

Robust Forecasting of Multiple Time Series with One-Sided Dynamic Principal Components



Daniel Peña and Víctor J. Yohai

Abstract Given a high-dimensional vector of time series, we define a class of robust forecasting procedures based on robust one-sided dynamic principal components. Peña et al. (*J Am Stat Assoc* 114(528):1683–1694, 2019) defined one-sided dynamic principal components as linear combinations of the present and past values of the series with optimal reconstruction properties. In order to make the estimation of these components robust to outliers, we propose here to compute the principal components by minimizing the sum of squares of the M-scales of the reconstruction errors of all the variables. The resulting robust components are called scale one-sided dynamic principal components (S-ODPC), and an alternating weighted least squares algorithm to compute them is presented. We prove that when both the number of series and the sample size tend to infinity, if the data follow a dynamic factor model, the mean of the squares of the M-scales of the reconstruction errors of the S-ODPC converges to the mean of the squares of the M-scales of the idiosyncratic terms, with rate $m^{1/2}$, where m is the number of dimensions. A Monte Carlo study shows that the S-ODPC introduced in this chapter can be successfully used for forecasting high-dimensional multiple time series, even in the presence of outlier observations.

Keywords Outliers · Dimensionality reduction · S-estimators · Dynamic factor models

D. Peña

Department of Statistics, Universidad Carlos III de Madrid, Madrid, Spain
e-mail: danielpena@uc3m.es

V. J. Yohai (✉)

Department of Mathematics and Instituto de Calculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina
e-mail: vyohai@dm.uba.ar

1 Introduction

High-dimensional sets of correlated time series are nowadays automatically generated in many fields, from engineering to environmental science and economics. These large data sets are often collected by using wireless sensor networks that may fail to record the data correctly due to depletion of batteries or environmental influence. These failures will produce outliers in the time series recorded that can modify strongly the forecasting generated from these contaminated data. Thus, using robust forecasting procedures based on robust estimation methods, which can deal with large outliers, is very important in this high-dimensional data sets.

Robust estimation of multivariate data sets generated by vector autoregressive (VAR) models was studied by Muler and Yohai (2013). They generalized to the multivariate case the robust estimation procedure proposed by Muler et al. (2009) for ARMA univariate models. However, the multivariate approach requires the estimation of a VAR/VARMA model for the vector of time series, and for these models, the number of parameters grows at least with the square of the number of series, turning their estimation unfeasible for high-dimensional sets of time series. Therefore, other alternatives for robust estimation in these situations have been explored. Dynamic factor models have been shown useful to model high-dimensional sets of time series, and some procedures have been proposed for the robust estimation of these models, see Fan et al. (2019), Alonso et al. (2020), Fan et al. (2021), and Trucíos et al. (2021).

Peña and Yohai (2016), following Brillinger's idea of dynamic principal components, see Brillinger (1964, 1981), proposed a new class of dynamic principal components that provide an optimal reconstruction of the observed set of series. These dynamic components are more general than those of Brillinger's, as they are computed without the assumption of being linear combinations of the data. In Peña and Yohai (2016) was also developed a robust estimation procedure for these principal components. However, although they are useful for reconstruction of the set of time series, these components are not expected to work well in forecasting problems, as their last values will be computed with a smaller number of observations than the central ones.

In order to have dynamic components useful for forecasting, (Peña et al. 2019) proposed the one-sided dynamic principal components (ODPC) that are defined as linear combinations of the observations based on a one-sided filter of past and present observations, instead of the double filter of past and future values, as proposed by Brillinger (1964). See also Forni et al. (2015, 2017) for a related approach to build one-sided filters for dynamic factor models.

Since the estimation procedure applied in Peña et al. (2019) minimizes the mean square error of the series reconstruction, it can be very sensitive to outliers. To overcome this drawback, in this chapter, we introduce a robust ODPC procedure that is based on the minimization of the sum of squares of M -scales of the reconstruction error of all the variables. Thus, the resulting forecasting procedure can be applied to automatic forecasting of large high-dimensional data sets of time series. The M -

scales introduced by Huber (1964) are robust estimators that measure how large in absolute value are the elements of a sample. These scales may have a 50 % breakdown point against outliers and inliers. Therefore, they protect for both types of anomalies. We call this procedure S-ODPC, and by means of a Monte Carlo procedure, we show that it produces accurate forecasts even with outlier contaminated data.

In Sect. 2 of this chapter, we review the one-sided dynamic principal components (ODPC) proposed by Peña et al. (2019). In Sect. 3, we introduce its robust version, the S-ODPC, and in Sect. 4, we describe an alternating weighted least squares to compute them. In Sect. 5, we discuss how to use the S-ODPC to forecast future values of a multiple time series. In Sect. 6, two possible robust strategies to determine the number of components and the number of lags to define each component and to reconstruct the time series are described. In Sect. 7, we show that asymptotically, when both the number of series and the sample size go to infinity, if the data follow a dynamic factor model, the reconstruction obtained with S-ODPC converges, in mean squared error, to the common part of a dynamic factor model. In Sects. 8 and 9, we illustrate with Monte Carlo simulations and with a real data example that in the presence of outliers the forecasting procedure based on S-ODPC performs much better than the one based on ODPC. Finally, Sect. 10 contains conclusions. An Appendix includes the mathematical proofs.

2 One-Sided Dynamic Principal Components

Consider a zero mean vector of stationary time series $\mathbf{z}_1, \dots, \mathbf{z}_T$, where $\mathbf{z}_t = (z_{t,1}, \dots, z_{t,m})'$. Let \mathbf{Z} be the data matrix of dimension $T \times m$ where each row is \mathbf{z}_t' . We will use \mathbf{E} for the expectation operator, $\|\cdot\|$ for the Euclidean norm of vectors and the spectral norm of matrices and $\|\cdot\|_F$ for the Frobenius norm of a matrix. Consider an integer number $k_1 \geq 0$, and let $\mathbf{a} = (\mathbf{a}'_0, \dots, \mathbf{a}'_{k_1})'$, where $\mathbf{a}'_h = (a_{h,1}, \dots, a_{h,m})$ is a vector of dimension m . Following Peña et al. (2019), the scores of the first one-sided dynamic principal component are of the form

$$f_t(\mathbf{a}) = \sum_{h=0}^{k_1} \mathbf{a}'_h \mathbf{z}_{t-h}, \quad t = k_1 + 1, \dots, T. \tag{1}$$

This component, built with k_1 lags, is used to reconstruct each observation $z_{t,j}$ using $k_2 \geq 0$ of its lags and the respective loading coefficients by

$$\widehat{z}_{t,j}(\mathbf{a}, \mathbf{B}) = \varphi_j + \sum_{h=0}^{k_2} b_{j,h} f_{t-h}(\mathbf{a}). \tag{2}$$

Let \mathbf{B} be the $m \times (k_2 + 2)$ loading matrix with row j equal to $(\varphi_j, b_{j,0}, \dots, b_{j,k_2})$. Then if

$$\widehat{\mathbf{z}}_t(\mathbf{a}, \mathbf{B}) = (\widehat{z}_{t,1}(\mathbf{a}, \mathbf{B}), \dots, \widehat{z}_{t,m}(\mathbf{a}, \mathbf{B}))'$$

and $\mathbf{F}_t(\mathbf{a}) = (1, f_t(\mathbf{a}), \dots, f_{t-k_2}(\mathbf{a}))$, (2) can be written as

$$\widehat{\mathbf{z}}_t(\mathbf{a}, \mathbf{B}) = \mathbf{B}\mathbf{F}_t(\mathbf{a}).$$

Note that we can consider the reconstruction (2) as the predictions of the common component in a dynamic factor model (DFM). This equation can be interpreted as the forecast of the common component of a DFM with one dynamic factor and k_2 lags. The loadings are given by the matrix \mathbf{B} . The dynamic factor is assumed to be a linear combination of the observations and their k_1 lags, defined by the \mathbf{a} weights in (1).

We suppose here that k_1 and k_2 are given, and in Sect. 6, we will propose a method to choose them. We will call $T^* = T - (k_1 + k_2)$ to the number of observations that can be reconstructed. The population optimal values of \mathbf{a} and \mathbf{B} were defined by Peña et al. (2019) as those that minimize the mean squared error in the reconstruction of the data

$$(\mathbf{a}^*, \mathbf{B}^*) = \arg \min_{\mathbf{a}, \mathbf{B}} \mathbf{E}(\|\mathbf{z}_t - \widehat{\mathbf{z}}_t(\mathbf{a}, \mathbf{B})\|^2).$$

Since if (\mathbf{a}, \mathbf{B}) is a solution of (2), then $(c\mathbf{a}, \mathbf{B}/c)$, for $c \neq 0$, will be one as well, we can normalize the vector \mathbf{a} , so that $\|\mathbf{a}\| = 1$, although, as in standard principal components $-\mathbf{a}, -\mathbf{B}$ works as well as \mathbf{a}, \mathbf{B} . Given a sample, $\mathbf{z}_1, \dots, \mathbf{z}_T$, the estimators $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{B}}$ of the optimal values $(\mathbf{a}^*, \mathbf{B}^*)$ are defined as

$$(\widehat{\mathbf{a}}, \widehat{\mathbf{B}}) = \arg \min_{\|\mathbf{a}\|=1, \mathbf{B}} \frac{1}{T^*} \sum_{t=(k_1+k_2)+1}^T \|\mathbf{z}_t - \widehat{\mathbf{z}}_t(\mathbf{a}, \mathbf{B})\|^2, \quad (3)$$

and the estimated first dynamic principal component is given by

$$\widehat{f}_t = f_t(\widehat{\mathbf{a}}) = \sum_{h=0}^{k_1} \widehat{\mathbf{a}}_h' \mathbf{z}_{t-h}, \quad k_1 + 1 \leq t \leq T, \quad (4)$$

and $\widehat{\mathbf{z}}_t = \widehat{\mathbf{z}}_t(\widehat{\mathbf{a}}, \widehat{\mathbf{B}})$ will provide an estimated optimal reconstruction of \mathbf{z}_t using k_2 of its lags at periods t , $(k_1 + k_2) + 1 \leq t \leq T$.

The second and higher orders of one-sided dynamic principal components are defined similarly. Let $k_1^{(h)}$ and $k_2^{(h)}$ be the number of lags to define the h -th component, and suppose that we have already computed the first l principal components. Denote by $\mathbf{r}_t^{(l)}$, $\max_{1 \leq h \leq l} (k_1^{(h)} + k_2^{(h)}) + 1 \leq t \leq T$, the residual

vector at time t using the first l components. Then, the $(l + 1)$ one-sided dynamic estimated component is a vector with components of the form

$$f_t(\mathbf{a}) = \mathbf{a}'_0 \mathbf{z}_t + \mathbf{a}'_1 \mathbf{z}_{t-1} + \dots + \mathbf{a}'_{k_1^{(l+1)}} \mathbf{z}_{t-k_1^{(l+1)}}, \quad \max_{1 \leq h \leq l+1} (k_1^{(h)} + k_2^{(h)}) + 1 \leq t \leq T, \quad (5)$$

where the vector $\mathbf{a} = (\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_{k_1})$ is chosen so that it optimizes the reconstruction of $\mathbf{r}_t^{(l)}$, $\max_{1 \leq h \leq l+1} (k_1^{(h)} + k_2^{(h)}) + 1 \leq t \leq T$. More precisely, consider reconstructions of $\mathbf{r}_t^{(l)}$ of the form $\widehat{\mathbf{r}}_t^{(l)}(\mathbf{a}, \mathbf{B}) = \mathbf{B} \mathbf{F}_t(\mathbf{a})$, where \mathbf{B} is a $m \times (k_2^{(l)} + 2)$ matrix, and then the $(l + 1)$ -th principal component has values $f_t(\widehat{\mathbf{a}}^{(l+1)})$, where $\widehat{\mathbf{a}}^{(l+1)}$ is defined so that there exists a $m \times (k_2^{(l+1)} + 2)$ matrix $\widehat{\mathbf{B}}^{(l+1)}$ such that

$$(\widehat{\mathbf{a}}^{(l+1)}, \widehat{\mathbf{B}}^{(l+1)}) = \arg \min_{\|\mathbf{a}\|=1, \mathbf{B}} \frac{1}{T^*} \sum_{t=k_1^{(l+1)}+k_2^{(l+1)}+1}^T \left\| \mathbf{r}_t^{(l)} - \widehat{\mathbf{r}}_t^{(l)}(\mathbf{a}, \mathbf{B}) \right\|^2.$$

More technical details can be found in Peña et al. (2019).

We will consider here only the estimating equations of the first component, and therefore, we will drop the superscript (l) . The estimating equations of higher order principal components can be found in Peña et al. (2019). Let \mathbf{Z}_h be the $T^* \times m$ data matrix of T^* consecutive observations

$$\mathbf{Z}_h = \begin{pmatrix} \mathbf{z}'_{h+1} \\ \mathbf{z}'_{h+2} \\ \vdots \\ \mathbf{z}'_{h+T^*} \end{pmatrix} \quad (6)$$

and we will consider this matrix for $h = 0, \dots, (k_1 + k_2)$. Merging these matrices, we can write in a compact way the data used in the estimation. Note that the matrix $\mathbf{Z}_{k_1+k_2}$ includes all the values of the series to be reconstructed.

Second, we will consider the larger matrix $\mathbf{Z}_{l,k_1} = [\mathbf{Z}_h, \mathbf{Z}_{h-1}, \dots, \mathbf{Z}_{h-k_1}]$ of dimension $T^* \times m(k_1 + 1)$ that includes the observations and also their k_1 lags required for the computation of the first component. The matrix of values of the components used for the reconstruction is the $T^* \times (k_2 + 2)$ matrix $\mathbf{F}_{k_1,k_2}(\mathbf{a})$, which has as rows $\mathbf{F}_t(\widehat{\mathbf{a}})$, $k_1 + k_2 + 1 \leq t \leq T$, and the reconstruction of the values $\mathbf{Z}_{k_1+k_2}$ is made with the $T^* \times m$ matrix $\widehat{\mathbf{Z}}_{k_1+k_2}$ computed as

$$\widehat{\mathbf{Z}}_{k_1+k_2} = \mathbf{F}_{k_1,k_2}(\widehat{\mathbf{a}}) \widehat{\mathbf{B}}'.$$

Third, let \mathbf{Z}^B be the matrix with dimensions $T^*(k_1 + 1) \times m(k_1 + 1)$ given by

$$\mathbf{Z}^B = \begin{pmatrix} \mathbf{Z}_{k_1+k_2,k_1} \\ \vdots \\ \mathbf{Z}_{k_1,k_1} \end{pmatrix}, \quad (7)$$

\mathbf{B}_1 the matrix \mathbf{B} with its first column deleted and \mathbf{I}_d the $d \times d$ identity matrix. Define the matrix of products of loadings and data values by

$$\mathbf{X}(\mathbf{B}) = (\mathbf{B}_1 \otimes \mathbf{I}_{T^*}) \mathbf{Z}^B \quad (8)$$

as a $mT^* \times m(k_1 + 1)$ matrix with rank $m(k_1 + 1)$. Then, in Peña et al. (2019), it is shown that $(\widehat{\mathbf{a}}, \widehat{\mathbf{B}})$ are values (\mathbf{a}, \mathbf{B}) satisfying

$$\mathbf{a} = (\mathbf{X}(\mathbf{B})' \mathbf{X}(\mathbf{B}))^{-1} \mathbf{X}(\mathbf{B})' \text{vec}(\mathbf{Z}_{k_1+k_2}), \quad (9)$$

and this vector is standardized to unit norm. On the other hand, $\widehat{\mathbf{B}}$ can also be computed by least squares by

$$\mathbf{B}' = (\mathbf{F}'_{k_1, k_2}(\mathbf{a}) \mathbf{F}_{k_1, k_2}(\mathbf{a}))^{-1} \mathbf{F}'_{k_1, k_2}(\mathbf{a}) \mathbf{Z}_{k_1+k_2}. \quad (10)$$

An alternating least squares algorithm to compute $(\widehat{\mathbf{a}}, \widehat{\mathbf{B}})$ can be carried out as follows. Given an initial value of $\widehat{\mathbf{a}}$, the matrix of values of the component $\mathbf{F}_{k_1, k_2}(\mathbf{a})$ is computed by (9), and the matrix $\widehat{\mathbf{B}}$ is obtained by (10). Then, this matrix allows to compute a new value of $\widehat{\mathbf{a}}$, by first applying (8) and then using (9). This alternating process is continued until convergence. A similar algorithm can be applied to obtain the i -th component for $i > 1$.

3 Robust One-Sided Dynamic Principal Components

Since the ODPC estimator described in Sect. 2 is based on the minimization of the reconstruction mean square error, this estimator is very sensitive to the presence of outliers in the sample. To address this problem, we are going to propose a class of robust one-sided principal components that will be called S-ODPC and that are based on a M-scale.

Given a random variable x , the M-scale $S_M(x)$ is defined by

$$\mathbf{E} \left(\rho \left(\frac{x}{S_M(x)} \right) \right) = b,$$

where $\rho: \mathbb{R} \rightarrow \mathbb{R}^+$ and ρ and b satisfy: (a) $\rho(0) = 0$, (b) $\rho(-x) = \rho(x)$, (c) $\rho(x)$ is non-decreasing for $x \geq 0$, (d) $\lim_{x \rightarrow \infty} \rho(x) = 1$, and (e) $0 < b < 1$. The scale $S_M(\mathbf{x})$ is a measure of how large are the values that x takes. Note that if $\rho(x) = x^2$ and $b = 1$, then $S_M(\mathbf{x})$ is the L_2 -scale given by $S_M^2(\mathbf{x}) = \mathbf{E}(x^2)$.

Given a sample $\mathbf{x} = (x_1, x_2, \dots, x_n)$ of x , $S_M(x)$ may be estimated by $s_M(\mathbf{x})$ satisfying

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{x_i}{s_M(\mathbf{x})}\right) = b. \tag{11}$$

Generally, it is required that $E_\Phi(\rho(x)) = b$, where Φ is the standard normal distribution. This condition implies that if \mathbf{x}_n is a sample of size n of a $N(0, \sigma^2)$ distribution, then $\lim_{n \rightarrow \infty} s_M(\mathbf{x}_n) = \sigma^2$.

A family of ρ functions satisfying these properties is the Tukey biweight family defined by

$$\rho_c^{BS}(x) = \left[1 - \left(1 - \left(\frac{x}{c} \right)^2 \right)^3 \right] I(|x| \leq c).$$

The breakdown point of a M-scale is $\varepsilon^* = \min(b, 1 - b)$ and is maximized when $\delta = 0.5$, and in this case, $\varepsilon^* = 0.5$. The consistency condition $E_\Phi(\rho_c^{BS}(x)) = 0.5$ is satisfied when $c = 1.547$. This is the function used in the simulations in Sect. 8 and in the example in Sect. 9.

The M-scales have been used to define robust estimators for many statistical problems. We will mention here two classes of estimators based on a M-scale: S-estimators for regression (see Rousseeuw and Yohai 1984) and S-estimators of the scatter matrix and multivariate location (see Davies 1987).

When the first component is used to reconstruct the series, the reconstruction error of the variable j at the period t is given by

$$r_{t,j}(\mathbf{a}, \mathbf{B}) = z_{t,j} - \widehat{z}_{t,j}(\mathbf{a}, \mathbf{B}), \quad k_1 + k_2 + 1 \leq t \leq T. \tag{12}$$

The first population S-ODPC is defined by the values $\mathbf{a}_S^*, \mathbf{B}_S^*$ given by

$$(\mathbf{a}_S^*, \mathbf{B}_S^*) = \arg \min_{\|\mathbf{a}\|=1, B} \mathcal{K}_M(\mathbf{a}, \mathbf{B}),$$

where

$$\mathcal{K}_M(\mathbf{a}, \mathbf{B}) = \sum_{j=1}^m S_M^2(r_{t,j}(\mathbf{a}, \mathbf{B})).$$

Given a sample \mathbf{z}_t , $1 \leq t \leq T$, $(\mathbf{a}_S^*, \mathbf{B}_S^*)$ can be estimated by

$$(\widehat{\mathbf{a}}_S, \widehat{\mathbf{B}}_S) = \arg \min_{\|\mathbf{a}\|=1, B} k_M(\mathbf{a}, \mathbf{B}),$$

where

$$k_M(\mathbf{a}, \mathbf{B}) = \sum_{j=1}^m s_M^2(\mathbf{r}_j(\mathbf{a}, \mathbf{B}))$$

and $\mathbf{r}_j(\mathbf{a}, \mathbf{B}) = (r_{t,j}(\mathbf{a}, \mathbf{B}))_{k_1+k_2+1 \leq t \leq T}$.

Remark 1 Observe that if the performance of the principal components is measured by the sum of the squares of the M-scales of the empirical reconstruction residuals, then the S-ODPC defined above is still optimal when dealing with nonstationary series. Therefore, it may be used for forecasting even in this case.

In the Appendix, we prove that differentiating $S(\mathbf{a}, \mathbf{B})$, $\hat{\mathbf{a}}$ and $\hat{\mathbf{B}}$ satisfy expressions similar to (9) and (10), respectively. In fact, it can be shown that $\hat{\mathbf{a}}_S$ and $\hat{\mathbf{B}}_S$ are values \mathbf{a} and \mathbf{B} satisfying the following weighted least squares relationships:

$$\begin{aligned} \mathbf{a}^* &= (\mathbf{X}(\mathbf{B})' \mathbf{W}(\mathbf{a}, \mathbf{B}) \mathbf{X}(\mathbf{B}))^{-1} \mathbf{X}(\mathbf{B})' \mathbf{W}(\mathbf{a}, \mathbf{B}) \text{vec}(\mathbf{Z}_{k_1+k_2}), \\ \hat{\mathbf{a}} &= \mathbf{a}^* / \|\mathbf{a}^*\|, \end{aligned} \tag{13}$$

where \mathbf{W} is a $mT^* \times mT^*$ diagonal matrix of weights. On the other hand, fixing $\mathbf{F}_{k_1,k_2}(\mathbf{a})$, the optimal \mathbf{B} also satisfies a weighted least squares expression. Then, $\mathbf{B}' = (\mathbf{b}_1, \dots, \mathbf{b}_m)$, where

$$\mathbf{b}_j = (\mathbf{F}_{k_1,k_2}(\mathbf{a})' \mathbf{W}_j(\mathbf{a}, \mathbf{B}) \mathbf{F}_{k_1,k_2}(\mathbf{a}))^{-1} \mathbf{F}_{k_1,k_2}(\mathbf{a})' \mathbf{W}_j(\mathbf{a}, \mathbf{B}) \mathbf{z}_j^*, \tag{14}$$

where \mathbf{z}_j^* is the j -th column of $\mathbf{Z}_{k_1+k_2}$ and \mathbf{W}_j is a $T^* \times T^*$ matrix of weights. The matrices \mathbf{W}_j , $1 \leq j \leq m$, are defined as follows: let $\psi(u) = \rho'(u)$ and $w(u) = \psi(u)/u$ if $u \neq 0$ and $w(0) = \lim_{u \rightarrow 0} w(u)$. Given \mathbf{a} and \mathbf{B} , let $\sigma_j(\mathbf{a}, \mathbf{B}) = s_M(\mathbf{r}_j(\mathbf{a}, \mathbf{B}))$ and $\mathbf{w}_j(\mathbf{a}, \mathbf{B}) = (w_{t,j}(\mathbf{a}, \mathbf{B}))_{k_1+k_2+1 \leq t \leq T}$, $1 \leq j \leq m$, where

$$w_{t,j}(\mathbf{a}, \mathbf{B}) = \frac{w(r_{t,j}(\mathbf{a}, \mathbf{B})/\sigma_j(\mathbf{a}, \mathbf{B}))}{\frac{1}{T^*} \sum_{k=k_1+k_2+1}^T w(r_{k,j}(\mathbf{a}, \mathbf{B})/\sigma_j(\mathbf{a}, \mathbf{B}))}. \tag{15}$$

Then, $\mathbf{W}_j(\mathbf{a}, \mathbf{B})$ and $\mathbf{W}(\mathbf{a}, \mathbf{B})$ have as diagonals $\mathbf{w}_j(\mathbf{a}, \mathbf{B})'$ and $\mathbf{w}(\mathbf{a}, \mathbf{B}) = (\mathbf{w}_1(\mathbf{a}, \mathbf{B})', \dots, \mathbf{w}_m(\mathbf{a}, \mathbf{B})')'$, respectively. These weights penalize outliers reducing or removing their influence on the estimators.

Observe that \mathbf{w}_j , \mathbf{W} , \mathbf{F}_{k_1,k_2} , and \mathbf{X} depend on \mathbf{a} and \mathbf{B} , and therefore, Eqs. (13) and (14) cannot be used to directly compute them. In the next section, we propose an alternating weighted least squares algorithm that overcomes this problem.

The second and higher order S-ODPC can be defined in a similar way as that in the non-robust ODPC.

4 Computing Algorithm for the S-ODPC

We propose here an alternating weighted least squares algorithm for computing the first S-ODPC. Let $\mathbf{a}^{(i)}$, $\mathbf{B}^{(i)}$, and $\mathbf{f}^{(i)}$ be the values of \mathbf{a} , \mathbf{B} , and $\mathbf{f} = (f_{k_1+1}, \dots, f_T)'$ corresponding to the i -th iteration and $\delta \in (0, 1)$, a tolerance parameter to stop the iterations. In order to define the algorithm, it is enough to give: (1) initial values $\mathbf{a}^{(0)}$ and $\mathbf{B}^{(0)}$, (2) a rule that given the values of the i -th iteration $\mathbf{a}^{(i)}$ and $\mathbf{B}^{(i)}$, $1 \leq j \leq m$, establishes how to compute $\mathbf{a}^{(i+1)}$ and $\mathbf{B}^{(i+1)}$, $1 \leq j \leq m$, and (3) a stopping rule. Then, the iterated algorithm is as follows:

1. To obtain the initial values, we first compute a standard (non- dynamic) robust principal component $\mathbf{f}^{(0)} = (f_t^{(0)})_{1 \leq t \leq T}$, for example using the proposal of Maronna (2005). Then, we take $\mathbf{B}^{(0)} = (\mathbf{b}_1^{(0)}, \dots, \mathbf{b}_m^{(0)})$, where $\mathbf{b}_j^{(0)}$, $1 \leq j \leq m$, is an S-regression estimator (see Rousseeuw and Yohai 1984) using as dependent variable $\mathbf{z}_{j,k_1+k_2} = (z_{k_1+k_2+1,j}, \dots, z_{T,j})$ and as matrix of independent variables $\mathbf{F}_{k_1,k_2}^{(0)}$ with rows $\mathbf{F}_t^{(0)} = (1, f_t^{(0)}, f_{t-1}^{(0)}, \dots, f_{t-k_2}^{(0)})$, $k_1 + k_2 + 1 \leq t \leq T$. Once obtained $\mathbf{B}^{(0)}$, we compute the matrix of residuals $\mathbf{R} = \mathbf{Z}_{k_1+k_2} - \mathbf{F}_{k_1,k_2}^{(0)} \mathbf{B}^{(0)'}'$. This matrix is used to define the weights $w_{t,j}$ as in (15) and the diagonal matrix $\mathbf{W}^{(0)}$. We compute $\mathbf{a}^{(0)} = \mathbf{a}^* / \|\mathbf{a}^*\|$, where

$$\mathbf{a}^* = (\mathbf{X}(\mathbf{B}^{(0)})' \mathbf{W}^{(0)} \mathbf{X}(\mathbf{B}^{(0)}))^{-1} \mathbf{X}(\mathbf{B}^{(0)})' \mathbf{W}^{(0)} \text{vec}(\mathbf{Z}_{k_1+k_2})$$

and $\mathbf{X}(\mathbf{B})$ is defined as in (8).

2. Given $\mathbf{a}^{(i)}$ and $\mathbf{B}^{(i)}$, we compute the reconstruction residuals matrix $\mathbf{R} = \mathbf{Z}_{k_1+k_2} - \mathbf{F}_{k_1,k_2}^{(i)} \mathbf{B}^{(i)'}'$ and the corresponding new M- scales $\sigma_j^{(i)}$, $1 \leq j \leq m$. Using these residuals and scales, we obtain new weights $w_{t,j}(\mathbf{a}^{(i)}, \mathbf{B}^{(i)})$, $k_1 + k_2 + 1 \leq t \leq T$, and the corresponding diagonal matrices $\mathbf{W}_j(\mathbf{a}^{(i)}, \mathbf{B}^{(i)})$, $1 \leq j \leq m$ and $\mathbf{W}(\mathbf{a}^{(i)}, \mathbf{B}^{(i)})$. Then $\mathbf{B}^{(i+1)} = (\mathbf{b}_1^{(i+1)}, \dots, \mathbf{b}_m^{(i+1)})$ is defined by (14) with $\mathbf{a} = \mathbf{a}^{(i)}$. To compute $\mathbf{a}^{(i+1)}$, we use (13) with $\mathbf{B} = \mathbf{B}^{(i+1)}$.
3. The stopping rule is as follows: stop when

$$\frac{S(\mathbf{a}^{(i)}, \mathbf{B}^{(i)}) - S(\mathbf{a}^{(i+1)}, \mathbf{B}^{(i+1)})}{S(\mathbf{a}^{(i)}, \mathbf{B}^{(i)})} \leq \delta.$$

As in Salibian-Barrera and Yohai (2006), it can be shown at each step the MSE decreases, and therefore, it converges to a local minimum. To obtain a global minimum, the initial value $\mathbf{f}^{(0)}$ should be close enough to the optimal one.

Note that, since the matrix $\mathbf{X}(\mathbf{B}) = (\mathbf{B}_1 \otimes \mathbf{I}_{T^*}) \mathbf{Z}^{\mathbf{B}}$ has dimensions $mT^* \times m(k_1 + 1)$, solving the associated least squares problem can be time- consuming for high-dimensional (large m) problems. The iterative nature of the algorithm we propose implies that this least squares problem will have to be solved several times for different \mathbf{B} matrices. However, as the matrix $\mathbf{B}' \otimes \mathbf{I}_{T^*}$ is sparse, it can be stored

efficiently, and multiplying it with a vector is relatively fast. We found that for problems with a moderately large m , the following modification of our algorithm works generally faster: instead of finding the optimal $\mathbf{a}^{(i+1)}$ corresponding to $\mathbf{B}^{(i+1)}$, just do one iteration of coordinate descent for $\mathbf{a}^{(i+1)}$.

5 Forecasting Using the S-ODPC

Suppose that we have computed estimators of Q robust dynamic principal components and that the lags used for the q component were $(k_1^{(q)}, k_2^{(q)})$. Let $\widehat{\mathbf{f}}^{(q)} = (\widehat{f}_t^{(q)})_{k_1^{(q)}+1 \leq t \leq T}$, $1 \leq q \leq Q$, be the estimated S-ODPC's and $\widehat{\mathbf{B}}^q$ the estimated reconstruction matrices. We will show now how we can predict the values of $\mathbf{z}_{T+1}, \dots, \mathbf{z}_{T+h}$ for some $h \geq 1$. For that purpose, fit a time series model for each component $\widehat{\mathbf{f}}^{(q)}$, $1 \leq q \leq Q$ (e.g., an ARMA model), using a robust procedure, and with this model, obtain predictions $\widehat{f}_{T+l}^{(q)}$ of $f_{T+l}^{(q)}$, $1 \leq q \leq Q$, $1 \leq l \leq h$. We fit in the simulations AR models for each component, and these models were estimated using the filtered τ -estimation procedure described in Chapter 8 of Maronna et al. (2019). This procedure selects automatically the order of the AR model, gives robust estimators of its coefficients, and provides a filtered series $\widetilde{\mathbf{f}}^{(q)} = (\widetilde{f}_t^{(q)})_{k_1^{(q)}+1 \leq t \leq T}$, $1 \leq q \leq Q$ cleaned of the detected outliers. With the help of these filtered series, we can obtain robust predictions as follows. Let $\widetilde{\mathbf{F}}_{T+l}^{(q)} = (1, \widetilde{f}_{T+l}^{(q)}, \widetilde{f}_{T+l-1}^{(q)}, \dots, \widetilde{f}_{T+l-k_2^{(q)}}^{(q)})'$; then a robust prediction \mathbf{z}_{T+l} given the first T observations is

$$\widehat{\mathbf{z}}_{T+l|T} = \sum_{q=1}^Q \widehat{\mathbf{B}}^{(q)} \widetilde{\mathbf{F}}_{T+l}^{(q)}, \quad 1 \leq l \leq h.$$

The filtered τ -estimation procedure is implemented in the function `arima.rob` of the R package `robarima`.

6 Selecting the Number of Lags and the Number of Components

An important problem is the selection of the number of dynamics components Q to use for prediction, and the number of lags $k_1^{(q)}$ and $k_2^{(q)}$, $1 \leq q \leq Q$, required to define each component. In order to simplify the presentation, we will assume that $k_1^{(q)} = k_2^{(q)} = k^{(q)}$. We can use two possible methods to select these values: (a) an information criterion and (b) cross-validation. Simulations performed for the ODPC in Peña et al. (2019) show that both procedures have similar efficiencies; However,

(a) is much faster than (b). For that reason, we propose to use (a). In what follows, we describe implementations of both procedures.

6.1 Selection Using an Information Criterion

In Peña et al. (2019), an adaptation of the Bai and Ng (2002) criterion for factor model was used to choose $k^{(q)}$ and Q for ODPC. Here, we modify this procedure for its use in S-ODPC. We should start given a maximum value K for $k^{(q)}$. We compute the first S-ODPC with $k^{(q)} = k$ for all values of k such that $0 \leq k \leq K$. For each of these values of k , we compute the residuals $(r_{t,j}^{(1,k)})_{2k+1 \leq t \leq T, 1 \leq j \leq m}$ and compute the M-scale $S_M(\mathbf{r}_j^{(1,k)})$ of each vector $\mathbf{r}_j^{(1,k)} = (r_{2k+1,j}^{(1,k)}, \dots, r_{T,j}^{(1,k)})$, $1 \leq j \leq m$. Let $\hat{\sigma}_{1,k} = ((1/m) \sum_{j=1}^m S_M(\mathbf{r}_j^{(1,k)}))^2$. Let $T_{1,k}^* = T - 2k$; then we choose as $k^{(1)}$ the value of k among $0, \dots, K$ that minimizes

$$BNG_{1,k} = \log(\hat{\sigma}_{1,k}^2) + (k + 1) \frac{\log(\min(T_{1,k}^*, m))}{\min(T_{1,k}^*, m)}.$$

Suppose we have already computed $q - 1$ dynamic principal components, where the component i uses $k^{(i)}$ lags. Then we compute the q component with i lags for each $0 \leq i \leq K$ and the corresponding residuals matrix $(r_{t,j}^{(q,k)})_{h_{q,k}+1 \leq t \leq T, 1 \leq j \leq m}$, where $h_{q,k} = 2(k + \sum_{i=1}^{q-1} k^{(i)})$. Let $T_{q,k}^* = T - h_{q,k}$, $\mathbf{r}_j^{(q,k)} = (r_{h_{q,k}+1,j}^{(q,k)}, \dots, r_{T,j}^{(q,k)})$ and $\hat{\sigma}_{q,k} = ((1/m) \sum_{j=1}^m S_M(\mathbf{r}_j^{(q,k)}))^2$; then the value of $k^{(q)}$ is the value of k , $0 \leq k \leq K$, which minimize the following robustification of the Bai and Ng criterion

$$BNG_{q,k} = \log(\hat{\sigma}_{q,k}^2) + \left(\sum_{i=1}^{q-1} ((k^{(i)} + 1) + k + 1) \right) \frac{\log(\min(T_{q,k}^*, m))}{\min(T_{q,k}^*, m)}.$$

The selected number of components is $Q = q - 1$, where q is the minimum value q such that $BNG_{q,k^{(q)}} \geq BNG_{q,k^{(q-1)}}$.

6.2 Selection Using Robust Cross-validation

Suppose that we are interested in using the S-ODPC to predict $\mathbf{z}_{T+1}, \dots, \mathbf{z}_{T+h}$. We can apply the following robust cross-validation procedure for selecting the number of components, Q , and $k^{(q)}$, $1 \leq q \leq Q$, the number of lags used for each component. Suppose that the first $T_1 < T$ observations are chosen as the

training set, and the last $T - T_1$ observations as testing set. Then the training set will be used to compute all the loading vectors $\mathbf{a}^{(q,k)}$ for the q component with k lags and the testing set to evaluate the prediction power of any choice of the number of lags $k^{(q)}$ for each component q and the number of components Q .

The cross-validation procedure starts choosing k_1 as follows. For $0 \leq d \leq T - T_1 - i$, $1 \leq j \leq m$, $1 \leq i \leq h$ and $k \geq 0$, let $\widehat{z}_{T_1+d+i,j|T_1+d}^{(1,k)}$ be the prediction of $z_{T_1+d+i,j}$ using the first component with k lags and loading vector $\mathbf{a}^{(1,k)}$ corresponding to the periods $t \leq T_1 + d$, and let $\widehat{r}_{T_1+d+i,j|T_1+d}^{(1,k)} = \widehat{z}_{T_1+d+i,j|T_1+d}^{(1,k)} - z_{T_1+d+i,j}$ the corresponding prediction error. We evaluate the quality of the predictions up to h periods ahead using the first component with k lags by

$$SS_k^{(1)} = \sum_{j=1}^m \sum_{i=1}^h S_M(\widehat{\mathbf{r}}_j^{(1,k,i)}), \tag{16}$$

where $\widehat{\mathbf{r}}_j^{(1,k,i)} = (\widehat{r}_{T_1+i,j|T_1}^{(1,k)}, \dots, \widehat{r}_{T-h+i,j|T-h}^{(1,k)})$ is the vector of all the i periods ahead predictions. We select as $k^{(1)}$ the first k , such that $SS_k^{(1)} \leq SS_{k+1}^{(1)}$. Suppose now that we have already computed $q - 1$ components with lags $k^{(1)}, \dots, k^{(q-1)}$. To obtain $k^{(q)}$, we proceed as when computing $k^{(1)}$, that is, for each k , we compute

$$SS_k^{(q)} = \sum_{j=1}^m \sum_{i=1}^h S_M(\widehat{\mathbf{r}}_j^{(q,k,i)}), \tag{17}$$

where $\widehat{\mathbf{r}}_j^{(q,k,i)} = (\widehat{r}_{T_1+i,j|T_1}^{(q,k)}, \dots, \widehat{r}_{T-h+i,j|T-h}^{(q,k)})$, $\widehat{r}_{T_1+d+i,j|T_1+d}^{(q,k)} = \widehat{z}_{T_1+d+i,j|T_1+d}^{(q,k)} - z_{T_1+d+i,j}$, and $\widehat{z}_{T_1+d+i,j|T_1+d}^{(q,k)}$ is the prediction of $z_{T_1+d+i,j}$ assuming that $\mathbf{z}_1, \dots, \mathbf{z}_{T_1+d}$ are known, using the first $q - 1$ components with lags $k^{(1)}, \dots, k^{(q-1)}$ and the q component with k lags. The number of lags $k^{(q)}$ for the q component is defined as the first value of k , such that $SS_k \leq SS_{k+1}$. Similarly, the number of components Q is chosen as the first q such that $SS_{k^{(q)}}^q \leq SS_{k^{(q+1)}}^{q+1}$. The robustness of this procedures follows from the fact that all the options selected by the procedure are evaluated using a robust scale, see (16) and (17). The procedure to make the forecasting is described in Sect. 5. The case where k_1 may be different of k_2 can be treated similarly but with more computational effort.

The forecasting of a particular time series can be improved if we add a specific or idiosyncratic component that explains the residuals of the series. For that purpose, we may fit for each variable an ARMA model for the respective S-ODPC reconstruction residuals.

7 Asymptotic Behavior of the S-ODPC in Factor Models

Let $z_{t,j}^{(m)}, 1 \leq j \leq m, m > 1$, be observations generated by a dynamic one-factor model with k lags, that is, they satisfy

$$z_{t,j}^{(m)} = \varphi_j^{(m)} + c_{j,0}^{(m)} f_t + \dots + c_{j,k}^{(m)} f_{t-k} + u_{t,j}, 1 \leq j \leq m, \quad (18)$$

where f_t and $u_{t,j}, 1 \leq j \leq m$, are independent stationary process. We also have $\mathbf{E}(u_{t,j}) = \mathbf{E}(f_t) = 0$, $\text{var}(f_t) = \tau^2$, and $\text{var}(u_{t,j}) = \sigma_j^2$.

In this section, we study the behavior of the first population S-ODPC when m tends to infinite. This is stated more precisely in Theorem 1, which is the analogous of Theorem 3 of Peña et al. (2019), but using S-ODPC instead of ODPC. Consider the following Assumptions:

- A1. There exist $\varepsilon > 0$ and A_1 such that $0 < \varepsilon < \sigma_j^2 < A_1 < \infty$ for all j .
- A2. Let $s_j = S_M(u_{t,j})$; then

$$0 < s_j \leq A_2. \quad (19)$$

- A3. The function ρ has a derivative ψ that is continuous and bounded. Then $A_3 = \sup \psi < \infty$.
- A4. $A_4 = \inf_j \mathbf{E}(u_{t,j} \psi(u_{t,j})) > 0$.
- A5. There exists C such that $\sup_m \sup_{1 \leq j \leq m, 0 \leq i \leq k} |c_{j,i}^{(m)}| \leq C$ and $\sup_m \sup_{1 \leq j \leq m} |\varphi_j^{(m)}| \leq C$.
- A6. Let $\mathbf{c}_i^{(m)} = (c_{1,i}^{(m)}, \dots, c_{m,i}^{(m)})$, $0 \leq i \leq k$, and $E^{(m)}$ the subspace of \mathbb{R}^m generated by $\mathbf{c}_i^{(m)}, 1 \leq i \leq m$. Then, we can write $\mathbf{c}_0^{(m)} = \mathbf{d}^{(m)} + \mathbf{e}^{(m)}$, where $\mathbf{d}^{(m)}$ is orthogonal to $E^{(m)}$ and $\mathbf{e}^{(m)} \in E^{(m)}$. Then, there exists δ such that $\|\mathbf{d}^{(m)}\|^2 \geq m\delta$ for all m . This condition implies that the common part of $\mathbf{z}_t^{(m)} = (z_{t,1}^{(m)}, \dots, z_{t,m}^{(m)})$ does not get close to the $k - 1$ -dimensional subspace $E^{(m)}$ when m increases.

Theorem 1 Assume A1–A6. Let $\mathbf{z}_t^{(m)} = (z_{t,1}^{(m)}, \dots, z_{t,m}^{(m)})$ generated as the dynamic one-factor model given in (18) and $\mathbf{z}_t^{(m)*} = (z_{t,1}^{(m)*}, \dots, z_{t,m}^{(m)*})$ its population optimal reconstruction using the first S-ODPC with k_1 equal to any nonnegative integer and k_2 equal to k , the number of lags of the factor model. Then, there exists a constant K independent of m such that

$$\frac{1}{m^{1/2}} \left(\sum_{j=1}^m S_M^2(z_{t,j}^{(m)} - z_{t,j}^{(m)*})^2 - \sum_{j=1}^m S_M^2(u_{t,j}) \right) \leq K. \quad (20)$$

Remark 2 This theorem can be generalized to a model with k factors, if the first k S-ODPC are defined simultaneously instead of sequentially so that they minimize

the sum of the squares of the population M-scales. The proof is similar to the case of one factor.

8 Simulation Results

We generate matrices $(z_{t,j})$ of $T = 102$ observations and $m = 50$ time series using the following dynamic factor model:

$$z_{t,j} = d_{j,1}f_t + d_{j,2}f_{t-1} + d_{j,3}f_{t-2} + 0.2u_{t,j}, \quad 1 \leq t \leq 102, 1 \leq j \leq 50,$$

where $u_{t,j}$ are i.i.d. $N(0, 1)$ and f_t follows the autoregressive model

$$f_t = 0.85f_{t-1} + v_t$$

and v_t are i.i.d. $N(0, 1)$. The coefficients $d_{j,i}$ are generated in each replication as i.i.d. with uniform distribution in $[0, 1]$.

The first 100 observations are used to obtain the one-sided dynamic components (ODPC and S-ODPC), and the last two to evaluate the prediction performance of these methods.

These values $z_{t,j}$ are contaminated as follows: for $t = kT/H, k = 1, 2, \dots, H - 1$, the values of the observed series are

$$z_{t,j}^* = z_{t,j} + K.$$

Three values of H are considered, 0, 5, and 10, that is, 0%, 4%, and 9% of outliers, respectively, and the values of K are 3, 5, 10, and 15.

The number of replications is 500. For each case, we compute the first ODPC and the first S-ODPC with $k_1 = 1$ and $k_2 = 2$. The performance of both procedures for predicting the values of the series one and two periods ahead is evaluated by the sum of squares of the M-scales of the respective prediction error. Let $s_{M,j}$, $1 \leq j \leq 50$, be the M-scale of the prediction errors of the j -th variable. Then the performance of each procedure is evaluated by

$$\sum_{j=1}^{50} s_{M,j}^2.$$

The results are shown in Table 1.

We observe that under outlier contamination the forecasting error using the S-ODPC is much smaller than that using ODPC.

Table 1 Sum of the squares of the M-scales of the prediction errors. Pr1 and Pr2 stand for one and two steps ahead, respectively

%out	K	Pr1 ODPC	Pr1 S-ODPC	Pr2 ODPC	Pr2 S-ODPC
0		21.23	22.69	74.19	77.20
4	3	36.63	23.90	108.73	83.27
	5	53.63	23.13	128.54	81.92
	10	85.54	23.52	157.62	79.94
	15	133.15	23.33	202.20	78.43
9	3	51.14	24.79	127.61	98.63
	5	66.69	24.33	138.69	82.47
	10	147.56	23.34	208.53	79.51
	15	254.02	24.17	299.28	82.75

9 Example with a Real Data Set

We consider a multiple time series $z_{t,j}$, $1 \leq t \leq 678$, $1 \leq j \leq 24$, the electricity price in the Connecticut region, New England, during the Thursdays of 676 weeks for each of the 24 hours of the day. Then we have 24 series of 676 observations. The data can be obtained at www.iso-ne.com and were previously considered for clustering time series by Alonso and Peña (2019).

Figure 1 shows plots of 12 of these 24 series. The series appear to be highly correlated, and therefore, dimension reduction is expected to be useful. We observe that at every hour of the day there are outliers especially between the weeks 500 and 600, and therefore, a robust procedure seems to be appropriate for these data.

For each d , $1 \leq d \leq 177$, we apply the ODPC and S-ODPC procedures to the set $z_{t,j}$, $1 \leq t \leq 500 + d$, and predict the values of $z_{500+d+1,j}$, $1 \leq j \leq 24$.

Figure 2 shows that in general the M-scales of the prediction errors of the S-ODPC are smaller than those of the ODPC specially between 10 am and 8 pm.

As indicated by one of the referees, the large values around weeks 500 and 600 may be due to some interesting facts that, if considered, can provide important insights in the data analysis. We have not investigated this important issue for a rigorous analysis of these data, as our objective here is to illustrate the performance of our robust procedure. Also, the plot of the series suggests that they are not stationary. However, as is mentioned in Remark 1, the S-ODPC may be a useful tool to predict future values of the series even in this case.

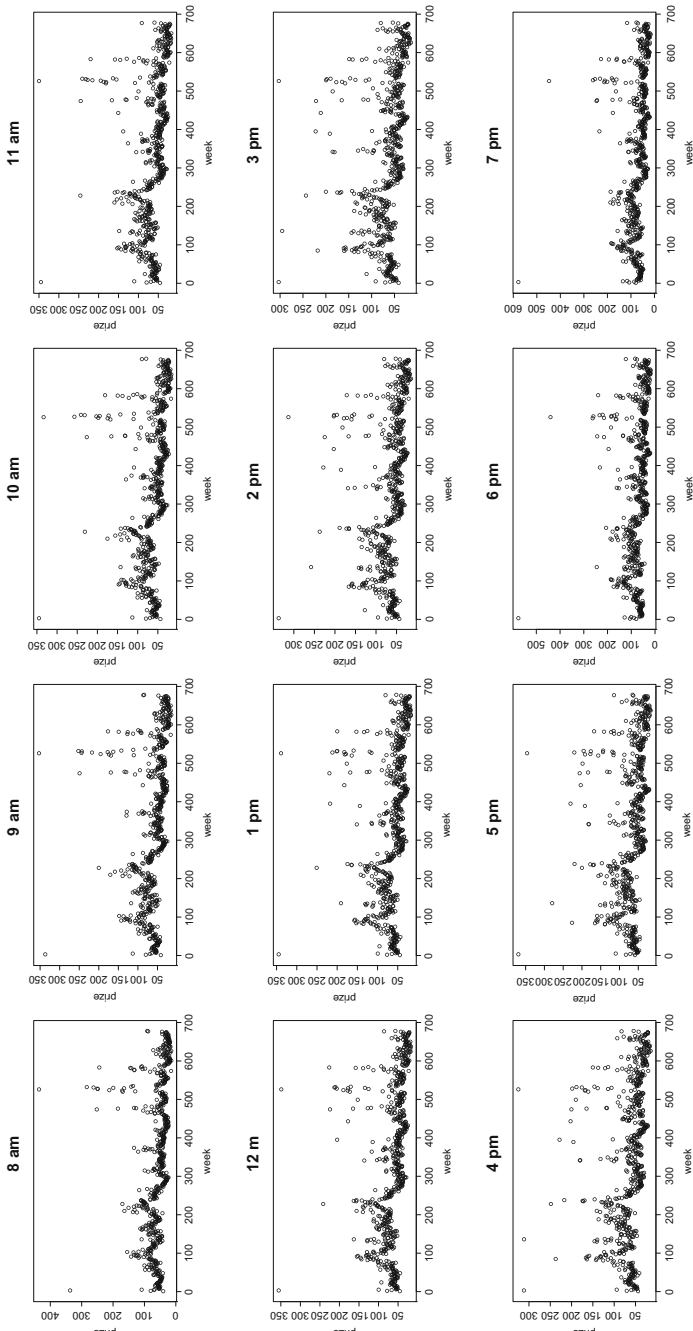


Fig. 1 Electricity data

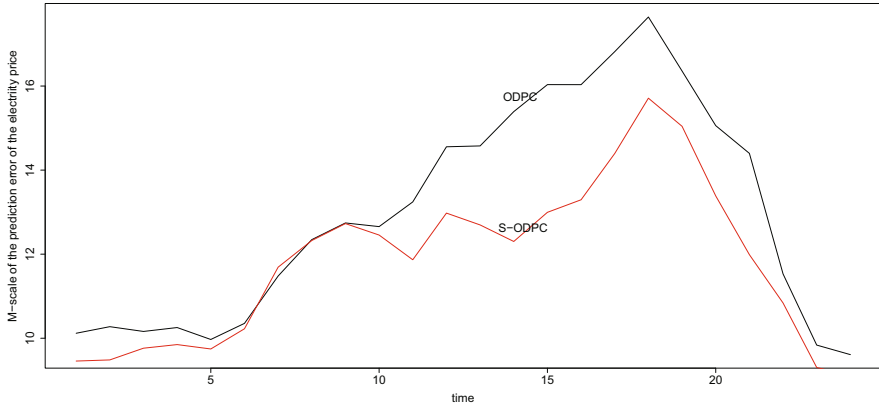


Fig. 2 M-scales of the one step ahead prediction errors

10 Conclusions

Given a vector series $\mathbf{z}_t, \leq t \leq T$, we have introduced the S one-sided dynamic principal components (S-ODPC) $f_t, k_1 + 1 \leq t \leq T$, defined as a linear combination of $\mathbf{z}_t, \mathbf{z}_{t-1} \dots \mathbf{z}_{t-k_1}$ that have the following properties:

- It allows the reconstruction $\widehat{\mathbf{z}}_t$ of the series \mathbf{z}_t for $k_1 + k_2 - 1 \leq t \leq T$ as a linear combination of $f_t, \dots f_{t-k_2}$.
- It allows the forecasting of \mathbf{z}_{T+h} .
- The reconstruction of the series and the forecasting of future values of \mathbf{z}_t can be improved taking higher order components.
- The values of k_1, k_2 and the number of components q can be chosen by: (a) a robust version of the Bai and Ng (2002) criterion for factor models or (b) cross-validation.
- The procedure S-ODPC is robust, that is, it can be applied successfully even in the presence of outliers.

Acknowledgments Daniel Peña was partially supported by the Spanish Agencia Nacional de Evaluación under grant PID2019-109196GB-I00, and Víctor J. Yohai was partially supported by Grants 20020170100330BA from the Universidad de Buenos Aires and PICT-201-0377 from ANPYCT.

Appendix

Derivation of the Estimating Equations

Call $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{B})$; then the value of $\boldsymbol{\theta}$ for the S-ODPC estimator is obtained minimizing

$$k_M(\boldsymbol{\theta}) = \sum_{j=1}^m s_M^2(\mathbf{r}_j(\boldsymbol{\theta})), \quad (21)$$

where $\mathbf{r}_j(\boldsymbol{\theta}) = (r_{k_1+k_2+1,j}(\boldsymbol{\theta}), \dots, r_{T,j}(\boldsymbol{\theta}))$, $r_{t,j}(\boldsymbol{\theta}) = z_{t,j} - \widehat{z}_{t,j}(\boldsymbol{\theta})$ satisfies

$$\frac{1}{T^*} \sum_{t=k_1+k_2+1}^T \rho\left(\frac{r_{t,j}(\boldsymbol{\theta})}{s_M(\mathbf{r}_j(\boldsymbol{\theta}))}\right) = b, \quad (22)$$

with $T^* = T - k_1 - k_2$. Differentiating (21), we get that the optimal $\boldsymbol{\theta}$ for the S-ODPC satisfies

$$\sum_{j=1}^m s_M(\mathbf{r}_j(\boldsymbol{\theta})) \frac{\partial s_M(\mathbf{r}_j(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = 0. \quad (23)$$

Let us differentiate (22). We have

$$\sum_{t=k_1+k_2+1}^T \psi\left(\frac{r_{t,j}(\boldsymbol{\theta})}{s_M(\mathbf{r}_j(\boldsymbol{\theta}))}\right) \frac{s_M(\mathbf{r}_j(\boldsymbol{\theta})) \partial r_{t,j}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} - r_{t,j}(\boldsymbol{\theta}) \partial s_M(\mathbf{r}_j(\boldsymbol{\theta})) / \partial \boldsymbol{\theta}}{s_M^2(\mathbf{r}_j(\boldsymbol{\theta}))} = 0$$

and

$$\frac{\partial s_M(\mathbf{r}_j(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = \sum_{t=k_1+k_2+1}^T \psi(r_{t,j}(\boldsymbol{\theta})/s_M(\mathbf{r}_j(\boldsymbol{\theta}))) \frac{s_M(\mathbf{r}_j(\boldsymbol{\theta})) \partial r_{t,j}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}}{\sum_{t=k_1+k_2+1}^T \psi(r_{t,j}(\boldsymbol{\theta})/s_M(\mathbf{r}_j(\boldsymbol{\theta}))) r_{t,j}(\boldsymbol{\theta})}.$$

This can also be written as

$$\begin{aligned} \frac{\partial s_M(\mathbf{r}_j(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} &= \frac{s_M(\mathbf{r}_j(\boldsymbol{\theta}))}{\sum_{t=k_1+k_2+1}^T w(r_{t,j}(\boldsymbol{\theta})/s_M(\mathbf{r}_j(\boldsymbol{\theta}))) r_{t,j}^2(\boldsymbol{\theta})} \\ &\times \sum_{t=k_1+k_2+1}^T w\left(\frac{r_{t,j}(\boldsymbol{\theta})}{s_M(\mathbf{r}_j(\boldsymbol{\theta}))}\right) r_{t,j}(\boldsymbol{\theta}) \frac{\partial r_{t,j}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \end{aligned}$$

where $w(u) = \psi(u)/u$. Then the estimating Eq. (23) satisfies

$$\sum_{t=k_1+k_2+1}^T w_{tj}^*(\boldsymbol{\theta}) r_{t,j}(\boldsymbol{\theta}) \frac{\partial r_{t,j}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0, \quad (24)$$

where

$$w_{t,j}^*(\boldsymbol{\theta}) = \frac{s_M^2(\mathbf{r}_{\cdot j}(\boldsymbol{\theta}))}{\frac{1}{T^*} \sum_{t=k_1+k_2+1}^T w(r_{t,j}(\boldsymbol{\theta})/s_M(\mathbf{r}_{\cdot j}(\boldsymbol{\theta}))) r_{t,j}^2(\boldsymbol{\theta})} w(r_{t,j}(\boldsymbol{\theta})).$$

Note that when $w_{t,j}^*(\boldsymbol{\theta}) = 1$ for all (t, j) , (24) is the estimating equation of the ODPC. Therefore, the only difference between the estimating equation of the RODPC and the one of the ODPC is that the least squares solutions to obtain the optimal values of \mathbf{a} and \mathbf{B} for the RODPC give weight $w_{t,j}^*(\boldsymbol{\theta})$ to the observation (t, j) . Then we obtain (13) and (14).

Proof of Theorem 1

Lemma 1 Suppose that $z_{t,j}$ satisfies A1, A4, A5, and A6; then there exist $\mathbf{a}^{(m)} \in \mathbb{R}^m$ and a $m \times (k + 2)$ matrix $\mathbf{B}^{(m)}$ such that if $g_t^{(m)} = \mathbf{a}^{(m)'} \mathbf{z}_t$ and $\mathbf{F}_t^{(m)} = (1, g_t^{(m)}, \dots, g_{t-k}^{(m)})$, the reconstruction $\tilde{\mathbf{z}}_t^{(m)} = \mathbf{B}^{(m)} \mathbf{F}_t^{(m)}$ satisfies

$$\mathbf{E}(z_{tj}^{(m)} - \tilde{z}_{t,j}^{(m)} - u_{t,j})^2 \leq \frac{K_1}{m} \quad (25)$$

for some constant K_1 .

Proof Let $\mathbf{u}_t^{(m)} = (u_{t,1}, \dots, u_{t,m})$ and $\boldsymbol{\varphi}^{(m)} = (\varphi_1^{(m)}, \dots, \varphi_m^{(m)})$; then

$$\mathbf{z}_t^{(m)} = \boldsymbol{\varphi}^{(m)} + f_t(\mathbf{d}^{(m)} + \mathbf{e}^{(m)}) + f_{t-1} \mathbf{c}_1^{(m)} + \dots + f_{t-k} \mathbf{c}_k^{(m)} + \mathbf{u}_t^{(m)},$$

where $\mathbf{d}^{(m)}$ and $\mathbf{e}^{(m)}$ are as in A6. Let $\mathbf{a}^{(m)} = \mathbf{d}^{(m)} / \|\mathbf{d}^{(m)}\|^2$; then by A6,

$$\|\mathbf{a}^{(m)}\| \leq 1/(m^{1/2} \delta^{1/2}). \quad (26)$$

Define $g_t^{(m)} = \mathbf{a}^{(m)'} \mathbf{z}_t^{(m)}$, and observe that

$$g_t^{(m)} = f_t + p^{(m)} + \eta_t^{(m)},$$

where $\eta_t^{(m)} = \mathbf{a}^{(m)'} \mathbf{u}_t^{(m)}$ and $p^{(m)} = \mathbf{a}^{(m)'} \boldsymbol{\varphi}^{(m)}$. Then, by (26) and A1, we have

$$\mathbf{E}(\eta_t^{(m)2}) \leq D/m \quad (27)$$

with $D = A_1/\delta$. Let us reconstruct $\mathbf{z}_t^{(m)}$ using $g_t^{(m)}$ as follows:

$$\tilde{\mathbf{z}}_t^{(m)} = (\boldsymbol{\varphi}^{(m)} - p^{(m)}(\mathbf{c}_0 + \dots + \mathbf{c}_k)) + g_t^{(m)} \mathbf{c}_0^{(m)} + g_{t-1}^{(m)} \mathbf{c}_1^{(m)} + \dots + g_{t-k}^{(m)} \mathbf{c}_k^{(m)}.$$

That is, if $\mathbf{B}^{(m)}$ is the $m \times (k + 2)$ with columns $\boldsymbol{\varphi}^{(m)} - p^{(m)}(\mathbf{c}_0 + \dots + \mathbf{c}_k)$ and $\mathbf{c}_i, 0 \leq i \leq k$, we have

$$\tilde{\mathbf{z}}_t^{(m)} = \mathbf{B}^{(m)} \mathbf{F}_t^{(m)}$$

and

$$\begin{aligned} z_{t,j}^{(m)} - \tilde{z}_{t,j}^{(m)} &= -c_{j,0} \eta_t^{(m)} - c_{1,j} \eta_{t-1}^{(m)} - \dots - c_{j,k} \eta_{t-k}^{(m)} + u_{t,j} \\ &= v_{t,j}^{(m)} + u_{t,j}, \end{aligned}$$

where

$$v_{t,j}^{(m)} = -c_{j,0} \eta_t^{(m)} - c_{1,j} \eta_{t-1}^{(m)} - \dots - c_{j,k} \eta_{t-k}^{(m)}. \tag{28}$$

Then by (27) and A5, there exists K_1 such that for all $1 \leq j \leq m$

$$\mathbf{E}(v_{t,j}^{(m)2}) \leq \frac{K_1}{m},$$

and this proves the Lemma.

Lemma 2 Suppose that $z_{t,j}$ satisfies A1–A6, and let $\mathbf{a}^{(m)}, \mathbf{B}^{(m)}$, and $\tilde{\mathbf{z}}_t^{(m)}$ as in Lemma 1, and then there exists K_2 independent of m such that

$$S_M(z_{tj}^{(m)} - \tilde{z}_{t,j}^{(m)}) \leq S_M(u_{t,j}) + \frac{K_2}{m} \tag{29}$$

for some constant K_2 .

It is enough to show that there exists m_0 such that for $m \geq m_0$ (29) holds. Let for $k > 0$ and v

$$L_{tj}(v, k) = \rho\left(\frac{v + u_{tj}}{s_j + k}\right),$$

where $s_j = S_M(u_{t,j})$.

Using the mean value theorem at $(0, 0)$, we get

$$L_{tj}(v, k) = L_{tj}(0, 0) + v \frac{\psi\left(\frac{v^* + u_{tj}}{s_j + k^*}\right)}{s_j + k^*} - k \frac{\psi\left(\frac{v^* + u_{tj}}{s_j + k^*}\right) \frac{v^* + u_{tj}}{s_j + k^*}}{s_j + k^*}, \tag{30}$$

where $|v^*| \leq |v|$ and $0 \leq k^* \leq k$. Put $k = K_2/m^{1/2}$. Since $z_{t,j}^{(m)} - \tilde{z}_{t,j}^{(m)} = v_{t,j}^{(m)} + u_{t,j}$, where $v_{t,j}^{(m)}$ is defined in Eq. (28) of Lemma 1, then using (30), we can write

$$\begin{aligned} \mathbf{E} \left(\rho \left(\frac{z_{t,j}^{(m)} - \tilde{z}_{t,j}^{(m)}}{s_j + K_2/m^{1/2}} \right) \right) &= \mathbf{E}_{tj} (L(v_{t,j}^{(m)}, K_2/m^{1/2})) \\ &= b + \frac{\mathbf{E} \left(v_{t,j}^{(m)} \psi \left(\frac{v^* + u_{t,j}}{s_j + K^*/m^{1/2}} \right) \right)}{s_j + K^*/m^{1/2}} \\ &\quad - \frac{K_2}{m^{1/2}} \frac{\mathbf{E} \left(\psi \left(\frac{v^* + u_{t,j}}{s_j + K^*/m^{1/2}} \right) \frac{v^* + u_{t,j}}{s_j + K^*/m^{1/2}} \right)}{s_j + K^*/m^{1/2}}, \end{aligned} \tag{31}$$

where $|v^*| \leq |v_{t,j}|$ and $K^* \leq K_2$.

Put

$$K_2 = 2A_3 K_1^{1/2} / A_4. \tag{32}$$

Since by Lemma 1 $\mathbf{E}(v_{t,j}^{(m)2}) \leq K_1/m$ and by A3 $A_3 = \max \psi(u)$, we have

$$\begin{aligned} \left| \frac{\mathbf{E} \left(v_{t,j}^{(m)} \psi \left(\frac{v^* + u_{t,j}}{s_j + K^*/m^{1/2}} \right) \right)}{s_j + K^*/m^{1/2}} \right| &\leq \frac{\mathbf{E}(v_{t,j}^{(m)2})^{1/2} \max_u \psi(u)}{s_j + K^*/m^{1/2}} \\ &\leq \frac{A_3 K_1^{1/2} / m^{1/2}}{s_j + K^*/m^{1/2}}. \end{aligned} \tag{33}$$

Take

$$K_2 > A_3 K_1^{1/2} / A_4; , \tag{34}$$

then since $K^* \leq K_2$ and $|v^*| \leq |v_{t,j}|$ by Lemma 1 and A4, there exists m_0 such that for $m \geq m_0$

$$\begin{aligned} &\frac{K_2}{m^{1/2} s_j + K^*/m^{1/2}} \mathbf{E} \left(\psi \left(\frac{v^* + u_{t,j}}{s_j + K^*/m^{1/2}} \right) \frac{v^* + u_{t,j}}{s_j + K^*/m^{1/2}} \right) \\ &\geq \frac{K_2 A_4 / 2}{m^{1/2} s_j + K^*/m^{1/2}}. \end{aligned} \tag{35}$$

Then by (31), (33), and (35), we have

$$\left(\rho \left(\frac{z_{t,j}^{(m)} - \tilde{z}_{t,j}^{(m)}}{s_j + K_2/m^{1/2}} \right) \right) < b \quad ,$$

and therefore, $S_M(z_{t,j}^{(m)} - \tilde{z}_{t,j}^{(m)}) \leq s_j + K_2/m^{1/2}$. This proves the Lemma.

Proof of Theorem 1

From Lemma 2, it can be derived that if $z_{t,j}^{*(m)}$ is the reconstruction with the optimal first S-ODPC with any k_1 and $k_2 = k$, we should have

$$\begin{aligned} \frac{1}{m^{1/2}} \sum_{j=1}^m S_M^2(z_{t,j}^{(m)} - z_{t,j}^{*(m)}) &\leq \frac{1}{m^{1/2}} \sum_{j=1}^m S_M^2(z_{t,j}^{(m)} - \tilde{z}_{t,j}^{(m)}) \\ &\leq \sum_{j=1}^m \left(s_j^2 + \frac{K_2^2}{m} + 2 \frac{K_2}{m^{1/2}} s_j \right) \\ &\leq \frac{1}{m^{1/2}} \sum_{j=1}^m \left(s_j^2 + \frac{K}{m^{1/2}} \right) \\ &= \frac{1}{m^{1/2}} \sum_{j=1}^m s_j^2 + K, \end{aligned}$$

where $K = \max(K_2^2, K_3)$ and $K_3 = 2K_2 \max_j s_j$. This proves Theorem 1.

References

Alonso, A. M., Galeano, P., & Peña, D. (2020). A robust procedure to build dynamic factor models with cluster structure. *Journal of Econometrics*, 216(1), 35–52.

Alonso, A. M. & Peña, D. (2019). Clustering time series by linear dependency. *Statistics and Computing*, 29(4), 655–676.

Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.

Brillinger, D. R. (1964). The generalization of the techniques of factor analysis, canonical correlation and principal components to stationary time series. In *Invited Paper at the Royal Statistical Society Conference in Cardiff, Wales*.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

Davies, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Annals of Statistics*, 15, 1269–1292.

Fan, J., Wang, K., Zhong, Y., & Zhu, Z. (2021). Robust high dimensional factor models with applications to statistical machine learning. *Statistical Science*, 36(2), 303.

Fan, J., Wang, W., & Zhong, Y. (2019). Robust covariance estimation for approximate factor models. *Journal of Econometrics*, 208(1), 5–22.

- Forni, M., Hallin, M., Lippi, M., & Zaffaroni, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, 185(2), 359–371.
- Forni, M., Hallin, M., Lippi, M., & Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *Journal of Econometrics*, 199(1), 74–92.
- Huber, P. J. (1964). *Robust statistics*. New York: Wiley.
- Maronna, R. (2005). Principal Components and Orthogonal Regression Based on Robust Scales. *Technometrics*, 47(3), 264–273.
- Maronna, R., Martin, D., Yohai, V., & Salibian-Barrera, M. (2019). *Robust Statistics*. New York: Wiley.
- Muler, N., Peña, D., & Yohai, V. J. (2009). Robust estimation for ARMA models. *The Annals of Statistics*, 37(2), 816–840.
- Muler, N. & Yohai, V. J. (2013). Robust estimation for vector autoregressive models. *Computational Statistics & Data Analysis*, 65, 68–79.
- Peña, D., Smucler, E., & Yohai, V. J. (2019). Forecasting Multiple Time Series with One-Sided Dynamic Principal Components. *Journal of the American Statistical Association*, 114(528), 1683–1694.
- Peña, D. & Yohai, V. J. (2016). Generalized dynamic principal components. *Journal of the American Statistical Association*, 111(515), 1121–1131.
- Rousseeuw, P. J. & Yohai, V. J. (1984). Robust regression by means of S-estimators. In J. Franke, W. Härdle, & D. Martin (Eds.), *Robust and Nonlinear Time Series*. Lecture Notes in Statistics (vol. 26, pp. 256–272). New York: Springer.
- Salibian-Barrera, M. & Yohai, V. J. (2006). A Fast Algorithm for S-Regression Estimates. *Journal of Computational and Graphical Statistics*, 15(2), 414–427.
- Trucíos, C., Mazzeu, J. H., Hotta, L. K., Pereira, P. L. V., & Hallin, M. (2021). Robustness and the general dynamic factor model with infinite-dimensional space: identification, estimation, and forecasting. *International Journal of Forecasting*, 37(4), 1520–1534.