DOI: 10.1016/j.jeconom.2020.01.004

A Robust Procedure to Build Dynamic Factor Models with Cluster Structure

Andrés M. Alonso* Pedro Galeano † Daniel Peña‡

Abstract

Dynamic factor models provide a useful way to model large sets of time series. These data often have heterogeneity and cluster structure and the formulation and estimation of dynamic factor models should be adapted to these features. This article presents a procedure to fit Dynamic Factor Models with Cluster Structure (DFMCS), where some of the factors are global and others group-specific, to heterogeneous data that may include multivariate additive outliers and level shifts. The procedure starts with an initial cleaning of the times series from outlying effects. Then a first estimation of the possible factors is applied to the cleaned data and these factors are used to build the common component of each series. The groups are found by studying the joint dependency of these common components. Then additional factors are estimated by using the series in each cluster and, finally, all the factors found are classified as global or group-specific. We show in a Monte Carlo study that the procedure works well and seems to be better than other alternatives in terms of estimation of factors and loadings as well as in terms of misclassification rates for the series. An example of an electricity market is presented to illustrate the advantages of cleaning for outliers and taking into account the cluster structure for understanding and forecasting.

Keywords: Clustering time series; Dependency measures; Multivariate additive outliers; Multivariate level shifts; Principal Components.

JEL Classification: C32; C51.

*Department of Statistics and Institute Flores de Lemus, Universidad Carlos III de Madrid, email: andres.alonso@uc3m.es

†Department of Statistics and Institute UC3M-BS of Financial Big Data, Universidad Carlos III de Madrid, email: pedro.galeano@uc3m.es.

‡Department of Statistics and Institute UC3M-BS of Financial Big Data, Universidad Carlos III de Madrid, email: daniel.pena@uc3m.es

1 Introduction

The study of multiple related time series is an important area of research where George Tiao has made seminal contributions. Among them, Box and Tiao (1977) presented a method to order a set of time series for predictability and found linear combinations of integrated time series that were stationary. This result prepared the way for the development ten years later of the concept of cointegration by Engle and Granger (1987). Tiao and Box (1981) and Tiao and Tsay (1989) developed general procedures to fit VARMA and VAR models that have been very useful for modelling small sets of related time series. However, as the dimension of the parameter matrices in these models grows with the squared of the number of series, new procedures are required for large data sets and Dynamic Factor Models (DFM) have proved to be useful in these cases. DFM may assume a contemporaneous relationship between the factors and the series, see, for instance, Peña and Box (1987), Stock and Watson (1988, 2002), Bai and Ng (2002), Lam and Yao (2012), Onatski (2012), and recently, Chen, Tsay and Chen (2019), among others, but also allow for lags in the factor effects, as in Forni et al. (2000, 2005, 2015). Large sets of series often contain outliers and cluster structure and the DFM should be adapted to these features. The presence of groups in panel data has been well documented. For instance, Lin and Ng (2012) estimated the allocation of the series to the groups by threshold panel regression, Bonhomme and Manresa (2015) minimized a least squares criterion for all possible groups and cross-sectional units, and Su et al. (2016) used a new variant of Lasso in the estimation. In the same way, the dynamic evolution of time series may be affected by some global factors, that reflect the evolution of the global economy, and some specific factors, that are group-dependent.

Dynamic Factor Models with Cluster Structure (DFMCS) have been studied, first, assuming that the classification of the observed series into the clusters is known. Wang (2010) proposed a multifactor model where the series in each group are affected by global and specific factors and derived conditions for identification of these models. Hallin and Liška (2011) proposed a two clusters model where the factors define four orthogonal subspaces: (1) strongly common variables that are common factors in the two clusters; (2) strongly idiosyncratic variables for both clusters; and (3) and (4) where the variables are common factors for a group, and idiosyncratic variables for the other. This model is very flexible, but the number of subspaces grows to 2^k for k groups. It is again assumed that the number of groups and the allocation to the series to the groups is known. Ando and Bai (2017) proposed a more general model assuming unknown membership. The main weakness of their procedure is the allocation of the series to the groups, which can avoid a correct estimation of the factors, as we will show in this work. Barnichon and Mesters (2018) presented a DFMCS to separate aggregate labor market forces and demographic-specific trends and decomposed the factors into a common component and several demographic specific components. These results are a good example of the importance of taking into account the clusters when they exist, because by ignoring them, we may miss some specific factors. On the other hand, when the group structure is included the specific factors are better estimated increasing our understanding of the problem and the forecast precision, as we will show in this work. As a simple illustration, suppose we generate a set of $m = 300$ time series with two groups and with one factor in each group. The first group includes 25% of the data, i.e., $m_1 = 75$ time series, and a specific factor formed by independent $N(0, 1/3)$; the second group has $m_2 = 225$ series and a specific factor formed by independent $N(0, 1)$. The group-specific factor loadings are drawn from a $N(0, 1)$ as well as the error term, that is white noise. Then, if we fit the standard dynamic factor model using the Ahn and Horenstein (2013) test to identify the number of factors, only one factor is usually found (70% in 100 cases) whereas using the groups structure identified by the procedure proposed in this article, the two factors are found 100% of the times.

An important problem when dealing with time series is outlier detection, a field in which George Tiao has also carried out pioneer work (Chang, Tiao and Chen, 1988). Large data sets are often recorded by controlling devices that automatically collect them using wireless sensor networks. However, sensor nodes sometimes fail to record the data correctly and these failures will produce outliers in the time series. Outliers in DFM have been studied, among others, by Baragona and Battaglia (2007) and Galeano and Peña (2019). These last authors proposed a procedure to clean the series from outliers based only on linear projections of the data, that can be applied in large data sets. This article presents a new procedure to fit DFMCS that has some advantages over previous methods. First, clusters are found by a powerful method that builds clusters of time series by linear dependency, proposed by Alonso and Peña (2018). Second, a set of rules to classify the estimated factors as global or specific is presented. Third, the procedure is made robust to additive outliers and level shifts using the results of Peña and Galeano (2019) for outlier detection in large data sets and by a new method to find outlying series that do not follow the DFMCS. We show in a Monte Carlo study that the procedure seems to work better than previous proposals and illustrate the advantages of taking into account outliers and groups in understanding and forecasting time series from an electricity market.

The rest of the paper is organized as follows. Section 2 introduces the Dynamic Factor Model with Cluster Structure. Section 3 proposes a procedure to fit these models. Section 4 generalizes the proposed procedure to make it robust to outliers and outlying time series. Section 5 illustrates the performance of the proposed procedure with a simulation study. Section 6 shows an application to series of electricity demand in New England. Finally, Section 7 concludes.

2 Dynamic factor models with cluster structure

Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{mt})'$ be an m-dimensional vector of stationary time series with zero mean. We assume that each component of the vector of observed series can be written as a linear combination of global and specific factors in k groups plus idiosyncratic noise. Then, we have the Dynamic Factor Model with Cluster Structure (DFMCS), that can be written as:

$$\mathbf{x}_t = \mathbf{P}_0\mathbf{f}_{0t} + \sum_{i=1}^{k} \mathbf{P}_i\mathbf{f}_{it} + \mathbf{n}_t, \tag{1}$$

where $\mathbf{f}_{0t} = (f_{01t}, \ldots, f_{0r_0t})'$ is an r_0-dimensional vector of global factors, $\mathbf{P}_0 = \left[\mathbf{P}'_{0,1}|\cdots|\mathbf{P}'_{0,k}\right]'$ is an $m \times r_0$ global factor loading matrix **and $P_{0,i}$, for $i = 1, \ldots, k$, is the $m_i \times r_0$ loading matrix for the m_i series of the i-th group and $k \geq 1$ is the number of clusters**. The vectors $\mathbf{f}_{it} = (f_{i1t}, \ldots, f_{ir_it})'$ are r_i-dimensional vectors of specific factors corresponding to the i-th cluster, and $\mathbf{P}_i = \left[\mathbf{0}'_{i,1}|\cdots|\mathbf{P}'_{i,i}|\cdots|\mathbf{0}'_{i,k}\right]'$ is the $m \times r_i$ matrix of specific factor loadings, that only affect to the m_i time series in the i-th group. We have assumed, without loss of generality and to simplify the exposition, that the series are ordered, so that the first m_1 series correspond to the first group and the last m_k to the last group, and $\sum_{i=1}^{k} m_i = m$. It is well known that in DFM we have to impose identification restrictions and the usual assumptions are orthonormal columns in the loading matrix and diagonal covariance matrix of the factors **or vice versa**. The identification of the DFMCS has been studied by Wang (2010) and the conditions **needed** can be written as: (1) $\mathbf{P}'_0\mathbf{P}_0 = \mathbf{I}_{r_0}$, where $\mathbf{I}_{r_0}$ is the identity matrix of order r_0; (2) $\mathbf{P}'_i\mathbf{P}_i = \mathbf{P}'_{i,i}\mathbf{P}_{i,i} = \mathbf{I}_{r_i}$ for $i = 1, \ldots, k$; (3) $\mathbf{P}'_0\mathbf{P}_i = \mathbf{P}'_{0,i}\mathbf{P}_{i,i} = \mathbf{0}_{r_0 \times r_i}$; and (4) the covariance matrix of the $r = \sum_{j=0}^{k} r_j$ factors is diagonal. Note that also $\mathbf{P}'_i\mathbf{P}_j = \mathbf{0}_{r_i \times r_j}$, for $i \neq j$. We can write this model as an standard factor model calling $\mathbf{f}_t = (\mathbf{f}'_{0t}, \mathbf{f}'_{1t}, \ldots, \mathbf{f}'_{kt})'$, and $\mathbf{P} = [\mathbf{P}_0|\mathbf{P}_1|\cdots|\mathbf{P}_k]$, we have

$$\mathbf{x}_t = \mathbf{P}\mathbf{f}_t + \mathbf{n}_t, \tag{2}$$

and the previous conditions implied the usual identification restriction $\mathbf{P}'\mathbf{P} = \mathbf{I}_r$.

The idiosyncratic term or noise, $\mathbf{n}_t = (n_{1t}, \ldots, n_{mt})'$, is a general sequence of stationary time series with mean $\mathbf{0}_m$, and weak dependency as stated in Bai and Ng (2002) or Ahn and Horenstein (2013). The global and specific factors are orthogonal and follow a diagonal vector autoregressive moving average, VARMA(p,q), model $\mathbf{\Phi}(B)\,\mathbf{f}_t = \mathbf{\Theta}(B)\,\mathbf{u}_t$, where the polynomials $\mathbf{\Phi}(B) = \mathbf{I}_r - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p$ and $\mathbf{\Theta}(B) = \mathbf{I}_r - \mathbf{\Theta}_1 B - \cdots - \mathbf{\Theta}_q B^q$ have diagonal parameter matrices, B is such that $B\mathbf{f}_t = \mathbf{f}_{t-1}$, the roots of the determinantal equation $|\mathbf{\Phi}(B)| = 0$ as well as those of $|\mathbf{\Theta}(B)| = 0$ are outside the unit circle, and $\mathbf{u}_t$ is a sequence of uncorrelated and identically distributed (i.i.d.) random vectors with mean $\mathbf{0}_r$ and diagonal covariance matrix $\mathbf{\Sigma}_{\mathbf{u}}$. Additionally, we assume that both noise processes appearing in the factor model are uncorrelated for all **lags, i.e.,** $E\left[\mathbf{n}_t\mathbf{u}'_{t-h}\right] = 0_{m\times r}$**, for all** $h = 0, \pm 1, \pm 2, \ldots$ We assume that the number of clusters and the allocation of the series to the clusters are unknown.

The estimation of a DFMCS requires obtaining the following parameters: (1) The number of global factors, r_0, the number of groups, k, and the number of specific factors in each of the k groups, $r_1, \ldots, r_k$; (2) The label variables $g_i \in \{1, \ldots, k\}$ for each of the time series **which indicate** to which group the series belongs. We call G the $m \times 1$ vector with components g_i, for $i = 1, \ldots, m$; (3) The loading matrices of the global and specific factors, $\mathbf{P}_0, \mathbf{P}_1, \ldots, \mathbf{P}_r$ and the time series of these factors, $f_{01,t}, \ldots, f_{0r_0,t}$ and $f_{11,t}, \ldots, f_{1r_1,t}, \ldots, f_{k1,t}, \ldots, f_{kr_k,t}$, respectively. Given the estimated factors and noises, where $\widehat{\mathbf{n}}_t = \mathbf{x}_t - \widehat{\mathbf{P}}\widehat{\mathbf{f}}_t$, estimators of the parameters of the univariate ARMA models for the factors and noises can be computed. In Section 3 we will describe the proposed procedure to find the groups and estimate the factors, assuming that we have a sample that is free from outlier effects. In Section 4 we will present the procedure to remove the effects of multivariate additive outliers and level shifts **as well as possible outlying time series**.

3 Fitting the DFMCS

The procedure we propose has the following four steps: (1) The observed time series are cleaned from additive outliers and level shifts. Additionally, outlying time series are removed; (2) An initial set of factors and their loadings is estimated from the time series obtained in step 1 and they are used to build the common component of each time series. Then, the clustering algorithm proposed by Alonso and Peña (2018) for finding groups of time series with similar linear dependency is applied to these common components to find the groups; (3) A set of factors and their loadings are estimated in each group of time series and all the factors found **now as well as** those found in step 2 are compared and classified as global or specific; (4) The effect of the **global factors is removed** from each of the time series and the group-specific residuals obtained are then used to re-estimate the specific factors. Finally, groups are checked for possible recombination.

Remark 2. Step 4 is included because, first, it usually improves the estimation of the specific factors and, second, the specific factors obtained in step 4 are orthogonal to the global factors as they are computed from residuals from the global factors. This latter property is not guaranteed for step 3.

To simplify the exposition, we present next steps $2-4$ of the procedure assuming that the time series **have already been** cleaned. The cleaning process of Step 1 will be explained in Section 4. We start with an initial estimation of **the factors** and factor loadings. This is carried out by minimizing

$$SE_1 = \sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{P}\mathbf{f}_t\|^2 , \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean vector norm. **The solution is equivalent to using** PCA. Let

$$\widehat{\mathbf{\Gamma}}_{\mathbf{x}}(0) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'. \tag{4}$$

be the sample covariance matrix of the time series $\mathbf{x}_t$. Then, the eigenvectors associated to the r_c largest eigenvalues of the matrix $\widehat{\mathbf{\Gamma}}_{\mathbf{x}}(0)$ provide us with an estimate of the factor loading matrix $\widehat{\mathbf{P}}$, with columns $\widehat{\mathbf{p}}_1, \ldots, \widehat{\mathbf{p}}_{r_c}$. The number of factors, r_c, is determined by using the test proposed by Ahn and Horenstein (2013) based on the ratios of consecutive eigenvalues of the matrix $\widehat{\mathbf{\Gamma}}_{\mathbf{x}}(0)$ in (4). Note that the matrix $\widehat{\mathbf{P}}$ is expected to include all the global factors and some (or all) of the group-specific factors, so that the number of factors in this matrix, r_c, will be in general larger than the true number of global factors, r_0. Also, in practice, when the factors are of different degree of strength, the results of these type of tests in real time series are usually better if they are applied twice, as suggested by Lam and Yao (2012). Then, the factors are estimated by $\widehat{\mathbf{f}}_{i,t} = \widehat{\mathbf{p}}_i'\mathbf{x}_t$, for $i = 1, \ldots, r_c$, **and the common component by** $\mathbf{c}_t = \widehat{\mathbf{P}}\widehat{\mathbf{P}}'\mathbf{x}_t$.

The groups are now built applying the algorithm proposed by Alonso and Peña (2018) for clustering time series by dependency to the m **time series of common components**, $\mathbf{c}_t$. The measure of linear dependency used is the Generalized Cross Correlation (GCC), defined for two stationary time series, z_t and s_t, as:

$$GCC(z_t, s_t) = 1 - \left(\frac{|\mathbf{R}_{zs,p}|}{|\mathbf{R}_{zz,p}|\,|\mathbf{R}_{ss,p}|}\right)^{1/(p+1)},$$

where $\mathbf{R}_{zs,p}$ is the $2\,(p+1)$ **squared** symmetric non negative definite matrix that corresponds to the correlation matrix of the vector stationary process $(z_t, z_{t-1}, \ldots z_{t-p}, s_t, \ldots, s_{t-p})'$, and $\mathbf{R}_{zz,p}$ and $\mathbf{R}_{ss,p}$ are, respectively, the $(p+1)$ squared and positive definite correlation matrices of the vectors $(z_t, z_{t-1}, \ldots, z_{t-p})'$ and $(s_t, s_{t-1}, \ldots, s_{t-p})'$. The GCC measure is non negative, reaches its largest value, equal to 1, if one of the series is a linear combination of its past and the values of the other series, and it is equal to 0, if the two series are uncorrelated. Therefore, we **compute** the estimators $\widehat{GCC}\,(c_{i,t}, c_{j,t})$ for all pairs of time series $(c_{i,t}, c_{j,t})$, with $i \neq j$, and build the $m \times m$ dissimilarity matrix with elements $d_{ij} = 1 - \widehat{GCC}\,(c_{i,t}, c_{j,t})$. Then, we apply a hierarchical clustering **algorithm** with single linkage to the dissimilarity matrix. The number of clusters is obtained by the a modification of the Silhouette algorithm proposed by Rousseeuw (1987), adding the restriction that the clusters must have a minimum size. We implement this restriction by omitting time series in the dendrogram analysis that have a relatively small dependency with the rest (for instance, the 90% percentile of the dendrogram's unions). Once the groups are formed, the omitted time series are assigned to the closer cluster in the single linkage sense. In this way, we obtain an estimated value of k, the number of groups, and an estimator of the vector G that gives the allocation of the series **to the** k **groups**.

In the third **step we use the series in the groups to estimate new sets of factors and their loadings.** Let $r_1^s, \ldots, r_k^s$ be the number of factors found in each group by using the Ahn and Horenstein (2013) test **applied to the eigenvalues of the sample covariance matrices of the time series in each group**. **The specific loading matrices** $\widehat{\mathbf{P}}_i$ **of dimension** $m \times r_i^s$ **and columns** $\widehat{\mathbf{p}}_{i1}, \ldots, \widehat{\mathbf{p}}_{ir_i^s}$ **are built by adding to the eigenvectors corresponding to the largest** r_i^s **eigenvalues in the** i**-th group, a set of zero values for the observations not included in the group**. The factors in each group are estimated by $\widehat{\mathbf{f}}_{ij,t}^s = \widehat{\mathbf{p}}_{ij}'\mathbf{x}_t$, with $j = 1, \ldots, r_i^s$. These group factors are expected to include all the specific factors and some (or all) of the global factors.

Now, in order to decide whether a factor is global or specific, we compare the set of r_c factors found in step 2 and the set of $\sum_{i=1}^k r_i^s$ factors found in step 3. Note that the factors contained in the first set may be a rotation of the factors contained in the second set and, therefore, it is not evident which ones should be classified as global and which as specific. Consequently, we first decide if each factor f in the first set of r_c factors is global or specific by applying the following three simple rules:

1. If f does not belong to **any of the second set of factors** then f is a global factor.

2. If f belongs **to only one of the sets of the second set of factors then** f **is a specific factor** in this group.

3. If f belongs **to more than one of the second sets of factors then** f **is a global factor**.

We decide if a factor, f, belongs to a set of specific factors by computing the empirical canonical correlation between the factor, f, and the ones in the set, $\widehat{\mathbf{f}}_{i1,t}^s, \ldots, \widehat{\mathbf{f}}_{ir_i^s,t}^s$, with $i = 1, \ldots, k$. When the empirical canonical correlation of factor f with elements of the set is higher than some threshold value, ρ_0, we say that f belongs to this set. The threshold value of $\rho_0 = 0.9$ seems to work well in the Monte Carlo exercise. Afterwards, we check if any of the groups with $r_1^s, \ldots, r_k^s$ factors include any factor that does not belong to the set of factors found in step 2. If this is the case, the factor is classified as specific factor in the corresponding group.

In step four, we compute the residuals $\mathbf{v}_t = \mathbf{x}_t - \widehat{\mathbf{P}}_0\widehat{\mathbf{f}}_{0t}$, where $\widehat{\mathbf{f}}_{0t}$ is the vector of estimated global factors obtained in step 3 and $\widehat{\mathbf{P}}_0$ is the loading matrix corresponding to these factors, and **the specific factors are re-estimated using the series v_{it} corresponding to each group.** As the clustering is based on the dependency among the series we have to check that this dependency is generated by different specific factors and not for different loadings of the global factors. To illustrate this problem, consider the very simple model

$$\begin{bmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \end{bmatrix} = \begin{bmatrix} a\mathbf{1}_{m_1} \\ b\mathbf{1}_{m_2} \end{bmatrix} f_t + \begin{bmatrix} \mathbf{n}_{1t} \\ \mathbf{n}_{2t} \end{bmatrix},$$

where $\mathbf{x}_{it}$ is $m_i \times 1$, with $m = m_1 + m_2$, and $\mathbf{1}_{m_i} = (1, \ldots, 1)'$ is also $m_i \times 1$. In this model, there are two groups of time series of similar dependency. For simplicity, let us assume that the noises are i.i.d. with the same variance σ^2 and let us call $s = var\left(f_t\right)/\sigma^2$ the signal to noise ratio. Then, the cross-correlations of any two variables in the first and second group of series will be respectively $r_1 = a^2 s/\left(1 + a^2 s\right)$ and $r_2 = b^2 s/\left(1 + b^2 s\right)$, while the correlation between series in different groups will be $r_{12} = abs/\sqrt{\left(1 + a^2 s\right)\left(1 + b^2 s\right)}$. Consequently, if a is very different from b, the clustering should detect two groups of time series. Thus, we must verify that the groups obtained are due to different specific factors and not due to differences between factor loadings in a global factor. **Therefore, we check whether all the groups have at least one specific factor**. **We may face the following cases**: (1) All the k groups found include at least one **specific factor, and we conclude that we have a DFMCS with** k **groups;** (2) k_1 **groups,** ($1 \leq k_1 < k$ **) contain specific factors, and** $k_2 = k - k_1$ **groups only contain global factors, then we have a DFMCS with** $k_1 + 1$ **groups; and** (3) **All the groups only contain global factors, then we have the standard DFM.**

Given the estimated factors, groups and loadings $\left(\widehat{\mathbf{P}}_0\widehat{\mathbf{f}}_{0t}, \widehat{\mathbf{P}}_1\widehat{\mathbf{f}}_{1t}, \ldots, \widehat{\mathbf{P}}_k\widehat{\mathbf{f}}_{kt}\right)$, we can compute the residuals from all the factors, or idiosyncratic component, $\widehat{\mathbf{n}}_t = \mathbf{x}_t - \widehat{\mathbf{P}}_0\widehat{\mathbf{f}}_{0t} - \sum_{i=1}^{k} \widehat{\mathbf{P}}_i\widehat{\mathbf{f}}_{it}$ and fit $AR(p)$ models to the idiosyncratic time series by

$$SE_2 = \sum_{t=1}^{T} \left\| \widehat{\mathbf{n}}_t - \sum_{i=1}^{p} \mathbf{\Phi}_i \widehat{\mathbf{n}}_{t-i} \right\|^2 + Tg\left(\boldsymbol{\phi}\right), \tag{5}$$

where $\mathbf{\Phi}_i$ are diagonal matrices and $g\left(\boldsymbol{\phi}\right)$ is a penalty function on the number of fitted AR parameters. The minimization of SE_2 requires choosing the penalty function. In this article we will concentrate in finding the common component of the DFMCS, as the estimation of (5) has well established standard solutions. Of course, it is possible to put both problems together and find the global minimizer of all the parameters, but this requires an iterative procedure which for large time series will be very slow. Therefore, in this article we will concentrate in estimating the common component.

4 Robustification of the procedure

The fitting of the DFMCS is made robust in two ways. First, we clean the observed time series from the effects of (univariate or multivariate) additive outliers and level shifts. Second, we eliminate from the set of observed time series those outlying time series that do not follow the DFMCS.

4.1 Cleaning of additive outliers and level shifts

In practice, the observed time series can be affected by univariate and/or multivariate additive outliers and/or level shifts. The effects of these outliers can be severe, see Tsay, Peña and Pankratz (2000), Galeano, Peña and Tsay (2006), and Galeano and Peña (2019). **For instance, they can produce strong bias in the sample estimates of autocovariance matrices and model** parameters. **In our case** we have to eliminate their effects on the observed time series before fitting the DFMCS for two main reasons: (i) the dependency between two series might be created, or hidden, due to the effect of large outliers; and (ii) we need a robust estimate of the covariance matrix to estimate the factors.

Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{mt})'$ be a vector of time series following the DFMCS model (1). Then, a multivariate additive outlier (MAO) appears at time $t = a$, if instead of $\mathbf{x}_t$ we observe the vector time series $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})'$, given by

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{w}^{(a)} I_t^{(a)}, \tag{6}$$

where $\mathbf{w}^{(a)} = \left(w_1^{(a)}, \ldots, w_m^{(a)}\right)'$ is the size of the MAO and $I_t^{(a)}$ is a dummy variable such that $I_t^{(a)} = 1$, if $t = a$, and $I_t^{(a)} = 0$, if $t \neq a$. Consequently, only the observation at time $t = a$ is affected by the presence of the MAO, **and** the effect in each time series x_{it}, $1 \leq i \leq m$, **depends** on the weights $w_i^{(a)}$. On the other hand, a multivariate level shift (MLS) appears at time $t = l$, if instead of $\mathbf{x}_t$ we observe $\mathbf{y}_t$ given by

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{w}^{(l)} S_t^{(l)}, \tag{7}$$

where $\mathbf{w}^{(l)} = \left(w_1^{(l)}, \ldots, w_m^{(l)}\right)'$ is the size of the MLS and $S_t^{(l)}$ is a step variable such that $S_t^{(l)} = 1$, if $t \geq l$, and $S_t^{(l)} = 0$, if $t < l$. Therefore, all the observations, from time $t = l$ onwards, are affected by the presence of the MLS depending on the weights $w_i^{(l)}$. **In general, the time series can be affected** by several MAOs and MLSs, as follows

$$\mathbf{y}_t = \mathbf{x}_t + \sum_{i=1}^{A} \mathbf{w}^{(a_i)} I_t^{(a_i)} + \sum_{i=1}^{L} \mathbf{w}^{(l_i)} S_t^{(l_i)}, \tag{8}$$

for $t = 1, \ldots, T$, where $\mathbf{w}^{(a_1)}, \ldots, \mathbf{w}^{(a_A)}$ are the sizes of the MAOs at locations $a_1, \ldots, a_A$, and $\mathbf{w}^{(l_1)}, \ldots, \mathbf{w}^{(l_L)}$ are the sizes of the MLSs at locations $l_1, \ldots, l_L$. The method to clean the observed time series, $\mathbf{y}_1, \ldots, \mathbf{y}_T$, defined in (8), from the effects of MAOs and MLSs proceeds following the next four steps:

1. Find the direction that **maximizes** the kurtosis coefficient of the projected time series, as explained in Galeano, Peña and Tsay (2006), and project the observed time series in this direction (see **remark** 3 below). Then, search for additive outliers and level shifts in the projected time series using the algorithm proposed by Chen and Liu (1993) (see remarks 4 and 5 below). If no additive outliers and level shifts are found, go to step 2. Otherwise, let $T^{(l)} = \left\{\widetilde{l}_1, \ldots, \widetilde{l}_{\widetilde{L}}\right\}$ be the vector of sorted locations of the level shifts found and let $\overline{\mathbf{y}}_1, \ldots, \overline{\mathbf{y}}_{\widetilde{L}+1}$ be the sample mean vectors of the time series in the subintervals $\left(1, \widetilde{l}_1 - 1\right), \left(\widetilde{l}_1, \widetilde{l}_2 - 1\right), \ldots, \left(\widetilde{l}_{\widetilde{L}-1}, \widetilde{l}_{\widetilde{L}} - 1\right), \left(\widetilde{l}_{\widetilde{L}}, T\right)$. Then, define $\widetilde{l}_0 = 1$ and $\widetilde{l}_{L+1} = T+1$ and build the series corrected by the **detected level shifts:**

   $$\mathbf{y}_t^* = \mathbf{y}_t - (\overline{\mathbf{y}}_{i+1} - \overline{\mathbf{y}}_1), \text{when } \widetilde{l}_i \leq t < \widetilde{l}_{i+1} - 1, \text{for } i = 0, \ldots, L.$$

   Let $T^{(a)} = \left\{\widetilde{a}_1, \ldots, \widetilde{a}_{\widetilde{A}}\right\}$ be the vector of locations of the additive outliers found. Then, the observations $\mathbf{y}_{\widetilde{a}_1}^*, \ldots, \mathbf{y}_{\widetilde{a}_{\widetilde{A}}}^*$ are replaced with their interpolated values using Exponentially Weighted Moving Average (EWMA) smoothing with moving average window width equal to 4 (see remark 6 below). Repeat this step with the time series $\mathbf{y}_t^*$ until no more additive outliers and level shifts are found. For simplicity, the time series obtained at the end of this step is denoted by $\mathbf{y}_t^*$.

2. Detect and clean MAOs and MLSs as in step 1 with directions that minimize the kurtosis coefficient **of the projected time series**. **The vector of time series obtained** is still denoted by $\mathbf{y}_t^*$.

3. Let n_{dir} be the total number of directions obtained in steps 1 and 2. Then, compute a random direction as explained in Peña and Prieto (2007) (see remark 3 below), and project the time series $\mathbf{y}_t^*$ in this direction. Afterwards, search for additive outliers and level shifts in the projected time series and

correct them as in steps 1 and 2. Then, repeat this step n_{dir} times until n_{dir} random directions have been generated. The time series obtained after this process is still denoted by $\mathbf{y}_t^*$.

4. Search and clean for additive outliers and level shifts in the m univariate **time series in y_t^* obtained at the end of step** 3.

Several remarks on this cleaning procedure are in order:

Remark 3. The methods to obtain the directions in Galeano, Peña and Tsay (2006) and in Peña and Prieto (2007), make use of the covariance and the precision matrices for standardizing the time series to be projected to have zero mean vector and identity covariance matrix. On the one hand, if $T > m$ and the covariance matrix is well conditioned, the standardization can be done even if the number of variables is large, because both algorithms are affine equivariant, so that, they are independent of the standardization used. On the other hand, if $T \leq m$ or if the covariance matrix is ill-conditioned, the projected time series are obtained using the most important **principal components of the observed time series.**

Remark 4. We use a modification of the procedure proposed by Chen and Liu (1993) for detecting additive outliers and level shifts in univariate time series (see Galeano and Peña, 2019) rather than the one proposed in Bianco et al. (2001) that appears to work better, but is slower. As the procedure for detecting univariate outliers is run a large number of times it is important that the algorithm used is fast.

Remark 5. With multiple testing, as in this case, it is necessary to control the detection of false MAOs and MLSs by selecting appropriate critical values. In the Chen and Liu procedure used in steps 1, 2 or 3 for detecting additive outliers and level shifts in projected time series, we use the $(1-\alpha)^{1/(2Tn_{pro})}$ quantile of the standard half-normal distribution, where n_{pro} is the number of projections already generated in the corresponding step. This quantile corresponds to the distribution of the maximum of all the likelihood ratio test statistics for an additive outlier and a level shift in all the existing projections, assuming that these statistics are independent. Similarly, in step 4, we use the $(1-\alpha)^{1/(2Tm)}$ quantile of the standard half-normal distribution, because the Chen and Liu procedure is run independently on every observed time series. Nonetheless, it is important to note that the proportion of false detection of MAOs and MLSs is expected to be smaller than α, because the likelihood ratio test statistics for detecting additive outliers and level shifts are dependent. Consequently, α is an upper bound of the proportion of false MAOs and MLSs detected by the cleaning method. This is confirmed in the simulations shown in Section 5.

Remark 6. We clean MAOs by using a fast interpolator. The optimal interpolator of a linear time series is a linear combinations of the values at both sides of the point to be interpolated with weights given by the inverse autocorrelation function (see Peña and Maravall, 1991). Thus, computing it many times would heavily increase the computational cost of the procedure. Instead, we interpolate by Exponentially Weighted Moving Average (EWMA) smoothing and use a window of four observations before and after the point to be interpolated. In this way, the probability of including outliers in the computation of the interpolator is small. The weights used are $(4/15, 2/15, 1/15, 1/30)$ that result from using the exponentially decreasing weights $(1/2, 1/4, 1/8, 1/16)$ divided by the sum of twice all the them. These weights ensure that the interpolation is unbiased.

Remark 7. In steps 1, 2 and 3, we clean the detected MAOs and MLSs in the m time series components. It is true that not all the time series might be affected by these effects, but the proposed cleaning should not have important effects in good observations. Therefore, although we may loss efficiency, we win robustness.

Remark 8. As noted in Tsay, Peña and Pankratz (2000), we cannot assume that by removing the effects of the outliers in each observed time series they become outlier-free. To illustrate the importance of searching for multivariate outliers in steps 1, 2 and 3, a small simulation is conducted in the supplementary material of the paper.

Finally, Section 5 shows some simulations that illustrate the good performance of the proposed procedure for cleaning the effects of MAOs and MLSs.

4.2 Detecting outlying time series

In practice, there exists the possibility that a few of the observed time series are not generated by a DFMCS. **We will call this series outlying time series**. **To detect them,** we can rely on the information provided by the clustering procedure used to build the groups. If we have outlier series that do not follow a DFM they will have small correlation with the rest of the series and **they will not be included in the initial allocation to the groups**, as explained in Section 3. Let $\mathbf{y}_t^s = (y_{1t}^s, \ldots, y_{st}^s)'$ be the set of s suspicious time series that are omitted in the dendrogram analysis in step 2. The goal is to test whether any of the series in $\mathbf{y}_t^s$ are independent from the set of factors $\mathbf{f}_t = (f_{1t}, \ldots, f_{rt})'$. In other words, we are interested in testing the null hypothesis $H_0 : y_{it}^s$ is independent of $\mathbf{f}_t = (f_{1t}, \ldots, f_{rt})'$ versus the alternative hypothesis $H_1 : y_{it}^s$ is not independent of $\mathbf{f}_t = (f_{1t}, \ldots, f_{rt})'$, for $i = 1, \ldots, s$. For that, we use 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. The statistic, denoted by Λ_p, where p is the order of the maximum autocorrelations used, 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), we have checked that 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, we first compute the log-determinant statistics for each y_{it}^s, for $i = 1, \ldots, s$, and each factor f_{jt}, for $j = 1, \ldots, r$, as well as their associated p-values, denoted by p_{ij}, and then, we label as non-outlying time series to those suspicious time series such that $\min_{1 \le j \le p} p_{ij} < c_\alpha$, 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. Finally, the outlying time series are eliminated from the fitting of the DFMCS. The simulations shown in Section 5 suggest that this is a simple, efficient and powerful method to detect outlying time series.

Remark 9. An alternative way to label non-outlier series is to use the False Discovery Rate (FDR). Following Cuesta-Albertos and Febrero-Bande (2010), we can define a global p-value of the test of independence for the suspicious time series y_{it}^s as follows:

$$p_i = inf \left\{ \frac{r}{j} p_{i(j)} : j = 1, \ldots, r \right\},$$

where $p_{i(1)}, \ldots, p_{i(r)}$ are the sorted p-values for the tests corresponding to the time series y_{it}^s in increasing order. Then, we can apply the Benjamini-Hochberg method to control the FDR and label as non-outlier series those time series such that $p_{(i)} \le \frac{i}{s}\alpha$,where $p_{(1)}, \ldots, p_{(s)}$ are the sorted p-values of the set of suspicious time series. Note that the Benjamini-Hochberg method is valid even if the statistics are dependent, because the distribution of these statistics is χ^2_1 (see, Benjamini and Yekuteli, 2001). The simulation analysis in Section 5 suggests that the results with this method are similar to the one proposed before. Therefore, for simplicity, we use the method described previously.

5 Monte Carlo Results

In this section we present Monte Carlo results to: (1) show the performance of **the proposed procedure** with clean time series; (2) compare its performance to the **one** proposed by Ando and Bai (2017)**; and** (3) show its robustness for outlier detection. As measures of performance we use: (*i*) for finding clusters, the Adjusted Rand Index (Hubert and Arabie, 1985); (*ii*) for loading estimates, the discrepancy measure proposed by Gao and Tsay (2019a,b); (*iii*) for factor estimates,the trace ratio suggested by Stock and Watson

(2002), and used by Bai and Li (2006) and Poncela and Ruiz (2016), among others.

In **the study 12** scenarios (DGP) have been considered. They include different number of groups ($k = 2$ and 3), number of factors ($r = 6$ and 10), structure for the factors (AR(1) and i.i.d.), structure for the noises (i.i.d., heteroscedastic with cross-sectional dependence and serial and cross-sectional dependence) and signal to noise ratio (**noise variances** $\sigma_{\mathbf{n}} = 1$, 2 and 3). Here we present the results for the six scenarios with factors generated by an AR(1) model, **those for i.i.d. factors are in the supplementary material.**

The first three DGPs are similar to the ones proposed by Ando and Bai (2017): $k = 3$ groups, with one global factor, $r_0 = 1$, and thee group-specific factors in each group, $r_1 = r_2 = r_3 = 3$, with $m_1 = m_2 = m_3$ for $m = 300$ and 600, and $T = 200$ and 400. In **this DGP1-DGP3 scenarios**, **the global factor** is a generated by an AR(1) model with autoregressive parameter $\phi = 0.75$ and the global factor noise, $\mathbf{u}_t$, is a vector of uniform $[0, 1]$ variables and the **elements of the** factor-loading matrix $\mathbf{P}_0$ follow an uniform $[-2, 2]$. **Each specific factor of group** i**,** $f_{i,t}$**, is** generated by an AR(1) with $\phi = 0.75$ and the corresponding noises and the factor loadings follow a $N(0, 1)$. Three structures are considered for the idiosyncratic component. In DGP1 the $\mathbf{n}_{i,t}$ are assumed $N(0, \sigma_{\mathbf{n}})$. In Ando and Bai (2017), $\sigma_{\mathbf{n}}$ was fixed to 1, but here, we consider three values $\{1, 2, 3\}$ in order **to evaluate the effect of different signal to noise ratios**. In DGP2 the errors are heteroscedastic **with cross-sectional** dependence, $\mathbf{n}_{i,t} = 0.2\mathbf{n}^1_{i,t} + \delta_t \mathbf{n}^2_{i,t}$, where $\delta_t = 1$, if t is odd, and zero otherwise, and the vectors $\mathbf{n}^1 = (\mathbf{n}^1_{1,t}, \mathbf{n}^1_{2,t}, \ldots, \mathbf{n}^1_{m,t})'$ and $\mathbf{n}^2 = (\mathbf{n}^2_{1,t}, \mathbf{n}^2_{2,t}, \ldots, \mathbf{n}^2_{m,t})'$ are independent and follows a multivariate $\mathrm{N}(\mathbf{0}_m, \mathbf{\Sigma}_{\mathbf{n}})$ with elements $0.3^{|i-j|}\sigma^2_{\mathbf{n}}$, for $i, j = 1, \ldots, m$. In DGP3 the errors have serial and cross-sectional dependence: $\mathbf{n}_{i,t} = 0.2\mathbf{n}_{i,t-1} + e_{i,t}$ where the vector $\boldsymbol{e}_t = (e_{1,t}, e_{2,t}, \ldots, e_{m,t})'$ follows a multivariate $\mathrm{N}(\mathbf{0}_m, \mathbf{\Sigma}_{\mathbf{n}})$ with elements $0.3^{|i-j|}\sigma^2_{\mathbf{n}}$, for $i, j = 1, \ldots, m$.

The next three scenarios are more difficult since first, the number of global factors can be bigger than the number of group-specific factors and, second, **we do not have symmetry** in the number of group-specific factors **and in the size of the clusters**. In all of them $k = 2$, the number of global factors is $r_0 = 2$, the numbers of group-specific factors are $r_1 = 1$ and $r_2 = 3$, respectively, and the number of elements in the groups are $m_1 = m/3$ and $m_2 = 2m/3$. Again m is 300 and 600, and T is 200 and 400. These three scenarios (DGP4, DGP5 and DGP6) assume that the factors follows an AR(1) model with autoregressive parameter $\phi = 0.75$ **and the idiosyncratic components follow** the same structure as DGP1-DGP3, respectively.

5.1 Performance of the procedure

We present first the results for the true number of groups and **factors. Table** 1 **gives the selected number of clusters**. Alonso and Peña (2018) showed that finding groups by applying directly the GCC to the set of series works well and we have compared our proposal, denoted by Proposal in this table, that use the estimated common component of the series to **the method, denoted by TS-GCC, that use the observed time series**. In the first row of each **method we report** the mean of the selected number of clusters and, in the second raw, the number of iterations out of 100 where the true number of clusters was selected. For DGP 1–3 and all values of $\sigma_{\mathbf{n}}$, both methods **select** around three clusters and work well. However, when we increase the noise and decrease the signal to noise ratio, the TS-GCC method fails, whereas the proposed method continues to obtains very good results. **See scenarios 4–6, where the true number of clusters is two.** For those scenarios when $\sigma_{\mathbf{n}} = 3$, the TS-GCC method obtains a mean number of clusters in the range $(2.07, 9.65)$, whereas the **proposed one** obtain values in the range $(1.24, 2.75)$. The results improve for larger T and m. In particular, for $T = 400$ and $m = 600$, **our proposal** selects the correct number of clusters in almost all **the iterations**.

Table 1 around here.

Table 2 shows **the results for the classification of the factors**. NG denotes the means of factors classified as global and NSi with $i = 1, 2, 3$ the mean **of factors** classified as specific. For DGP 1–3, we expect that NG is around one and **the NSi are** around three. For DGP 4–6, we expect that NG is around two, **NS1 around one, and NS2 around three**. **Table** 2 shows that the proposed procedure properly identifies the type of the factors even when the signal is weak, as in the case $\sigma_{\mathbf{n}} = 3$. As expected, the classification **improves** for larger T and m.

Table 2 around here.

5.2 Comparison with other procedures

We now compare our procedure to the one proposed by Ando and Bai (2017). In Table 3, we report the mean of the Adjusted Rand Index (ARI) of three different methods: ABCi corresponds to the first step of Ando and Bai clustering, ABCf to their final solution and Proposal to the **proposed procedure.** The ARI, proposed **by Hubert and Arabie (1985),** compares two different cluster partition, $C = (C_1, \ldots, C_k)$ and

$C' = (C'_1, \ldots, C'_{k'})$ using the following formulas:

$$ARI(C, C') = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k'}\binom{\#(C_i \bigcap C'_j)}{2} - \sum_{i=1}^{k}\binom{\#(C_i)}{2}\sum_{j=1}^{k'}\binom{\#(C'_j)}{2}/\binom{n}{2}}{\left(\sum_{i=1}^{k}\binom{\#(C_i)}{2} + \sum_{j=1}^{k'}\binom{\#(C'_j)}{2}\right)/2 - \sum_{i=1}^{k}\binom{\#(C_i)}{2}\sum_{j=1}^{k'}\binom{\#(C'_j)}{2}/\binom{n}{2}}.$$

The closer this index to one, the higher is the agreement between the two partitions. The results in Table 3 imply that the proposed clustering procedure improves **the ABC one**, that does not have a good performance at scenarios 4–6 for all values of $\sigma_{\mathbf{n}}$.

Table 3 around here.

Table 4 reports the mean of the discrepancy measure proposed by Gao and Tsay (2019a,b) that compares the linear space spanned by the columns of the theoretical loadings, $\mathcal{M}(\mathbf{P})$, with the linear space spanned by the columns of the estimated loadings, $\mathcal{M}\left(\widehat{\mathbf{P}}\right)$, using the following expression

$$D\left(\mathcal{M}(\mathbf{P}), \mathcal{M}\left(\widehat{\mathbf{P}}\right)\right) = \sqrt{1 - \frac{1}{\min(r_1, r_2)} tr\left(\mathbf{H}_P \mathbf{H}_{\widehat{P}}\right)},$$

where $\mathbf{P}$ is the theoretical loading matrix having rank r_1, $\widehat{\mathbf{P}}$ is the estimated loading matrix having rank r_2, and $\mathbf{H}_P = \mathbf{P}\left(\mathbf{P}'\mathbf{P}\right)^{-1}\mathbf{P}'$ and $\mathbf{H}_{\widehat{P}} = \widehat{\mathbf{P}}\left(\widehat{\mathbf{P}}'\widehat{\mathbf{P}}\right)^{-1}\widehat{\mathbf{P}}'$. This measure is equal to 0 if and only if either $\mathcal{M}\left(\widehat{\mathbf{P}}\right) \subset \mathcal{M}(\mathbf{P})$ or $\mathcal{M}(\mathbf{P}) \subset \mathcal{M}\left(\widehat{\mathbf{P}}\right)$, and it is equal to 1 if and only if $\mathcal{M}(\mathbf{P}) \perp \mathcal{M}\left(\widehat{\mathbf{P}}\right)$. For factor estimates evaluation, we use the trace ratio measure suggested by Stock and Watson (2002), and given by:

$$\frac{tr\left(\mathbf{F}'\widehat{\mathbf{F}}\left(\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\right)^{-1}\widehat{\mathbf{F}}'\mathbf{F}\right)}{tr(\mathbf{F}'\mathbf{F})},$$

where F **represents the theoretical factors and $\widehat{\mathbf{F}}$ the estimated ones**. The closer to one the measure, the higher is the canonical correlation between the estimated and the true factors. For reasons of space, we present the results with the trace ratio measure results in the supplementary material. In these tables, ABCf corresponds to final solution of Ando and Bai clustering **and Proposal to the proposed procedure**. In these tables, CF denotes the discrepancy measure (trace ratio measure) between the generated global loading (factors) and the estimated ones with the two procedures and GFi denote these statistics for the specific factors. The results point out that the accuracy of the estimation decreases with the decrease of the signal to noise ratio, due to the increases of the variance of the idiosyncratic component, and **increases with the variance of the factors**. This explains why the specific factors, that have large variance than the global factor, are estimated better. At scenarios DGP 1 – 3, with ten factors, the increase in precision in the estimation of the global factor goes from $20-45\%$ and for the specific factors is around 20%. Similar results are obtained at scenarios DGP 4 – 6, where **our proposal** generally improves the ABC in factors estimation, specially for the global factors. The worst results for **our proposal** are in the difficult cases DGP4 – DGP6 when $\sigma_{\mathbf{n}} = 3$, and T and m are small, but **they are** still clearly better **than those of** ABCf.

Table 4 around here.

5.3 Robustification

To analyze the performance of the time series cleaning method proposed in Section 4.1, sets of 100 time series are generated following one of the six DGPs. For each generated time series, we introduce four MAOs at locations $a = \left[\frac{T}{5}\right]$, $\left[\frac{2T}{5}\right]$, $\left[\frac{3T}{5}\right]$, and $\left[\frac{4T}{5}\right]$, and two MLSs at locations $l = \left[\frac{T}{3}\right]$ and $\left[\frac{2T}{3}\right]$. In all cases, the sizes of MAOs and MLSs is given by $\mathbf{w} = \omega \times diag\left(\widehat{\mathbf{\Gamma}}_{\mathbf{x}}(0)\right)$, where $\widehat{\mathbf{\Gamma}}_{\mathbf{x}}(0)$ is the sample covariance matrix of the series without the MAOs and MLSs, and ω is chosen such that the size of the corresponding AO or LS in the projected time series on the direction of the multivariate outlier is equal to 10 times the standard deviation of the projected time series without the outlier. This might look **as a large size** but **note that as the optimal direction is unknown, the size of the univariate outlier for any other projection** will be much smaller. Indeed, in most of the possible projections, the size of the AO or LS will be **close to** 0.

Then, we apply the time series cleaning method to each generated time series. The critical values used are those described in **remark** 5 for $\alpha = 0.05$. Table 5 shows two measures of the performance of the method **averaged over the** 100 **generated time series:** (1) **the proportion of true detection of MAOs and MLSs; and** (2) **the proportion of false detection of MAOs and MLSs**. **Note, first,** that the procedure does a remarkable job in cleaning the time series from the effects of MAOs and MLSs, as most of the proportions of true detection of MAOs and MLSs are over the 90%. Nevertheless, it seems that the procedure is able to detect slightly more precisely the MLSs. This is reasonable because a MLS may have a larger effect in the observed time series than a MAO. Second, the larger $\sigma_{\mathbf{n}}$, the better the performance. This is because the size of the MAOs and the MLSs generated increases slightly when the variance of the noise increases. Third, as noted in remak 5, the procedure is able to control the detection of false MAOs and MLSs. Anyway, the critical values selected look conservative, as expected.

Finally, we run a final Monte Carlo experiment to analyze the performance of the procedure for detecting outlying time series. **For that, we generate sets of** 100 **time series following one of the six DGPs and** $\left[\frac{m}{10}\right]$ **independent time series following the same model as the factors. Then, we fit a DFMCS to obtain the estimated factors and use the proposed method to find outlying time series. The significance level in the test** is $\alpha = 0.05$. Table 6 shows the proportion of true detection of outlying time series averaged over the 100 generated time series data sets. As it can be seen, the procedure detects very well the time series that are independent of the estimated factors. Indeed, all the proportions are over the 95%. Note that an increment of the number of time series does not improve the **results, but an increase of** the number of observations leads to a **larger frequency of detection**.

Tables 5 – 6 around here.

6 Analysis of Electricity Demands in New England

We analyse a dataset of hourly day-ahead demand for the ISO New England electricity market from January 2004 to December 2016. The data set is available at `www.iso-ne.com`. The New England region is divided in eight load zones: Connecticut (CT), Maine (ME), New Hampshire (NH), Rhode Island (RI), Vermont (VT), Northeastern Massachusetts and Boston (NEMA), Southeastern Massachusetts (SEMA) and Western/Central Massachusetts (WCMA). Each of the time series, $D_{t,i}$, for $1 \leq i \leq 192$ and $1 \leq t \leq 4749$, corresponds to the demand of electricity in one of the eight regions at one of the 24 hours in a given day, that is, we have 192 time series. The number of points in each series is 365 (or 366) days and 13 years, making a total of 4749 data points. As an example, Figure 1 represents the 24 series of demands for Connecticut. The series require a seasonal difference and a logarithm transformation to become stationary (see García-Martos and Conejo, 2013), so that the series analyzed are $X_{t,i} = \nabla_7 \log D_{t,i}$.

Figure 1 around here.

In order to illustrate the relevance of outlier detection we will present the cluster solution with the original time series, $X_{t,i}$ and the cluster solution with the outlier corrected time series, $X^c_{t,i}$. First, **three factors**

are selected for the series $X_{t,i}$ using the two-steps Ahn and Horenstein's procedure with loadings shown in Figure 2. It is clear that hour 02:00 **has** a different behaviour that the **others**: the first factor is essentially measuring the effect of **this second hour** and the second and third factors also have a peak **at this hour**.

Figure 2 around here.

When we apply the cleaning procedure to $X_{t,i}$, in the first step, projections in directions of maximum kurtosis, the procedure identifies 70 MAOs in seven time series projections, an **average of** 10 **outliers** per projected time series. Once the outlier effects are removed, as explained in Section 4, the second step, projections in directions of minimum kurtosis, identifies 23 MAOs in five time series projections, i.e., 4.6 outliers in average per projected time series. After removing these outliers, the third step, random projections, finds 20 MAOs in twelve random projections, i.e., 1.66 outliers in average per projected time series. Finally, the fourth step, univariate search, discovers 59 additive outliers in 47 out of the 192 components, i.e., 1.25 outliers in average in those components with detected outliers. In summary, the procedure detects 113 MAOs and 59 univariate additive outliers and cleans 2.38% of the total number of data points of all the time series.

Table 7 shows the number of outliers detected by day of the week. The largest number of MAOs appears at Sundays and Mondays. Note that 24 out of the 56 MAOs detected on **Sundays correspond** to the daylight saving time days, where the demand is set to zero at the second hour. In the period considered, at the beginning of the sample daylight saving time days occur on April, but after the application of the Energy Policy Act in 2007, the daylight saving time days are on the second Sunday in March. Of course, when we take a week seasonal difference, the outlier appears twice in a seven days interval. These 24 data points represent the 21.18% of the total number of data points cleaned. Moreover, the number of MAOs detected in festive days is 15 **and** 11 of them are on Mondays. Therefore, the number of MAOs on Mondays that are not festive days is 11 that is a **similar number** to the MAOs detected in other days. **The univariate outliers detected look scattered uniformly on the seven days of the week.**

Table 7 around here.

With all the outlier corrected series, the two-steps Ahn and Horenstein's test finds two factors, **that**

explain 77.1% and 8.8% of the total variability, respectively. The loadings of these two factors are shown in Figure 4. The first **one** is almost an average of the series **with** "similar" weights (in the range 0.037−−0.100) to all of them. Therefore, it reproduces the **global** dynamic of the differentiated series. The second **factor gives** negative weights to series of 1th – 11th hours and positive to the 12th – 24th hours. Note that these factors also differentiate across regions, since the loadings for the second (ME) and seventh (SEMA) regions are different to the weights of the remaining regions. Thus, if we do not consider the possible presence of clusters in this data **we may conclude that a DFM with two factors will be appropriate for these data**.

Figure 4 around here.

We search for clusters using the GCC of the series. First, to see the effect of the outliers, we apply this measure to the **original series.** Figure 3(a) **shows the dendrogram obtained.** The demand series of the second hour (blue cluster) **appear at** the top, revealing that these series are far away from the others. **No other groups are found**. However, the **dendrogram for the outlier corrected time series** in Figure 3(b) shows clearly two clusters. **The first one (in red) contains the time series of demand for hours 11:00 – 24:00 and the second one (in green) contains the ones for hour 01:00 – 10:00.** The Silhouette statistics also indicates two clusters. Thus, we conclude that the series corrected from outliers form two groups: the **first one** broadly including daylight hours and the **second one the** night hours.

Figure 3 around here.

When the two-steps Ahn and Horenstein's procedure is applied to series at the two groups, seven factors **are obtained in each cluster**. **These seven factors** explain the 96.8% and 97.6% of the variability of the series at the first and the second group, respectively. As some of these factors may be global and others may be specific, **we compare the two factors found with all the series and the fourteen factors found in the two groups applying the rules presented in Section** 3. The two initial factors are classified as global factors. The first has canonical correlations with the factors in the two groups of 0.984 and 0.967. The second has weaker correlations, 0.673 and 0.799, respectively, but its canonical correlation with the set of all the specific factors is almost one. This implies that its effect is distributed among several factors found in the groups. Now, we apply step 4 of the procedure and obtain the $R_{t,i} = X^c_{t,i} - \widehat{\mathbf{P}}_0\widehat{\mathbf{f}}_{0t}$, where $\widehat{\mathbf{P}}_0$ and $\widehat{\mathbf{f}}_{0t}$

are the estimated loadings and factors, respectively for the two global factors. The Ahn and Horenstein's test applied to $R_{t,i}$, for series at each group, obtains six and five factors for the first and the second clusters, respectively. These factors are clearly specific and orthogonal to the two global factors. Figures 5 and 6 show the loadings for these specific factors. Note that two extreme zones from the geographical point of view, Maine (ME) and Southeastern Massachusetts (SEMA), have the largest effect in almost all the specific factors in both **groups,** whereas for the global factors the situation was just the opposite: these zones have the smallest weights in the two global factors in Figure 4. Regarding the effect of each hour, in the two global factors (see Figure 4) the **first one** gives more relative weight to the afternoon (13:00–19:00), whereas the **second one** differentiate between the night (1:00–7:00) and the rest of the hours, with a peak in 17:00–19:00. A richer picture appears in **the structure of the group factors**. In group one the three first factors give more weight to hours from 11:00–18:00 than those from 19:00–24:00, and factors four and six account for a peak in electricity demand when most people return home, hours 17:00-18:00. In the second group the first **two factors** have opposite peaks of demand around 1:00-2:00 and 7:00-8:00. The other three factors have small variability in the hours but they differentiate strongly among the eight zones.

Figures 5 and 6 around here.

Finally, we perform an out of sample prediction exercise to compare the model fitted with and without cluster effects. Thus we consider: (1) the fitted DFM with two factors (M1); and (2) the fitted DFMCS model with two global factors and eleven specific factors (M2). The first ten years of data were used as training period for the estimation of the models and the last three years (1095 days) as testing period. **Seasonal ARIMA models (as in García-Martos and Conejo, 2013, and Alonso et al, 2016) were fitted to the factors in models M1 and M2.** For simplicity, **in both models the idiosyncratic** terms are assumed to be white noise. We perform a one-day ahead prediction exercise using a rolling windows across the testing period. We calculate the mean absolute prediction **errors** as well as the root mean squared prediction **errors** using the following expressions:

$$MAE = \frac{1}{192}\frac{1}{1095}\sum_{i=1}^{192}\sum_{t=1}^{1095}|X_{t,i} - \widehat{X}_{t,i}|$$

and

$$RMSE = \left(\frac{1}{192}\frac{1}{1095}\sum_{i=1}^{192}\sum_{t=1}^{1095}(X_{t,i} - \widehat{X}_{t,i})^2\right)^{1/2},$$

where $X_{t,i}$ corresponds to the i-th series at day t and $\widehat{X}_{t,i}$ is its **prediction.** The MAE and RMSE for model M1 were 0.0526 and 0.1131, while for model M2 were 0.0487 and 0.1100, respectively. These results point out that introducing the cluster structure and the specific factors produce a moderate improvement on the **overall out of sample prediction**. The MAE and RMSE calculated at the time series of cluster 1 (cluster 2) are 0.0545 and 0.0734 (0.0501 and 0.1522) for model M1, and 0.0502 and 0.0677 (0.0465 and 0.1505) for model M2, an improvement of 7.88% in MAE and 7.86% in RMSE (7.04% and 1.10%), respectively. That is, the improvement is observed in both clusters **and is higher on the first one**.

7 Conclusions

We have presented a robust and efficient way to estimate a Dynamic Factor Model with Cluster Structure. The **proposed procedure** can be used as a starting point for other estimation methods for these models or **applied** by itself for large sets of time series. A Monte Carlo study have shown its good performance with respect to some alternatives. This research can be extended in several ways. It is easy to show that **our procedure** will provide a consistent estimation of the factors under general assumptions (Stock and Watson, 2002, Bai and Ng, 2002) when m and T go to infinity, if we assume a fixed known value of k, and $m_i = \alpha_i m$, with $\alpha_i > 0$, and $\sum_{i=1}^{k} \alpha_i = 1$. However, a general proof of the consistency of the method will require a consistent estimate of the groups labels. This result **could be obtained** by a similar analysis to the one made by Bonhomme and Manresa (2015) and Ando and Bai (2017), but more research is needed to define the set of assumptions required **for the proof.** Also, the model can be extended to a hierarchical structure in the grouping, so that calling m the set of data and g_i the data belonging to group i-th so that $\cup_{i=1}^{k} g_i = m$ and $g_i \cap g_j = 0\ \forall i, j$ we have factors affecting groups g_i but also factors that affect to groups $C_1, \ldots, C_h$, with $h < k$ and where $\cup_{i=1}^{h} C_i = m$ and $C_i \cap C_j = 0\ \forall i, j$. These topics will be the subject of further research.

Acknowledgements

The authors want to acknowledge the very important contribution of Prof. George C. Tiao to Multivariate Time Series and thank to the Editors of this issue for their initiative to produce it and their help in the revision of the paper. We are also very grateful to the two referees for their useful comments and to professor Tomohiro Ando for making available their code

and kindly answering questions regarding its implementation. This research has been supported by Grant ECO2015-66593-P of MINECO/FEDER/UE.

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81, 1203–1227.

Alonso, A. M., Bastos, G. and García-Martos, C. (2016). Electricity price forecasting by averaging dynamic factor models. *Energies*, 9, 600, 21pp.

Alonso, A. M. and Peña, D. (2018). Clustering time series by linear dependency. *Statistics and Computing*, https://doi.org/10.1007/s11222-018-9830-6.

Ando, T. and Bai J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictor and factor structures. *Journal of the American Statistical Association*, 112, 1182–1198.

Bai, J. and Li, K. (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension. *Review of Economics and Statistics*, 98, 298–309.

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

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

Barnichon, R. and Mesters, G. (2018). On the demographic adjustment of unemployment. *Review of Economics and Statistics*, 100(2), 219–231.

Bianco, A. M., Garcia Ben, M., Martinez, E. L. and Yohai, V. J. (2001). Outlier detection in regression models with ARIMA errors using robust estimates. *Journal of Forecasting*, 20, 565–579.

Benjamini, Y. and Yekuteli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165–1188.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83, 1147–1184.

Box, G. E. P. and Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika*, 64(2), 355–365.

Chang, I., Tiao, G. C. and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2), 193-204.

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

Chen, E. Y., Tsay, R. S. and Chen, R. (2019). Constrained factor models for high-dimensional matrix-variate time series. *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2019.1584899.

Cuesta-Albertos, J. A. and Febrero-Bande, M. (2010). A simple multiway ANOVA for functional data. *Test*, 19, 537–557.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55, 251–276.

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

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

Forni, M., Hallin, M. Lippi, M., and Zaffaroni, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, 185, 359–371.

Gao, Z. and Tsay, R. S. (2019a). A Structural–factor approach to modeling high–dimensional time series and space–time data. *Journal of Time Series Analysis*, 40, 343–362.

Gao, Z. and Tsay, R. S. (2019b). Structural–factor modeling of high–dimensional time series: Another look at approximate factor models with diverging eigenvalues. *arXiv:1808.07932*.

Galeano, P., Peña, D. and Tsay, R. S. (2006). Outlier detection in multivariate time series by Projection Pursuit. *Journal of American Statistical Association*, 101, 654–669.

Galeano, P. and Peña, D. (2019). Outlier detection in high dimensional time series. Submitted.

García-Martos, C. and Conejo, A. J. (2013). Price forecasting techniques in power system. In J. Webster (ed.), Wiley Encyclopedia of Electrical and Electronics Engineering. John Wiley & Sons, Inc.

Hallin, M. and Liška, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics*, 163, 29–41.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2, 193–218.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Annals of Statistics*, 40, 694–726.

Lin, C. and Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, 1(1), 42–55.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168, 244–258.

Peña, D. and Box, G. E. P. (1987). Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82, 836–43.

Peña, D. and Maravall, A. (1991) Interpolation, outliers and inverse autocorrelation function. *Communications in Statistics (Theory and Methods)*, 20, 3175–3186.

Peña, D. and Prieto, F. J. (2007). Combining random and specific directions for outlier detection and robust estimation of high-dimensional multivariate data. *Journal of Computational and Graphical Statistics*, 16, 228–254.

Poncela, P. and Ruiz, E. (2016). Small- versus Big-Data factor extraction in dynamic factor models: An empirical assessment. In Eric Hillebrand, Siem Jan Koopman (ed.) *Dynamic Factor Models* (Advances in Econometrics, Volume 35) Emerald Group Publishing Limited, 401– 434.

Robbins, M. W. and Fisher, T. J. (2015). Cross-Correlation matrices for tests of independence and causality between two multivariate time series. *Journal of Business & Economic Statistics*, 33, 459–473.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics*, 20, 53–65.

Su, L., Shi, Z. and Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, 84, 2215–2264.

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

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

Tiao, G. C. and Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76(376), 802–816.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B*, 51(2), 157–195.

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

Wang, P. (2010). Large dimensional factor models with a multi-level factor structure. Working paper, Department of Economics, HKUST.

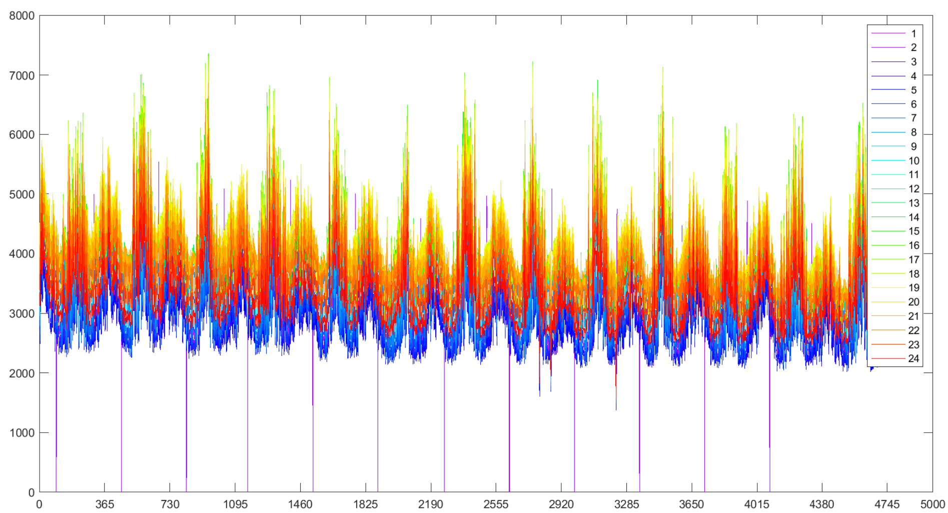

Figure 1: Demands at 01:00 – 24:00 for Connecticut from January 2004 to December 2016.

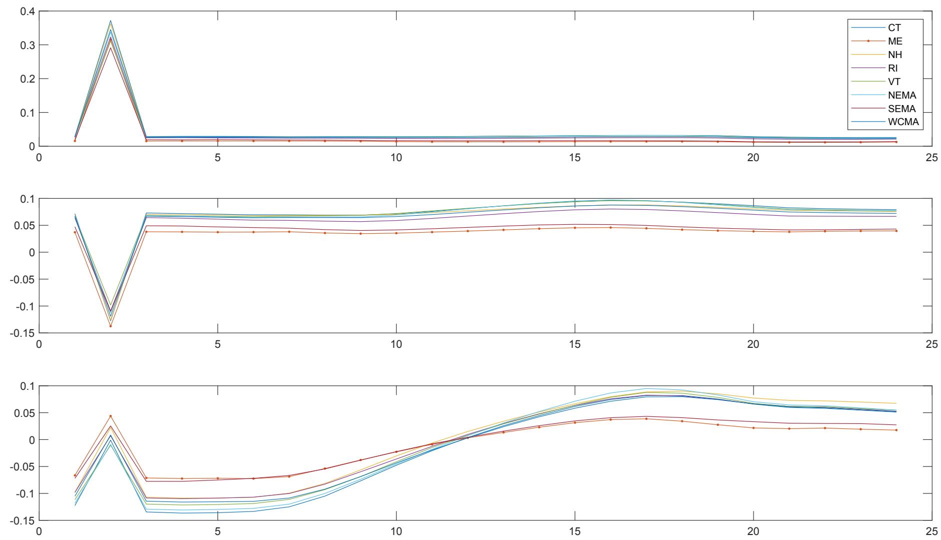

Figure 2: Estimated loadings of the three initial factors using all (log differentiated) original series, $X_{t,i}$.

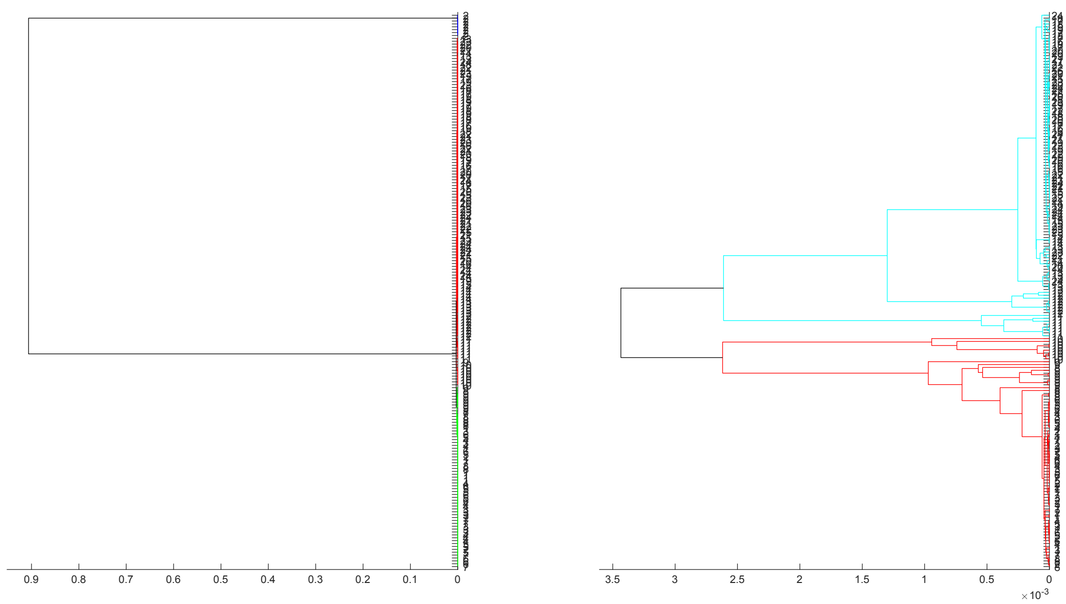

Figure 3: (a) Dendrogram obtained with the original series, $X_{t,i}$. (b) Dendrogram obtained with the outlier corrected series, $X^c_{t,i}$.

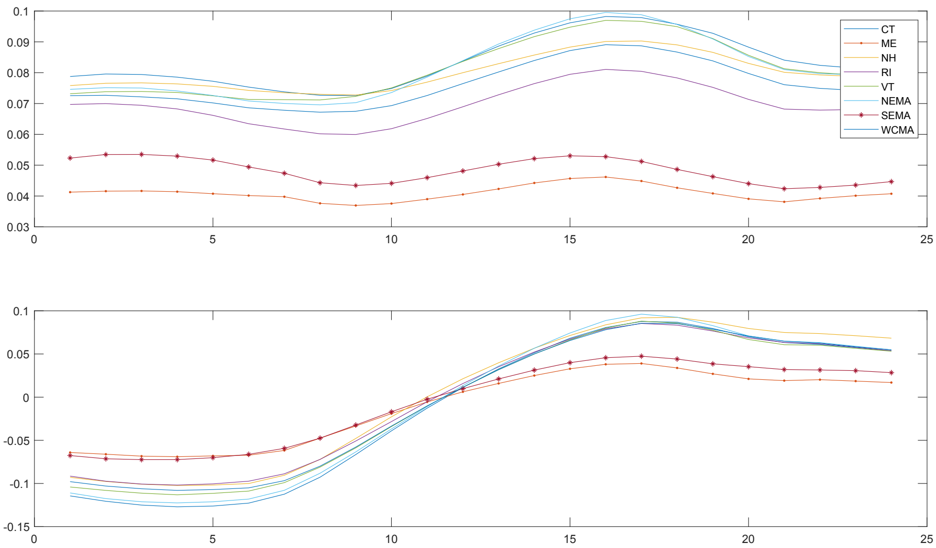

Figure 4: Estimated loadings of the two initial factors using all outlier corrected series.

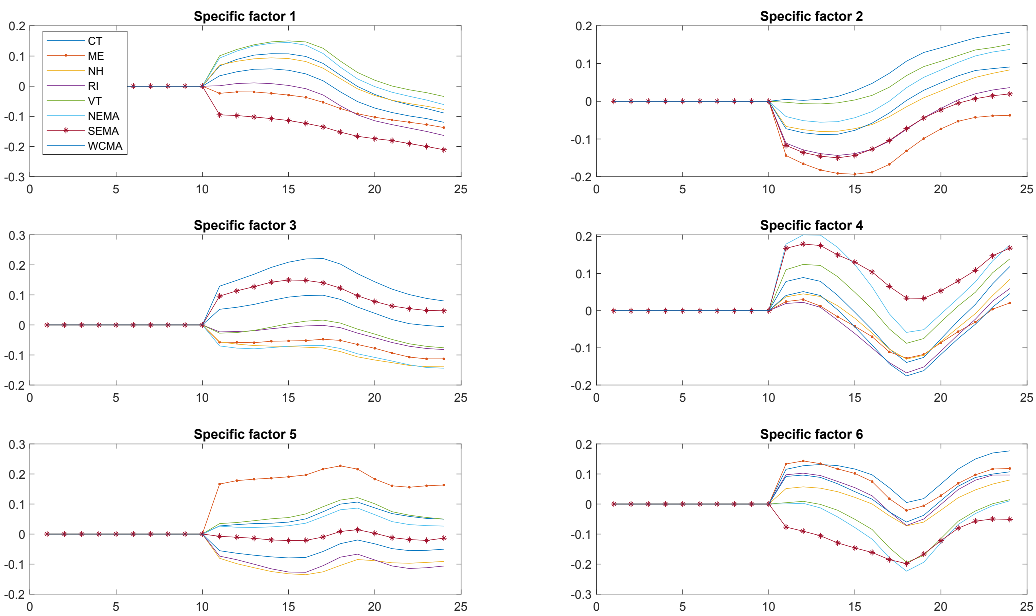

Figure 5: Estimated loadings for six specific factors using the outlier corrected series at group 1 (Hours 11:00–24:00).

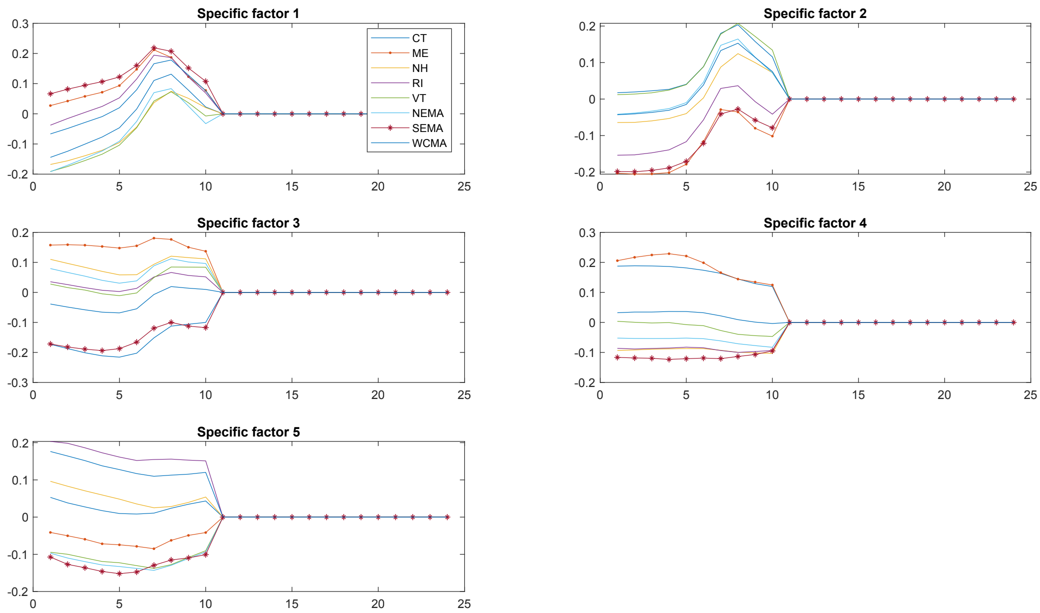

Figure 6: Estimated loadings for five specific factors using the outlier corrected series at group 2 (Hours 1:00–10:00).

			$\sigma_{\mathbf{n}}=1$			$\sigma_{\mathbf{n}}=2$			$\sigma_{\mathbf{n}}=3$		
Method	T	m	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3
TS-GCC	200	300	3.00	3.00	3.00	3.13	3.00	3.00	3.96	3.02	3.94
			100	100	100	94	100	91	47	99	62
Proposal			3.01	3.03	3.02	3.00	3.04	3.00	3.01	3.01	3.00
			99	97	98	100	96	100	99	99	100
TS-GCC	400	300	3.00	3.00	3.00	3.00	3.00	3.03	3.13	3.00	3.30
			100	100	100	100	100	97	97	100	93
Proposal			3.08	3.17	3.08	3.04	3.05	3.03	3.02	3.06	3.02
			94	88	93	97	96	97	99	96	98
TS-GCC	200	600	3.00	3.00	3.00	3.00	3.00	3.03	2.99	3.00	2.86
			100	100	100	91	100	88	51	100	51
Proposal			3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
			100	100	100	100	100	100	100	100	100
TS-GCC	400	600	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	2.99
			100	100	100	100	100	100	100	100	99
Proposal			3.00	3.02	3.00	3.00	3.00	3.00	3.00	3.00	3.00
			100	99	100	100	100	100	100	100	100
Method	T	m	DGP 4	DGP 5	DGP 6	DGP 4	DGP 5	DGP 6	DGP 4	DGP 5	DGP 6
TS-GCC	200	300	4.21	2.60	3.61	5.97	3.29	5.34	5.11	2.39	3.42
			61	84	66	44	66	44	9	37	27
Proposal			2.01	2.00	2.02	2.00	2.02	2.04	1.24	1.25	1.28
			99	100	98	100	99	98	57	58	56
TS-GCC	400	300	2.53	2.24	2.77	7.81	2.57	4.49	9.65	3.22	6.16
			84	92	83	21	85	63	3	80	33
Proposal			2.01	2.01	2.00	2.00	2.00	2.00	2.00	2.03	2.05
			99	99	100	100	100	100	100	99	98
TS-GCC	200	600	3.26	2.46	3.19	2.43	2.68	2.69	2.07	2.75	2.29
			76	89	78	93	87	88	99	85	93
Proposal			2.00	2.00	2.00	2.00	2.00	2.01	2.00	2.01	2.03
			100	100	100	100	100	99	100	99	97
TS-GCC	400	600	2.83	2.27	2.49	3.30	2.46	2.22	4.63	2.49	2.33
			80	91	88	75	90	92	61	87	87
Proposal			2.00	2.00	2.00	2.00	2.00	2.00	2.01	2.00	2.00
			100	100	100	100	100	100	99	100	100

Table 1: Mean of selected number of clusters at scenarios 1–6 (global and group-specific factors are AR(1)), $T = 200$ (400) and $m = 300$ (600). TS-GCC denotes the clustering procedure that computes the GCC among the observed time series; Proposal denotes the clustering procedure that computes the GCC among the estimated common component of the series. First row reports the mean of the selected number of clusters and the second raw reports the number of iterations out of 100 where the true number of clusters was selected.

			$\sigma_{\mathbf{n}} = 1$			$\sigma_{\mathbf{n}} = 2$			$\sigma_{\mathbf{n}} = 3$		
Factor	T	m	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3
NG	200	300	1.00	1.00	1.00	0.99	0.96	0.96	0.95	0.96	0.83
NS1			3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
NS2			3.00	3.00	3.00	3.00	3.00	3.00	2.98	3.00	2.99
NS3			3.00	3.00	3.00	3.00	3.00	3.00	3.00	2.98	3.00
NG	400	300	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.01	0.96
NS1			3.00	3.00	3.00	3.00	3.00	3.00	3.00	2.98	3.00
NS2			3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
NS3			3.00	3.00	3.00	3.00	3.00	2.98	3.00	3.00	3.00
NG	200	600	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.97	0.95
NS1			3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
NS2			3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
NS3			3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
NG	400	600	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NS1			3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
NS2			3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
NS3			3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
Factor	T	m	DGP 4	DGP 5	DGP 6	DGP 4	DGP 5	DGP 6	DGP 4	DGP 5	DGP 6
NG	200	300	1.98	2.01	2.00	1.81	1.89	1.67	1.01	1.01	0.46
NS1			0.98	1.00	0.98	0.98	0.98	0.96	0.94	0.96	0.91
NS2			3.02	2.99	3.01	3.06	3.02	3.02	3.07	3.00	3.19
NG	400	300	2.00	2.00	2.00	2.00	2.00	1.95	1.89	1.89	1.47
NS1			1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.98	0.95
NS2			3.00	3.00	3.00	3.00	3.00	3.02	2.99	3.03	3.09
NG	200	600	2.00	2.00	2.00	1.94	2.00	1.79	1.84	1.79	1.35
NS1			1.00	1.00	1.00	0.98	1.00	0.95	0.97	0.96	0.88
NS2			3.00	3.00	3.00	3.06	3.00	3.12	3.06	3.09	3.23
NG	400	600	2.00	2.00	2.00	2.00	2.00	2.00	1.97	2.00	1.90
NS1			1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	0.99
NS2			3.00	3.00	3.00	3.00	3.00	3.00	3.03	3.00	3.03

Table 2: Results for factor classification for scenarios 1–6 (global and group-specific factors are AR(1)), $T = 200$ (400) and $m = 300$ (600). NG denotes the means of factors classified as global factor and NSi with $i = 1, 2, 3$ denote the mean of factors classified as specific factors.

			$\sigma_{\mathbf{n}} = 1$			$\sigma_{\mathbf{n}} = 2$			$\sigma_{\mathbf{n}} = 3$		
Method	T	m	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3
ABCi	200	300	.044	.042	.043	.044	.048	.040	.037	.039	.039
ABCf			.831	.823	.814	.770	.822	.783	.779	.789	.710
Proposal			.990	.991	.987	.983	.985	.974	.974	.984	.957
ABCi	400	300	.055	.059	.051	.051	.059	.048	.051	.053	.042
ABCf			.807	.847	.803	.794	.827	.808	.803	.824	.740
Proposal			.987	.979	.985	.989	.987	.987	.988	.986	.983
ABCi	200	600	.036	.039	.034	.037	.037	.034	.034	.033	.029
ABCf			.864	.822	.780	.835	.781	.785	.787	.864	.757
Proposal			.991	.994	.990	.985	.990	.976	.976	.987	.962
ABCi	400	600	.049	.052	.049	.049	.047	.044	.046	.045	.043
ABCf			.799	.838	.796	.854	.790	.809	.779	.842	.775
Proposal			.996	.994	.994	.993	.996	.990	.989	.993	.987
Method	T	m	DGP 4	DGP 5	DGP 6	DGP 4	DGP 5	DGP 6	DGP 4	DGP 5	DGP 6
ABCi	200	300	-.004	-.006	-.005	-.004	-.000	-.000	-.003	-.004	-.001
ABCf			.121	.213	.094	.086	.149	.115	.030	.052	.029
Proposal			.929	.955	.917	.877	.921	.824	.460	.504	.403
ABCi	400	300	-.002	-.006	.002	-.002	.000	-.003	-.003	-.002	-.000
ABCf			.181	.209	.152	.096	.146	.052	.036	.140	.056
Proposal			.962	.968	.959	.945	.955	.922	.927	.935	.846
ABCi	200	600	-.002	-.003	-.003	-.003	.001	-.001	.000	-.004	-.002
ABCf			.193	.295	.167	.153	.258	.126	.095	.131	.063
Proposal			.948	.960	.939	.897	.947	.859	.845	.892	.710
ABCi	400	600	-.003	.004	-.002	-.004	-.005	-.005	-.006	-.002	-.006
ABCf			.327	.369	.276	.158	.226	.134	.132	.198	.105
Proposal			.963	.970	.961	.953	.964	.940	.929	.957	.909

Table 3: Clustering performance evaluation using the Adjusted Rand Index at scenarios 1 – 6 (Common and group-specific factors are AR(1)), $T = 200$ (400) and $m = 300$ (600). ABCi denotes the first step of Ando and Bai clustering, ABCf their final solution and Proposal denotes the clustering procedure that computes the GCC among the estimated common part of the series.

				$\sigma_{\mathbf{n}} = 1$			$\sigma_{\mathbf{n}} = 2$			$\sigma_{\mathbf{n}} = 3$		
Method	Factor	T	m	DGP1	DGP2	DGP3	DGP1	DGP2	DGP3	DGP1	DGP2	DGP3
ABCf	CF	200	300	.900	.900	.908	.923	.899	.918	.924	.902	.944
	GF1			.238	.226	.248	.320	.242	.290	.299	.278	.443
	GF2			.246	.211	.277	.307	.247	.282	.275	.267	.398
	GF3			.220	.252	.255	.332	.234	.295	.301	.297	.393
Proposal	CF			.361	.333	.387	.432	.391	.462	.477	.430	.517
	GF1			.148	.117	.171	.205	.159	.235	.250	.197	.289
	GF2			.150	.119	.170	.207	.163	.234	.246	.192	.283
	GF3			.153	.122	.176	.206	.165	.238	.243	.191	.283
ABCf	CF	400	300	.893	.894	.900	.910	.898	.897	.920	.902	.931
	GF1			.250	.205	.248	.300	.246	.252	.262	.242	.332
	GF2			.229	.165	.250	.254	.204	.247	.279	.253	.334
	GF3			.248	.212	.255	.252	.234	.259	.295	.225	.355
Proposal	CF			.323	.310	.336	.368	.335	.391	.400	.355	.434
	GF1			.111	.094	.126	.146	.115	.167	.173	.136	.201
	GF2			.110	.091	.123	.147	.116	.168	.173	.136	.200
	GF3			.114	.095	.128	.145	.116	.167	.172	.135	.200
ABCf	CF	200	600	.899	.901	.918	.904	.903	.914	.918	.902	.941
	GF1			.228	.222	.249	.221	.294	.264	.341	.212	.355
	GF2			.165	.230	.294	.252	.251	.300	.341	.210	.373
	GF3			.203	.218	.297	.239	.282	.328	.261	.216	.345
Proposal	CF			.306	.265	.329	.372	.318	.406	.429	.370	.470
	GF1			.143	.109	.165	.196	.151	.226	.240	.185	.276
	GF2			.144	.111	.164	.200	.153	.229	.236	.184	.271
	GF3			.143	.110	.163	.199	.153	.229	.236	.184	.273
ABCf	CF	400	600	.903	.887	.889	.893	.909	.907	.910	.891	.921
	GF1			.244	.209	.274	.217	.236	.263	.277	.212	.323
	GF2			.246	.189	.246	.196	.265	.256	.264	.193	.292
	GF3			.247	.202	.241	.224	.275	.252	.330	.219	.289
Proposal	CF			.255	.233	.273	.306	.272	.332	.339	.296	.375
	GF1			.104	.082	.118	.137	.107	.159	.164	.126	.191
	GF2			.101	.081	.117	.137	.106	.158	.165	.127	.192
	GF3			.102	.080	.117	.139	.108	.160	.166	.128	.192
Method	Factor	T	m	DGP4	DGP5	DGP6	DGP4	DGP5	DGP6	DGP4	DGP5	DGP6
ABCf	CF	200	300	.871	.855	.877	.883	.861	.874	.894	.898	.894
	GF1			.392	.325	.406	.449	.333	.415	.542	.428	.489
	GF2			.598	.556	.667	.683	.624	.627	.635	.706	.689
Proposal	CF			.303	.265	.320	.361	.310	.408	.433	.355	.490
	GF1			.165	.138	.187	.213	.171	.253	.256	.230	.291
	GF2			.145	.126	.149	.171	.146	.193	.191	.161	.209
ABCf	CF	400	300	.879	.873	.869	.863	.877	.891	.893	.874	.894
	GF1			.310	.355	.314	.398	.298	.372	.722	.299	.403
	GF2			.595	.551	.612	.576	.624	.696	.721	.609	.644
Proposal	CF			.255	.229	.267	.289	.258	.323	.335	.293	.372
	GF1			.141	.116	.144	.162	.131	.178	.192	.154	.225
	GF2			.119	.110	.126	.133	.123	.154	.155	.137	.169
ABCf	CF	200	600	.870	.859	.863	.879	.869	.892	.883	.873	.898
	GF1			.348	.334	.294	.422	.349	.468	.461	.324	.525
	GF2			.562	.501	.578	.606	.548	.631	.652	.618	.677
Proposal	CF			.261	.223	.288	.334	.275	.372	.392	.324	.451
	GF1			.149	.121	.171	.196	.156	.226	.247	.189	.292
	GF2			.119	.100	.132	.153	.121	.172	.181	.143	.207
ABCf	CF	400	600	.862	.848	.867	.879	.871	.883	.882	.876	.885
	GF1			.293	.257	.251	.357	.310	.379	.328	.314	.367
	GF2			.495	.433	.529	.612	.564	.649	.634	.609	.658
Proposal	CF			.206	.181	.226	.255	.214	.285	.296	.247	.335
	GF1			.115	.098	.130	.145	.118	.164	.171	.140	.196
	GF2			.095	.083	.104	.115	.094	.128	.136	.110	.151

Table 4: Loadings estimates evaluation using the discrepancy measure proposed by Gao and Tsay (2019) at scenarios $1-6$ (Common and group-specific factors are AR(1)), $T = 200$ (400) and $m = 300$ (600). ABCf denotes the final solution of Ando and Bai clustering and Proposal denotes the clustering procedure that computes the GCC among the estimated common part of the series.

	Proportion of true detection						Proportion of false detection					
	$\sigma_{\mathbf{n}} = 1$		$\sigma_{\mathbf{n}} = 2$		$\sigma_{\mathbf{n}} = 3$		$\sigma_{\mathbf{n}} = 1$		$\sigma_{\mathbf{n}} = 2$		$\sigma_{\mathbf{n}} = 3$	
DGP	MAO	MLS	MAO	MLS	MAO	MLS	MAO	MLS	MAO	MLS	MAO	MLS
1	.960 (.187)	.950 (.207)	.992 (.075)	.995 (.050)	.975 (.135)	.965 (.177)	.001 (.002)	.002 (.002)	.001 (.002)	.001 (.002)	.001 (.002)	.001 (.002)
2	.962 (.175)	.965 (.177)	.962 (.178)	.955 (.202)	.995 (.035)	1 (0)	.001 (.002)	.001 (.003)	.001 (.002)	.000 (.002)	.002 (.004)	.001 (.002)
3	.977 (.142)	.965 (.177)	.955 (.199)	.955 (.202)	.987 (.102)	.990 (.100)	.001 (.002)	.000 (.001)	.001 (.002)	.001 (.002)	.001 (.002)	.001 (.002)
4	.977 (.118)	.970 (.156)	.995 (.035)	.990 (.070)	.980 (.140)	.980 (.140)	.001 (.002)	.000 (.001)	.001 (.002)	.000 (.001)	.002 (.003)	.001 (.003)
5	.970 (.171)	.970 (.171)	.980 (.126)	.985 (.111)	.990 (.100)	.990 (.100)	.002 (.004)	.001 (.002)	.019 (.015)	.001 (.003)	.034 (.021)	.001 (.003)
6	.972 (.158)	.970 (.171)	.977 (.142)	.980 (.140)	1 (0)	1 (0)	.000 (.001)	.000 (.001)	.001 (.002)	.001 (.002)	.001 (.002)	.001 (.002)
1	.980 (.140)	.980 (.140)	.990 (.100)	.990 (.100)	1 (0)	1 (0)	.014 (.003)	.000 (.000)	.001 (.001)	.000 (.000)	.000 (.001)	.001 (.001)
2	.955 (.199)	.940 (.227)	.995 (.050)	1 (0)	1 (0)	1 (0)	.009 (.006)	.000 (.000)	.011 (.007)	.000 (.001)	.021 (.010)	.001 (.002)
3	.960 (.196)	.960 (.196)	.990 (.100)	.990 (.100)	1 (0)	1 (0)	.012 (.005)	.000 (.001)	.000 (.000)	.000 (.001)	.000 (.001)	.000 (.001)
4	.975 (.148)	.965 (.177)	1 (0)	1 (0)	1 (0)	1 (0)	.012 (.004)	.000 (.000)	.001 (.001)	.000 (.000)	.001 (.002)	.000 (.001)
5	.965 (.177)	.960 (.196)	.995 (.053)	1 (0)	.985 (.111)	.990 (.100)	.010 (.005)	.000 (.000)	.013 (.008)	.000 (.001)	.023 (.011)	.001 (.002)
6	.980 (.116)	.960 (.183)	.990 (.100)	.990 (.100)	.990 (.100)	.990 (.100)	.011 (.005)	.000 (.000)	.001 (.002)	.000 (.000)	.000 (.001)	.000 (.000)
1	.942 (.200)	.970 (.156)	.982 (.124)	.980 (.140)	.997 (.025)	.990 (.070)	.001 (.002)	.001 (.002)	.001 (.002)	.000 (.001)	.001 (.002)	.001 (.003)
2	.895 (.291)	.895 (.303)	.970 (.151)	.965 (.177)	.990 (.100)	.990 (.100)	.001 (.001)	.001 (.003)	.001 (.002)	.001 (.002)	.001 (.002)	.001 (.003)
3	.927 (.239)	.920 (.253)	.960 (.187)	.955 (.202)	.962 (.185)	.955 (.202)	.000 (.001)	.001 (.002)	.001 (.002)	.001 (.002)	.001 (.002)	.001 (.002)
4	.952 (.193)	.950 (.207)	.972 (.146)	.965 (.177)	.955 (.199)	.955 (.202)	.001 (.002)	.000 (.001)	.001 (.002)	.000 (.001)	.001 (.002)	.001 (.002)
5	.960 (.180)	.950 (.207)	.932 (.240)	.915 (.256)	.990 (.100)	.990 (.100)	.003 (.004)	.000 (.001)	.012 (.012)	.001 (.002)	.030 (.021)	.002 (.003)
6	.920 (.253)	.925 (.250)	.980 (.140)	.980 (.140)	.990 (.100)	.990 (.100)	.000 (.001)	.001 (.002)	.001 (.001)	.001 (.002)	.001 (.001)	.000 (.001)
1	.942 (.215)	.945 (.223)	.975 (.144)	.970 (.156)	1 (0)	1 (0)	.000 (.001)	.000 (.000)	.000 (.001)	.000 (.000)	.000 (.001)	.000 (.001)
2	.962 (.185)	.970 (.156)	.990 (.100)	.990 (.100)	1 (0)	1 (0)	.001 (.002)	.000 (.000)	.008 (.006)	.000 (.000)	.017 (.010)	.001 (.001)
3	.967 (.165)	.965 (.162)	.960 (.196)	.960 (.196)	.990 (.100)	.990 (.100)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.001)	.000 (.000)	.000 (.001)
4	.995 (.050)	.990 (.100)	.990 (.100)	.990 (.100)	.980 (.140)	.980 (.140)	.000 (.000)	.000 (.000)	.001 (.003)	.000 (.001)	.001 (.002)	.000 (.000)
5	.965 (.177)	.960 (.183)	.995 (.050)	1 (0)	1 (0)	1 (0)	.002 (.003)	.000 (.000)	.013 (.008)	.000 (.001)	.020 (.010)	.001 (.001)
6	.972 (.158)	.970 (.171)	.995 (.050)	.995 (.050)	.990 (.100)	.990 (.100)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.001 (.001)	.000 (.000)

Table 5: Proportion of true detection of MAOs and MLSs and proportion of false detection of MAOs and MLSs (in parenthesis, standard deviation).

	(T,m)			
DGP	$(200,300)$	$(400,300)$	$(200,600)$	$(400,600)$
1	.992 (.019)	.998 (.005)	.994 (.017)	.995 (.010)
2	.993 (.016)	.997 (.007)	.988 (.022)	.996 (.008)
3	.994 (.019)	.995 (.010)	.990 (.023)	.996 (.009)
4	.995 (.015)	.996 (.008)	.994 (.017)	.995 (.009)
5	.992 (.023)	.996 (.009)	.991 (.021)	.994 (.009)
6	.987 (.024)	.998 (.006)	.991 (.018)	.994 (.009)

Table 6: Proportion of outlier series that are detected by the proposed procedure (in parenthesis, standard deviation).

	Day of the week							
Outlier type	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Total
MAO	22	10	5	3	9	8	56	113
Univariate AO	14	9	7	11	10	7	4	59

Table 7: Number of outliers detected by day of the week.