A Review of Outlier Detection and Robust Estimation Methods for High Dimensional Time Series Data

Daniel Peña [a,*], Víctor J. Yohai [b]

[a] *Department of Statistics, Universidad Carlos III de Madrid, Getafe, 28903, Madrid, Spain*
[b] *Facultad de Ciencias Exactas, Universidad de Buenos Aires, Buenos Aires, Argentina*

ARTICLE INFO

ABSTRACT

Diagnostic procedures for finding outliers in high dimensional multivariate time series and robust estimation methods for these data are reviewed. First, methods for searching for outliers assuming that the data have been generated by a Dynamic Factor Model are presented. Then, other existing methods for detecting different types of multivariate time series outliers are analyzed. They include identifying outlying series and, also, looking for segments, or periods of time, where the series have unusual dynamics. Second, robust estimation methods are considered. Dynamic Principal Components, as a very general procedure to estimate the dynamic in a high dimensional data set, is introduced and different types of robust estimation of these components are reviewed. Dynamic Principal Components can be applied for robust estimation of Generalized Dynamic Factor models and some results are given. Finally, other methods proposed for robust estimation of high dimensional VAR models and other multivariate time series problems are also briefly discussed.

1. Introduction

High-dimensional sets of time series are often collected in many fields, as multimedia, health sciences, engineering and economics and finance. These data are often recorded in an automatic way with wireless sensor networks, including computer vision, electrodes for Electroencephalogram (EEG) or Functional Magnetic Resonance Imaging (fMRI). The devices used may fail to record the observations correctly for different reasons, producing outliers in the data. As it is well known that outliers may lead to inefficient estimation and poor forecasts, robust estimation and outlier detection are very important tools in the analysis of these high-dimensional (HD) multivariate time series (MTS) data sets.

Outliers may appear in time series in different ways. The first, and most studied case, corresponds to some unusual event at a given time point, so that the observed value is heterogeneous with the others. In particular, this value may have been recorded with an error. For instance, we have a set of monitoring devices that are controlling the pollution in a city and a power failure occurs at a given time. In this situation an isolated multivariate outlier will appear at the time of the

---

* Corresponding author.
  E-mail address: daniel.pena@uc3m.es (D. Peña).

unusual event. A more complex situation occurs when in a period of time the observations of the MTS are made under different conditions from the others, or, in particular, they are badly recorded. For instance, the number of admissions in the hospitals of Argentina or Spain when the COVID started was unusual, given the past hospital experience. In these cases, we have a patch of outliers, or a time segment that is an outlier in the MTS data.

In a third situation, a few of the time series in the MTS set are not generated in the same way as the others and, therefore, they show a different dynamic behavior. As a consequence, these series will be outliers in the set. For instance, a hospital is recording the time evolution of a certain disease in each of a set of patients and a few of them have a very unusual performance. Finally, we may have many sets of MTS data and some of them, the entire set, may be an outlier with respect to the other sets. For instance, we measured over time the performance of groups of persons under some stressing situations and some set is heterogeneous with the others, due to different environmental conditions.

Detection of individual outliers in MTS models was introduced by Tsay et al. (2000). They generalized the four types of outliers proposed for univariate series (see Fox (1972); Tsay (1986) and Chen and Liu (1993)) for vectors of time series. These authors also investigated the outlier effects on parameter estimation and forecasting in VARMA models and proposed an iterative detection method to find outliers, based on individual and joint likelihood ratio statistics. Galeano et al. (2006) introduced a different approach based on projections, that can be used with larger number of time series. In this procedure, outliers are found by projecting the vector of time series is some useful directions and applying univariate detection methods to the projected univariate time series. The projection directions chosen are those that lead to extreme values of the kurtosis coefficient of the projected univariate series. It was proved that these directions have some optimal properties. Cucina et al. (2014) assumed that the data can be approximated by a Vector Autoregressive (VAR) model and proposed a genetic algorithm to find additive outliers.

As the number of parameters grows with the squared of the number of time series, procedures for finding outliers in VARMA models are less useful when the dimension of the data set is large. A flexible structure for representing high dimensional time series are dynamic factor models (DFM) and this representation have been used to identify multivariate outliers. The first approach in this direction was made by Baragona and Battaglia (2007b), that proposed a procedure to identify the times of occurrence of multivariate additive outliers in DFM. An alternative procedure was proposed by Galeano and Peña (2020), and Chou et al. (2020) developed a penalized least squares method to estimate the parameters of DFMs subject to large outliers.

The identification of outlier time series has been studied more recently, often related to cluster analysis. For instance, Alonso et al. (2020) presented an outlier detection procedure for DFM with cluster structure and, in a natural way, they also searched for series in the set that are outliers. Methods for finding outlier series and segments of unusual behavior in a MTS have been also been proposed in the computer science and machine learning literature. See Aggarwal (2017); Blázquez-García et al. (2021) and Meng et al. (2021). However, the connections between procedures developed in these fields and those in statistics have been small. Statistical methods have been mostly concentrated on finding outlier detection procedures with some optimal properties, whereas computer science and machine learning methods have relied more in the practical performance in simulations and specific applications. Thus, very few comparative studies between approaches proposed in different fields have been carried out, see López-Oriona and Vilar (2021a) for one exception of this rule and a comparison of several methods.

An alternative approach to deal with outliers in MTS is to use a robust estimation method. We should note that in the statistical literature two different criteria are used to qualify a statistical procedure as robust. Some authors describe as robust a method that is suitable for estimating models where the errors have heavy tails distributions. On the other hand, a stronger criterion qualify as robust a method that is little sensitive to the presence of an arbitrary positive fraction of large (even extreme) outliers in the sample. These two criteria may not coincide. For example, in regression models the L_1 estimator is robust according to the first criterion, but not to the second (see for instance Sections 5.3.1.1 and 5.13.1 of Maronna et al. (2017)). Instead, the least median of squares estimator proposed by Rousseeuw (2005), or the S estimators proposed by Rousseeuw and Yohai (1984)), are robust according to both criteria.

To quantify the degree of robustness of an statistical procedure according to the second criterion, several measures have been proposed. One of the most important ones is the finite sample breakdown point of an estimator, proposed by Donoho and Huber (1983). Let $x_1, \ldots, x_n$ be a random sample of a distribution that depends on a parameter $\theta \in R^p$ and let $\widehat{\theta}$ be an estimator of it. The finite sample breakdown point of $\widehat{\theta}$ is the minimum fraction of outliers in the sample that may cause the estimator to take arbitrarily large values. More information about the finite sample breakdown point and other measures of robustness can be found in the books by Hampel et al. (1986) and Maronna et al. (2017).

Robust estimation of MTS that follow a Vector Autoregressive (VAR) model was studied by Muler and Yohai (2013). However, for high dimensional time series, as indicated before, other alternative models are needed. An overview of robust ideas in signal processing can be seen in Zoubir et al. (2012) and in neural networks in Su et al. (2019).

A dimension reduction technique for high dimensional time series was proposed by Peña and Yohai (2016), following Brillinger's idea of dynamic principal components (DPC), see Brillinger (1964, 1981). They proposed a more general formulation, where the DPC are computed without the assumption of being linear combinations of the data. These authors also developed a robust estimation procedure for these principal components. However, although these DPC are useful for reconstruction of the set of time series they are not oriented to forecasting. Peña et al. (2019a) introduced the one-sided dynamic principal components (ODPC), that are defined as linear combination 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). A robust

estimation method for these ODPC have been recently proposed by Peña and Yohai (2023), that is shown to provide a consistent way to estimate a Dynamic Factor model. Other alternative robust estimation for Dynamic factor models proposed by Fan et al. (2019b), Fan et al. (2021a) and Trucíos et al. (2021), will be briefly reviewed.

The organization of this article is as follows. In Section 2.2 we discuss finding outliers in Dynamic factor models (DFM). First, these models are described as a general representation for high dimensional multivariate time series and, second, some of the procedures proposed to identify outliers in them are discussed. In Section 3 other methods to find outliers, including outlying time series and outlier segments in MTS data sets, are presented. Section 4 reviews some static and dynamic principal components methods as a general tool for dimension reduction in MTS. Robust estimation methods for DPC are discussed in Section 5, where the scale one-sided dynamic principal component (SODPC) robust estimation method is described. Section 6 presents a robust forecasting procedure using the scale one-sided DPC (SODPC), and Section 7 a robust criterion to choose the number of components and the number of lags for each component. Section 8 deals with robust estimation of DFM. It is shown that the SODPC can be used for this objective and, also, other alternative procedures are presented. Section 9 discusses other robust approaches for MTS estimation. Section 10 reviews some real data examples of the methods covers in this article and Section 11 contains the main conclusions of this survey.

2. Outliers in Dynamic Factor Models

In this section the dynamic factor model (DFM) is reviewed, outliers in this model are defined and different outlier identification procedures are presented. The dynamic factor model can be considered as a general representation of a high dimensional vector of time series in the following way. Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{mt})'$ be an m-dimensional vector time series with strong cross correlated dynamic dependency. We can always decompose this vector in the projection on two orthogonal subspaces. Thus, for any orthogonal matrices $\mathbf{P}$, of dimension $m \times r$, and $\mathbf{V}$, of dimension $m \times m - r$, and such that $\mathbf{V}'\mathbf{P} = \mathbf{0}$,

$$\mathbf{x}_t = \mathbf{P}\mathbf{P}'\mathbf{x}_t + \mathbf{V}\mathbf{V}'\mathbf{x}_t = \mathbf{c}_t + \mathbf{u}_t. \tag{1}$$

The matrix $\mathbf{P}$ can be chosen so that the vector $\mathbf{c}_t = \mathbf{P}\mathbf{P}'\mathbf{x}_t$ includes most of the variability of the original series. If we use the condition that the sum of the variances of the components of this vector is as large as possible, we end up selecting the first r principal components of the covariance matrix of the data. Let us call $\mathbf{p}_t = \mathbf{P}'\mathbf{x}_t$, then the residual, or remaining, term, $\mathbf{u}_t = \mathbf{V}\mathbf{V}'\mathbf{x}_t$, includes the $m - r$ principal components with minimum sum of variances, and we have the decomposition,

$$\mathbf{x}_t = \mathbf{P}\mathbf{p}_t + \mathbf{u}_t. \tag{2}$$

Equation (2) is completely general and can be applied to any vector of time series. It implies that the vector $\mathbf{x}_t$ is generated by a small number of factors, r, but with the strong assumption that $\mathbf{c}'_t\mathbf{u}_t = \mathbf{p}'_t\mathbf{P}'\mathbf{u}_t = 0$ for all t. A more flexible decomposition, but also quite general, is to assume that the noises are uncorrelated with the factors in an asymptotic way, and allow for noises of dimension m, instead of $m - r$. These generalizations leads to the dynamic factor model.

2.1. Dynamic Factor Models

We assume that the vector of observed time series, $\mathbf{x}_t$, can be written as a linear combination of a small number of dynamic factors plus an idiosyncratic noise. Assuming, first, contemporaneous factors, the model is

$$\mathbf{x}_t = \mathbf{P}\mathbf{f}_t + \mathbf{n}_t, \tag{3}$$

where $\mathbf{P}$ is a $m \times r$ factor loading matrix, $\mathbf{f}_t = (f_{1t}, \ldots, f_{rt})'$ is an r-dimensional vector of contemporaneous dynamic factors and $\mathbf{n}_t = (n_{1t}, \ldots, n_{mt})'$ is a sequence of noises. Several conditions for the identification of this model have been proposed (see, for instance, Peña and Tsay (2021)), and here we assume that $\mathbf{P}'\mathbf{P} = \mathbf{I}_r$, where $\mathbf{I}_r$ is the r-dimensional identity matrix and $r < m$. The dynamic factors follow a diagonal VARMA(p, q) model,

$$\boldsymbol{\Phi}(B)\mathbf{f}_t = \boldsymbol{\Theta}(B)\mathbf{u}_t, \tag{4}$$

where $\boldsymbol{\Phi}(B) = \mathbf{I}_r - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$, and $\boldsymbol{\Theta}(B) = \mathbf{I}_r - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$, are polynomial matrices in the back-shift operator B with diagonal parameter matrices. The roots of $|\boldsymbol{\Phi}(B)| = 0$ are on or outside the unit circle, while those of $|\boldsymbol{\Theta}(B)| = 0$ are outside the unit circle, and $\mathbf{u}_t$ is a sequence of i.i.d. random vectors with mean $\mathbf{0}_r$ and full rank diagonal covariance matrix $\boldsymbol{\Sigma}_\mathbf{u}$. Consequently, the dynamic factors, i.e., $f_{1t}, \ldots, f_{rt}$, can be either stationary or integrated. When the dynamic factors are stationary, the model defined in (3) and (4) has been studied under different assumptions by many authors, including Peña and Box (1987); Stock and Watson (2002) and Bai and Ng (2002).

We assume that the vector of noises $\mathbf{n}_t$ in (3) is stationary and may include both weakly correlated and cross correlated random variables, as defined in Stock and Watson (2002) and Bai and Ng (2002), with mean vector $\mathbf{0}_m$ and full rank covariance matrix. Additionally, we assume that both noises appearing in the model are uncorrelated for all lags, i.e., $E[\mathbf{n}_t\mathbf{u}'_{t-h}] = \mathbf{0}_{m \times r}$, for all $h = 0, \pm 1, \pm 2, \ldots$ This assumption can be relaxed allowing some weak cross-correlation between factors and idiosyncratic noises.

When $\mathbf{f}_t$ is stationary, $\mathbf{x}_t$ will also be stationary, and we can compute autocovariance matrices, $\boldsymbol{\Gamma}_\mathbf{x}(h) = E\left[(\mathbf{x}_t - \boldsymbol{\mu}_\mathbf{x})(\mathbf{x}_{t-h} - \boldsymbol{\mu}_\mathbf{x})'\right]$, where $\mu_\mathbf{x} = E[\mathbf{x}_t]$, with terms $\gamma_{ij}(h)$ for $h = 0, \pm 1, \pm 2, \ldots$, and $1 \leq i, j \leq m$. These autocovariance

matrices are given by

$$\mathbf{\Gamma_x}(h) = \mathbf{P}\mathbf{\Gamma_f}(h)\mathbf{P}' + \mathbf{\Gamma_n}(h),\tag{5}$$

where $\mathbf{\Gamma_f}(h)$, for $h = 0, \pm1, \pm2, \ldots$, are the square r-dimensional diagonal autocovariance matrices of the factors, and $\mathbf{\Gamma_n}(h)$, for $h = 0, \pm1, \pm2, \ldots$, are the square m-dimensional autocovariance matrices of the idiosyncratic components. Note that for any column $\mathbf{p}_{\cdot k}$ of the loading matrix $\mathbf{P}$, where $k \in \{1, \ldots, r\}$, we have $\mathbf{\Gamma_x}(0)\mathbf{p}_{\cdot k} = \left(\sigma_{\mathbf{f},k}^2 \mathbf{I}_m + \mathbf{\Gamma_n}(0)\right)\mathbf{p}_{\cdot k}$, where $\sigma_{\mathbf{f},k}^2$ is the variance of the k-th factor, whereas for any m-dimensional vector $\mathbf{v}$ such that $\mathbf{P}'\mathbf{v} = \mathbf{0}_r$, we have $\mathbf{\Gamma_x}(0)\mathbf{v} = \mathbf{\Gamma_n}(0)\mathbf{v}$. In the particular case in which the idiosyncratic noises are white noise with the same variance, so that $\mathbf{\Gamma_n}(0) = \sigma_{\mathbf{n}}^2 \mathbf{I}_m$, then the matrix $\mathbf{\Gamma_x}(0)$ will have r eigenvalues of sizes $\sigma_{\mathbf{f},1}^2 + \sigma_{\mathbf{n}}^2, \ldots, \sigma_{\mathbf{f},r}^2 + \sigma_{\mathbf{n}}^2$ and $m - r$ eigenvalues of size $\sigma_{\mathbf{n}}^2$. Also, $\mathbf{\Gamma_x}(h)$ will have r eigenvalues of size $\gamma_{\mathbf{f},1}(h), \ldots, \gamma_{\mathbf{f},r}(h)$ linked to the r columns of the loading matrix $\mathbf{P}$ and $m - r$ zero eigenvalues linked to $m - r$ orthonormal eigenvectors, denoted by $\mathbf{v}_{\cdot 1}, \ldots, \mathbf{v}_{\cdot m-r}$, that define the orthogonal complement matrix $\mathbf{V} = [\mathbf{v}_{\cdot 1} | \ldots | \mathbf{v}_{\cdot m-r}]$, where $\mathbf{V}'\mathbf{P} = \mathbf{0}_{(m-r) \times r}$. The subspace defined by $\mathbf{P}$ is the factor space, as $\mathbf{P}'\mathbf{x}_t = \mathbf{f}_t + \mathbf{P}\mathbf{n}_t$, and the subspace defined by $\mathbf{V}$ will be called the idiosyncratic space, as $\mathbf{V}'\mathbf{x}_t = \mathbf{V}'\mathbf{n}_t$.

The estimation of the factors can be made by principal components using the intuition that, as

$$\mathbf{P}'\mathbf{x}_t = \mathbf{f}_t + \mathbf{P}'\mathbf{n}_t,\tag{6}$$

for large samples the second term will be close to its mean, zero, and the factors close to $\mathbf{P}'\mathbf{x}_t$. Assuming in (5) that in $\mathbf{\Gamma_x}(0)$ the signal $\mathbf{P}\mathbf{\Gamma_f}(0)\mathbf{P}'$ dominates the noise, $\mathbf{\Gamma_n}(0)$, $\mathbf{P}$ can be estimated by the eigenvectors corresponding to the largest r eigenvalues of the covariance matrix. This principal components estimation of the factors has been proved to be consistent under general conditions when T and m go to infinity (see Stock and Watson (2002)). For this estimation we have to determine the number of factors, r, that can be made using any of the several available procedures (see, for instance, Peña and Tsay (2021)). This PCA estimation of DFM will be considered in Section 4.1, but other possible estimation methods, including maximum likelihood, have been proposed.

Geweke (1977) introduced a factor model with lags in the factors, that has been generalized by Forni et al. (2000). This is the so-called generalized dynamic factor model (GDFM). Assuming a finite number of lags, L, its representation is:

$$\mathbf{x}_t = \sum_{l=0}^{L-1} \mathbf{P}_l \mathbf{f}_{t-l} + \mathbf{n}_t,\tag{7}$$

where $\mathbf{f}_t$ is a q-dimensional stationary vector of uncorrelated common factors, the L matrices $\mathbf{P}_l$ are $m \times q$ factor loading matrices and $\mathbf{n}_t$ is the noise or idiosyncratic part, that is stationary and follow the conditions of the approximate DFM. The GDFM was proposed by allowing L to go to infinity but here we only discuss the case that L is finite, that is the most important in practice. The conditions for identification of the loading matrices and the factors are similar to the DFM, $\mathbf{P}_l'\mathbf{P}_l = \mathbf{I}_q$ for $l = 0, 1, \ldots, L - 1$, and $E(\mathbf{f}_t\mathbf{f}_t')$ is diagonal. Model (7) can also be written as

$$\mathbf{x}_t = \mathbf{P}(B)\mathbf{f}_t + \mathbf{n}_t,$$

where $\mathbf{P}(B) = \mathbf{P}_0 + \mathbf{P}_1 B + \ldots + \mathbf{P}_{L-1}B^{L-1}$ is a matrix polynomial in the lag operator. Assuming that the factors follow a VAR model with diagonal parameter matrices

$$\mathbf{A}(B)\mathbf{f}_t = \mathbf{a}_t,$$

and,

$$\mathbf{x}_t = \mathbf{\Psi}(B)\mathbf{a}_t + \mathbf{n}_t,$$

where $\mathbf{\Psi}(B) = \mathbf{P}(B)\mathbf{A}(B)$. Note that in finite samples there is an identification problem in the decomposition of $\mathbf{\Psi}(B)$ into $\mathbf{A}(B)$ and $\mathbf{P}(B)$, as none of them are observed. This problem can be solved by writing the model as

$$\mathbf{x}_t = \sum_{s=0}^{S-1} \mathbf{\Psi_s}\mathbf{a}_{t-s} + \mathbf{n}_t,\tag{8}$$

where $\mathbf{a}_t$ is a vector white noise sequence.

The identification of the GDFM has been studied in Forni et al. (2000) and requires that all the lag covariance matrices have q eigenvalues that diverge when m and T increase, whereas the largest eigenvalue of the covariance matrices of the idiosyncratic component are bounded for all lags. From an intuitive point of view, this means that the factors have some effect on almost all the series and only a few values of the loading parameters are zero. It is said that the factors have a pervasive effect on the observed series. See Fan et al. (2021a) for a good discussion of this point. However, the noises are concentrated on a finite number of series. These assumptions for the noises are similar to the ones presented for the approximate DFM. In other words, the noise-to-signal ratio must converge to zero as $m \to \infty$ and $T \to \infty$ for the model to be identifiable.

2.2. Types of Outliers in Dynamic Factor Models

Outliers in model (3) were first proposed by Baragona and Battaglia (2007b) as follows. Let $\mathbf{x}_t^*$ be the observed vector of time series that can be contaminated with outliers and let $\mathbf{x}_t$ the unobserved vector without outliers. Then,

$$\mathbf{x}_t^* = \mathbf{P}\mathbf{f}_t + \mathbf{n}_t + \mathbf{w}I_t^{(a)} = \mathbf{x}_t + \mathbf{w}I_t^{(a)}, \tag{9}$$

where $\mathbf{w} = (w_1, \ldots, w_m)'$ is the outlier size at location $t = a$ and $I_t^{(a)}$ a dummy variable that takes the value 1 when $t = a$ and zero otherwise. These authors argued that by imposing restrictions on $\mathbf{w}$, three different types of outliers can be considered: outliers only on the factors (when $\mathbf{w} = \mathbf{P}\mathbf{w_f}$ for some r-dimensional vector $\mathbf{w_f}$); outliers on the idiosyncratic component (when $\mathbf{P}'\mathbf{w} = \mathbf{0}_r$) and mixed outliers affecting both the factors and the idiosyncratic component (when $\mathbf{w}$ is not restricted). In this formulation the default are mixture outliers. However, as recognized by Baragona and Battaglia (2007b), these outliers cannot be identified directly by projections derived from the properties of the factor model.

To avoid this problem, Galeano and Peña (2020) considered first factor outliers, that affect to most of the series, and idiosyncratic outliers, that affect only to one of the series, but allowing that at the same time both types of basic outliers can happen. Therefore, in this approach mixture outliers are also possible but they will appear when both types of basic outliers, factor and idiosyncratic, occur at the same time. The factor outliers are defined as follows. Let $\mathbf{f}_t = (f_{1t}, \ldots, f_{rt})'$ be the vector of dynamic factors following the model in (4). Then, a factor outlier appears in the k-th dynamic factor at time $t = o$, if $\mathbf{f}_t$ is replaced by

$$\mathbf{f}_t^* = \mathbf{f}_t + \mathbf{w}_k^{(o)}I_t^{(o)}, \tag{10}$$

where $\mathbf{w}_k^{(o)} = \left(0, \ldots, 0, w_k^{(o)}, 0, \ldots, 0\right)'$ is the size of the factor outlier in the k-th dynamic factor and $I_t^{(o)}$ is a dummy variable such that $I_t^{(o)} = 1$, if $t = o$, and $I_t^{(o)} = 0$, if $t \neq o$. The factor outlier in (10) potentially affects to all the time series components, depending on the loading matrix. The observed time series will be,

$$\mathbf{x}_t^* = \mathbf{P}\mathbf{f}_t^* + \mathbf{n}_t = \mathbf{P}\left(\mathbf{f}_t + \mathbf{w}_k^{(o)}I_t^{(o)}\right) + \mathbf{n}_t = \mathbf{x}_t + \mathbf{P}\mathbf{w}_k^{(o)}I_t^{(o)},$$

where $\mathbf{P}\mathbf{w}_k^{(o)} = w_k^{(o)}\mathbf{p}_{\cdot k}$ is the effect of the factor outlier on $\mathbf{x}_t^*$. Therefore, a factor outlier in the k-th factor at time $t = o$ produces an outlier of size $w_k^{(o)}\mathbf{p}_{\cdot k}$ in the observed time series. Thus, the effect of the factor outlier will be larger in time series associated with bigger k-th factor loading weights.

Idiosyncratic outlier are defined by Galeano and Peña (2020) as additive outliers in a single idiosyncratic noise component. Consequently, if $\mathbf{n}_t = (n_{1t}, \ldots, n_{mt})'$ is a vector of idiosyncratic noise, as defined in (3), an idiosyncratic outlier appears in the l-th idiosyncratic noise component at time $t = i$ if, instead of $\mathbf{n}_t$, we have $\mathbf{n}_t^* = \left(n_{1t}^*, \ldots, n_{mt}^*\right)'$, given by

$$\mathbf{n}_t^* = \mathbf{n}_t + \mathbf{w}_l^{(i)}I_t^{(i)}, \tag{11}$$

where $\mathbf{w}_l^{(i)} = \left(0, \ldots, 0, w_l^{(i)}, 0, \ldots, 0\right)'$ is the size of the idiosyncratic outlier and $I_t^{(i)}$ is a dummy variable. The idiosyncratic outlier in (11) only affects to one of the time series components, because the observed time series will be

$$\mathbf{x}_t^* = \mathbf{P}\mathbf{f}_t + \mathbf{n}_t^* = \mathbf{P}\mathbf{f}_t + \left(\mathbf{n}_t + \mathbf{w}_l^{(i)}I_t^{(i)}\right) = \mathbf{x}_t + \mathbf{w}_l^{(i)}I_t^{(i)}, \tag{12}$$

and the idiosyncratic outlier in the l-th noise component at time $t = i$ produces an outlier of size $w_l^{(i)}$ in the l-th component of the observed time series.

In practice, the time series that we observe, $\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*$, may be affected by several factor and/or idiosyncratic outliers, as follows,

$$\mathbf{x}_t^* = \mathbf{P}\left(\mathbf{f}_t + \sum_{j=1}^{F}\mathbf{w}_{k_j}^{(o_j)}I_t^{(o_j)}\right) + \left(\mathbf{n}_t + \sum_{j=1}^{I}\mathbf{w}_{l_j}^{(i_j)}I_t^{(i_j)}\right) = \mathbf{x}_t + \sum_{j=1}^{F}\mathbf{P}\mathbf{w}_{k_j}^{(o_j)}I_t^{(o_j)} + \sum_{j=1}^{I}\mathbf{w}_{l_j}^{(i_j)}I_t^{(i_j)}, \tag{13}$$

for $t = 1, \ldots, T$, where $\mathbf{w}_{k_1}^{(o_1)}, \ldots, \mathbf{w}_{k_F}^{(o_F)}$ are the outlier sizes of the F factor outliers in the k_1-th,...,k_F-th dynamic factors at times $o_1, \ldots, o_F$, and $\mathbf{w}_{l_1}^{(i_1)}, \ldots, \mathbf{w}_{l_I}^{(i_I)}$ are the outlier sizes of the I idiosyncratic outliers in the l_1-th,...,l_I-th idiosyncratic noise components at times $i_1, \ldots, i_I$, respectively. Note that expression (13) is flexible enough to cover situations such as, at the same time: (1) a set of factor outliers affecting different dynamic factors; (2) a factor outlier and a idiosyncratic outlier; (3) different dynamic factors/noise components.

The effects of the outliers in sample estimates of the mean vector and autocovariance matrices of the time series can be severe (see Baragona and Battaglia (2007b) and Galeano and Peña (2020)). Let $\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*$ be a time series as in (13) and let

$$\mathbf{o}_t = \sum_{j=1}^{F}\mathbf{P}\mathbf{w}_{k_j}^{(o_j)}I_t^{(o_j)} + \sum_{j=1}^{I}\mathbf{w}_{l_j}^{(i_j)}I_t^{(i_j)},$$

for $t = 1, \ldots, T$, be the effect of the outliers in the time series. The sample mean vector of $\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*$ is given by $\overline{\mathbf{x}}^* = \overline{\mathbf{x}} + \overline{\mathbf{o}}$, where $\overline{\mathbf{x}}$ is the sample mean vector of the time series free of the effects of the outliers, and $\overline{\mathbf{o}}$ is the sample mean vector of the effects of the outliers, and $\overline{\mathbf{x}}^*$ can be very different from $\overline{\mathbf{x}}$ if the number of outliers is large compared with the sample size and/or the outlier sizes are large enough. Moreover, the sample autocovariance matrices of $\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*$, for $h = 0, 1, 2, \ldots$ are given by

$$\widehat{\mathbf{\Gamma}}_{\mathbf{x}^*}(h) = \widehat{\mathbf{\Gamma}}_{\mathbf{x}}(h) + \frac{1}{T - h - 1} \sum_{t=h+1}^{T} (\mathbf{x}_t - \overline{\mathbf{x}})(\mathbf{o}_{t-h} - \overline{\mathbf{o}})'$$

$$+ \frac{1}{T - h - 1} \sum_{t=h+1}^{T} (\mathbf{o}_t - \overline{\mathbf{o}})(\mathbf{x}_{t-h} - \overline{\mathbf{x}})' + \frac{1}{T - h - 1} \sum_{t=h+1}^{T} (\mathbf{o}_t - \overline{\mathbf{o}})(\mathbf{o}_{t-h} - \overline{\mathbf{o}})',$$

where $\widehat{\mathbf{\Gamma}}_{\mathbf{x}}(h)$, for $h = 0, 1, 2, \ldots$, are the sample autocovariance matrices of $\mathbf{x}_1, \ldots, \mathbf{x}_T$. Consequently, taking expected values,

$$\mathbf{\Gamma}_{\mathbf{x}^*}(h) = \mathbf{\Gamma}_{\mathbf{x}}(h) + \frac{1}{T - h - 1} \sum_{t=h+1}^{T} (\mathbf{o}_t - \overline{\mathbf{o}})(\mathbf{o}_{t-h} - \overline{\mathbf{o}})',$$

from which two main conclusions arise. First, it is easy to see that the sample covariance matrix is always biased, as $\mathbf{\Gamma}_{\mathbf{x}^*}(0)$ can be very different from $\mathbf{\Gamma}_{\mathbf{x}}(0)$ when the number of outliers is large compared with the sample size and/or the outlier sizes are large enough. Second, the sample autocovariance matrices $\widehat{\mathbf{\Gamma}}_{\mathbf{x}^*}(h)$, for $h = 1, 2, \ldots$ are also biased, except if none of the distances between the times of the outliers is h, in which case $\mathbf{\Gamma}_{\mathbf{x}^*}(h) = \mathbf{\Gamma}_{\mathbf{x}}(h)$. In other words, if two outliers are located at a distance h they will bias the matrices $\mathbf{\Gamma}_{\mathbf{x}^*}(0)$ and $\mathbf{\Gamma}_{\mathbf{x}^*}(h)$. Therefore, in order to use eigenvectors of sample estimates of the autocovariance matrices to identify the loading matrix $\mathbf{P}$ and to estimate the factors, as is standard in the literature on DFMs, we need to clean the observed time series from the presence of the outliers.

2.3. Procedures for Outlier Detection in Dynamic Factor models

Several procedures have been proposed for outlier detection in DFM. Here we will review the ones based on projections proposed by Baragona and Battaglia (2007b) and Galeano and Peña (2020) and the lasso estimation of the outlier sizes by Chou et al. (2020). We also discuss how these ideas can be extended to the detection of outlier in non stationary series.

2.3.1. Outlier Detection by Projections

As mixed outliers can not be detected by projections, Baragona and Battaglia (2007b) proposed a two step procedure in which first outliers are identified by using the univariate time series obtained by projecting the data in the direction of the eigenvector $\mathbf{v}_m$, linked to the smallest eigenvalue of the covariance matrix of the data. As then $\mathbf{v}_m'\mathbf{P} \approx \mathbf{0}$, the projected series

$$\mathbf{v}_m'\mathbf{x}_t^* \approx \mathbf{v}_m'\mathbf{n}_t + \mathbf{v}_m'\mathbf{w}I_t^{(a)},$$

contains a combination of the noises plus outlier effects. The time periods in which outliers occur can be identified by searching for them in this single univariate time series, and these outlier effects can be removed from the vector of time series by considering these points as missing values and interpolating the vector of series. The proposal for detection of factor outliers is more complex and several procedures were considered.

Galeano and Peña (2020) proposed first an initial cleaning of the data in order to use a data covariance matrix $\mathbf{\Gamma}_{\mathbf{x}}(0)$ with small influence of the possible outliers. This matrix is used to determine the number of factors r and to find m projection directions to be used for identification of the factors and idiosyncratic outliers. The detection is made searching on the projections on the columns of the estimated loading matrix $\mathbf{P}$ and its orthogonal complement $\mathbf{V}$, (where, $\mathbf{V}'\mathbf{P} = \mathbf{0}$). Therefore, the procedure requires having good estimates of the matrices $\mathbf{V}$ and $\mathbf{P}$. With this objective the authors use some small modification of the method proposed by Galeano et al. (2006), who showed that an outlier in a vector time series produces an outlier in almost every projected time series and that multivariate additive outliers can be found by maximizing the kurtosis coefficient of the projections.

Suppose that we have applied an initial cleaning procedure to the data and using this cleaned covariance matrix we obtain the $m \times r$ matrix $\widehat{\mathbf{P}}_R$ and its $m \times (m - r)$ orthogonal complement $\widehat{\mathbf{V}}_R$, (where, $\widehat{\mathbf{V}}_R'\widehat{\mathbf{P}}_R = \mathbf{0}$). The projections of $\mathbf{x}_t^*$ on the columns of $\widehat{\mathbf{V}}_R$, assuming that $\widehat{\mathbf{V}}_R'\mathbf{P} \simeq \mathbf{0}$, is, by (13),

$$\mathbf{z}_t = \widehat{\mathbf{V}}_R'\mathbf{x}_t^* \simeq \widehat{\mathbf{V}}_R'\mathbf{n}_t + \sum_{j=1}^{I} \widehat{\mathbf{V}}_R'\mathbf{w}_{l_j}^{(i_j)} I_t^{(i_j)} = \widehat{\mathbf{V}}_R'\mathbf{n}_t + \sum_{j=1}^{I} w_{l_j}^{(i_j)} \widetilde{\mathbf{v}}_{l_j}' I_t^{(i_j)}, \tag{14}$$

where $w_{l_j}^{(i_j)}$ denotes the size of the idiosyncratic noise in the l_j-th idiosyncratic noise component at time $t = i_j$ and $\widehat{\mathbf{v}}_{l_j}$. denotes the l_j-th row of the matrix $\widehat{\mathbf{V}}_R'$. Therefore, $\mathbf{z}_t$ is an $(m - r)$-dimensional vector time series that results from adding the transformed idiosyncratic noise $\widehat{\mathbf{V}}_R'\mathbf{n}_t$, and the outlying vectors $w_{l_1}^{(i_1)}\mathbf{v}_{l_1}'., \ldots, w_{l_I}^{(i_I)}\mathbf{v}_{l_I}'.$. Thus, $\widehat{\mathbf{V}}_R'\mathbf{x}_t^*$ does not depend neither

on the dynamic factors, nor on the factor outliers. Consequently, we can detect the idiosyncratic outliers and estimate their sizes using the set of transformed series $\mathbf{z}_t = \widehat{\mathbf{V}}'_R \mathbf{x}^*_t$, as the effect of an idiosyncratic outlier in any specific l-th idiosyncratic noise component at any specific time $t = i$ in the series $\mathbf{x}^*_t$ is transmitted to the vector of new projected series $\mathbf{z}_t$ by (14)

$$\mathbf{z}_t = w_l^{(i)}\widehat{\mathbf{v}}'_{l.} + \widehat{\mathbf{a}}_t, \tag{15}$$

where $\widehat{\mathbf{a}}_t = \widehat{\mathbf{V}}'_R \mathbf{n}_t$. Note that the whole vector of transformed series $\mathbf{z}_i$ has information about the outlier and the effect on each component depend on the coefficients $\widehat{\mathbf{v}}_{l.}$, that define the relation among the components of $\mathbf{z}_i$ and the original idiosyncratic components $\mathbf{n}_i$. The outlier size $w_l^{(i)}$ can be estimated by the regression equation (15), that relates the outlier size in a single component of $\mathbf{n}_i$ to the transformed series $\mathbf{z}_i$. However, as the noises, $\widehat{\mathbf{a}}_t$, are heteroskedastic this estimation will be carried out by generalized least squares. Consequently, if we knew the time and the idiosyncratic component affected by the outlier we can estimate its size by the regression parameter. In this way, we will clean the original series from idiosyncratic outliers and in what follows we focus on detecting the factor outliers in a time series free from idiosyncratic outliers.

To identify factor outliers we multiply $\mathbf{x}^*_t$ in (13) by $\widehat{\mathbf{P}}'_R$, and assuming $\widehat{\mathbf{P}}'_R \mathbf{P} \simeq \mathbf{I}$, we get a r-dimensional vector time series, $\mathbf{y}^*_t$, given by,

$$\mathbf{y}^*_t = \widehat{\mathbf{P}}'_R \mathbf{x}^*_t = \mathbf{f}_t + \sum_{j=1}^{F} \mathbf{w}_{k_j}^{(o_j)} I_t^{(o_j)} + \widehat{\mathbf{P}}'_R \mathbf{n}_t \tag{16}$$

that results from adding the vector of dynamic factors $\mathbf{f}_t$, the factor outliers $\mathbf{w}_{k_1}^{(o_1)}, \ldots, \mathbf{w}_{k_F}^{(o_F)}$, and the transformed idiosyncratic noise $\widehat{\mathbf{P}}'_R \mathbf{n}_t$. Consequently, if we are able to detect the factor outliers, to estimate their sizes using the transformed series $\mathbf{y}^*_t$ in (16), and to clean their effects on the observed time series $\mathbf{x}^*_t$ in (13), we will obtain a time series free of any kind of effects of the outliers. To do this, note that (16) can be rewritten as

$$\mathbf{y}^*_t = \mathbf{y}_t + \sum_{j=1}^{F} \mathbf{w}_{k_j}^{(o_j)} I_t^{(o_j)}, \tag{17}$$

where $\mathbf{y}_t = \mathbf{f}_t + \mathbf{P}'\mathbf{n}_t$. In particular, the k-th component of $\mathbf{y}^*_t$ in (17), can be written as

$$y^*_{kt} = y_{kt} + \sum_{j=1}^{F_k} w_k^{(o^k_j)} I_t^{(o^k_j)}, \tag{18}$$

where y_{kt} is the k-th component of $\mathbf{y}_t$ that follows an ARMA model because it is the sum of processes that follow ARMA models. Therefore, the factor outliers can be identified by detecting additive outliers in the time series $y^*_{1t}, \ldots, y^*_{rt}$.

2.3.2. Lasso Estimation of Outlier Sizes

Chou et al. (2020) consider estimating an approximate factor model with outliers that will be used to forecast a given time series and assume the DFM,

$$\mathbf{x}^*_t = \mathbf{P}\mathbf{f}_t + \mathbf{n}_t + \mathbf{J}_t,$$

where $\mathbf{J}_t = (j_{1t}, \ldots, j_{mt})$ in the vector of outlier sizes at time t. It is assumed that the sequence of vectors $\mathbf{J}_t$ is sparse, so that most of the components are equal to zero. Calling $\mathbf{X}$, $\mathbf{F}$ and $\mathbf{J}$ to the $T \times m, T \times r$ and $T \times m$ matrices of time series, factors and outlier sizes, they proposed to solve the following lasso problem

$$\min \left\| \mathbf{X} - \mathbf{F}\mathbf{P}' - \mathbf{J} \right\|_F^2 + \delta \|\mathbf{J}\|_1,$$

where $\|\mathbf{A}\|_F$ is the Frobenius norm of the matrix $\mathbf{A}$. The optimization is made with the restriction $\mathbf{F}'\mathbf{F}/T = \mathbf{I}_r$. They derive an algorithm to solve this problem as follows: (1) Start with $\mathbf{J}^{(0)} = 0$; (2) given an estimated value for $\mathbf{J}^{(h)}$ compute the corrected data values by $\mathbf{C}^{(h-1)} = \mathbf{X} - \mathbf{J}^{(h-1)}$ and estimate with this data $\mathbf{F}^{(h)}$ and $\mathbf{P}^{(h)}$ by PCA; (3) given $\mathbf{L}^{(h)} = \mathbf{X} - \mathbf{F}^{(h)}\mathbf{P}'^{(h)}$ update the i-th element of the vector $\mathbf{J}_t^{(i)}$ as

$$J_{it}^{(h)} = ST(L_{it}^{(h)}, \delta/2),$$

where ST is the soft-thresholding function, $ST(x, y) = sign(x)(|x| - y)$. The authors shows that this algorithm can be seen as choosing at each step a low-rank matrix and a sparse matrix that minimize the objective function. The value of the penalization δ is chosen as $\delta = \widehat{\sigma}\sqrt{8\ln T}$, where $\widehat{\sigma}$ is the average sample standard deviation of the residuals from using the PCA method. This selection seems to work well in the simulations.

2.3.3. Outliers in Non-stationary Factor Models

We consider first the case of integrated processes, studied by Galeano and Peña (2020). Suppose that some of the factors are $I(d)$ and the others are stationary. Then, following Peña and Poncela (2006), the projection procedure on the extreme

kurtosis directions can be applied after replacing the robust autocovariance matrix, $\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}^*}^R(0)$, with the robust generalized autocovariance matrix given by

$$\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}^*}^{G,R}(0) = \frac{1}{T^{2d}} \sum_{t=1}^{T} \left(\mathbf{x}_t^{*R} - \overline{\mathbf{x}}^{*R}\right)\left(\mathbf{x}_{t-h}^{*R} - \overline{\mathbf{x}}^{*R}\right)',$$

where $\overline{\mathbf{x}}^{*R}$ is the sample mean vector of the time series $\mathbf{x}_1^{*R}, \ldots, \mathbf{x}_T^{*R}$, after cleaning the data. In this case, the number of non-stationary and stationary factors is obtained using the IC_1 model selection criterion proposed by Bai and Ng (2002) after replacing $\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}^*}^R(0)$ with $\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}^*}^{G,R}(0)$. The method identifies first the non-stationary factors and, after eliminating their effect on the time series $\mathbf{x}_1^{*R}, \ldots, \mathbf{x}_T^{*R}$, the stationary factors.

On the other hand, when some of the factors are seasonally integrated, others are integrated, and the rest are stationary, as in Nieto et al. (2016), the projection procedure based on extreme kurtosis can be applied after replacing the robust lagged autocovariance matrices, $\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}^*}^R(h)$, for $h = 0, 1, 2, \ldots$ with the robust seasonal generalized lagged autocovariance matrices given by

$$\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}^*}^{S,R}(h) = \frac{S^{2d}}{T^{2d}} \sum_{t=h+1}^{T} \left(\mathbf{x}_t^{*R} - \overline{\mathbf{x}}^{*R}\right)\left(\mathbf{x}_{t-h}^{*R} - \overline{\mathbf{x}}^{*R}\right)',$$

where S is the period of the seasonal factors. The number of seasonally integrated, integrated and stationary factors is obtained using, for instance, the method proposed by Lam and Yao (2012) after replacing $\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}^*}^R(h)$ with $\widehat{\boldsymbol{\Gamma}}_{\mathbf{x}^*}^{S,R}(h)$. In this case, the procedure is expected to identify first the seasonal factors, then, after eliminating the effect of the seasonal factors, the integrated ones and, after eliminating the effect of the the non-stationary factors on the time series, the stationary ones.

2.3.4. Outliers and Misspecification of Factor Models

An important topic that has not been much explored in the literature is the effect of using a factor model when the data have been generated in other ways. For this objective is useful to compare the results obtained by a model free procedure to those given under the assumption of a factor model. A model free procedure to search for outliers is to analyze the univariate time series obtained by projecting the data in directions of maximum and minimum kurtosis, as proposed by Galeano et al. (2006). In fact, we will see that this method can be applied for some initial cleaning of the series before estimating any model in order to obtain an initial robust covariance matrix. This topic will be discussed in the next section.

The interpretation of the DFM as a general representation for high dimensional time series of the common and idiosyncratic information in the set of time series introduced in Section 2.2 suggests that the procedures for finding outliers under the DFM framework can be quite general. Common outliers, with large effect on the series variability, are expected to appear in the principal components of the data with large variance, whereas specific outliers in the original series, with small effect on the aggregate variability, are expected to appear in the series formed by the smallest eigenvectors of the covariance matrix. Thus, the searching for outliers in the univariate time series formed by the principal components of the data is expected to detect these two types of outliers. These ideas can be used without any reference to any specific DFM.

3. Other Methods for Finding Outliers in MTS

In this section some other proposals for searching for outliers in MTS are first reviewed. Second, the problem of identifying atypical segments or periods, mainly studied outside the statistical literature, is considered. Third, methods for detecting outlying series and outlier sets are analyzed.

3.1. Multivariate Outliers

Several procedures have been proposed for finding multivariate outliers in high dimensional data set without using a model to find them. A related problem is to understand when an outlier is expected to be identified given its size and context. See Battaglia and Cucina (2020). Model free outlier detection can be made by building a set of linear projections of the time series and searching for outliers in the projected univariate time series. As indicated before, this approach was explored by Galeano et al. (2006), using projection pursuit, and Baragona and Battaglia (2007a), with independent component analysis.

Other model free procedures are based on graphical methods. Gonzalez and Dang (2021) proposed a dual view framework for finding outliers in MTS. The first view is a parallel coordinate chart of some relevant features that includes an outlier score rank for each time series. The second view uses a multi-line chart to represent how each time series changes over time. Parallel coordinates charts were introduced by Inselberg and Dimsdale (1990) (see also Inselberg (1998)), although their origin can be found in the statistical literature in the 60's to visualize multivariate data. The idea is to represent the values of each variable in a vertical line and connect the values of the variables at each time point. In this way, each vector of series is represented by a sequence of straight lines and sequences very different from the others indicate outlier points.

For large T the outliers must be very large to stand out from the rest of the points. This information is used with the multi-line chart that shows changes over time and indicates outlying series. Gonzalez and Dang (2021) argue that their proposal is able to identify outlying series in some examples, but this method has limitation for large values of both T and m.

A complementary approach to identify outlying series by graphical plotting are the extreme empirical dynamic quantiles proposed by Peña et al. (2019b). These quantiles are defined as those observed series that, over the available time span, have properties as close as possible as the series formed by the timewise marginal quantiles. For instance, the dynamic empirical median is the series that minimizes the L_1 distance to the marginal median of all the series over the whole period. Extreme dynamic quantiles will identify series that are very different from the others in the MTS data set.

Model based approaches often assume a Vector Autoregressive Structure (VAR) for the time series to search for outliers. For instance, Bellini (2016) use this structure and a forward-search approach to find outliers, whereas Cucina et al. (2014) apply a genetic algorithm for multiple additive outliers under a VAR model for the data. This last procedure seems to be able to detect patches of outliers, reducing masking and swamping effects, but requires to fix some unknown parameter values.

Outliers is MTS have been also analyzed from the machine learning perspective of deep neural networks (NN). Vishwakarma et al. (2021) proposed to consider each series in term and use a simple forward NN to identify pairs of outliers in a univariate time series, that is transformed into bivariate by using the observation and one of its lags, to find outliers. This approach does not seem to be very useful for multivariate outliers, as the joint information is not used. Other approaches in this line are presented in the review article by Blázquez-García et al. (2021). In our opinion the available methods for MTS outlier detection develop in the computer science and machine learning literature seem to be less efficient for multivariate outliers than the alternative statistical procedures.

3.2. Outliers Segments

Cheng et al. (2009) proposed to find outlier segments by defining a measure of the similarity of two segments of length s, and choosing the extreme segments with that measure. For $s \geq 0$ and two segments defined by the vectors $(\mathbf{x}_{t_1}, ..., \mathbf{x}_{t_{1+s-1}})$ and $(\mathbf{x}_{t_2}, ..., \mathbf{x}_{t_{2+s-1}})$ they proposed the measure

$$K(t_1, t_2, s) = \frac{\exp(-\sum_{j=0}^{s-1} (\mathbf{x}_{t_1+j} - \mathbf{x}_{t_2+j})'(\mathbf{x}_{t_1+j} - \mathbf{x}_{t_2+j}))}{\sigma^2},$$

where σ^2 is an estimate of the average variability of the series. Note that $K(t_1, t_2, 0)$ is a measure of the similarity between $\mathbf{x}_{t_1}, \mathbf{x}_{t_2}$. With this function we can construct a non-negative, symmetric kernel matrix that shows the pairwise similarity between the pair of selected segments (or sub sequences) or single data points. The kernel matrix is associated to a Markov chain with a transition matrix S, whose (i, j) element, that denotes the transition probability from node i to node j, is obtained by normalizing each column of the kernel matrix. Then, outliers are defined as those nodes with small connectivity with others, or small transition probabilities.

Wang et al. (2019b) studied the problem of finding time series segments with heterogeneous behavior. They assume that the MTS data follow in different segments of time different models. That is, it is assumed that changes in the model are not smooth and that at a set of time periods, $T_1, ..., T_k$ the model parameter change. The problem is to identify the number and location of the points of change in the MTS sequence. The authors solved the problem by finding a partition of the sample $P = (1, ...T_1 - 1), (T_1, ..., T_2 - 1), ..., (T_k, ..., T)$, that maximizes

$$P = \arg\max(L(P) + \delta k),$$

where $L(P)$ is a loss function that is chosen as the sum of squares of the fit when the VAR model is estimated in the period with a Lasso fit. This problem can be solved by a dynamic programming algorithm that is described in the paper. Meng et al. (2021), also extend methods for isolated outlier detection to find anomalous sequences using a complex procedure based on a similarity metric to measure the coupling between segments. These works do not include any comparison of the performance of their proposals with respect to alternative procedures.

3.3. Outlier Series

Yang and Shahabi (2007) and Weng and Shen (2008) address the problem of finding outlying series and outlying sets of MTS. Suppose we have several sets of multivariate time series (MTS) and, instead of finding individual observations that are outliers, we are interested in detecting the series that are outliers with respect to the others in the set, or sets that are outliers from the others sets. In this last case we have matrices of data, $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, ..., \mathbf{W}$ with T_i observations and m variables. The similarity they propose between these matrices is the Eros (Extended Frobenius norm) similarity, due to Yang and Shahabi (2007), that is defined as follows. Given two sets of time series data represented by matrices $\mathbf{X}, \mathbf{Y}$, with the same number of variables in columns, m, and possible different sample sizes, T_1, T_2, summarize the sets by their covariance matrices. Then, the eigenvectors of the $m \times m$ covariance matrices of the two sets, $\mathbf{v}_i^{(x)}, \mathbf{v}_i^{(y)},$ are computed for $i = 1, ..., p$, where p is a parameter to be selected, so that a large part of the variability in the matrices is covered. The similarity between these sets is based on the sum of absolute values of the scalar product of the eigenvectors. These scalar products

are weighted by their importance, as measured by their relative eigenvalues. Thus, given a set of weights, $\sum_{i=1}^{p} w_i = 1$, the Eros similarity is given by

$$Eros(\mathbf{X}, \mathbf{Y}, \mathbf{w}) = \sum_{i=1}^{p} w_i \left| (\mathbf{v}_i^{(x)\prime} \mathbf{v}_i^{(y)}) \right|.$$

Note that the Eros similarity is similar to a sum of the weighted absolute values of the correlation coefficient between the principal components of the two sets. As the euclidean distance between two norm one vectors is $D_{Eu} = \sqrt{2 - 2(\mathbf{v}^{(x)\prime} \mathbf{v}^{(y)})}$, the Eros distance between two matrices is defined with the Eros similarity as follows

$$D_{Eros} = \sqrt{2 - 2 \sum_{i=1}^{p} w_i \left| (\mathbf{v}_i^{(x)\prime} \mathbf{v}_i^{(y)}) \right|}.$$

This measure is bounded and takes the maximum value when the eigenvectors are orthogonal and the minimum when they have correlations equal to one.

Algorithms to apply these ideas to find outlying series have been proposed in Weng and Shen (2008). They defined the outlier score of a series in a given data set as the sum of the Eros distances from its k nearest neighbors. Then, outlier series are those having the largest outlier scores. However, a test to decide if the series with the largest outlier score is really an outlier is not given. Similar ideas can be applied to the comparison of MTS sets with the Eros distance and a k nearest neighbor approach (kNN). The limitation of all these methods is to use just zero lag correlation between the principal components of the series and ignore all the other lag correlations between them.

Finding outlying series can be done by summarizing each series in a vector of features, or variables, of dimension p and applying a cluster method to find groups of outlier series. If the individual features are related, they can be replaced by a few principal components and then apply the clustering method to these selected principal components. For instance, Hyndman et al. (2015) define a large set of univariate time series characteristics. as Mean, Variance, First order autocorrelation, Trend Strength, Linearity Strength, Curvature Strength, Season Strength, and many others. Then, all these characteristics are summarized by the first two principal components and a cluster method is applied to the series represented by these two new features. The result, that depends on the cluster method applied, the selection of the variables and the number of component used, provides the most extreme data of the set. However, no statistical test is applied to check if these series are really heterogeneous from the others.

Some of these problems are solved in Talagala et al. (2020). The selection of features and their summary in two principal components is kept, but now a density estimation using a Gaussian kernel is applied to the feature space to identify outliers that are tested with extreme value theory. Once the problem is solved with a training set of time series, detecting outlying time series in streaming can be made by applying the same ideas to the on line data.

Another method for finding outlying time series has been suggested by López-Oriona and Vilar (2021a) using a functional approach. Each MTS is described by an estimator of its quantile cross-spectral density, that was introduced as a clustering tool for the same authors in López-Oriona and Vilar (2021b). Then, this transformation is treated as multivariate functional data and an outlier score is assigned to each series by using the half space depth proposed by Tukey. The performance of the procedure seems to be better in a simulation study that the feature extraction method by Hyndman et al. (2015) and the eros distance method due to Weng and Shen (2008). This better performance is to be expected as the functional approach uses much more information about the dynamics of the series than the other proposals.

Outlying series can also be found by clustering the series with a measure of similarity that takes into account all the relevant cross correlations for all lags: the generalized cross correlation, introduced by Alonso and Peña (2019). The idea was proposed by Alonso et al. (2020) in the framework of outliers in dynamic factor models with cluster structure. The detection is made by using the information provided by the clustering procedure when using the generalized cross correlation as measure of similarity. Suppose we apply a hierarchical clustering to the MTS set and find a set of series that are quite dissimilar to the others, according to the dendrogram distances. These series are candidates to be outlying time series. Let $\mathbf{x}_t^s = (x_{1t}^s, \ldots, x_{st}^s)'$ be the set of s suspicious time series. For testing the null hypothesis $H_0 : x_{it}^s$ is independent of $\mathbf{f}_t = (f_{1t}, \ldots, f_{rt})'$ these authors apply the statistic proposed in Robbins and Fisher (2015) to test for independence between two sets of time series. This statistic is the log of the determinant of the matrix formed by residual autocorrelation matrices after fitting VARMA models to the two sets of series, and follows a χ^2 distribution with degrees of freedom equal to the product of the dimensions of the two sets of series to test for independence. Therefore, one option would be to test directly H_0 versus H_1 with the whole set of factors. However, as noted in the simulations in Robbins and Fisher (2015), the log-determinant statistic is more powerful when the dimensions of the two sets of time series to test for independence are small. For this reason, first the log-determinant statistics for each y_{it}^s, for $i = 1, \ldots, s$, is computed and each factor f_{jt}, for $j = 1, \ldots, r$, as well as their associated p-values, denoted by p_{ij}. Then, we label as non-outlying time series to those suspicious time series such that $\min_{1 \leq j \leq p} p_{ij} < c_\alpha$, where $c_\alpha = \chi^2_{1,(1-\alpha)^{1/(sr)}}$ is the $(1-\alpha)^{1/(sr)}$ quantile of the χ^2_1 distribution. This quantile corresponds to the $1 - \alpha$ quantile of the distribution of the minimum of $r \times s$ independent χ^2_1 distributions. The authors show by simulations that this is a simple, efficient and powerful method to detect outlying time series.

4. Dimension Reduction by Principal Components

Dimension reduction without assuming any model can be made by principal components (PC). PC deals with the following problem: given a large set of time series, to find a small number of another time series, called the principal components, that explain a large part of the variability of the whole set of original time series. We consider two types of PC, the static PC, where, as in DFM, the values of the series at one period are explained using only values of the principal components at the same period, and dynamic principal components, (DPC), where, as in GDFM, the reconstruction of each variable at a given period depends on the values of the DPC's at that period and also on their values at several lags. In this section we describe classic and robust versions of static PC, while in Section 5 some robust versions of DPC will be presented.

4.1. Static Principal Components

Let $\mathbf{X}$ be a $m \times T$ data matrix where each column correspond to a different time series. In the static approach, the classical case proposed by Hotelling (1933), the principal components are defined as follows. Let $\bar{\mathbf{x}}$ be the vector of row means and

$$C = \frac{1}{n}\mathbf{X}\mathbf{X}' - \bar{\mathbf{x}}'\bar{\mathbf{x}},$$

the sample covariance matrix of $\mathbf{X}$. Let $\mathbf{v}_1, ..., \mathbf{v}_m$ be the eigenvectors of $\mathbf{C}$ corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$. The i-th static principal component is defined by

$$\mathbf{f}_i^* = \mathbf{X}'\mathbf{v}_i. \tag{19}$$

These principal components have the following property: Take $I \leq m$ and consider the problem of finding vectors $\mathbf{f}_i = (f_{i,t})_{1 \leq t \leq T}$, $1 \leq i \leq I$, a $m \times I$ matrix $\mathbf{B} = (b_{j,i})$, $1 \leq j \leq m$, $1 \leq i \leq I$ and a vector $\mathbf{d} = (d_1, ..., d_m)'$ minimizing

$$\sum_{t=1}^{T}\sum_{j=1}^{m}\left(x_{j,t} - d_j - \sum_{i=1}^{I}b_{j,i}f_{it}\right)^2. \tag{20}$$

Okamoto and Kanazawa (1968) showed that the solution to this problem is $\mathbf{f}_i = \mathbf{f}_i^*$, $d_j = \bar{x}_j$ and $\mathbf{B}$ the $m \times I$ matrix whose i-th row is $\mathbf{v}_j$, and the optimal reconstruction of $\mathbf{X}$ using linear combinations of the first I principal components is the $m \times T$ matrix $\widetilde{\mathbf{X}} = (\widetilde{x}_{j,t})$,

$$\widetilde{x}_{j,t} = \bar{x}_j + \sum_{i=1}^{I}b_{j,i}f_{i,t}^*.$$

Observe that, according to (19), each $\mathbf{f}_i^*$ is a linear combination of the rows of $\mathbf{X}$ even though this condition was not required in the formulation of the problem.

As in most of the problems which have least squares solutions, these principal components are not robust. Unrestricted robust static principal components can be obtained taking the number of lags $k = 0$ in the robust procedure for DPC described in Section 5.1 and taking $k_1 = k_2 = 0$ in the robust DPC procedure of Section 5.2, where robust static principal components which are linear combination of the observations are obtained. Unlike what happens for classical static PC, these two procedures do not agree, even for $k = 0$. Several other robust statistical principal components procedures have been proposed. We can cite, among others, Li and Chen (1985); Croux and Haesbroeck (2000); Maronna (2000); Maronna and Yohai (2008) and Salibián-Barrera et al. (2006).

4.2. Brillinger's Dynamic Principal Components

Since in a time series the values corresponding to two close periods of time are generally correlated, to reconstruct the values of the m series at a period t not only may be useful the values at this period but also at others. This leads Brillinger (see Brillinger (1964), Brillinger (1981)), to define the dynamic principal components as follows. Suppose that $\mathbf{x}_t$, $-\infty < t < \infty$ is a stationary multiple time series of dimension m such that $E(\mathbf{x}_t) = \mathbf{0}$, and $E(\mathbf{x}_t\mathbf{x}_t') = \Sigma$. Given two sequences of m dimensional vectors $\mathbf{a}_h$ and $\mathbf{b}_h$, $-\infty < h < \infty$, let

$$F_t = \sum_{h=-\infty}^{\infty} \mathbf{a}_h'\mathbf{x}_{t-h}, \quad -\infty < t < \infty,$$

and $\widetilde{\mathbf{x}}_t$ be the reconstructed vector,

$$\widetilde{\mathbf{x}}_t = \sum_{h=-\infty}^{\infty} \mathbf{b}_h F_{t-h}.$$

Let $\mathbf{a}_h^*$ and $\mathbf{b}_h^*$, $-\infty < h < \infty$, be the sequences of vectors that minimize the L_2 norm of the reconstruction error vector, $\mathbf{x}_t - \widetilde{\mathbf{x}}_t$, that is, that minimize

$$E((\mathbf{x}_t - \widetilde{\mathbf{x}})'(\mathbf{x}_t - \widetilde{\mathbf{x}}_t),$$

Brillinger defines the first dynamic principal component as $F^* = (F_t^*)_{-\infty < t < \infty}$, where $F_t^* = \sum_{h=-\infty}^{\infty} \mathbf{a}_h^{*\prime} \mathbf{x}_{t-h}$. He found that the optimal vectors $\mathbf{a}_h^*$ are the inverse Fourier transform of the principal component of the cross spectral matrices for each frequency and $\mathbf{b}_h$ the inverse Fourier transform of the conjugates of the same principal components. Higher order DPC can be defined using the successive reconstruction residuals.

Some of the drawbacks of these definition are : (1) The large amount of coefficients $\mathbf{a}_h$ and $\mathbf{b}_h$ that should be estimated; (2) It is defined only for stationary time series; (3) The value of F_t depends also on the future values $\mathbf{x}_{t+h}$ with $h > 0$, and this makes them not suitable for forecasting; (4) It is difficult to robustify.

To overcome these difficulties, Peña and Yohai (2016) and Peña et al. (2019a) proposed two different ways to define dynamic principal components. In the first reference is presented an unrestricted DPC and, in the second, a DPC which is a linear combination of the current and past observations.

4.3. Unrestricted Dynamic Principal Components

In this section we describe a non robust unrestricted DPC: the generalized DPC (GDPC). Consider a $m \times T$ matrix $\mathbf{X}$ where its t-th column $\mathbf{x}_t$ is an observation of a multiple time series of dimension m. Let $k \geq 0$ be the number of lags of the DPC used for reconstructing the data. Let $\mathbf{f} = (f_t)_{1 \leq t \leq T}$ be a possible value of the first DPC, and a $m \times (k+1)$ matrix $\mathbf{B}$ with coefficients $(b_{j,i})$ and a m dimensional vector $\mathbf{d} = (d_1, ..., d_m)'$. Then, the reconstruction of the data matrix $\mathbf{X}$ is defined as

$$\widetilde{x}_{j,t}(\mathbf{f}, \mathbf{B}, \mathbf{d}) = d_j + \sum_{i=0}^{k} b_{j,i} f_{t-i}, \quad 1 \leq j \leq m, \ k+1 \leq t \leq T.$$

Let

$$(\mathbf{f}^*, \mathbf{B}^*, \mathbf{d}^*) = \arg \min_{\mathbf{f}, \mathbf{B}, \mathbf{d}} \sum_{j=1}^{m} \sum_{t=k+1}^{T} (x_{j,t} - \widetilde{x}_{j,t}(\mathbf{f}, \mathbf{B}, \mathbf{d}))^2. \tag{21}$$

Then, $\mathbf{f}^*$ is the first GDPC proposed in Peña and Yohai (2016). The higher order DPC are obtained similarly taking successive residuals, e.g., the second unrestricted DPC is the first DPC of the residual $m \times (T-k)$ matrix $\mathbf{R} = (r_{j,t})$ where $r_{j,t} = x_{j,t} - \widetilde{x}_{j,t}(\mathbf{f}^*, \mathbf{B}^*, \mathbf{d}^*)$, $1 \leq j \leq m$, $k+1 \leq t \leq T$. Observe than when $k = 0$, the GDPC coincides with the classical PC.

Since the GDPC is defined by the minimization of the sum of the squares of the reconstruction errors, it is not robust. In Section 5.1 we show a robust version of this procedure.

4.4. Linearly Constrained DPC Estimators

One shortcoming of the GDPC is that the last principal components are computed using a few number of observations and, therefore, they are not much stable. As a consequence, these DPC are not apt for forecasting. For this reason, Peña et al. (2019a) proposed a lineally constrained class of DPC which do not have this inconvenience: the one-sided DPC (ODPC).

Let $k_1 \geq 0$ be the number of lags of the variables used to define the principal components and $k_2 \geq 0$ the number of lags of the principal components used to reconstruct the variables. More precisely given $\mathbf{a} = (\mathbf{a}_1, ..., \mathbf{a}_{k_1})$ where $\mathbf{a}_i \in R^m$, the first principal component is an element in R^{T-k_1} of the form $\mathbf{f}(\mathbf{a}) = (f_{k_1+1}(\mathbf{a}), ..., f_T(\mathbf{a}))'$ where $f_t(\mathbf{a}) = \sum_{i=0}^{k_1} \mathbf{a}_i' \mathbf{x}_{t-i}$. Besides, given a $m \times (k_2 + 1)$ matrix $\mathbf{B} = (b_{j,i})$ and a vector $\mathbf{d} = (d_1, ..., d_m)'$, the elements of the $m \times (T - k_1 - k_2)$ reconstructed data matrix are defined as

$$\widetilde{x}_{j,t}(\mathbf{a}, \mathbf{B}, \mathbf{d}) = d_j + \sum_{i=0}^{k_2} b_{j,i} f_{t-i}(\mathbf{a}), \ k+1 \leq t \leq T, \ 1 \leq j \leq m.$$

Then, the optimal reconstruction is given by

$$(\mathbf{a}^*, \mathbf{B}^*, \mathbf{d}^*) = \arg \min_{\mathbf{f}, \mathbf{B}, \mathbf{d}} \sum_{j=1}^{m} \sum_{t=k+1}^{T} (x_{j,t} - \widetilde{x}_{j,t}(\mathbf{a}, \mathbf{B}, \mathbf{d}))^2, \tag{22}$$

and $\mathbf{f}^* = \mathbf{f}(\mathbf{a}^*)$ is the first ODPC.

Observe that this definition of the first ODPC is similar to Brillinger DPC, with the difference that only a finite number of lags, and not leads, are used in both its definition as well as in the reconstruction of the observations. Thus, this ODPC is apt for forecasting.

The second order ODPC can be defined as in (22) replacing the $x_{j,t}$'s by the residuals $r_{j,t} = x_{j,t} - \widetilde{x}_{j,t}(\mathbf{a}^*, \mathbf{B}^*, \mathbf{d}^*)$. Higher order ODPC's can be defined using successive residual matrices.

As in the case of the GDPC, since the objective function is to minimize the sum of squares of the reconstruction errors of the data matrix, this estimator is not robust. In Section 5.2 a robust version of the ODPC will be presented.

5. Robust Dynamic Principal Components

In this section we define two different robust DPC. In Section 5.1 a robust version of the GDPC and in Section 5.2 a robust version of the ODPC

5.1. Robust GDPC

In this section we describe a robust version of the GDPC based on an M scale. Huber (1964) defined the class of robust M-scales. Given a sample $\mathbf{x} = (x_1, ..., x_n)$, a M scale $s_M(\mathbf{x})$ is a value s satisfying,

$$\frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{x_i}{s}\right) = \delta,$$

where $\rho : R \rightarrow R_{\geq 0}$ is even and non decreasing for non-negative values. Observe that if $\rho(x) = x^2$ and $\delta = 1$, the M-scale is $s_M(\mathbf{x}) = \left((1/n) \sum_{i=1}^{N} x_i^2\right)^{1/2}$. The fact that x^2 is unbounded make this scale extremely sensitive to outliers. To define a robust scale is required that ρ be bounded. Without loss of generality we can assume that $\sup \rho = 1$ and $0 \leq \delta \leq 1$. Let x be a random variable, then $S_M(x)$ is defined by $E_F(\rho(x/S_M(x)) = \delta$. Thus, if $\mathbf{x}_n = (x_1, \ldots, x_n)$ is a random sample of the distribution of a random variable x, we have $S_M(x) = \lim s_M(\mathbf{x}_n)$ a.s.. Then, to obtain consistency of the M scale estimator to the standard error for normal samples is required that $E_\Phi(\rho(x)) = \delta$, where Φ is the standard normal distribution. In this case, if $\mathbf{x}_n$ is a random sample of size n of a $N(0, \sigma^2)$ distribution, then $\lim_{n \rightarrow \infty} s_M(\mathbf{x}_n) = \sigma^2$.

A popular family of functions satisfying these properties is the Tukey bisquare family given by

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

The finite sample breakdown point of an M-scale is $\varepsilon^* = \min(\delta, 1 - \delta)$ and is maximized when $\delta = 0.5$, and in this case $\varepsilon^* = 0.5$. The consistency condition for this value of δ is $E_\Phi(\rho_c^{BS}(x)) = .5$ which is satisfied when $c = 1.547$.

An unrestricted robust DPC, called scale GDPC (SGDPC), was proposed by Peña and Yohai (2016). The SGDPC can be obtained in a similar way that the GDPC but instead of minimizing the sum of squares of the errors, as in (21), the SGDPC minimizes the sum of squares of the M scales of the reconstruction errors of each of the m variables. More precisely, let the scale $s_j(\mathbf{f}, \mathbf{B}, \mathbf{d})$ of the j-th variable, $1 \leq j \leq m$, be defined by

$$\frac{1}{T - k} \sum_{t=k+1}^{T} \rho((x_{j,t} - \widetilde{x}_{j,t}(\mathbf{f}, \mathbf{B}, \mathbf{d}))/s_j(\mathbf{f}, \mathbf{B}, \mathbf{d})) = \delta.$$

Let

$$(\mathbf{f}^*, \mathbf{B}^*, \mathbf{d}^*) = \arg \min_{\mathbf{f}, \mathbf{B}, \mathbf{d}} \sum_{j=1}^{m} s_j^2(\mathbf{f}, \mathbf{B}, \mathbf{d}),$$

then, the $(T - k)$ dimensional vector $\mathbf{f}^* = (f_t^*)_{k+1 \leq t \leq T}$ is the first SGDPC. The values of $\mathbf{B}^*, \mathbf{d}^*$ can be used to give an approximated reconstruction of the last $T - k$ rows of $\mathbf{X}$ as

$$\widetilde{x}_{j,t}^* = d_j^* + \sum_{i=0}^{k} b_{j,i}^* f_{t-i}^*, \quad k + 1 \leq t \leq T, 1 \leq j \leq m.$$

Since $(\mathbf{f}^*, \mathbf{B}^*, \mathbf{d}^*)$ have been defined by the minimization of a robust scale, they are robust too. The second SGDPC can be obtained by computing the first SGDPC of the residual $m \times (T - k)$ matrix $\mathbf{R} = (r_{j,t})$ with $r_{j,t} = x_{j,t} - \widetilde{x}_{j,t}^*$. The higher order SGDPC are computed similarly.

Note that for $k = 0$ the SGDPC procedure can be used to obtain static principal components. This case was studied in detail in Maronna and Yohai (2008).

5.2. Robust ODPC

A robust version of the ODPC using a M scale, called scale ODPC (SODPC) was proposed in Peña and Yohai (2023). Similarly to the SGDPC definition, this estimate is defined by replacing in the goal function of the ODPC the sum of the squares of the reconstruction errors by the sum of the squares of the scales of the reconstruction errors of each variable.

Let $s_j(\mathbf{a}, \mathbf{B}, \mathbf{d})$ be the reconstruction errors of the j-th variable, $1 \leq j \leq m$, that is

$$\frac{1}{T - k} \sum_{t=k+1}^{T} \rho((x_{j,t} - \widetilde{x}_{j,t}(\mathbf{f}(\mathbf{a}), \mathbf{B}, \mathbf{d}))/s_j(\mathbf{a}, \mathbf{B}, \mathbf{d})) = \delta. \tag{23}$$

Then, the optimal values of $\mathbf{a}$, $\mathbf{B}$ and $\mathbf{d}$ are given by

$$(\mathbf{a}^*, \mathbf{B}^*, \mathbf{d}^*) = \arg\min_{\mathbf{a}, \mathbf{B}, \mathbf{d}} \sum_{j=1}^m s_j(\mathbf{a}, \mathbf{B}, \mathbf{d})^2,$$

and the SODPC is $\mathbf{f}^* = (f_{k_1+1}^*, \ldots, f_T^*)$, where $f_t^* = f_t(\mathbf{a}^*)$.

In the case that $\mathbf{x}_t$ is stationary, we can also define the population first ODPC as follows. Define $S_j(\mathbf{a}, \mathbf{B}, \mathbf{d})$ by

$$E\rho((x_{j,t} - \widetilde{x}_{j,t}(\mathbf{f}(\mathbf{a}), \mathbf{B}, \mathbf{d}))/S_j(\mathbf{a}, \mathbf{B}, \mathbf{d})) = \delta, \tag{24}$$

$$(\mathbf{a}_P^*, \mathbf{B}_P^*, \mathbf{d}_P^*) = \arg\min_{\mathbf{f}, \mathbf{B}, \mathbf{d}} \sum_{j=1}^m S_j(\mathbf{a}, \mathbf{B}, \mathbf{d})^2. \tag{25}$$

Observe that $S_j(\mathbf{a}_P^*, \mathbf{B}_P^*, \mathbf{d}_P^*)$ is the population M scale of the residuals of the j-th variable and satisfies

$$\widetilde{r}_{P,j,t}^* = x_{j,t} - \widetilde{x}_{j,t}(\mathbf{a}_P^*, \mathbf{B}_P^*, \mathbf{d}_P^*), \tag{26}$$

that is, it is defined by

$$E(\rho(\widetilde{r}_{P,j,t}^*/S_j(\mathbf{a}_P^*, \mathbf{B}_P^*, \mathbf{d}_P^*)) = \delta. \tag{27}$$

In Peña and Yohai (2023) is described an iterative algorithm to compute the SODPC. Each step requires to compute $m + 1$ regressions using weighted least squares. In one of the regressions a new value of $\mathbf{a}$ is computed and in each of the m remaining regressions new estimators of each row of $\mathbf{B}$ and the element of $\mathbf{d}$ of the same row. The weights penalize observations that at the current step have a large reconstruction residual and, therefore, are possible outliers. The computing algorithm for the ODPC is similar, but with all the weights equal one.

Note that when $k_1 = 0$ and $k_2 = 0$, the SODPC procedure provides robust static principal components. This case was extensively studied by Maronna (2000).

The SODPC procedure has some of the properties that characterize robust estimators. Observe that the robustness of a forecasting will depend, in a long measure, on the robustness of the estimators of $(\mathbf{a}^*, \mathbf{B}^*, \mathbf{d}^*)$.

The *finite sample breakdown point* of an estimator defined in the Introduction quantifies its resistance to the presence of outliers. In general, the finite sample breakdown point of non robust estimators based on the minimization of the sum of the squared residuals (e.g. the ODPC) is $1/n$, where n is the sample size. That is, one atypical observation may have an unlimited effect on the estimator. It can be shown that if δ in (23) is 0.5, then the finite sample breakdown point of the SODPC estimator of $(\mathbf{a}, \mathbf{B}, \mathbf{d})$ is $0.5/(k_1 + k_2 + 1)$. The reason why in the denominator appears $(k_1 + k_2 + 1)$ is because when k_1 and k_2 lags are used, one outlier affects to $(k_1 + k_2 + 1)$ residuals.

Another property, introduced by Hampel (1971), to characterize robust estimators is *qualitative robustness.* Informally, an estimator is qualitatively robust if a small fraction of outliers or small changes in the data have a small effect on the estimators. Again for a more formal definition see Maronna et al. (2017). The SODPC estimator of $(\mathbf{a}^*, \mathbf{B}^*, \mathbf{d}^*)$ is qualitatively robust too.

As we will see in the next section $(\mathbf{a}^*, \mathbf{B}^*, \mathbf{d}^*)$ can be used to forecast future values of the m variables. These forecasted values will inherit its robustness properties.

6. Forecasting With SODPC

Suppose that we have computed estimators of the first Q SODPC's and the q-th principal component uses $(k_1^{(q)}, k_2^{(q)})$ lags, $1 \le q \le Q$. Here we describe how to use this information to predict the values of all the variables for the next h periods.

Let $\widehat{\mathbf{f}}^{(q)} = (\widehat{f}_t^{(q)})_{k_1^{(q)}+1 \le t \le T}$ be the estimated SODPC's and $\widehat{\mathbf{B}}^{(q)}$ and $\widehat{\mathbf{d}}^{(q)}$, $1 \le q \le Q$, the estimated matrices and intercept vectors to use for the reconstruction. Robust predictions, $\widetilde{f}_{t+l}^{(q)}$ of $f_{t+l}^{(q)}$, $1 \le q \le Q$, $1 \le l \le h$, can be obtained by fitting a time series model for each component, $\widehat{\mathbf{f}}^{(q)}$, using a robust procedure. This can be done applying to each of the Q components the filtered τ-estimation procedure (see Section 8.8.3 of Maronna et al. (2017)). This procedure chooses automatically the order of an AR model, gives robust estimators of its coefficients, and predicts future values of each component. Moreover it provides a filtered series, $\widetilde{\mathbf{f}}^{(q)} = (\widetilde{f}_t^{(i)})_{k_1^{(q)}+1 \le t \le T+h}$, $1 \le q \le Q$, cleaned of the detected outliers. Using this information, the robust predictions of the futures values of the Q series are obtained as follows. Define $\widetilde{\mathbf{F}}_{T+l}^{(q)} = (1, \widetilde{f}_{T+l}^{(q)}, \widetilde{f}_{T+l-1}^{(q)}, ..., \widetilde{f}_{T+l-k_2}^{(q)})'$, then a robust prediction $\mathbf{x}_{T+l}$ given the first T observations is

$$\widehat{\mathbf{x}}_{T+l|T} = \sum_{q=1}^Q \left(\widehat{\mathbf{B}}^{(q)}\widetilde{\mathbf{F}}_{T+l}^{(q)} + \widehat{\mathbf{d}}^{(q)}\right), 1 \le l \le h.$$

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

The robust forecasts presented above will have optimal properties only if the reconstruction residuals of the series with the DPC are white noise. Otherwise, there is idiosyncratic information in the series that can be used to improve the forecast by making robust univariate models for the residuals of each series, in the same way that has been explained for the factors.

7. Selection of the Number of Lags and the Number of Components for the SODPC

In Bai and Ng (2007) is given an information criterion to choose the number of lags for each factor and the number of factors of a DFM. In Peña and Yohai (2023) this criterion was adapted to choose the number of lags $k_1^{(q)}$ and $k_2^{(q)}$ for the q component and the total number of components Q for the SODPC. We should start given maximum values K_1 and K_2 for $k_1^{(q)}$ and $k_2^{(q)}$ respectively. We compute the first SODPC with $k_1^{(q)} = k_1$ and $k_2^{(q)} = k_2$ for all values of k_1 and k_2 such that $0 \leq k_1 \leq K_1, 0 \leq k_2 \leq K$. For each of these values of k_1 and k_2 the residuals $(r_{t,j}^{(1,k_1,k_2)})_{k_1+k_2+1 \leq t \leq T}, 1 \leq j \leq m$ using the first SODPC with k_1 and k_2 lags are obtained and the corresponding M scales $S_M(\mathbf{r}_{.j}^{(1,k_1,k_2)})$ of the vector $\mathbf{r}_{j.}^{(1,k_1,k_2)} = (r_{j,k_1+k_2+1}^{(1,k_1,k_2)}, ..., r_{j,T}^{(1,k_1,k_2)})$, $1 \leq j \leq m$, are computed. Let $\widehat{\sigma}_{1,k_1,k_2} = \left((1/m)\sum_{j=1}^m s_M^2(\mathbf{r}_{.j}^{(1,k_1,k_2)})\right)^{1/2}$ and $T_{1,k_1,k_2}^* = T - k_1 - k_2$, then

$$(k_1^{(1)}, k_2^{(1)}) = \arg \min_{0 \leq k_1 \leq K_1, 0 \leq k_2 \leq K_2} \text{BNG}_{1,k_1,k_2},$$

where BNG_{1,k_1,k_2} is the following robust version of the Bai and Ng type criterion

$$\text{BNG}_{1,k_1,k_2} = \log(\widehat{\sigma}_{1,k_1,k_2}^2) + (k_1 + k_2 + 1)\frac{\log(\min(T_{1,k_1,k_2}^*, m))}{\min(T_{1,k_1,k_2}^*, m)}.$$

Once we have computed the $q-1$ dynamic principal components, where the component i, $1 \leq i \leq q-1$, is obtained with $k_1^{(i)}$ and $k_2^{(i)}$ lags, the q component is computed using as lags each of the values $0 \leq k_1 \leq K_1$ and $0 \leq k_2 \leq K_2$. Let $(r_{j,t}^{(q,k_1,k_2)})_{h_{q,k_1,k_2}+1 \leq t \leq T, 1 \leq j \leq m}$ the corresponding residuals matrix, where $h_{q,k_1,k_2} = \left(k_1 + k_2 + \sum_{i=1}^{q-1}(k_1^{(i)} + k_2^{(i)})\right)$, $T_{q,k_1,k_2}^* = T - h_{q,k_1,k_2}$, $\mathbf{r}_{j.}^{(q,k_1,k_2)} = (r_{h_{q,k_1,k_2}+1,j}^{(q,k_1,k_2)}, ..., r_{T,j}^{(q,k_1,k_2)})$ and let $\widehat{\sigma}_{q,k_1,k_2} = ((1/m)\sum_{j=1}^m S_M(\mathbf{r}_{.j}^{(q,k_1,k_2)}))^{1/2}$. Then, the values $k_1^{(q)}$ and $k_2^{(q)}$ are given by

$$(k_1^{(q)}, k_2^{(q)}) = \arg \min_{0 \leq k_1 \leq K_1, 0 \leq k_2 \leq K_2} \text{BNG}_{q,k_1,k_2},$$

where

$$\text{BNG}_{q,k_1k_2} = \log(\widehat{\sigma}_{q,k_1,k_2}^2) + \left(\left(\sum_{i=1}^{q-1}(k_1^{(i)} + k_2^{(i)} + 1)\right) + k_1 + k_2 + 1\right)\frac{\log(\min(T_{q,k_1,k_2}^*, m))}{\min(T_{q,k_1,k_2}^*, m)}.$$

The selected number of components is $Q = q_0 - 1$, where q_0 is the minimum value q such that $\text{BNG}_{q,k_1^{(q)},k_2^{(q)}} \geq \text{BNG}_{q-1,k_1^{(q-1)},k_2^{(q-1)}}$.

8. Robust Estimation Procedures for Dynamic Factor Models

Generalized dynamic factor models can be estimated in a robust way by using the SODPC. Let $z_{j,t}^{(m)}, 1 \leq j \leq m, m > 1$ be observations generated by a stationary generalized dynamic one-factor model with k lags, that is, they satisfy

$$z_{j,t}^{(m)} = d_j^{(m)} + c_{j,0}^{(m)}f_t + ... + c_{j,k}^{(m)}f_{t-k} + u_{j,t}^{(m)}, \quad 1 \leq j \leq m, 1 \leq t \leq T, \tag{28}$$

where the $u_{j,t}$, $1 \leq j \leq m$, are independent stationary processes. Moreover the processes $u_{j,t}$, $1 \leq j \leq m$, are independent of the stationary process f_t. We assume that the $m + 1$ processes have finite variance and without loss of generality mean 0.

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 $s_j^{(m)} = S_M(u_{tj}^{(m)})$, then

$$0 < s_j \leq A_2. \tag{29}$$

A3. The function ρ has a derivative ψ which is continuous and bounded. Then $A_3 = \sup \psi < \infty$.

A4. $A_4 = \inf_j \mathbf{E}(u_{t,j}\psi(u_{t,j}^{(m)})) > 0$.

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 $\sup_m \sup_{1 \leq j \leq m} |d_j^{(m)}| \leq C$.

A6. Let $\mathbf{c}_i^{(m)} = (c_{1,i}^{(m)}, ..., c_{m,i}^{(m)})$, $0 \leq i \leq k$, and $E^{(m)}$ the subspace of $\mathbb{R}^m$ generated by $\mathbf{c}_i^{(m)}, 1 \leq i \leq m$. Then, we can write $\mathbf{c}_0^{(m)} = \mathbf{g}^{(m)} + \mathbf{e}^{(m)}$, where $\mathbf{d}^{(m)}$ is orthogonal to $E^{(m)}$ and $\mathbf{e}^{(m)} \in E^{(m)}$. Then, there exists δ such that $||\mathbf{g}^{(m)}||^2 \geq m\delta$ for all m. This condition implies that the common part of $\mathbf{z}_t^{(m)} = (z_{t,1}^{(m)}, ..., z_{t,m}^{(m)})$ does not get close to the $k - 1$ dimensional subspace $E^{(m)}$ when m increases.

ARTICLE IN PRESS

Theorem 1 of Peña and Yohai (2023) proves that the SODPC satisfies the following property.

Theorem 1. *Assume A1-A6. Let* $\mathbf{z}_t^{(m)} = (z_{t,1}^{(m)}, ..., z_{t,m}^{(m)})$ *generated as the dynamic one factor model given in (28) and let* $(\mathbf{a}_P^{*(m)}, \mathbf{B}_P^{*(m)}, \mathbf{d}_P^{*(m)})$ *be the asymptotical optimal values 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*

$$m^{1/2}\left(\frac{1}{m}\sum_{j=1}^{m}S_j^2(\mathbf{a}_P^{*(m)}, \mathbf{B}_P^{*(m)}, \mathbf{d}_P^{*(m)})^2 - \frac{1}{m}\sum_{j=1}^{m}S_M^2(u_{j,t}^{(m)})\right) = O(1). \tag{30}$$

That is, when $m \to \infty$, the mean of the squares of the population M scales of the reconstruction errors using the first population ODPC tends to the mean of the squares of the population M-scales of the idiosyncratic terms with rate $m^{1/2}$. In particular, for a dynamic factor model with $k = 1$, that is, when $z_{j,t} = d_j + c_j f_t + u_{j,t}, 1 \le j \le m, 1 \le t \le T$, this property holds with $k_1 = k_2 = 0$. Observed that this property holds regardless of the value of k_1. Then the first principal components of a GDFM with one factor and k lags can be carry out with $k_1 = 0$ and $k_2 = k$.

An intuitive reason for the validity of this theorem is similar to the one given to justify the ordinary principal components to estimate factorial models without lags (see Eq. (6)).

Theorem 1 can be generalized for a model with q factors, if the first q S-ODPC are obtained simultaneously, so that they minimize the sum of the squares of the M-scales of the reconstruction errors.

Other procedures have been proposed for robust estimation of DFM. Kristensen (2014) considered least absolute deviations (LAD) to estimate DFM. However, this method does not have a high finite sample breakdown point in this case and may fail with a moderate proportion of outliers. Fan et al. (2019b) studied the robust estimation of the covariance matrix in factor models, although they assume that the factors are known and, therefore, the model analyzed is, in fact, a multivariate regression model. They suppose that the distribution of the observations, as well as the one of the factors, have long tails but with finite fourth moments and apply a Huber M-estimate to obtain the terms of the estimation matrix. Note that their procedure do not use any time series structure and is develop for independent data. Thus, we expect that more powerful and efficient estimation methods for time series can be obtained when the time series dynamics are considered.

Fan et al. (2021a) includes an overview of factor models and its robustness. When observations are heavy tailed they review the proposal of Fan et al. (2021b) of a shrinkage estimate of the covariance matrix. Suppose zero mean variables, then the shrinkage robust covariance matrix estimate is defined as follow:

$$\widehat{\mathbf{\Sigma}}_x^R(\tau) = \frac{1}{n}\sum_{i=1}^{n}\widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i',$$

where,

$$\widetilde{\mathbf{x}}_i = \mathbf{x}_i, \qquad if\|\mathbf{x}_i\|_4 \le \tau,$$

and

$$\widetilde{\mathbf{x}}_i = \frac{\tau}{\|\mathbf{x}_i\|_4}\mathbf{x}_i, \qquad if\|\mathbf{x}_i\|_4 > \tau.$$

Thus, observations with $4th$ order norm larger than some threshold τ are shrinked. These authors derive the statistical error rate of $\widehat{\mathbf{\Sigma}}_x^R(\tau)$ as estimate of $\mathbf{\Sigma}_x$ in terms of the spectral norm. The results can be generalized to non zero means by using the proposal in Fan et al. (2019a).

Trucíos et al. (2021) consider the robust estimation of dynamic factors with possibly infinite dimensional factor spaces and propose robust versions of identification, estimation, and forecasting procedures. The main idea is to robustify the whole method by using robust estimates of the lag covariance matrices based on the proposal of Raymaekers and Rousseeuw (2019). In this way, instead of using the standard Hallin and Liška (2007) criterion, the number of factor is decided by inserting in the criterion a robust estimate of the covariance matrices. The same idea applies to all the steps of the standard procedure.

9. Robust Estimation of Other Multivariate Time Series Models

The literature on this field is not large. For VAR models we have only found a few references apart from the already indicated work by Muler and Yohai (2013). Qiu et al. (2015) develop a framework for estimating heavy-tailed VAR processes. They use an elliptical VAR model to accommodate heavy-tailed time series and propose a quantile-based robust estimator for the transition matrix of the VAR process. Wu and Wu (2016) studied the properties of Lasso and penalized estimators for a VAR model with weakly correlated and heavy-tailed noises. Wang and Tsay (2021) proposed a robust estimation for high-dimensional VAR models including sparse, reduced-rank, banded, and network VAR models. It is assumed that the noise in the VAR model may have heavy-tailed contamination, and conditional heteroskedasticity, as well as some weak correlation structure. The suggested methods is based on two ingredients. First a constrained multivariate Yule-Walker estimator assuming either a latent low-dimensional structure or coefficients subject to some linear restrictions. Second, a robust covariance

matrix estimator, based on truncation of the time series data. Assuming bounded moment condition the consistency of the proposed estimators is established.

A few other approaches have been proposed for MTS in other problems. The robust estimation of the covariance matrices have been studied by Kim and Lee (2013), using minimum density power divergence estimator (MDPDE), as proposed by Basu et al. (1998), and employing a multivariate normal mixture family. Zhang (2021) considers the robust estimation of stationary vector processes with distributions with a small number of finite moments and dimension p of order $log(p) = o(nc)$, where n is the sample size and $c > 0$. This work derives the convergence rates for Huber estimators of the mean vector, covariance matrix, and precision matrix for a large class of time series.

Croux et al. (2010) developed a robust version of multivariate exponential smoothing that chooses the smoothing parameter matrix by minimizing a robust loss function, and Lanius and Gather (2010) considered a robust regression-based online filter for signal extraction in multivariate time series. Observations are trimmed and the estimates of the local level and slope of the signal are obtained by means of multivariate LS regression from the trimmed observation.

10. Empirical Studies of Outlier Detection and Robust Estimation in HDMTS

Gupta et al. (2014) discussed different areas in which searching for time series outliers have been useful, as in environmental, industrial, computer network, biological, astronomy, web, information network and economics applications. In this section we will summarize some real data examples of outlier detection or robust estimation in multivariate time series data that have been published. Most of these works deal with time series of small dimension and, given the orientation of this review to high dimensional data sets, we will only comment on time series data sets in which the dimension of the vector of series is larger than 20.

From the point of view of outlier detection, Baragona and Battaglia (2007b) considered 27 quarterly economic series from 3 countries with $T = 156$, fit a DFM with 6 factors and found, by using projections on the idiosyncratic space, 5 outliers. Three of these 5 multivariate outliers were classified as international, whereas the other two where national, as they affect only to one of the countries. Alonso et al. (2020) analyzed a dataset of hourly day-ahead demand for the ISO New England electricity market from January 2004 to December 2016. The number of series is 192 and $T = 4749$ days. They use projection methods to clean from outliers and 113 MAOs and 59 univariate additive outliers are detected, including all the values corresponding to the daylight saving time days, where the demand is set to zero at the second hour.

He et al. (2020) analyzed three public technical datasets to test a proposed machine learning method for multivariate outlier detection. Two of them are public datasets of NASA-U2122s spacecraft with 55 and 25 variables such as telemetrydata, radiation, temperature, power and so on, and the third one includes five weeks of server data in a large Internet company, with 38 series of CPUload, network-usage, memory-usage and so on. No comparison with other alternative methods is made.

Regarding robust methods, Kristensen (2014) considered the database of the Federal Reserve Economic Data (FRED), available by the Federal Reserve of St. Louis that was also used by Stock and Watson (2002). A total of 111 monthly US macroeconomic variables were included, with data spanning the period 1971:2 to 2012:10. When applying estimation using the LAD (least absolute deviation) estimator and robust scale using MAD a gain in forecasting performance is obtained compared to the traditional principal components estimation. Fan et al. (2019b) used daily returns of 393 stocks of S&P 500 index from 2005 to 2013 and showed that the robust approach proposed in this paper has smaller risk errors and better empirical performance in estimating portfolio risks. Trucíos et al. (2021) analyzed again the 115 US macroeconomic and financial time series at monthly frequency from January 1959 through August 2014 and compared the forecasting results to previous no robust methods. Again, no comparison is made with alternative robust estimation procedures. Wang and Tsay (2021) compared a proposed robust estimation procedure for HD VAR models to some alternative methods in a real data set of 40 US quarterly macroeconomic variables from the third quarter of 1959 to fourth quarter of 2007, with 194 observations for each variable.

Finally, Peña and Yohai (2023) compared the predicions of a non-robust ODPC procedure and a robust SODPC one in 24 electricity series (one for every hour) of New England, and showed that the prediction errors achieved with the robust SODPC method are much smaller than those obtained with the non robust ODPC, specially between 10 am and 8 pm.

11. Conclusions

Outlier detection and robust estimation of high dimensional set of MTS have been often studied in the statistics, computer science and machine learning literature. However, as we have seen in the previous section, very few works have tried to compare different robust estimation methods. Our first conclusion is that large scale studies comparing approaches are really needed, especially for the high-dimensional case with sample size not much larger than the number of time series. It is to be expected that methods based on more relevant information will be more useful. Thus, procedures that use the time series structure of the data should be more powerful than those that assume independent observations, and those that are based on the whole dynamic of the series, including lag autocorrelations and cross-correlations, should work better than those based on simple instantaneous correlations.

An important problem that has not been explored in the literature is the consequence of model misspecification on the procedures of outlier detection. For instance, to assume an DFM in data that have been generated by a different, possible

non linear, dynamic model. These analyses may lead to better model free outlier detection methods for very HD MTS and this is an important topic of future research.

A common framework in MTS is to assume that the data have a vector structure. These vectors are often assumed to be approximately generated by a Dynamic Factor Model. However, more complex types of MTS sets contaminated by outliers should be considered in the future, as matrix and tensor time series. Note that high dimensional data is often the result of disaggregating the information. For instance, instead of considering the total sales in a set of d_1 cities the data are split into sales in d_2 types of shops and sales in d_3 different products. In this way, we end up with $N = d_1 \times d_2 \times d_3$ time series classified by city, type of shop and product, the three dimensions of the array, or tensor time series. The statistical analysis for this type of data started with the matrix form, see Wang et al. (2019a), but is has been generalized as tensor data with many dimensions, see Chen et al. (2022) and Peña (2022).

These disaggregated time series data may appear in all fields and, therefore, the need of new procedures to find outliers in these tensor sets of high dimensional time series classified in several dimensions will be a very important problem in the future.

Acknowledgments

References

Aggarwal, C.C., 2017. An Introduction to Outlier Analysis. Springer.

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. doi:10.1111/1468-0262.00273.

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

Baragona, R., Battaglia, F., 2007. Outliers detection in multivariate time series by independent component analysis. Neural computation 19 (7), 1962–1984.

Baragona, R., Battaglia, F., 2007. Outliers in dynamic factor models. Electronic Journal of Statistics 1, 392–432.

Basu, A., Harris, I.R., Hjort, N.L., Jones, M., 1998. Robust and efficient estimation by minimising a density power divergence. Biometrika 85 (3), 549–559.

Battaglia, F., Cucina, D., 2020. Outlier identifiability in time series. Stat 9 (1), e281.

Bellini, T., 2016. The forward search interactive outlier detection in cointegrated var analysis. Advances in Data Analysis and Classification 10 (3), 351–373.

Blázquez-García, A., Conde, A., Mori, U., Lozano, J.A., 2021. A review on outlier/anomaly detection in time series data. ACM Computing Surveys (CSUR) 54 (3), 1–33.

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

Brillinger, D.R., 1981. Time Series: Data Analysis and Theory. Society for Industrial and Applied Mathematics.

Chen, C., Liu, L.-M., 1993. Joint estimation of model parameters and outlier effects in time series. Journal of the American Statistical Association 88 (421), 284–297.

Chen, R., Yang, D., Zhang, C.-H., 2022. Factor models for high-dimensional tensor time series. Journal of the American Statistical Association 117 (537), 94–116.

Cheng, H., Tan, P.-N., Potter, C., Klooster, S., 2009. Detection and characterization of anomalies in multivariate time series. In: Proceedings of the 2009 SIAM international conference on data mining. SIAM, pp. 413–424.

Chou, R.Y., Yen, T.-J., Yen, Y.-M., 2020. Macroeconomic forecasting using approximate factor models with outliers. International Journal of Forecasting 36 (2), 267–291.

Croux, C., Gelper, S., Mahieu, K., 2010. Robust exponential smoothing of multivariate time series. Computational Statistics & Data Analysis 54 (12), 2999–3006.

Croux, C., Haesbroeck, G., 2000. Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. Biometrika 87 (3), 264–273.

Cucina, D., Di Salvatore, A., Protopapas, M.K., 2014. Outliers detection in multivariate time series using genetic algorithms. Chemometrics and Intelligent Laboratory Systems 132, 103–110.

Donoho, D.L., Huber, P.J., 1983. The notion of breakdown point. In: P.J. Bickel, K.D., Hodges, J. (Eds.), Festschrift for Erich L.Lehmann. Springer, Belmont Ca..

Fan, J., Ke, Y., Sun, Q., Zhou, W.-X., 2019. Farmtest: Factor-adjusted robust multiple testing with approximate false discovery control. Journal of the American Statistical Association 114 (528), 1880–1893.

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.

Fan, J., Wang, W., Zhu, Z., 2021. A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery. Annals of Statistics 49 (3), 1239.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: Identification and estimation. The Review of Economics and Statistics 82 (4), 540–554. doi:10.1162/003465300559037.

Fox, A.J., 1972. Outliers in time series. Journal of the Royal Statistical Society B 34 (3), 350–363.

Galeano, P., Peña, D., 2020. Outlier detection in dynamic factor models. Working paper, Universidad Carlos III de Madrid.

Galeano, P., Peña, D., Tsay, R.S., 2006. Outlier detection in multivariate time series by projection pursuit. Journal of the American Statistical Association 101 (474), 654–669.

Geweke, J., 1977. The dynamic factor analysis of economic time series models. In: Aigner, D.J., Goldberger, A.S. (Eds.), Latent Variables in Socio-Economic Models. North-Holland Publishing Company, Amsterdam, pp. 365–383.

Gonzalez, J., Dang, T., 2021. Outviz: Visualizing the outliers of multivariate time series. In: The 12th International Conference on Advances in Information Technology, pp. 1–5.

Gupta, M., Gao, J., Aggarwal, C.C., Han, J., 2014. Outlier detection for temporal data: A survey. IEEE Transactions on Knowledge and data Engineering 26 (9), 2250–2267.

Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association 102 (478), 603–617.

Hampel, F.R., 1971. A general qualitative definition of robustness. Annals of Mathematical Statistics 42 (6), 1887–1896.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: The Approach Based on Influence Functions. Wiley.

He, Q., Zheng, Y., Zhang, C., Wang, H.-Y., 2020. Mtad-tf: Multivariate time series anomaly detection using the combination of temporal pattern and feature pattern. Complexity 2020.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24 (6). doi:10.1037/h0071325.

Huber, P.J., 1964. Robust estimation of a location parameter. The Annals of Mathematical Statistics 35 (1), 73–101. doi:10.1214/aoms/1177703732.

Hyndman, R.J., Wang, E., Laptev, N., 2015. Large-scale unusual time series detection. In: 2015 IEEE international conference on data mining workshop (ICDMW). IEEE, pp. 1616–1619.

Inselberg, A., 1998. A survey of parallel coordinates. Mathematical Visualization: Algorithms, Applications and Numerics 167–179.

Inselberg, A., Dimsdale, B., 1990. Parallel coordinates. VIS'90: Proceedings of the 1st Conference on Visualization.

Kim, B., Lee, S., 2013. Robust estimation for the covariance matrix of multivariate time series based on normal mixtures. Computational Statistics & Data Analysis 57 (1), 125–140.

Kristensen, J.T., 2014. Factor-based forecasting in the presence of outliers: Are factors better selected and estimated by the median than by the mean. Studies in Nonlinear Dynamics & Econometrics 18 (3), 309–338.

Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: Inference for the number of factors. The Annals of Statistics 40 (2), 694–726. doi:10.1214/12-AOS970.

Lanius, V., Gather, U., 2010. Robust online signal extraction from multivariate time series. Computational Statistics & Data Analysis 54 (4), 966–975.

Li, G., Chen, Z., 1985. Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and monte carlo. Journal of the American Statistical Association 80 (391), 759–766. doi:10.1080/01621459.1985.10478181.

López-Oriona, Á., Vilar, J.A., 2021. Outlier detection for multivariate time series: A functional data approach. Knowledge-Based Systems 233, 107527.

López-Oriona, Á., Vilar, J.A., 2021. Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series. Expert Systems with Applications 185, 115677.

Maronna, R.A., 2000. Principal components and orthogonal regression based on robust scales. Technometrics 47 (3), 603–618. doi:10.11198/004017005000000166.

Maronna, R.A., Martin, D.R., Yohai, V.J., Salibián-Barrera, M., 2017. Robust Statistics: Theory and Methods (with R), edition 2nd. Wiley.

Maronna, R.A., Yohai, V.J., 2008. Robust low-rank approximation of data matrices with elementwise contamination. Technometrics 50 (3), 295–304. doi:10.1198/004017008000000190.

Meng, F., Gao, Y., Wang, H., Liu, Y., Wang, H., 2021. Tslod: a coupled generalized subsequence local outlier detection model for multivariate time series. International Journal of Machine Learning and Cybernetics 1–12.

Muler, N., Yohai, V.J., 2013. Robust estimation for vector autoregressive models. Computational Statistics & Data Analysis 65, 68–79.

Nieto, F.H., Peña, D., Saboyá, D., 2016. Common seasonality in multivariate time series. Statistica Sinica 26 (1), 1389–1410.

Okamoto, M., Kanazawa, M., 1968. Minimization of eigenvalues of a matrix and optimality of principal componentss. Annals of Mathematical Statistics 39 (3). doi:10.1214/aoms/1177698308. 727–164

Peña, D., 2022. Comment on "factor models for high-dimensional tensor time series". Journal of the American Statistical Association 117 (537), 118–123.

Peña, D., Box, G.E.P., 1987. Identifying a simplifying structure in time series. Journal of the American Statistical Association 82 (399), 836–843. doi:10.1080/01621459.1987.10478506.

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., Tsay, R.S., 2021. Statistical Learning for Big Dependent Data. John Wiley & Sons.

Peña, D., Tsay, R.S., Zamar, R., 2019. Empirical dynamic quantiles for visualization of high-dimensional time series. Technometrics 61 (4).

Peña, D., Yohai, V.J., 2016. Generalized dynamic principal components. Journal of the American Statistical Association 111 (515), 1121–1131. doi:10.1080/01621459.2015.1072542.

Peña, D., Yohai, V.J., 2023. Robust forecasting of multiple time series with one-sided dynamic principal components. In: Yi, M., Nordhausen, K. (Eds.), Festschrift for David E. Tyler. Springer, New York.

Qiu, H., Xu, S., Han, F., Liu, H., Caffo, B., 2015. Robust estimation of transition matrices in high dimensional heavy-tailed vector autoregressive processes. In: International Conference on Machine Learning. PMLR, pp. 1843–1851.

Raymaekers, J., Rousseeuw, P., 2019. A generalized spatial sign covariance matrix. Journal of Multivariate Analysis 171, 94–111.

Rousseeuw, P.J., 2005. Least median of squares. Journal of the American Statistical Association 79 (388), 871–880. doi:10.1080/01621459.1984.10477105.

Rousseeuw, P.J., Yohai, V.J., 1984. Robust regression by means of S-estimators. In: Franke, J., Härdle, W., Martin, D. (Eds.), Robust and Nonlinear Time Series. In: Lecture Notes in Statistics, Vol. 26. Springer, New York, US, pp. 256–272.

Salibián-Barrera, M., Van Aelst, S., Willems, G., 2006. Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and monte carlo. Journal of the American Statistical Association 101 (475), 1198–2011. doi:10.1198/016214506000000096.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97 (460), 1167–1179.

Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D., 2019. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 2828–2837.

Talagala, P.D., Hyndman, R.J., Smith-Miles, K., Kandanaarachchi, S., Muñoz, M.A., 2020. Anomaly detection in streaming nonstationary temporal data. Journal of Computational and Graphical Statistics 29 (1), 13–27.

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.

Tsay, R.S., 1986. Time series model specification in the presence of outliers. Journal of the American Statistical Association 81 (393), 132–141.

Tsay, R.S., Peña, D., Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika 87 (4), 789–804.

Vishwakarma, G.K., Paul, C., Elsawah, A.M., 2021. A hybrid feedforward neural network algorithm for detecting outliers in non-stationary multivariate time series. Expert Systems with Applications 184, 115545.

Wang, D., Liu, X., Chen, R., 2019. Factor models for matrix-valued high-dimensional time series. Journal of Econometrics 208, 231–248.

Wang, D., Tsay, R.S., 2021. Robust estimation of high-dimensional vector autoregressive models. arXiv preprint arXiv:2107.11002.

Wang, D., Yu, Y., Rinaldo, A., Willett, R., 2019. Localizing changes in high-dimensional vector autoregressive processes. arXiv preprint arXiv:1909.06359.

Weng, X., Shen, J., 2008. Detecting outlier samples in multivariate time series dataset. Knowledge-based systems 21 (8), 807–812.

Wu, W.-B., Wu, Y.N., 2016. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. Electronic Journal of Statistics 10 (1), 352–379.

Yang, K., Shahabi, C., 2007. An efficient k nearest neighbor search for multivariate time series. Information and Computation 205 (1), 65–98.

Zhang, D., 2021. Robust estimation of the mean and covariance matrix for high dimensional time series. Statistica Sinica 31 (2), 797–820.

Zoubir, A.M., Koivunen, V., Chakhchoukh, Y., Muma, M., 2012. Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. IEEE Signal Processing Magazine 29 (4), 61–80.