan. We use monthly data from
Figure 1: Industrial production Index of six countries 1991-2012
January 1991 to December 2012 and the data are taken from Eurostat. The seven series are plotted in Figure 1.

Let $f_k, k \geq 0$, the first DPC$_k$. In Table 2 we show the percentage of variability explained by $f_0$ and $f_k$ using $k$ lags, computed as $EV_{j,k} = \min_{\beta, \alpha} MSE_k(f_j, \beta, \alpha)/\sum_{i=1}^{6} V_i$, for $j = 0, k$ where $V_i$ is the variance of the series $i$.

| $k$ | $EV_{0,k}$ | $EV_{k,k}$ |
|-----|------------|------------|
| 0   | 63.07      | 63.07      |
| 1   | 66.19      | 82.47      |
| 5   | 76.66      | 90.05      |
| 10  | 77.98      | 94.81      |
| 12  | 80.00      | 96.67      |

Table 2: Explained variability of the IPI series using the OPC and DPC with different number of lags

We note that the reconstruction of the series using the DPC is notably better than the one obtained by means of the OPC 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. With 12 lags the reconstruction error with the first DPC is smaller than 3.5%. Table 3 includes the coefficients of the six IPI series in the ordinary PC and in the first DPC with $k = 1$. 
For the OPC the coefficients in the first column in Table 3 coincide with the weights given to each country in the definition of the OPC. Thus, the first OPC gives the largest weight to Italy and then France, because of the strong seasonality of these series which have the largest variability. The second and third columns show that for reconstructing the original variables including the lag of the OPC is practically irrelevant. The fourth and fifth columns show that the DPC with one lag is almost equivalent to using the first difference of the DPC in the reconstruction of the series.

Table 3: Coefficients to reconstruct the IPI series by using OPC and DPC with one lag

| PC  | PC(0) | PC(1) | DPC(0) | DPC(1) |
|-----|-------|-------|--------|--------|
| -0.456 | -0.456 | -0.001 | -3.951 | 3.965  |
| -0.285 | -0.275 | -0.034 | -1.509 | 1.492  |
| -0.719 | -0.750 | 0.099  | -6.548 | 6.577  |
| -0.298 | -0.269 | -0.092 | -2.114 | 2.111  |
| -0.241 | -0.198 | -0.138 | -0.787 | 0.760  |
| -0.212 | -0.212 | -0.001 | -1.885 | 1.894  |

Figure 2 shows the original and reconstructed values using the first OPC and the first DPC, both with one lag. We can see that the reconstruction obtained with the DPC is clearly better than the one obtained with the OPC for Germany and USA. In the other cases the reconstruction with the DPC is still better but the differences are smaller and therefore more difficult to detect in the plots.

Figure 3 is similar to figure 2 but with twelve lags. Note that the reconstruction errors are significantly smaller than in the case of one lag, and that there is an important improvement of the reconstruction series when using the DPC instead.
Figure 2: Values of the original and reconstructed series of Example 1 with the OPC (o) and DPC (*) with one lag
of the OPC.

4.2 Example 2.

In this example the data set is composed of 31 daily stock prices in the stock market in Madrid corresponding to the 251 trading days of the year 2004. These 31 series are the main components of the IBEX (general index of the Madrid stock market). The source of the data is the Ministry of Economy, Spain. In Table 4 we show the explained variability of the reconstructed series using the DPC and OPC with different lags.

| $k$ | $EV_{0,k}$ | $EV_{k,k}$ |
|-----|------------|------------|
| 0   | 0.598      | 0.598      |
| 1   | 0.602      | 0.822      |
| 5   | 0.610      | 0.873      |
| 10  | 0.620      | 0.881      |

Table 4: Explained variability of the OPC and DPC for the stock prices series with different number of lags.

In Figure 4 we show the first four series in alphabetic order out of the thirty one and their reconstruction obtained by the first OPC and DPC with one lag. As shown in Table 4 including one lag in the OPC does not make much difference in the results, but it has a deep effect when using the DPC. In fact, in the case of the DPC, the coefficient of the one lag variable is very close but with opposite sign to the instantaneous coefficient and therefore the reconstruction is similar.
Figure 3: Values of the original and reconstructed series with the OPC (o) and DPC (*) with twelve lags
Figure 4: Values of the original and reconstructed of the first four stocks chosen in alphabetic orders. The reconstruction was made with the OPC (o) and DPC (*) using one lag
Figure 5: First OPC and DPC for the stock prices series
to the one obtained using the first difference of the first DPC without lags. Figure
presents the first OPC and the DPC. The dynamic principal components seems
to be very useful to represent the general trend of the set of time series.
5 Robust Dynamic Principal Components

As most of the procedures minimizing the mean square error, the DPC defined by (4) is not robust. In fact a very small fraction of outliers may have an unbounded influence on \((f, \alpha, \beta)\). For this reason we are going to study a robust alternative. One of the standard procedures 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 other 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 extend the S-estimators for the case of the DPC.

Special care is required for time series with strong seasonality. The reason is that a robust procedure may take the values corresponding to a particular season which is very different to the others as outliers, and therefore downweight these values. As a consequence, the reconstruction of these observations may be affected by large errors. Thus, the procedure we present here assumes that the series have been adjusted by seasonality and therefore this problem is not present.

5.1 S-Dynamic Principal Components

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

\[
\frac{1}{n} \sum_{i=1}^{n} \rho_0 \left( \frac{x_i}{s} \right) = b.
\]  

(15)

If \(\rho_0\) is bounded, then the breakdown point to \(\infty\) of \(S(\mathbf{x})\), that is, the minimum
fraction of outliers than can take \( S(x) \) to \( \infty \) is \( b/\max \rho_0 \). Moreover, the breakdown point to 0, that is, the minimum fraction of inliers that can take \( S(x) \) 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 \( \max \rho_0 = 1 \). We also assume that \( b = 0 \) so that both breakdowns are equal 0.5. Moreover \( \rho_0 \) is chosen so that \( E_{\phi}(\rho_0(x)) = b \), where \( \phi \) is the standard normal distribution. This condition guarantees that for normal samples \( S(x) \) is a consistent estimator of the standard deviation. One very popular family of \( \rho \) functions is the Tukey biweight family defined by

\[
\rho^T_c(x) = \begin{cases} 
1 - (1 - (x/c)^2)^3 & \text{if } |x| \leq c \\
1 & \text{if } |x| > c
\end{cases}
\]

Then, we can define the first S-DPC as follows: for \( 1 \leq j \leq m \), let \( r_j(f, \beta_j, \alpha_j) = (r_{j,t}(f, \beta_j, \alpha_j))_{1 \leq t \leq T} \), where \( r_{j,t}(f, \beta_j, \alpha_j) = z_{j,t} - \sum_{i=0}^{k} \beta_{j,i} f_{t+i} - \alpha_j \). Define

\[
SRS(f, \beta, \alpha) = \sum_{j=1}^{m} S^2(r_j(f, \beta_j, \alpha_j)),
\]

where

\[
\hat{f}, \hat{\beta}, \hat{\alpha} = \arg \min_{f, \beta, \alpha} SRS(f, \beta, \alpha),
\]

then \( \hat{f} \) is the first S-DPC and \( \hat{\beta} \) and \( \hat{\alpha} \) are the coefficients to reconstruct the \( z_{j,t} \)'s from \( \hat{f} \).

Note that the only difference with the definition given in (14) 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 \( \psi = \rho' \), \( w(u) = \psi(u)/u \),

\[
s_j = s_j(f, \beta_j, \alpha_j) = S(r(f, \beta_j, \alpha_j)).
\]

Note that \( s_j \) satisfies

\[
\frac{1}{T} \sum_{t=1}^{T} \rho \left( \frac{z_{v-t} - \sum_{i=0}^{k} \beta_{j,i+1} f_{t+i} - \alpha_j}{s_j} \right) = b.
\]
Define the weights
\[ w_{j,t} = w_{j,t}(f, \beta_j, \alpha_j) = w_0 \left( \frac{r_{j,t}(f, \beta_j)}{s_j} \right), \quad 1 \leq j \leq m, \quad 1 \leq t \leq T \] (20)
and
\[ W_{j,t,v} = W_{j,t,v}(f, \beta, \alpha, s) = \frac{s_j^2 w_{j,v}(f, \beta_j, \alpha_j, s_j)}{\sum_{h=(t-k)\vee 1}^{t\wedge T} w_{j,h}(f, \beta, \alpha, s_j)r_{j,h}^2}, \] (21)
where \( s = (s_1, \ldots s_m) \). Let \( C_j(f, \beta_j, s) = \left( c_{j,t,q}(f, \beta_j, s) \right)_{1 \leq t \leq T+k, 0 \leq q \leq k} \) be the \((T+k) \times (k+1)\) matrix defined by
\[
c_{j,t,q}(f, \beta, \alpha, s) = \begin{cases} 
W_{j,t-q+1}(f, \beta, \alpha, s)(z_{j,t-q+1} - \alpha_j) & \text{if } 1 \vee (t - T + 1) \leq q \leq (k+1) \wedge t \\
0 & \text{otherwise}
\end{cases},
\]
(22)
\[ D_j(f, \beta, \alpha, s) = (d_{j,t,q}(f, \beta, \alpha, s)) \] the \((T+k) \times (T+k)\) matrix with elements
\[
d_{j,t,q}(f, \beta, \alpha, s) = \begin{cases} 
\sum_{v=(t-k)\vee 1}^{t\wedge T} W_{j,t,v}\beta_{j,q-v+1}\beta_{j,t-v+1} & \text{if } (t-k)\vee 1 \leq q \leq (t+k) \wedge (T+k) \\
0 & \text{otherwise}
\end{cases}
\]
and
\[ D(f, \beta, \alpha, s) = \sum_{j=1}^{m} D_j(f, \beta, \alpha, s). \] (23)

Differentiating (19) with respect to \( f_t \) we get the following equation
\[ f = D(f, \beta, \alpha, s)^{-1} \sum_{j=1}^{m} C_j(f, \beta, \alpha, s)\beta_j. \] (24)

Let \( F(f) \) be the \( T \times (k+2) \) matrix with \( t \)-th row \( (f_t, f_{t+1}, \ldots, f_{t+k}, 1) \) and \( W_j(f, \beta, s) \) be the diagonal matrix with diagonal equal to \( w_{j,1}(f, \beta_j, s), \ldots, w_{j,T}(f, \beta_j, s) \). Then differentiating (19) with respect to \( \beta_{j,i} \) and \( \alpha_j \) we get
\[ \begin{pmatrix} \beta_j \\ \alpha_j \end{pmatrix} = (F(f)^t W_j(f, \beta_j, s) F(f))^{-1} F(f) W_j(f, \beta_j, s)^t z^{(i)}. \] (25)
Then the first S-PDC is determined by equation (18), (24) and (25). Note that the estimator defined by (4) is an S-estimate corresponding to $\rho_0^2(u) = u^2$ and $b = 1$. In this case $w(u) = 2$ and then we have $w_{j,v} = 1$ and $W_{j,v} = T$ for all $j$ and all $v$. Then for this case (24) and (25) become (7) and (8) respectively.

The second S-DPC is defined as the first S-DPC of the residuals $r_{j,t}(f, \beta)$. Higher order S-DPC are defined in a similar manner.

One important point is the choice of $b$. At first sight, $b = .5$ 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 $r_{j,t}(f, \beta_j, \alpha_j)$s are zero the scale $S(r_j(f, \beta_j, \alpha_j))$ is 0 no matter the value of the remaining values. Moreover, if 50% of the $|r_{j,t}(f, \beta_j, \alpha_j)|$ are small the scale $S(r_j(f, \beta_j, \alpha_j))$ is small too. Then when $b = 0.5$, the procedure may choose $f, \beta$ and $\alpha$ 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 $S(r_j(f, \beta_j, \alpha_j)) = 0$, it is required that 90% of the $r_{j,t}(f, \beta_j, \alpha_j)$s be 0.

One may wonder why for regression is common to use $b = 0.5$ and the 50% exact fitting property does not bring 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 algorithms for the S-dynamic principal components

The compute the first S-DPC we propose to use an iterative algorithm. We start the computing algorithm in step 0, and denote by $f^{(h)}$, $\beta^{(h)}$, $\alpha^{(0)}$, and $s$ the values computed in step $h$.

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

Then to define the algorithm is enough to describe how to compute $(f^{(h+1)}, \beta^{(h+1)}, s^{(h+1)})$ once $(f^{(h)}, \beta^{(h)}, s^{(h)})$ is known. This is done in the following three steps:

**step 1** According to (24), compute

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

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

**step 2** By (24), calling $W_j^{(h)} = W_j(f^{(h)}, \beta^{(h)}, \alpha^{(h)}, s^{(h)})$ compute the $j$-th row by

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

for $1 \leq j \leq m$.

**step 3** Compute $s_j^{(h+1)} = S(r_j(f^{(h+1)}, \beta, \alpha_{h+1}))$.  

26
The procedure is stopped when

\[
\frac{\text{SRS}(f^{(h)}, \beta^{(h)}, \alpha^{(h)}) - \text{SRS}(f^{(h+1)}, \beta^{(h+1)}, \alpha^{(h+1)})}{\text{SRS}(f^{(h)}, \beta^{(h)}, \alpha^{(h)})} < \varepsilon,
\]

where \(\varepsilon\) is a fixed small value.

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 7781 values composing the dataset was modified with 5% probability adding 20 to the true value. In Table 5 we include MSE in the reconstruction of the series with the DPC. Since the DPC is very sensitive to the presence of outliers, we also compute the S-DPC. Since the MSE is very sensitive to outliers, we evaluate the performance of the principal components to reconstruct the series by using the SRS criterion. We take as \(\rho\) 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 5 gives the MSE of the non DPC\(_k\) and the SRS for the DPC\(_k\) and S-DPC\(_k\) for \(k = 1, 5\) and 10.

Figure 6 shows the reconstruction of the four stock prices by using the DPC and the S-DPC. It can be seen, as expected, that the robust methods has a better performance.
Figure 6: Contaminated Stock prices series and their reconstruction by DPC (o) and by S-DPC (x)
| $k$ | MSE of the DPC$_k$ | SRS of the DPC$_k$ | SRS of the S-DPC$_k$ |
|-----|------------------|-------------------|---------------------|
| 1   | 309.70           | 106.69            | 39.84               |
| 5   | 295.84           | 119.03            | 37.81               |
| 10  | 274.74           | 111.33            | 31.95               |

Table 5: MSE and SRS of the DPC$_k$ and S-DPC$_k$ for the contaminated stock prices series

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 based on a robust scale. These procedures, in contrast to previous ones, can also be applied for nonstationary time series. A Monte Carlo study shows that the proposed dynamic principal component based on the MSE criterion can improve considerably the reconstruction obtained by both ordinary principal components and a finite sample version of Brillinger approach. 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 in order to balance accuracy in the series reconstruction and economy in the number of values stored for that purpose.
7 Appendix

7.1 Proof of (24)

Differentiating MSE\( (f, \beta, \alpha) \) with respect to \( f_t \) for \( t = 1, ..., T + k \) we get
\[
\sum_{j=1}^{m} \sum_{v=(t-k)\wedge 1}^{t\wedge T} (z_{j,v} - \sum_{i=0}^{k} \beta_{j,i+1} f_{v+i}) \beta_{j,t-v+1} = 0,\]
where \( a \wedge b \) denote minimum of \( a \) and \( b \) and \( a \vee b \) maximum. Then, we have

\[
\sum_{j=1}^{m} \sum_{v=(t-k)\wedge 1}^{t\wedge T} (z_{j,v} - \alpha_j) \beta_{j,t-v+1} = \sum_{j=1}^{m} \sum_{v=(t-k)\wedge 1}^{t\wedge T} \sum_{i=0}^{k} \beta_{j,i+1} \beta_{j,t-v+1} f_{v+i} \tag{26}
\]

that can be written as

\[
a_t(\beta) = b_t(f, \beta), \tag{27}
\]

where \( a_t(\beta) \) and \( b_t(f, \beta) \) are the left and right side of (26) respectively. Putting \( q = t - v + 1 \) we have

\[
a_t(\beta) = \sum_{j=1}^{m} \sum_{v=(t-k)\wedge 1}^{t\wedge T} (z_{j,v} - \alpha_j) \beta_{j,t-v+1} \tag{28}
\]

\[
= \sum_{j=1}^{m} \sum_{q=1\vee(t-T+1)}^{(k+1)\wedge t} (z_{j,t-q+1} - \alpha_j) \beta_{j,q}. \tag{29}
\]

and calling \( a(\beta) = (a_1(\beta), ..., a_{T+k}(\beta))' \)

\[
a(\beta) = \sum_{j=1}^{m} \mathcal{C}_j(\alpha_j) \beta_j. \tag{30}
\]

where \( \mathcal{C}_j \) is given by (5)

Now we will get an expression for \( b_t(f, \beta) \). Putting \( q = v + i \) we get

\[
b_t(f, \beta) = \sum_{j=1}^{m} \sum_{v=(t-k)\wedge 1}^{t\wedge T} \sum_{q=v}^{v+k} \beta_{j,q-v+1} \beta_{j,t-v+1} f_q. \tag{28}
\]
Then, calling $b(f, \beta) = (b_1(f, \beta), \ldots, b_{T+k}(f, \beta))'$

$$b(f, \beta) = D(\beta)f, \quad (31)$$

where $D$ is given in (6). Then, from (30) and (31), equation (27) can be also written as $\sum_{j=1}^{m} C_j(\alpha_j)\beta_j = D(\beta)f$. Then (7) follows.

### 7.2 Proof of (10)

To prove (10) it is enough to show that we can find $\lambda$ such that

$$\lambda(1 + a_1/a_2) - 1 = (\lambda b/a_2)^2 \quad (32)$$

and

$$|\lambda|(1 + a_1/a_2) < 1. \quad (33)$$

In this case (10) holds with

$$c = |\lambda|(1 + a_1/a_2). \quad (34)$$

According to (32) $\lambda$ should satisfy

$$(b^2/a_2^2)\lambda^2 - (1 + a_1/a_2)\lambda + 1 = 0. \quad (35)$$

A necessary and sufficient condition for the existence of a real solution of this equation is that $(1 + a_1/a_2)^2 - 4b^2/a_2^2 \geq 0$ which is equivalent to

$$a_1 + a_2 \geq 2|b|. \quad (36)$$

To prove this is enough

$$\sum_{j=1}^{m} \beta_{j,0}^2 + \sum_{j=1}^{m} \beta_{j,1}^2 \geq 2 \sum_{j=1}^{m} \beta_{j,0}\beta_{j,1}$$
which is always true. Solving [35] we get that one of the roots is

$$\lambda = \frac{a_2(a_1 + a_2)}{2b^2} - \frac{a_2^2}{2b^2} \left( \frac{(a_1 + a_2)^2}{a_2^2} - \frac{4b^2}{a_2^2} \right)^{1/2}$$

$$= \frac{a_2^2}{2b^2} \left( \left(1 + \frac{a_1}{a_2}\right) - \left(1 + \frac{a_1}{a_2}\right)^2 - 4\right)^{1/2}$$

and therefore $|\lambda| < a_2(a_1 + a_2)/2b^2$ and using [34] and [36] we get $|c| = |\lambda| (1 + (a_1/a_2)) < (a_1 + a_2)^2/(2b^2) \leq 1$, proving (33).

### 7.3 Derivation of (24) and (25)

Differentiating (19) with respect to $f_t$ and using (20) we get

$$\delta s_j(\beta, \alpha, f_t) \frac{\partial f_t}{\partial f_t} = -s \sum_{v=(t-k)\vee 1}^{t\wedge T} W_{j,t,v}(z_r - \alpha) \beta_{j,t-v+1} + s \sum_{v=(t-k)\vee 1}^{t\wedge T} \sum_{i=0}^{k} w_{j,v} \beta_{j,i+1} \beta_{j,t-v+1} f_{v+i} \right) \sum_{h=(t-k)\vee 1}^{t\wedge T} w_{j,h} r_{j,h}^2.$$  \tag{37}

Differentiating (17) with respect to $f_t$ we get

$$\sum_{j=1}^{m} s_j \delta s_j(\beta, \alpha, f) \frac{\partial f_t}{\partial f_t} = 0, \tag{38}$$

and then, from (37) and (38) we get

$$\sum_{j=1}^{m} \sum_{v=(t-k)\vee 1}^{t\wedge T} W_{j,t,v}(z_r - \alpha) \beta_{j,t-v+1} = \sum_{j=1}^{m} \sum_{r=(t-k)\vee 1}^{t\wedge T} \sum_{i=0}^{k} W_{j,t,v} \beta_{j,i+1} \beta_{j,t-v+1} f_{v+i},$$

where $W_{j,t,v}$ is given by (21). This equation can also be written as

$$a_t(f, \beta) = b_t(f, \beta), \tag{39}$$

where

$$a_t(f, \beta, \alpha) = \sum_{j=1}^{m} \sum_{v=(t-k)\vee 1}^{t\wedge T} W_{j,t,v}(z_r - \alpha) \beta_{j,t-v+1}$$
and

\[ b_t(f, \beta) = \sum_{j=1}^{m} \sum_{r=(t-k)^\wedge 1}^{(k+1)^\wedge T} \sum_{i=0}^{k} W_{j,t,v} \beta_{j,i+1} \beta_{j,t-v+1} f_{v+i} \]

Putting \( q = t - v + 1 \) we get

\[ a_t(f, \beta, \alpha) = \sum_{j=1}^{m} \sum_{q=1\vee(t-T+1)}^{(k+1)^\wedge t} W_{j,t-q+1}(f, \beta_j, \alpha, s) (z_{j,t-q+1} - \alpha_j) \beta_{j,q} \]

\[ = \sum_{j=1}^{m} C_j(f, \beta_j, s) \beta_j, \quad (40) \]

where \( C_j(f, \beta_j, s) \) is the \((T+k) \times (k+1)\) defined in (22). Putting \( v + i = q \) we get

\[ b_t(f, \beta, \alpha) = \sum_{j=1}^{m} \sum_{v=(t-k)^\wedge 1}^{v+k} \sum_{q=v}^{v+k} W_{j,t,v} \beta_{j,q-v+1} \beta_{j,t-v+1} f_q \]

\[ = D(f, \beta, s) f, \quad (41) \]

where \( D(f, \beta, \alpha, s) \) is the \((T+k) \times (T+k)\) matrix defined in (23) and \( s = (s_1, \ldots s_m) \).

Then from (39), (40) and (41) we derive (24). Differentiating (19) with respect to \( \beta_{j,i} \) and \( \alpha_j \), we get

\[ \frac{1}{T} \sum_{t=1}^{T} \psi \left( z_{j,v} - \sum_{i=0}^{k} \beta_{j,i+1} f_{v+i} - \alpha_j \right) \left( -s_j f_{v+i-1} - r_j f_{v+i-1} \frac{\partial s_j}{\partial \beta_{j,i}} \right) = 0 \]

\[ \frac{1}{T} \sum_{t=1}^{T} \psi \left( z_{j,v} - \sum_{i=0}^{k} \beta_{j,i+1} f_{v+i} - \alpha_j \right) \left( -s_j f_{v+i-1} - r_j f_{v+i-1} \frac{\partial s_j}{\partial \alpha_j} \right) = 0. \]

Then putting \( \partial s_j / \partial \beta_{j,i} = 0, 1 \leq i \leq k+1 \) and \( \partial s_j / \partial \alpha_j = 0 \) by (21) we get the following equations

\[ \sum_{t=1}^{T} w_{j,v} \left( z_{j,v} - \sum_{i=0}^{k} \beta_{j,i+1} f_{v+i} - \alpha_j \right) f_{v+i-1} = 0, 1 \leq i \leq k+1 \quad (42) \]
and

$$
\sum_{t=1}^{T} w_{j,v} \left( z_{j,v} - \sum_{i=0}^{k} \beta_{j,i+1} f_{v+i} - \alpha_j \right) = 0.
$$

(43)

From (42) and (43) equation (25) follows immediately.

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.

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.

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.

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.

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, 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.

Seber, G.A.F. (1984) Multivariate observations, Wiley, New York.

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

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

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.