On the robustness of the general dynamic factor model with infinite-dimensional space: identification, estimation, and forecasting

Carlos Trucíos
São Paulo School of Economics

João H. G. Mazzeu
University of Campinas

Luiz K. Hotta
University of Campinas

Pedro L. Valls Pereira
São Paulo School of Economics

Marc Hallin
ECARES, Université libre de Bruxelles

December 2019

ECARES working paper 2019-32

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

On the robustness of the general dynamic factor model with infinite-dimensional space: identification, estimation, and forecasting

Carlos Trucíos[1], João H. G. Mazzeu[2], Luiz K. Hotta[2],
Pedro L. Valls Pereira[1], and Marc Hallin[3*]

[1]São Paulo School of Economics, FGV, Brazil
[2]Department of Statistics, University of Campinas, Brazil.
[3]ECARES, Université libre de Bruxelles, Belgium

Abstract

General dynamic factor models have demonstrated their capacity to circumvent the curse of dimensionality in time series and have been successfully applied in many economic and financial applications. However, their performance in the presence of outliers has not been analysed yet. In this paper, we study the impact of additive outliers on the identification, estimation and forecasting performance of general dynamic factor models. Based on our findings, we propose robust identification, estimation and forecasting procedures. Our proposal is evaluated via Monte Carlo experiments and in empirical data.

Keywords: Dimension reduction, Forecast, Jumps, Large panels.

JEL classifications:

*The first and fourth authors acknowledge financial support from São Paulo Research Foundation (FAPESP) grants 2016/18599-4 and 2018/03012-3. The second author acknowledges financial support from Coordination for the Improvement of Higher Level Personnel (CAPES) grant 88882.305837/2018-01. The third author acknowledges financial support from São Paulo Research Foundation (FAPESP) grant 2018/04654-9 as well as the Brazilian National Council for Scientific and Technological Development (CNPq) grant 313035/2017-2. All authors acknowledge support of the Centre for Applied Research on Econometrics, Finance and Statistics (CAREFS), Centre of Quantitative Studies in Economics and Finance (CEQEF) and European Centre for Advanced Research in Economics and Statistics (ECARES).

1 Introduction

In recent years, the analysis of high-dimensional time series data has become one of the most active subjects of modern statistical and econometric studies, bringing significant challenges, both from the statistical and the numerical points of view. The most successful procedures so far, particularly for the analysis of economic and financial data, are based on high-dimensional factor models, which allow both the sample size and the dimension of the time series under study to go to infinity.

Above all, factor models do not suffer from the so-called curse of dimensionality as the number of assets grows; see the surveys of Barhoumi et al. (2014) and Bai and Wang (2016) for more details. These models can be used to summarize the information contained in a large number of economic and financial variables into a small number of factors or shocks common to the set of variables. The factors, being estimated from the high dimensional data, can be used for either descriptive or predictive purposes. Applications include: forecasting macroeconomic time series (Stock and Watson, 2002a,b; Forni et al., 2005; Bai and Ng, 2008); excess returns in stock and bond markets (Ludvigson and Ng, 2007, 2009); construction of business cycle indicators and nowcasting (Cristadoro et al., 2005; Giannone et al., 2008; Altissimo et al., 2010); structural macroeconomic analysis and monetary policy (Bernanke and Boivin, 2003; Favero et al., 2005; Stock and Watson, 2005; Eickmeier, 2007; Forni et al., 2009; Forni and Gambetti, 2010); prediction of conditional variance-covariance matrix (Alessi et al., 2009; Aramonte et al., 2013; Trucíos et al., 2019b), to quote only a few.

However, applications are based on a *static* factor-loading scheme (Bai and Ng, 2002; Stock and Watson, 2002a,b), the main advantage of which is to allow for estimation methods based on traditional principal components. Although this approach is easy to implement and widely used, the assumption of a *static* factor-loading scheme, as pointed out by Forni and Lippi (2011) and Forni et al. (2015), is quite restrictive and rules out some very simple and plausible cross-correlation patterns leading to infinite-dimensional factor spaces . To overcome this issue, Forni et al. (2000) introduced the so-called *generalized* or *general dynamic factor model* (GDFM), in which factors (equivalently, common shocks) are loaded through filters

rather than matrices. As shown in Hallin and Lippi (2013), the GDFM arises as a representation result that, besides second-order stationary and the existence of spectral densities, essentially does not place any restriction on the data-generating process, and therefore, encompasses all other high-dimensional factor models considered in the literature. An information criterion for determining the number of common shocks and one-sided filters for a consistent non-parametric estimation of the GDFM are provided by Hallin and Liška (2007) and Forni et al. (2015, 2017), respectively. The Forni et al. (2015, 2017) procedure has been successfully used to forecast inflation and financial returns; see Della Marra (2017), Forni et al. (2018), Giovannelli et al. (2018). It also has been used in the prediction of the conditional variances of financial returns, the extraction of market shocks (Barigozzi and Hallin, 2016, 2017, 2018), and the prediction of conditional variance-covariance matrices (Trucíos et al., 2019b).

Nevertheless, the estimation of the GDFM, including the identification of the number of common shocks, does not take into account the existence of possible outliers. It is known that principal components and likelihood-based estimates are quite sensitive to outliers, especially outliers to the additive type, which are the most common ones in practice. Several methods for outlier detection in time series are available. Most methods, however, apply to univariate time series and little attention has been given to robustness issues in the context of factor model. A method for detecting and estimating the size of outliers in the dynamic factor model is proposed by Baragona et al. (2007), based on linear transformations of the observed data. Kristensen (2014) shows that the performance of predictors in static factor models can be improved by replacing principal components with a robust alternative based on least absolute deviations. A similar idea has been investigated previously by Croux and Exterkate (2011). In their paper a number of alternatives to principal components are examined including LAD-based approaches, but they obtain mixed results as to which approach to be preferred from the point of view of forecasting performance.

We claim that the problem lies in the non-robustness of the estimation *and* prediction procedures as well. As discussed by Baragona et al. (2007), both the tra-

ditional (static) PCA methods[1] and the more general dynamic PCA methods[2] yield biased estimates in the presence of outliers. Given the good forecasting performance of the GDFM model evidenced in the literature, we propose a robust version of the criterion introduced by Hallin and Liška (2007) to estimate the number of common shocks and a robust version of the estimation procedure of Forni et al. (2015, 2017) in order to obtain robust estimates of common shocks, impulse-response functions, and forecasts.

This paper contributes to the literature in three ways. First, we show through Monte Carlo experiments that the identification, estimation, and forecasting of the GDFM are strongly affected by the presence of outliers. In particular, the criterion of Hallin and Liška tends to overestimate the number of common shocks. These results are in agreement, for instance, with those obtained by Kristensen (2014), who finds that the commonly used information criteria of Bai and Ng (2002) (estimating the number of static factors) is severely inflated by outliers. Second, we propose robust procedures for the identification, estimation, and prediction of the GDFM. Third, an empirical application indicates that the best performance of our robust prediction procedure, relative to the non-robust procedure, is achieved during crisis periods, i.e, in the presence of outliers.

The structure of the paper is as follows. In Section 2, we present the GDFM model with the estimation and prediction procedures and the identification criterion for the number of common shocks. Section 3 presents Monte Carlo experiments evaluating the performance of the GDFM in the presence of additive outliers. Because the results indicate that the existing procedures are highly non-robust to additive outliers, Section 4 presents a robust alternative to circumvent the problem and simulations showing that the suggested alternative presents a substantially better performance. In Section 5 an empirical application is conducted to assess the pseudo real-time forecasting performance of our robust procedure. We employ the same large monthly dataset of macroeconomic and financial time series for the US economy used in Forni et al. (2018). Concluding remarks are presented in Section 6.

[1] Based on the contemporary covariance matrix of the observations

[2] Based on the spectral density matrix of the observations

2 The general dynamic factor model (GDFM)

2.1 General dynamic factor model with infinite-dimensional factor space

Let $\{\mathbf{X}_t = (X_{1t}\ X_{2t}\ \dots)', \quad t \in \mathbb{Z}\}$, be a double-indexed zero-mean second-order stationary stochastic process, where the first index stands for the series and t for time. The GDFM introduced in Forni et al. (2000) is based on a dynamic factor representation of the form

$$\begin{aligned} X_{it} &= \chi_{it} + \xi_{it} \\ &= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + ... + b_{iq}(L)u_{qt} + \xi_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}, \end{aligned} \tag{1}$$

where L stands for the lag operator and the unobservable χ_{it}, ξ_{it}, and u_{jt} for the common components, idiosyncratic components, and common shocks, respectively. We assume the following.

1. The vector process $\{\mathbf{u}_t = (u_{1t}\ u_{2t}\ ...\ u_{qt})', \quad t \in \mathbb{Z}\}$ is an unobservable q-dimensional orthonormal white noise process: the common shocks.

2. The idiosyncratic process $\{\boldsymbol{\xi}_t = (\xi_{1t}\ \xi_{2t}\ \dots)', \quad t \in \mathbb{Z}\}$ is zero-mean second-order stationary and, additionally, ξ_{kt} and $u_{k't'}$ are mutually orthogonal for any k, k', t and t'. Moreover, it is assumed that $\{\boldsymbol{\xi}_t\}$ is *weakly* cross-sectionally correlated, so that the comovements of the X_{it}'s are mainly accounted for by the q common shocks.

3. The filters $b_{ik}(L)$ are one-sided polynomials with square-summable coefficients for any $i = 1, 2, ...$ and any $k = 1, ..., q$.

4. The number q of common shocks is the smallest integer for which 1-3 hold.

The assumptions above define the GDFM, of which all other factor models in the econometric time series literature are particular cases; see Forni et al. (2015, 2017).

An additional assumption which is adopted by many authors is that the common components span a finite-dimensional space (Bai and Ng, 2002; Stock and Watson, 2002b; Forni et al., 2005, 2009; Alessi et al., 2010; Aramonte et al., 2013). Under

this assumption, we can rewrite the decomposition (1) in the *static* form

$$X_{it} = \lambda_{i1} F_{1t} + ... + \lambda_{ir} F_{rt} + \xi_{it}, \tag{2}$$

where the *static* factors $F_{1t}, ..., F_{rt}$ and the loadings $\lambda_{i1}, ..., \lambda_{ir}$, $i = 1, 2, ...$, can be estimated consistently using the first r standard principal components, $r \geq q$. However, as pointed out by Forni et al. (2000), Forni and Lippi (2011), Forni et al. (2015, 2017) and Forni et al. (2018), representation (2) rules out simple and quite plausible cases as

$$X_{it} = a_i \left(1 - d_i L\right)^{-1} u_t + \xi_{it}, \tag{3}$$

where the coefficients d_i are drawn, e.g., from a uniform distribution over the stationary region. In this case, the space spanned by the common components in model (3) is no longer finite-dimensional.

Forni et al. (2000) and Forni et al. (2004) propose to use Brillinger's (1981) concept of dynamic principal components, which is based on the spectral density of the X's, to estimate model (1). While this estimator does not require a finite-dimensional assumption on the space spanned by the common components, it involves the application of two-sided filters, which lead to poor forecasting performances.

Recently, Forni et al. (2015, 2017) showed how to obtain one-sided filters without assuming a finite-dimensional factor space and how to construct estimators for (1) by imposing the mild additional assumption that the common components have rational spectral density (Assumption A.3 of Forni et al. (2015)), that is, each filter $b_{if}(L)$ in (1) is a ratio of polynomials in L with unspecified, yet finite orders. Thus, they assume that the common component in (1) can be rewritten as

$$\chi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + ... + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z} \tag{4}$$

where

$$c_{if}(L) = c_{if,0} + c_{if,1} L + ... + c_{if,S_1} L^{S_1}, d_{if}(L) = 1 + d_{if,1} L + ... + d_{if,S_2} L^{S_2}; \quad f = 1, 2, ..., q,$$

the roots of each polynomial are outside the unit circle, and there are no common roots among $c_{if}(L)$ and $d_{if}(L)$ for any i and $f = 1, 2, ..., q$.

Then, under this mild assumption of a rational spectrum, Forni et al. (2015, 2017) derive a static factor model representation (2) for a block-diagonal autoregressive filtering of the observed process $\mathbf{X}$ satisfying (1). From Assumption A.3 of Forni et al. (2015), the $(q+1)$-dimensional vector $\boldsymbol{\chi}_t^{(k)} = (\chi_{(k-1)(q+1)+1,t} \ \cdots \chi_{k(q+1),t})'$ has the autoregressive representation

$$\mathbf{A}^{(k)}(L)\boldsymbol{\chi}_t^{(k)} = \mathbf{R}^{(k)}\mathbf{u}_t^{(k)}, \tag{5}$$

where $\mathbf{R}^{(k)}$ is $(q+1) \times q$, $\mathbf{A}^{(k)}(L)$ is a $(q+1) \times (q+1)$ polynomial matrix with finite degree, and $\mathbf{u}_t^{(k)} = (u_{1t} \ \ldots u_{qt})'$, $k \in \mathbb{N}$. Moreover, the filters $\mathbf{A}^{(k)}(L)$ are one-sided and fundamental, i.e. $\det(\mathbf{A}^{(k)}(z)) \neq 0$ for $z \in \mathbb{C}$ such that $|z| \leq 1$. That assumption, actually, is very mild, as it holds generically[3] under (4).

2.2 Estimation and forecasting

In practice, we have an observed $(n \times T)$-dimensional panel of time series. Therefore, assume, without loss of generality[4], that n factorizes into $n = m(q+1)$ for some $m \in \mathbb{N}$, and partition the vector $\boldsymbol{\chi}_{nt}$ as $\boldsymbol{\chi}_{nt} = (\boldsymbol{\chi}_t^{(1)} \ \boldsymbol{\chi}_t^{(2)} \cdots \ \boldsymbol{\chi}_t^{(m)})'$. In view of (5), the n-dimensional vector $\boldsymbol{\chi}_{nt}$ has a block-diagonal VAR representation of the form

$$\mathbf{A}_n(L)\boldsymbol{\chi}_{nt} = \begin{bmatrix} \mathbf{A}^{(1)}(L) & 0 & \ldots & 0 \\ 0 & \mathbf{A}^{(2)}(L) & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \ldots & \mathbf{A}^{(m)}(L) \end{bmatrix} \boldsymbol{\chi}_{nt} = \mathbf{R}_n\mathbf{u}_t = \begin{bmatrix} \mathbf{R}^{(1)} \\ \mathbf{R}^{(2)} \\ \vdots \\ \mathbf{R}^{(m)} \end{bmatrix} \mathbf{u}_t, \tag{6}$$

where $\mathbf{R}_n$ is an $n \times q$ matrix of *static* loadings.

From (6), letting $\boldsymbol{X}_{nt} = (\mathbf{X}_t^{(1)} \ldots \mathbf{X}_t^{(m)})'$ with $\mathbf{X}_t^{(k)} = (X_{(k-1)(q+1)+1,t} \ \ldots \ X_{k(q+1),t})'$, and filtering both sides of (1) by $\mathbf{A}_n(L)$, we obtain

$$\mathbf{Y}_{nt} = \mathbf{A}(L)\mathbf{X}_{nt} = \mathbf{R}_n\mathbf{u}_t + \mathbf{A}(L)\boldsymbol{\xi}_{nt}, \tag{7}$$

which has a factor model representation with finite-dimensional common space.

[3]Precisely, it holds for all values of the parameters $c_{if,j}$ and $d_{if,k}$, except for a subset with Lebesgue measure zero.

[4]This is taken care of at the estimation stage, by generating random permutations of the cross-section.

The common components $\boldsymbol{\chi}_{nt}$ in (6) can be recovered by inversion of the polynomial $\mathbf{A}_n^{(k)}(L)$:

$$\boldsymbol{\chi}_{nt} = [\mathbf{A}_n(L)]^{-1}\mathbf{R}_n\mathbf{u}_t = \mathbf{B}_n(L)\mathbf{u}_t = \mathbf{B}_{n0}\mathbf{u}_t + \mathbf{B}_{n1}\mathbf{u}_{t-1} + ... \tag{8}$$

where the common shocks $\mathbf{u}_t$ in (8) are the same as in (7) and (1), and

$$[\mathbf{A}_n(L)]^{-1} = \begin{bmatrix} [\mathbf{A}^{(1)}(L)]^{-1} & 0 & \dots & 0 \\ 0 & [\mathbf{A}^{(2)}(L)]^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & [\mathbf{A}^{(m)}(L)]^{-1} \end{bmatrix}. \tag{9}$$

The main advantage of representation (7) over (1) is that, after a simple filtering involving $(q+1)$-dimensional VARs, the GDFM can be estimated using one-sided filters.

As mentioned in Forni et al. (2017, 2018) and Barigozzi et al. (2018), the estimation of $\mathbf{A}_n(L)$ depends on the arbitrary cross-sectional ordering of the panel. Based on a Rao-Blackwell argument, Forni et al. (2017) propose to average the estimates over the $n!/m!\,[(q+1)!]^m$ possible $(q+1)$-tuples of n cross-sectional items or, equivalently, over the $n!$ possible permutations of the cross-section. Clearly, averaging over all $n!$ permutations or a number $n!/m!\,[(q+1)!]^m$ of $(q+1)$-tuples is unfeasible even for moderate n. Fortunately, simulations reported in Forni et al. (2017) reveal that the stabilization of the estimates is very fast, so that few permutations are sufficient to obtain the same performance as if the $n!$ possible ones were performed.

The estimation procedure is described as follows.

- **Step 1:** Determine the number q of common shocks in (1) applying, for instance, the Hallin and Liška criterion.

- **Step 2:** For a given permutation of $\mathbf{X}_{nt}$, start with a consistent estimator

$$\hat{\boldsymbol{\Sigma}}^X(\theta) = \frac{1}{2\pi}\sum_{k=-M_T}^{M_T} e^{-ik\theta}K\left(\frac{k}{B_T}\right)\hat{\boldsymbol{\Gamma}}_k^X$$

of the spectral density matrix of $\mathbf{X}_{nt}$, where $\theta \in [-\pi, \pi]$, $K(\cdot)$ is a kernel function, M_T is a truncation parameter, B_T is the bandwidth parameter, and $\hat{\boldsymbol{\Gamma}}_k^X$

is the estimated covariance matrix between $\mathbf{X}_{nt}$ and $\mathbf{X}_{n,t-k}$. In this paper, we use the triangular kernel with $B_T = \sqrt{T}$.

- **Step 3:** Using the first q dynamic principal components of $\hat{\mathbf{\Sigma}}^X(\theta)$, estimate, as in Forni et al. (2000), the spectral density matrix of the common components $\hat{\mathbf{\Sigma}}^\chi(\theta)$ and, by classical inverse Fourier transform, the corresponding autocovariance matrices, $\hat{\mathbf{\Gamma}}_k^\chi$.

- **Step 4:** For each of the m $(q+1)\times(q+1)$ diagonal blocks of $\hat{\mathbf{\Gamma}}_k^\chi$, estimate (after AIC or BIC order identification) the coefficients of $\mathbf{A}^{(i)}(L)$ via the Yule-Walker method for $i = 1, ..., m$ in (6). This yields an estimation of the block-diagonal operator $\mathbf{A}_n(L)$ and, therefore, $\hat{\mathbf{Y}}_{nt} = \hat{\mathbf{A}}_n(L)\mathbf{X}_{nt}$ is an estimate of the left-hand side of (7).

- **Step 5:** As $\hat{\mathbf{Y}}_{nt}$ (up to estimation errors) admits a static factor model representation, estimates $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{R}}$ of $\mathbf{u}_t$ and $\mathbf{R}$, respectively, can be obtained from the first q standard principal components of $\hat{\mathbf{Y}}_t$: see Stock and Watson (2002a) and Stock and Watson (2002b). Inverting the estimated polynomial matrix $\hat{\mathbf{A}}_n(L)$ yields the estimated impulse-response matrix

$$\hat{\mathbf{B}}_n(L) = [\hat{\mathbf{A}}_n(L)]^{-1}\hat{\mathbf{R}}_n.$$

- **Step 6:** Use $\hat{\mathbf{B}}_n(L)$ to obtain the estimated common factors:

$$\hat{\boldsymbol{\chi}}_{nt} = [\hat{\mathbf{A}}_n(L)]^{-1}\hat{\mathbf{R}}_n\hat{\mathbf{u}}_t = \hat{\mathbf{B}}_n(L)\mathbf{u}_t = \hat{\mathbf{B}}_{n0}\hat{\mathbf{u}}_t + \hat{\mathbf{B}}_{n1}\hat{\mathbf{u}}_{t-1} + ... + \hat{\mathbf{B}}_{ns}\hat{\mathbf{u}}_{t-s},$$

where s is a truncation threshold, large enough.

- **Step 7:** Repeat steps 2 - 6 for B different permutations. The estimated impulse-response matrix $\widehat{\widehat{\mathbf{B}}}_{\mathbf{n}}(\mathbf{L})$ and the estimated common components $\widehat{\widehat{\boldsymbol{\chi}}}_{nt}$, then, are obtained by averaging the B matrices $\hat{\mathbf{B}}_n(L)$ and $\hat{\boldsymbol{\chi}}_{nt}$. Note that before averaging, each $\hat{\mathbf{B}}_n(L)$ and $\hat{\boldsymbol{\chi}}_{nt}$, for $b = 1, ..., B$, must be rearranged in the original order of the panel. The averaging of each $\hat{\mathbf{B}}_n(L)$ also requires their identification and, as in Forni et al. (2017), we impose a Cholesky identification constraint on the first q variables.

To obtain the h-step-ahead common component forecast, an additional step should be added. For each permutation, the prediction equation for the common components at horizon h takes the form

$$\hat{\boldsymbol{\chi}}_{n,t+h|t} = \hat{\mathbf{B}}_{nh}\hat{\mathbf{u}}_t + \hat{\mathbf{B}}_{n,h+1}\hat{\mathbf{u}}_{t-1} + ...$$

Then, the h-step-ahead common component forecast $\widehat{\widehat{\boldsymbol{\chi}}}_{n,t+h|t}$ is obtained by averaging the B vectors $\hat{\boldsymbol{\chi}}_{n,t+h|t}$. Finally, putting $\widehat{\widehat{\boldsymbol{\xi}}}_{nt} = \mathbf{X}_{nt} - \widehat{\widehat{\boldsymbol{\chi}}}_{nt}$, each of the idiosyncratic variables $\widehat{\widehat{\xi}}_{it}$ can be predicted using univariate methods, yielding the h-step ahead predictor

$$\widehat{\widehat{X}}_{i,t+h|t} = \widehat{\widehat{\chi}}_{i,t+h|t} + \widehat{\widehat{\xi}}_{i,t+h|t}.$$

2.3 Determining the number of common shocks

A crucial step in the analysis of dynamic factor models is the identification of the number of common shocks. This number, beyond the economic interpretation, also plays an important role in estimation and forecasting; see, for instance, Forni et al. (2009), Aramonte et al. (2013), Della Marra (2017), Barigozzi et al. (2018) and Forni et al. (2018).

A formal information criterion to determine the number of common shocks was proposed by Hallin and Liška (2007) and achieves good performance, even in small samples. This procedure is based on the eigen-decomposition of the spectral density matrix and does not assume that the space spanned by the common components is finite.

For given n, T and a positive constant c, the criterion selects the number of common shocks that minimizes the contribution of the idiosyncratic components

$$\hat{q}_{n,T;c} = \arg\min_{0\leq k\leq q_{max}} IC_{n,T;c}(k), \quad 0 \leq k \leq q_{max}, \tag{10}$$

where q_{max} is a predefined upper bound and $IC_{n,T;c}(k)$ is a information criterion associated with the spectral density matrix $\boldsymbol{\Sigma}^X(\theta)$. In this paper, we use the logarithmic information criterion as in Forni et al. (2017), which is given by

$$IC_{n,T;c}(k) = \log\left(\frac{1}{n}\sum_{i=k+1}^{n}\frac{1}{2M_T+1}\sum_{l=M_t}^{M_T}\lambda_i^T(\theta_l)\right) + c\ k\ p(n,T), \tag{11}$$

where $p(n,T)$ is a penalty function such that $\min(n, M_T^2, M_T^{-1/2}T^{1/2})p(n,T) \to \infty$ and $p(n,T) \to 0$ when n, $T \to \infty$, $\theta_l = \pi l/(M_T + 1/2)$, $\lambda_i^T(\theta_l)$ is the i-th eigenvalue of the spectral density matrix $\mathbf{\Sigma}^X(\theta)$; c is an arbitrary positive real value and the estimator of $\mathbf{\Sigma}^X(\theta)$ is defined in Step 2; a maximal value q_{max} of q also has to be chosen. Hallin and Liška (2007) prove that $\hat{q}_{n,T;c}$ is consistent for any $c > 0$ as n and T tend to infinity. An optimal value of c, denoted by c^*, is selected as follows. Setting an upper bound C for the constant c, consider J subsamples of size (n_j, T_j), with $0 < n_1 < ... < n_J = n$ and $0 < T_1 \leq ... \leq T_J = T$, $j = 1, ..., J$. Although we can take $T_j < T_{j+1}$, choosing $T_j = T$ for all j is recommended and it is used in this paper. For each $c > 0$ and each subsample, the criterion yields a number $\hat{q}_{n_j,T_j;c}$ of common shocks. For each $c > 0$, the variability among the J values of $\hat{q}_{n_j,T_j;c}$ for $j = 1, ..., J$, is captured by

$$S_c = \frac{1}{J}\sum_{j=1}^{J}\left[\hat{q}_{n_j,T_j;c} - \frac{1}{J}\sum_{j=1}^{J}\hat{q}_{n_j,T_j;c}\right]^2 .$$

To select c^* we look for intervals of c over which S_c=0. Hereafter, such intervals are called *stability intervals*. Stability intervals are such that $\hat{q}_{n_j,T_j;c}$=$\hat{q}_{n,T;c}$ is constant for c ranging over such intervals. Starting in the neighbourhood of c=0 (no penalty at all), a first stability interval $(0, c_1^+)$ corresponds to $\hat{q}_{n,T;c}$=q_{max}. Disregarding this q_{max} which clearly is not a consistent solution, choose c^* as any point in the next stability interval (c_2^-, c_2^+). The selected number of factors is then $\hat{q}_{n,T}$=$q_{n,T;c^*}$.[5]

Summing up, in practice the identification method is performed as follows:

- choose M_T and a maximum number q_{max} of common shocks; we chose $M_T = 0.75\sqrt{T}$ and q_{max}=6;
- set a grid of values for the constant $c \in C \subset [C^-, C^+] \subset \mathbb{R}^+$; we chose $c = 0.01, 0.02, ..., 3.00$;
- for each value of c in that grid, (a) randomly choose subsamples of increasing dimension $0 < n_1 < ... < n_J = n$; we chose $n_j = n_1 + \lfloor (n - n_1)/10 \rfloor$, $j = 2, 3, ..., J$, with n_1 not too small[6]; (b) solve (10) to find $\hat{q}_{c;n_j,T}$

[5] See Hallin and Liška (2007) for an extensive explanation of the role of the constant c and other parameters.

[6] We set this value to $n_1 = 3n/4$.

for $j = 1, ..., J$; (c) using the sequence $\hat{q}_{n_j,T;c}$, $j = 1, ..., J$, compute the variance of S_c;

- identify q as $\hat{q} = \hat{q}_{n,T;c^*}$, where c^* belongs to the second stability interval of c.

Hallin and Liška (2007) in their Monte Carlo experiments use the following three penalty functions:

$$p_1(n,T) = (M_T^{1/2}T^{-1/2} + M_T^{-2} + n^{-1}) \times \log(\min[T^{1/2}M_T^{-1/2}; M_T^2; n]);$$
$$p_2(n,T) = (\min[T^{1/2}M_T^{-1/2}; M_T^2; n])^{-1/2};$$
$$p_3(n,T) = (\min[T^{1/2}M_T^{-1/2}; M_T^2; n])^{-1} \times \log(\min[T^{1/2}M_T^{-1/2}; M_T^2; n]).$$

In our estimations we used $p_1(n,T)$.

3 Monte Carlo experiments

In order to evaluate the performance of the GDFM in the presence of additive outliers we carry out Monte Carlo experiments to evaluate their effects on the number of common shocks identified by the Hallin and Liška (2007) criterion and on the ensuing estimation of the common shocks and impulse response functions using the procedure of Forni et al. (2015, 2017). Results are presented in Sections 3.1 and 3.2, respectively. Finally, in Section 3.3, we assess the impact of outliers on the one-step-ahead forecast procedure described in Forni et al. (2015, 2017) and Forni et al. (2018). We consider the same data-generating process (admitting no static factor representation) as in Forni et al. (2017), namely,

$$X_{it} = a_{i1}(1 - \alpha_{i1}L)^{-1}u_{1t} + a_{i2}(1 - \alpha_{i2}L)^{-1}u_{2t} + \xi_{it}, \tag{12}$$

where u_{jt} and ξ_{it} ($j = 1, 2$, $i = 1, ...n$, $t = 1, ..., T$) are generated as *i.i.d.* standard Gaussian variables; a_{ij} as *i.i.d.* uniform variables on the interval $[-1, 1]$; and α_{ij} as i.i.d. uniform variables on the interval $[-0.8, 0.8]$.

The (n,T)-dimensional panel is contaminated with two consecutive outliers either in the middle or at the end of the sample period, in 5%, 10% and 15% of the series. In all cases, outliers of size 10 times the standard deviation of the univariate uncontaminated processes were considered.

3.1 Estimation of the number of common shocks

Table 1 presents the percentage of times the Hallin and Liška criterion identified the correct number of common shocks from series contaminated by two consecutive outliers in the middle or at the end of the series. For the sake of comparison, we also include the results for uncontaminated series. We consider panel dimensions $n = 60, 120$ and 240, sample size T=120 and 500 replications.

Results show that the Hallin and Liška criterion never under identifies the right number of common shocks. When there are no outliers, the identification was incorrect in only 5 replications, all for panel dimension 60, which corresponds to 1% of the cases only for this panel dimension. However, in most cases, only two consecutive outliers in a few series are sufficient to produce an overestimation of the number of common shocks. The overestimation is larger when the outliers occur in the middle of the series and also when the percentage of outliers and sample size increase. When only 5% of the series are contaminated by outliers at the end of the series, no big differences with the uncontaminated case are observed. For outliers in the middle of the series, when we increase the panel dimension to 120, the overestimation increases to 28.2%. When 10% of the series are contaminated, for panel dimension 120, we already have overestimation in 96.4% of the cases when the outliers are at the end of the series, and 99.4% when the outliers are in the middle of the series. When we have contamination in 15% of the series, the overestimation percentage is already as high as 97.2%(97.4%) when the panel dimension is equal to 60 and the outliers occur at the end (in the middle) of the series. It is clear that the Hallin and Liška (2007) criterion under such contamination tends to overestimate the number of common shocks, and it can reach all cases as the proportion of contaminated series and the number of series increases. These results demonstrate the need for a more robust method, especially considering that the method is recommended to be used for high-dimensional data sets.

These results are in concordance with those obtained by Kristensen (2014) and Trucíos et al. (2019a) who, in a different but related context, found that the number of principal components (Peason, 1901; Hotelling, 1933) and principal volatility components (Hu and Tsay, 2014; Li et al., 2016) also tend to be over-identified when the series are contaminated by additive outliers.

Table 1: Percentage of common shocks selected by the Hallin and Liška criterion in uncontaminated and contaminated series. Panel dimension n equal to 60, 120, and 240 and sample size T=120. Pattern of contamination: two consecutive outliers of size 10 either in the middle or at the end of the sample period. The number of Monte Carlo replications is 500 and the correct number of common shocks is 2.

n	$\hat{q}$	No outlier	Percentage of series contaminated by **two consecutive** additive outlier:					
			5%		10%		15%	
			in the middle	at the end	in the middle	at the end	in the middle	at the end
60	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	99.0	93.4	98.0	20.2	34.6	1.2	2.4
	3	1.0	6.4	2.0	79.0	64.8	97.4	97.2
	4	0.0	0.2	0.0	0.8	0.6	1.4	0.2
	5	0.0	0.0	0.0	0.0	0.0	0.0	0.2
120	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	100	83.2	99.4	0.4	3.6	0.0	0.0
	3	0.0	16.8	0.6	99.4	96.4	99.8	100
	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	5	0.0	0.0	0.0	0.2	0.0	0.2	0.0
240	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	100	71.8	92.8	0.0	0.0	0.0	0.0
	3	0.0	28.2	7.2	100	100	100	99.8
	4	0.0	0.0	0.0	0.0	0.0	0.0	0.2

3.2 Estimation of common shocks and impulse-response function

We reproduce the Monte Carlo experiment of Forni et al. (2017) using (12) and compare the average and standard deviation of the normalized mean squared errors (MSE) in uncontaminated series with those obtained under different patterns of contamination. The number of common shocks is assumed to be known when computing the normalized MSEs. Furthermore, the comparison of the estimated shocks and impulse–response functions with the corresponding simulated quantities requires an identification rule. As in Forni et al. (2017), our exercise is based on a Cholesky identification scheme on the first q variables; see Forni et al. (2017) for more details. A superscript* is used for identified quantities. The normalized MSE

for the impulse-response functions is given by

$$\frac{\sum_{i=1}^{n}\sum_{f=1}^{q}\sum_{k=1}^{K}(\hat{b}^{*}_{if,k}-b^{*}_{if,k})^2}{\sum_{i=1}^{n}\sum_{f=1}^{q}\sum_{k=1}^{K}(b^{*}_{if,k})^2}, \tag{13}$$

where $b^{*}_{if,k}$ is the estimated impulse-response coefficient of variable i for shock f at lag k and the truncation lag K is set to 60. Similarly, the estimation error on the shocks is measured by

$$\frac{\sum_{f=1}^{q}\sum_{t=1}^{T}(\hat{u}^{*}_{ft}-u^{*}_{ft})^2}{\sum_{f=1}^{q}\sum_{t=1}^{T}(u^{*}_{ft})^2}. \tag{14}$$

Table 2 reports the results for different values of the panel dimension n and sample size T=120. Results confirm the intuition that the performance on the estimation of the impulse-response functions and structural shocks decreases as the proportion of contaminated series increases. A substantial increase in the average and standard deviation of the MSE is observed regardless of the outlier position. Note that, even with as little as 5% of series contaminated, a significant increase in the average and standard deviation of the MSE is observed.

To understand the effects of over-identification of the number of common shocks, Figure 1 plots the estimated common shocks of a single simulated panel with $n = 60$, T=120, and q=2 where 15% of the series are contaminated. We consider uncontaminated series (first and second columns), series contaminated at the end of the sample period (third and fourth columns) and series contaminated in the middle of the sample period (fifth and sixth columns). We either considered q as known or determined by the Hallin and Liška criterion, which yields $\hat{q} = 3$ in contaminated series and $\hat{q} = 2$ for the uncontaminated ones. For the sake of comparison, we have also considered an imposed value of $\hat{q} = 3$ in the uncontaminated case.

For uncontaminated series, over-identification of the number of common shocks does not seem to be a big concern. The first two estimated common shocks are quite similar, whether $\hat{q} = 2$ or $\hat{q} = 3$ and, for $\hat{q}$=3, the third estimated common shock is close to zero with small variability. On the other hand, for contaminated series, over-identification has a strong effect in the estimation of the common shocks. Note that, when we estimate three common shocks, as determined by the Hallin and Liška criterion, the results are worse than using the correct number of common shocks in the presence of outliers.

Table 2: Monte Carlo averages and standard deviations (in parentheses) of normalized MSE for estimated impulse-response functions (top panel) and structural shocks (bottom panel) in uncontaminated and contaminated series across 500 data sets. Panel dimensions n=60, 120, 240 and sample size T=120.

n	No outliers	Percentage of series contaminated by two consecutive additive outlier					
		5%		10%		15%	
		in the middle	at the end	in the middle	at the end	in the middle	at the end
		Impulse-response functions					
60	0.145	0.246	0.244	0.267	0.271	0.299	0.315
	(0.034)	(0.098)	(0.092)	(0.106)	(0.096)	(0.113)	(0.112)
120	0.157	0.261	0.258	0.272	0.281	0.297	0.315
	(0.034)	(0.089)	(0.096)	(0.090)	(0.100)	(0.098)	(0.104)
240	0.163	0.268	0.268	0.280	0.288	0.302	0.321
	(0.033)	(0.082)	(0.090)	(0.087)	(0.093)	(0.092)	(0.102)
		Structural shocks					
60	0.135	0.209	0.208	0.221	0.222	0.239	0.249
	(0.033)	(0.101)	(0.093)	(0.111)	(0.097)	(0.119)	(0.111)
120	0.093	0.171	0.169	0.177	0.181	0.193	0.202
	(0.031)	(0.092)	(0.106)	(0.096)	(0.110)	(0.106)	(0.116)
240	0.069	0.145	0.147	0.154	0.157	0.169	0.179
	(0.028)	(0.084)	(0.098)	(0.093)	(0.101)	(0.100)	(0.114)

3.3 Forecasting

In this section, we analyse the forecasting performance of the GDFM in the presence of outliers. For the sake of comparison, as in Forni et al. (2017), the accuracy of one-step-ahead forecasts is measured by

$$\frac{\sum_{i=1}^{n}(\hat{\chi}_{i,T+1} - \chi_{i,T+1})^2}{\sum_{i=1}^{N}(\chi_{i,T+1})^2}, \tag{15}$$

where $\hat{\chi}_{i,T+1} = \sum_{f=1}^{q}(\hat{b}_{if,1}\hat{u}_{fT} + \hat{b}_{if,2}\hat{u}_{f,T-1} + ...)$. We consider two cases. In the first case (top panel of Table 3), the number of common shocks is known, that is, there is no misidentification. In the second case (bottom panel of Table 3), the number of common shocks is determined by the Hallin and Liška criterion which, as shown in Section 3.1, is not robust in the presence of outliers. The normalized MSE of the one-step-ahead forecasts reported in Table 3 reveals the strong effect of outliers on the forecasting performance. The highest MSEs are observed when the number

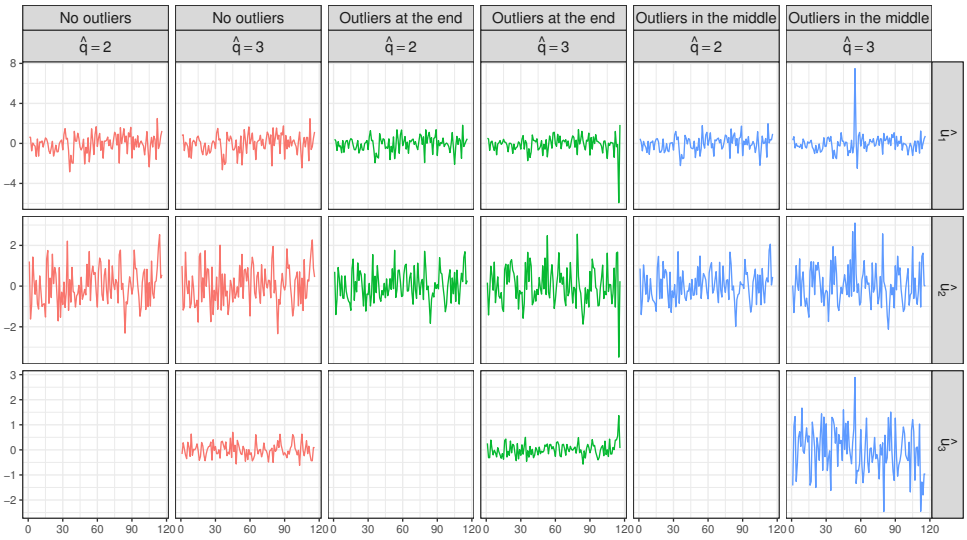

Figure 1: Estimated common shocks in a simulated panel with n=60, T=120, q=2. Uncontaminated series (first and second column), 15% of contamination at the end (third and fourth columns) and 15% of contamination in the middle (fifth and sixth columns) of the sample period.

of common shocks is determined by the Hallin and Liška criterion. These results show that, in the presence of outliers at the end of the sample period, identifying more common shocks than necessary has a strong impact on the forecasts. When q is known the normalized MSE decreases when the panel dimension increases, regardless the presence of outliers. When q is unknown, that also happens, except when the outliers occur at the end of the series with 5% and 10% of contamination. This possibly happens because the negative effect of the overestimation of the number of common shocks is stronger than the gain from the panel dimension increase.

In practice, we are in fact interested in forecasting the variables X_{it}s and we do not know the number of common shocks, which is the case in the application of Section 5.

4 Robustification

As reported in the previous section, Monte Carlo experiments show that the identification, estimation, and forecasting of the GDFM are strongly affected by the presence of outliers. In this section, we provide a robust alternative to circumvent these problems.

Table 3: Monte Carlo averages and standard deviations (in parentheses) of normalized MSE for the one-step-ahead forecasts in uncontaminated and contaminated series across 500 data sets. The panel dimensions are $n = 60, 120$ and 240 and sample size is T=120.

N	No outliers	Percentage of series contaminated by **two consecutive** additive outlier					
		5%		10%		15%	
		at the middle	at the end	at the middle	at the end	at the middle	at the end
				q is known			
60	0.414	0.527	0.560	0.565	0.704	0.636	0.955
	(0.246)	(0.215)	(0.298)	(0.240)	(0.402)	(0.332)	(0.658)
120	0.368	0.493	0.499	0.531	0.624	0.600	0.823
	(0.211)	(0.175)	(0.191)	(0.210)	(0.262)	(0.307)	(0.430)
240	0.344	0.487	0.484	0.520	0.593	0.583	0.779
	(0.146)	(0.154)	(0.163)	(0.189)	(0.198)	(0.260)	(0.315)
				q is determined by Hallin and Liška			
60	0.414	0.527	0.586	0.554	4.382	0.567	7.830
	(0.246)	(0.221)	(0.402)	(0.327)	(10.244)	(0.343)	(12.831)
120	0.368	0.492	0.605	0.488	4.919	0.498	6.408
	(0.211)	(0.193)	(1.244)	(0.305)	(7.039)	(0.262)	(10.450)
240	0.344	0.471	1.047	0.448	4.475	0.462	5.378
	(0.146)	(0.169)	(5.705)	(0.191)	(8.075)	(0.188)	(8.474)

4.1 Robust identification criterion

The correct identification of the number of common shocks is crucial for the estimation of the GDFM. In Section 3 we showed through Monte Carlo simulations the

non-robustness of the procedure proposed by Hallin and Liška (2007)[7] and its implications in the estimation and forecasting of the GDFM. To overcome the misidentification observed under contaminated data, we propose a robust version of the Hallin and Liška criterion.

As the Hallin and Liška criterion is based on the eigen-decomposition of $\hat{\mathbf{\Sigma}}^X(\theta)$, we propose to replace $\hat{\mathbf{\Sigma}}^X(\theta)$ by a robust estimator $\widetilde{\mathbf{\Sigma}}^X(\theta)$ of the spectral density matrix. This is achieved by using a robust estimator $\widetilde{\mathbf{\Gamma}}_k$ of the covariance matrix between $\mathbf{X}_t$ and $\mathbf{X}_{t-k}$, yielding the robust estimator

$$\widetilde{\mathbf{\Sigma}}^X(\theta) = \frac{1}{2\pi} \sum_{k=-M_T}^{M_T} e^{-ik\theta} K\left(\frac{k}{B_T}\right) \widetilde{\mathbf{\Gamma}}_k. \tag{16}$$

The GDFM is used in high-dimensional data to circumvent the curse of dimensionality. Because robust procedures with high computational costs make the estimation unfeasible in a high-dimensional context, a robust and fast procedure to estimate $\mathbf{\Gamma}_k$ is necessary. As mentioned in Maronna et al. (2006), a fast and robust alternative can be achieved via a robust estimation of pairwise covariances. We propose to use the robust estimator of Ma and Genton (2000), which is based on the scale parameter of Rousseeuw and Croux (1992, 1993) for each pair of variables. This estimator is fast to compute, location-free, and has shown a good trade-off between efficiency and robustness. Plenty of robust alternatives to Ma and Genton (2000) are available in the literature, but they are computationally more expensive and generally unfeasible in a high-dimensional framework.

We ran a Monte Carlo experiment with 500 replications considering the robust criterion in Table 4 with $n = 60$, 120, 240 and T =120. The minimum values n_1 (see Section 3.2) used in the robust procedure were $3n/4$, $n/2$ and $n/4$ for $n = 60$, 120, and 240, respectively. The performance of our procedure appears to be sensitive to the choice of n_1. As a rule of thumb, we suggest using $3n/4$ when the concentration ratio is smaller than one, $n/2$ when the concentration ratio is close to one, and $n/4$ when the concentration ratio is larger than one. These values yields good performances in our Monte Carlo experiments. As we observe in Table 4, the robustified procedure correctly identifies the number of common shocks almost 100% of times, whereas

[7] Actually, the Hallin and Liška criterion, for the GDFM, is the only consistent method available in the literature.

399 the unrobustified procedure of Hallin and Liška (2007) overestimates the number of
400 common shocks (see Table 1).

Table 4: Percentage of common shocks selected by the robust version of Hallin and Liška criterion in uncontaminated and contaminated series for dimensions n= 60 (top panel), 120 (middle panel), 240 (bottom panel), and sample size $T = 120$. The number of Monte Carlo replications is 500.

N	$\hat{q}$	No outlier	Percentage of series contaminated by additive outlier					
			5%		10%		15%	
			in the middle	at the end	in the middle	at the end	in the middle	at the end
60	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	99.6	99.6	99.4	99.6	99.6	99.6	99.6
	3	0.4	0.4	0.6	0.4	0.4	0.4	0.4
120	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	99.6	99.6	99.4	99.6	99.6	99.6	99.4
	3	0.4	0.4	0.6	0.4	0.4	0.4	0.6
240	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	100	100	100	100	100	100	100
	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0

401

402 Although the pairwise approach by Ma and Genton (2000) is fast and easy
403 to implement, it lacks the affine equivariance and the positive definiteness prop-
404 erties. Modifications to obtain positive definiteness and approximate equivariance
405 have been proposed in the literature; see Rousseeuw and Molenberghs (1993) and
406 Maronna and Zamar (2002). Nevertheless, the componentwise estimator without
407 any modification reported the best performance in our Monte Carlo experiments.
408 An even faster alternative is the procedure recently proposed by Raymaekers
409 and Rousseeuw (2018), which is based on a V-robust transformation of the data;
410 see Hampel et al. (1981) and Raymaekers and Rousseeuw (2018) for details. This
411 procedure consists in applying a transformation to all observations in the original
412 dataset and then computing the sample covariance matrix as usual. The transformed
413 observations are of the form

$$X_{it}^{*} = \hat{\mu}_i + \hat{\sigma}_i \Psi_{b,c}\left(\frac{X_{it} - \hat{\mu}_i}{\hat{\sigma}_i}\right), \tag{17}$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i$ are robust estimates of μ_i and σ_i, respectively, and

$$\Psi_{b,c}(x) = \begin{cases} x, & 0 \leq |x| \leq b, \\ d_1 \tanh\left(d_2\left(c - |x|\right)\right) \text{sign}(x), & b \leq |x| \leq c, \\ 0, & c \leq |x|, \end{cases}$$

with the constants d_1 and d_2 chosen such that $\Psi_{b,c}(\cdot)$ is a continuous function.[8] We use $b = 1.5$, $c = 4$, $d_1 = 1.540793$ and $d_2 = 0.8622731$ as in Raymaekers and Rousseeuw (2018). The robust scale estimator proposed by Rousseeuw and Croux (1992, 1993) is used to estimate σ_i, and $\hat{\mu}_i$ is obtained by an M-estimator using the function $\Psi_{b,c}(\cdot)$. Besides its cheaper computational cost, the robust estimator of Raymaekers and Rousseeuw (2018) satisfies the affine equivariance as well as the positive semidefiniteness properties, which makes its use more attractive.

Table 5 reports the results of the robust version of Hallin and Liška (2007) based on the robust estimators of Raymaekers and Rousseeuw (2018). Results are very similar to those obtained in Table 4, the number of common shocks is correctly identified almost 100% of times.

Despite the good finite-sample properties of both methods, we suggest using the last one because its computational time is much smaller, and also due to its desirable properties of invariance and positiveness. Unlike the robust alternative using the estimator of Ma and Genton (2000), where the value of n_1 plays an important role and needs to be chosen according to the concentration ratio, the Raymaekers and Rousseeuw (2018) method is not sensitive to the choice of n_1, and we set this value to $n_1 = 3n/4$.

[8] Details about how to obtain the constants can be found in the supplementary material of Raymaekers and Rousseeuw (2018).

Table 5: Percentage of common shocks selected by the Hallin and Liška criterion when robustified via Raymaekers and Rousseeuw (2018), in uncontaminated and contaminated series, for dimensions n=60 (top panel), 120 (middle panel), 240 (bottom panel), and sample size $T = 120$. The number of Monte Carlo replications is 500.

N	$\hat{q}$	No outlier	Percentage of series contaminated by additive outlier					
			5%		10%		15%	
			in the middle	at the end	in the middle	at the end	in the middle	at the end
60	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	99.0	99.2	99.6	99.0	99.0	99.0	99.2
	3	1.0	0.8	0.4	1.0	1.0	1.0	0.8
120	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	100	99.8	99.8	99.8	99.8	100	99.8
	3	0.0	0.2	0.2	0.2	0.2	0.0	0.2
240	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	100	100	100	100	100	100	100
	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0

4.2 Robust estimation and forecasting procedures

A robust estimator of the spectral density matrix alone is not enough to robustify the GDFM estimator, as it only ensures robust estimates $\tilde{\mathbf{A}}_n(L)$ for $\mathbf{A}(L)$. Indeed, $\tilde{\mathbf{Y}}_{nt} = \tilde{\mathbf{A}}_n(L)\mathbf{X}_{nt}$ in Step 3 of Section 2.2 still will be affected by the presence of outliers due to the contamination in $\mathbf{X}_{nt}$. To overcome this issue, we propose a slight modification in Steps 5 and 6 of Section 2.2. Once the number of common shocks is selected using the previously described robust procedure, we proceed as follows.

- **Step 5*:** Apply a robust principal component procedure to $\tilde{\mathbf{Y}}_{nt} = \tilde{\mathbf{A}}_n(L)\mathbf{X}_{nt}$, where $\tilde{\mathbf{A}}_n(L)$ is a robust estimate of $\mathbf{A}(L)$ based on $\widetilde{\mathbf{\Sigma}}^X(\theta)$. Then, the impulse-response matrix is given by $\tilde{\mathbf{B}}_n(L) = [\tilde{\mathbf{A}}_n(L)]^{-1}\tilde{\mathbf{R}}$, where $\tilde{\mathbf{R}}$ is the matrix of eigenvectors associated with the q largest eigenvalues obtained from the robust principal component procedure.

- **Step 6*:** Use $\tilde{\mathbf{B}}_n(L)$ to obtain a robust estimation

$$\tilde{\boldsymbol{\chi}}_{nt} = [\tilde{\mathbf{A}}_n(L)]^{-1}\tilde{\mathbf{R}}_n\tilde{\mathbf{u}}_t = \tilde{\mathbf{B}}_n(L)\tilde{\mathbf{u}}_t = \tilde{\mathbf{B}}_{n0}\tilde{\mathbf{u}}_t + \tilde{\mathbf{B}}_{n1}\tilde{\mathbf{u}}_{t-1} + ... + \tilde{\mathbf{B}}_{ns}\tilde{\mathbf{u}}_{t-s},$$

of the common components, where $\tilde{\mathbf{A}}_n(L)$ and $\tilde{\mathbf{R}}_n$ are as defined in the previous step. As $\tilde{\mathbf{u}}_t = \tilde{\mathbf{y}}_{nt}\tilde{\mathbf{R}}_n$ is still affected by outliers, the robust estimation of the common shocks is taken as $\tilde{\mathbf{u}}_t = \boldsymbol{\rho}(\tilde{\mathbf{y}}_t)\tilde{\mathbf{R}}_n$ with

$$\boldsymbol{\rho}(\tilde{\mathbf{y}}_t) = \begin{cases} (\tilde{\mathbf{y}}_{1t}, ..., \hat{\mu}_i^R, ..., \tilde{\mathbf{y}}_{nt}), & \text{if } (\tilde{\mathbf{y}}_{it} - \hat{\mu}_i^R)/\hat{\sigma}_i^R > c_1 \\ \hat{\boldsymbol{\mu}}^R, & \text{if } SD_t > c_2 \text{ and } OD_t > c_3 \\ \tilde{\mathbf{y}}_t, & \text{otherwise,} \end{cases} \tag{18}$$

where i stands for the ith series in the panel, $\hat{\mu}_i^R = (\hat{\mu}_{1t}, ..., \hat{\mu}_{nt})'$ is a robust location estimator of $\tilde{\mathbf{Y}}_{it}$, $\hat{\boldsymbol{\mu}}^R$ is a multivariate robust location estimator of $\tilde{\mathbf{Y}}_{nt}$ and SD_t and OD_t stand for the score distance and orthogonal distance associated to $\tilde{\mathbf{y}}_t$; see, for instance, Hubert et al. (2002, 2005, 2018) for more details about SD_t and OD_t. The first inequality in (18) can be valid for no series or even for all series, and eventually obtain a vector of the form $(\tilde{\mathbf{y}}_{1t}, \ldots, \mu_{j_1}, \tilde{\mathbf{y}}_{j_1+1,t}, \ldots, \mu_{j_k}, \ldots, \tilde{\mathbf{y}}_{nt})$.

Similarly to the non-robust version, the robust forecast of the common components at horizon h is obtained as

$$\tilde{\boldsymbol{\chi}}_{n,t+h|t} = \tilde{\mathbf{B}}_{nh}\tilde{\mathbf{u}}_t + \tilde{\mathbf{B}}_{n,h+1}\tilde{\mathbf{u}}_{t-1} + ... + \tilde{\mathbf{B}}_{n,t+h-1}\tilde{\mathbf{u}}_1.$$

Finally, the robust version of the final estimated impulse-response matrix $\tilde{\tilde{\mathbf{B}}}_{\mathbf{n}}(\mathbf{L})$, the common components $\tilde{\tilde{\boldsymbol{\chi}}}_{nt}$, and the h-step-ahead common component $\tilde{\tilde{\boldsymbol{\chi}}}_{n,t+h|t}$ are obtained by averaging their corresponding versions across B permutations as in Step 7 of Section 2.2.

There are a number of robust alternatives to classical principal components analysis; see, for instance, Croux and Haesbroeck (2000), Engelen et al. (2005) and Maronna (2005) for interesting comparative studies. Those approaches can be divided into two groups. The first group is based on a robust estimation of the covariance matrix and the second is based on projection pursuit. However, only few of the existing methods are feasible in a high-dimensional framework. In this paper, we have used the robust principal component procedure (ROBPCA) of Hubert et al. (2005) because its good performance in high dimensions. That procedure combines projection pursuit and robust estimation of the covariance matrix. We

have also used the robust procedure of Hubert et al. (2002), but the results using the ROBPCA procedure were much better.

Results comparing the robust and non-robust procedures when estimating the impulse-response functions, structural shocks and common components of the GDFM are reported in Figures 2 - 4. In the absence of outliers, the performance of the non-robust procedure is (not surprisingly so) slightly better. However, the advantage of the use of the robust procedure in the presence of outliers is clear in all cases. Note that, when the number of common shocks is estimated in a non-robust way, the differences between the robust and non-robust procedures are huge. Assuming that the true number of common shocks is known results in an improvement in the non robust procedure, although better results still are obtained with the robust approach. Whenever outliers are likely to be present in the observations, we suggest using our robust approach, as the consequences of neglecting the impact of those outliers may be quite dramatic.

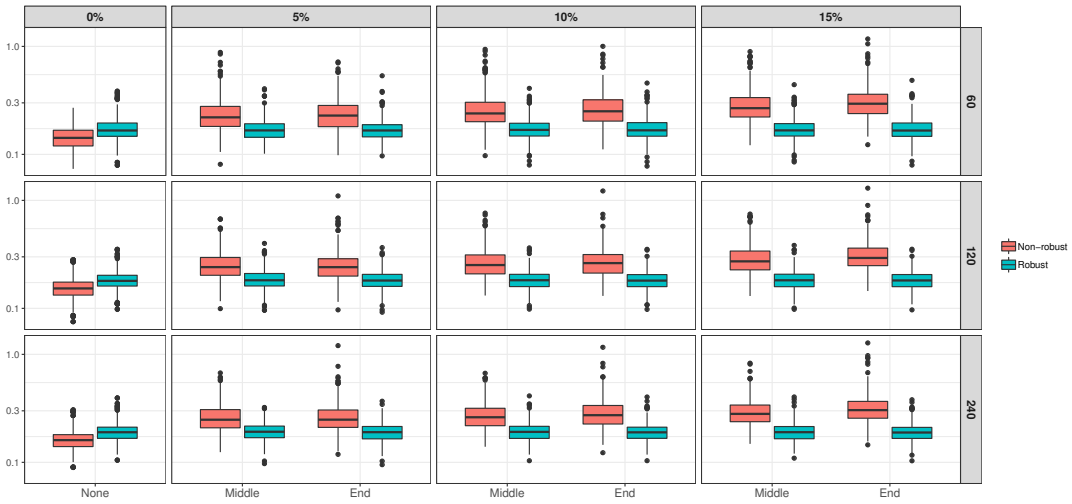

Figure 2: Boxplots of the normalized MSE in logarithmic scale for estimated impulse-response functions in uncontaminated and contaminated series, using the non-robust (red) and robust (blue) procedures. Dimension $n = 60$ (top panel), 120 (middle panel), and 240 (bottom panel). Sample size $T = 120$. The number of Monte Carlo replications is 500.

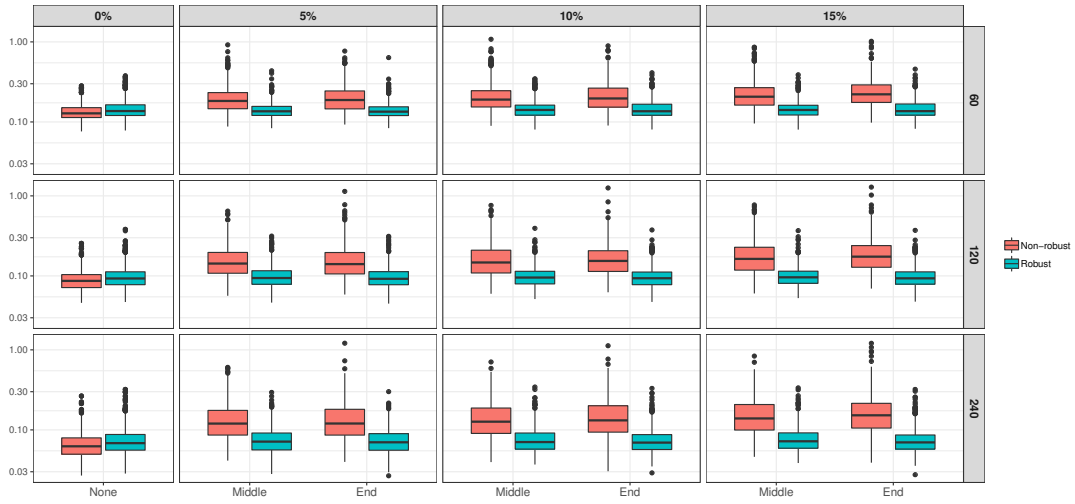

Figure 3: Boxplots of the normalized MSE in logarithmic scale for estimated structural shocks in uncontaminated and contaminated series, using the non-robust (red) and robust (blue) procedures. Dimension $n = 60$ (top panel), 120 (middle panel), and 240 (bottom panel). Sample size $T = 120$. The number of Monte Carlo replications is 500.

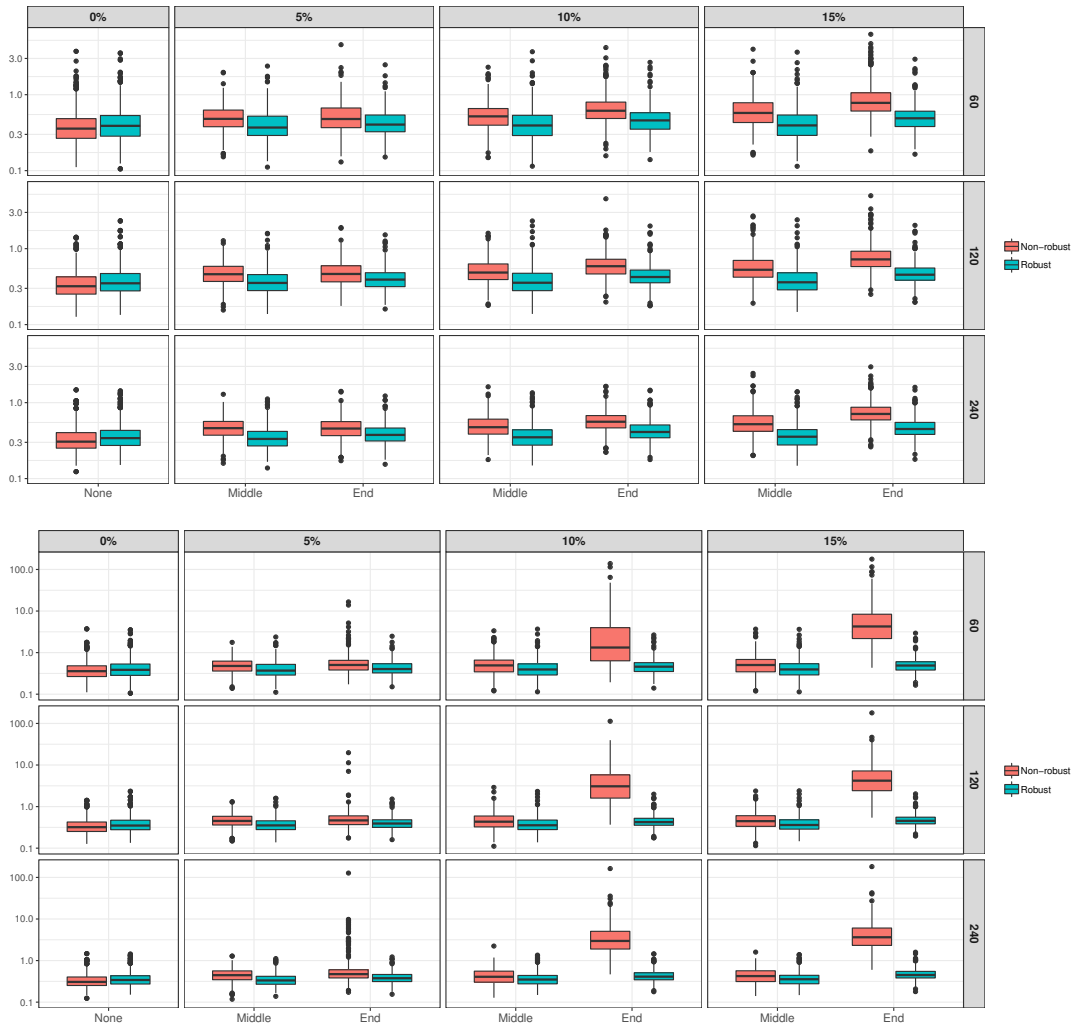

Figure 4: Boxplots of the normalized MSE in logarithmic scale for the estimated common components when the number of common shocks is known (top) and estimated (bottom). The Hallin and Liška criterion was used in the non-robust analysis while the robust analysis was based on the robust procedure in (16) based on the estimator of Raymaekers and Rousseeuw (2018). Dimension $n = 60$, 120, and 240 . Sample size $T = 120$. The number of Monte Carlo replications is 500.

5 Empirical application

We use the same dataset as Forni et al. (2018), which consists of 115 US macroeconomic and financial time series at monthly frequency between January and August 2014. Each series is transformed properly to achieve stationarity.[9]

Let $\mathbf{Z}_t = (Z_{1t} \;\; Z_{2t} \;\; ... \;\; Z_{nt})'$ be the raw dataset, and $\mathbf{X}_t = (X_{1t} \;\; X_{2t} \;\; ... \;\; X_{nt})'$ be the stationary result of the transformations of $\mathbf{Z}_t$.[10] Estimation is carried out using the marginal standardized version of $\mathbf{X}_t$ denoted as $\mathbf{x}_t$.

As in Forni et al. (2018), at time t we compute the h-step ahead forecasts for the i-th series $x_{i,t+h}$, $h = 6, 12, 24$. The forecasts are estimated using (m_1) the non-robust estimation procedure of Forni et al. (2015, 2017) (FHLZ); (m_2) our proposed robust estimation procedure (RFHLZ); (m_3) the standard principal component model introduced in Stock and Watson (2002a,b) with five factors (SW5); (m_4) the model based on generalized principal components introduced in Forni et al. (2005) (FHLR); and (m_5) n marginal univariate autoregressive models (AR) as the benchmark. The specifications of models FHLZ, FHLR, SW and AR and the calibration procedure are summarized in the Appendix, and they are the same as in Forni et al. (2018),

For all the methods we use a rolling 10-year window [t-119, t], and the models are re-estimated for each t. All the forecasts are obtained directly for each horizon h, not iterating one-step-ahead forecasts. The forecast of $X_{i,t+h}$ is, then, obtained by restoring the standard deviation and the mean. As in Forni et al. (2018), our first rolling window comprises February 1975 to January 1985 and the last forecast is August 2014 for all horizons. The period previous to February 1975 is used in the calibration procedure.

Our objective is to assess the performance of methods m_l when predicting the industrial production index (IP) and the consumer price index (CPI) ($i = 1, 77$). Figure 5 presents the plots of these series and their normalized versions. For the IP series the normalized series are their returns, while for the CPI series they are the seasonal difference of their returns. During the sub-prime crisis period, there

[9] We dropped the variable "US AVG OVERTIME HOURS - MANUFACTURING VOLA" (USHXPMANO) due to its lack of variation. Our final database, then, contains 114 monthly macroeconomic time series.

[10] We implement the same transformations to stationary as Forni et al. (2018).

is a continuously strong decrease of the IP series. There is also an increase in the volatility of IP and CPI series during the crisis period.

In order to check for the presence of outliers, in Figure (6) we apply to all the series the robust transformation procedure of Raymaekers and Rousseeuw (2018) given by Equation (17). We observe in Figure (6) many points outside the bands $|b|$=1.5 and $|c|$=4, with larger concentration in the crisis and post-crisis periods (from December 2007 on). For simplicity we call this period the *crisis period*.

As in Forni et al. (2018), we compare the predictor performance by the MSFE, the Diebold-Mariano test (Diebold and Mariano, 1995) for the null hypothesis of global equal performances between two predictors, and the fluctuation test of Giacomini and Rossi (Giacomini and Rossi, 2010) to compare locally the performance of two predictors.

For the variable i and the m_l prediction method, the MSFE forecasting performance evaluated as

$$MSFE_{i,h}^{m_l} = \frac{1}{(T_1 - h) - T_0 + 1} \sum_{\tau=T_0}^{T_1-h} [FE_{i,\tau,h}^{m_l}]^2,$$

where

$$FE_{i,t,h}^{m_l} = \frac{1}{h}((\hat{X}_{i,t+1|t}^{m_l} - X_{i,t+1}) + ... + (\hat{X}_{i,t+h|t}^{m_l} - X_{i,t+h})),$$

and $\hat{X}_{i,t+k|t}^{m_l}$, $k = 1, \cdots, h$, is the $k-$step-ahead prediction for the variable i given by method m_l, $l = 1, ..., 5$.

Denote by $MSFE_{i,h}^{m_5}$ the MSFE of the benchmark AR model. As in Forni et al. (2018), the relative performance of the m prediction method at horizon $h = 6, 12, 24$ for the variable i in relation to the benchmark AR model is defined as

$$RMSFE_{i,h}^{m_l} = \frac{MSFE_{i,h}^{m_l}}{MSFE_{i,h}^{m_5}}, \quad l = 1, ..., 4.$$

The results of the application is given in Table 6. As in Forni et al. (2018), we give results for the pre-crisis period, from February 1985 to November 2007, and for the full sample period, from February 1985 to August 2014. We also add the results for the crisis period.

Tables 7 presents the p-values of the two-sided Diebold-Mariano test for the null hypothesis of global equal performance between two predictors. Due to the presence

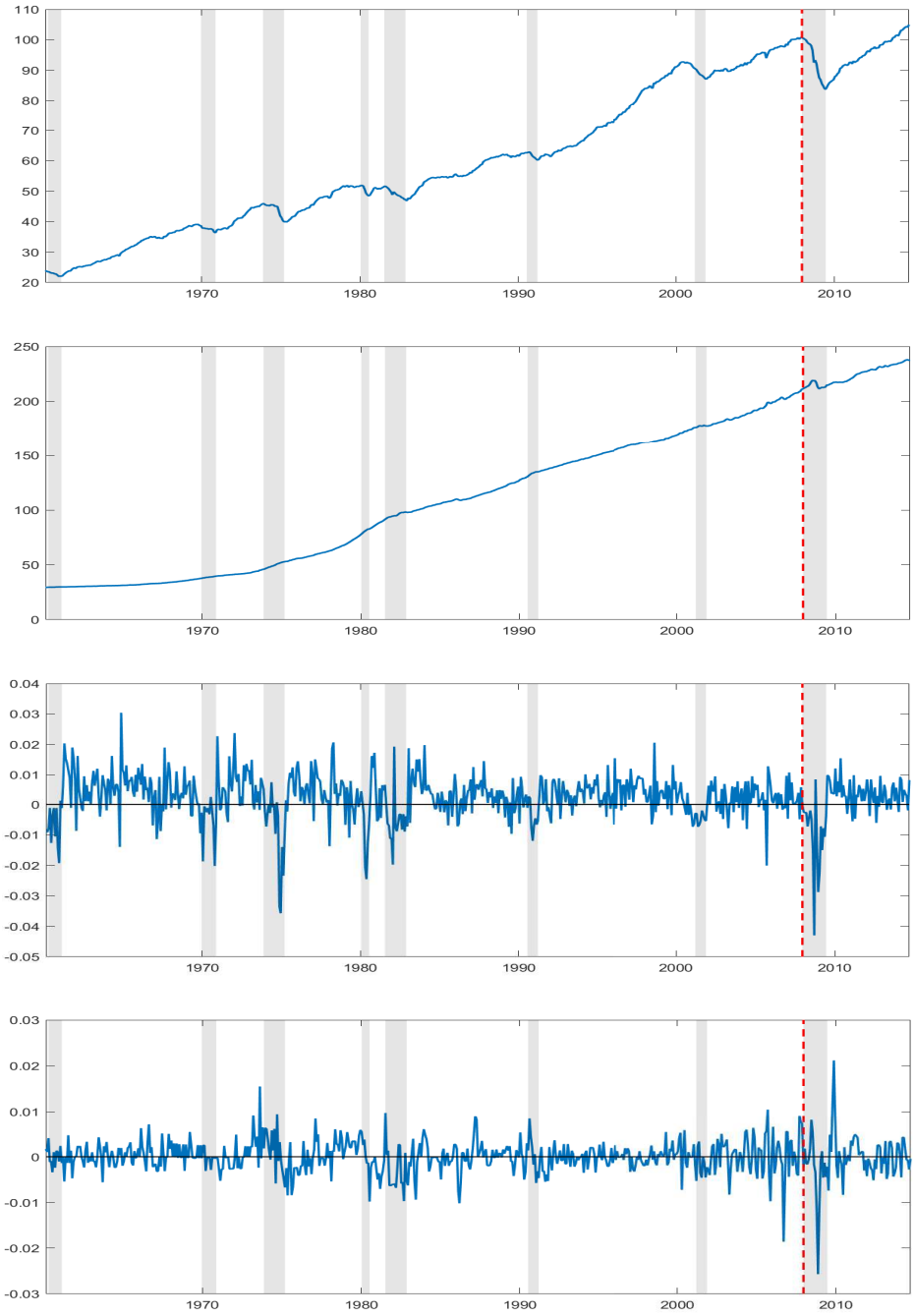

Figure 5: $Z_{1t} = IP_t$ (first panel); $Z_{77t} = CPI_t$ (second panel); $x_{1t} = (1-L)\log IP_t$ (third panel); $x_{77,t} = (1-L)(1-L)^{12}\log CPI_t$ (fourth panel). Dashed red line represents the crisis beginning (2007:12). The NBER recession periods are shaded in light gray.

Table 6: Relative MSFE for IP and CPI series for horizons $h = 6, 1, 24$. The best result for each horizon over all methods is in bold.

IP				
	RFHLZ	FHLZ	FHLR	SW(5)
h=6	[a] 0.8705	**0.8641**	0.9943	0.9998
	[b] 0.9465	1.0641	0.7079	**0.6945**
	[c] 0.9118	0.9716	0.8393	**0.8347**
h=12	[a] 0.8919	**0.8782**	0.9541	0.9813
	[b] 0.8595	1.0197	0.7588	**0.7475**
	[c] 0.8794	0.9645	**0.8398**	0.8427
h=24	[a] 0.9703	0.9587	**0.9383**	0.9912
	[b] **0.6393**	0.9604	0.7690	0.7879
	[c] **0.8756**	0.9732	0.9070	0.9304
CPI				
	RFHLZ	FHLZ	FHLR	SW(5)
h=6	[a] 0.9456	**0.9352**	1.0365	1.0413
	[b] 0.9298	**0.9099**	1.0779	1.0353
	[c] 0.9491	**0.9369**	1.0818	1.0586
h=12	[a] **0.7941**	0.8110	0.9284	0.9750
	[b] **1.0279**	1.0745	1.2046	1.1787
	[c] **0.9254**	0.9658	1.0920	1.1020
h=24	[a] **0.7676**	0.8475	0.8217	0.8925
	[b] 1.5578	1.8129	1.6656	**1.5936**
	[c] **0.9524**	1.0257	0.9748	1.0034

[a]pre-crisis (1985:2-2007:11); [b]crisis (2007:12-2014:8);
[c]full sample (1985:2-2014:8)

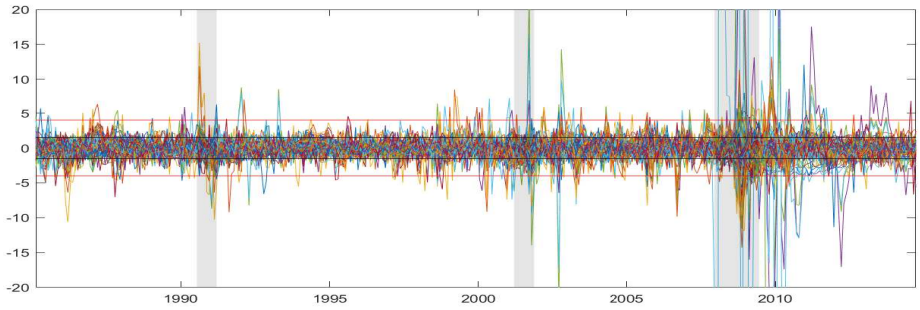

Figure 6: V-robust transformation (17) applied to the whole stationary data set. Black and red bands correspond to $|b|$=1.5 and $|c|$=4, respectively. The NBER recession dates are shaded in light gray. Out-of-sample period reported (1985:2-2014:8).

of structural breaks and outliers, as in Forni et al. (2018) we only present the results for the pre-crisis period.

Figures 7 and 8 present the equal local performance fluctuation test of Giacomini and Rossi (2010). We draw 5% critical values for the bilateral test. When testing Model 1 vs Model 2, values below (above) the critical values means that Model 1 (Model 2) is statistically better (worse) than Model 2 (Model 1).

The analysis of the results for the non-robust predictors are similar to the Forni et al. (2018) results. For the IP series, all predictors have smaller MSFEs than the AR model, except for FHLZ at horizons 6 and 12 in the second period. The null of equal performance with AR, in the pre-crisis period, is rejected for both RFHLZ and FHLZ at horizons 6 and 12 (see Tables 6 and 7). For the CPI series, compared to the AR model, SW(5) and FHLR have a poor performance while FHLZ and RFHLZ have smaller MSFES, except for horizons 12 and 24 in the second period. The null of equal performance with AR, in the pre-crisis period, is rejected for both RFHLZ and FHLZ at horizons 12 and 24 and for FHLR at horizon 24 (see Tables 6 and 7). Analyzing the same tables, comparing RFHLZ to FHLZ, we can see that, for the pre-crisis period, in general their performances are almost equivalent, while in the crisis period we have a better performance of RFHLZ. The comparison between the effect of robustification is better illustrated by the fluctuation test presented in

Table 7: p-values of Diebold-Mariano test comparing Model 1 vs Model 2 with null hypothesis of equal global prediction performance during the pre-crisis period. Two-sided test and horizons (h) equal to $6, 12$ and 24. In bold are the results significant at 10% level.

IP					
	RFHLZ vs AR	FHLZ vs AR	RFHLZ vs FHLZ	FHLR vs AR	SW(5) vs AR
h=6	**0.055**	**0.027**	0.757	0.958	0.999
h=12	**0.028**	**0.002**	0.488	0.623	0.852
h=24	0.248	0.113	0.674	0.341	0.890
CPI					
	RFHLZ vs AR	FHLZ vs AR	RFHLZ vs FHLZ	FHLR vs AR	SW(5) vs AR
h=6	0.322	0.265	0.549	0.855	0.826
h=12	**0.023**	**0.038**	0.452	0.632	0.834
h=24	**0.011**	**0.009**	**0.028**	0.373	0.657

Figures 7 and 8. We can see that robustification improves the performance of FHLZ and that, in fact, the improvements start some months before the outset of the crisis.

Summing up, the empirical application shows a better performance of the robust version of the FHLZ model, it increases the performance in periods with outliers and/or structural breaks, and there is no significant decrease of the performance in other periods.

6 Conclusions

In this paper, we addressed the identification, estimation, and forecasting procedures in the GDFM with infinite-dimensional factor space, showing that all procedures are badly affected by the presence of additive outliers, even when only a few outliers are present. We also illustrate the impact of neglecting this issue and propose a robust alternative to circumvent this problem.

Using a robust estimator of the covariance/spectral density matrix, we propose a robust version of the identification criterion of Hallin and Liška (2007) with good sample properties. The robust alternative has a good performance in contaminated as well as uncontaminated series.

Furthermore, based on robust estimators and robust filters, we also propose robust estimation and robust forecasting procedures in the context of GDFM. Our simulations indicate that our procedures are superior to the non-robust ones in the presence of outliers with little to no cost in the uncontaminated case.

The new procedures are also applied to macroeconomic and financial time series, where better results are observed in the crisis period (presence of outliers) compared to those of the non-robust procedures, and with comparable performance in periods without crisis.

Our findings are useful for practitioners interested in applying GDFM for forecasting purposes giving tips of better practices in the estimation and forecasting processes. Additionally, the results of our paper contribute to the literature of GDFM providing new insights and material for future theoretical results.

7 Appendix

The pre-sample period, February 1960 to January 1985, is used by Forni et al. (2018) to calibrate the FHLZ, FHLR, SW, and AR methods. To compare two specifications m_a and m_b, say, of method m at horizon h=6, 12, 24 for variable i, they use the ratio

$$RMSFE_{i,h}^{m_a/m_b} = \frac{MSFE_{i,h}^{m_a}}{MSFE_{i,h}^{m_b}}.$$

The calibration procedure is limited to IP and CPI ($i = 1$ and 77, respectively); see Forni et al. (2018) for details about the specifications of each model used in the calibration procedure. The resulting specification of FHLZ and FHRL uses the triangular kernel with $B = 30$ and $B = 40$, respectively. For each rolling window, the degrees of the VARs are determined by AIC with maximum lag 5, and q is determined by Hallin-Liška criterion. For FHLR, the number of static factors r is fixed and equal to 6 for IP and 5 for CPI, and the prediction equation of FHLR does not include lagged values of the generalized principal components and of the predicted variable. For SW, the selected specifications include a static factor model with 5 or 6 static factors for IP and a model with 5 static factors for CPI, and no lags of the static factors and of the predicted variable are included in the prediction equation. In our paper, we estimate SW with 5 static factors for both IP and CPI. For the AR model, the number of lags is determined at each rolling window, for each h, by BIC with maximum lag 13. Finally, we use the same specifications of FHLZ for its robust version.

References

Alessi, L., Barigozzi, M., and Capasso, M. (2009). Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors. Working paper series 1115, European Central Bank, Frankfurt am Main, Germany.

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters*, 80(23-24):1806–1813.

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., and Veronese, G. (2010). New eurocoin: tracking economic growth in real time. *The Review of Economics and Statistics*, 92(4):1024–1034.

Aramonte, S., del Giudice Rodriguez, M., and Wu, J. (2013). Dynamic factor value-at-risk for large heteroskedastic portfolios. *Journal of Banking & Finance*, 37(11):4299–4309.

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

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

Bai, J. and Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8:53–80.

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

Barhoumi, K., Darné, O., and Ferrara, L. (2014). Dynamic factor models: A review of the literature. *Journal of Business Cycle Research*, 2013(2):73.

Barigozzi, M. and Hallin, M. (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks. *The Econometrics Journal*, 19(1):C33–C60.

Barigozzi, M. and Hallin, M. (2017). Generalized dynamic factor models and volatilities: estimation and forecasting. *Journal of Econometrics*, 201(2):307–321.

Barigozzi, M. and Hallin, M. (2018). Generalized dynamic factor models and volatilities: Consistency, rates, and prediction intervals. *arXiv preprint:1811.10045*.

Barigozzi, M., Hallin, M., and Soccorsi, S. (2018). Identification of global and local shocks in international financial markets via general dynamic factor models. *Journal of Financial Econometrics*.

Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525–546.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, volume 36. Siam.

Cristadoro, R., Forni, M., Reichlin, L., and Veronese, G. (2005). A core inflation indicator for the euro area. *Journal of Money, Credit and Banking*, pages 539–560.

Croux, C. and Exterkate, P. (2011). Robust and sparse factor modelling. *Available at SSRN: https://ssrn.com/abstract=1967424*.

Croux, C. and Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, 87(3):603–618.

Della Marra, F. (2017). A forecasting performance comparison of dynamic factor models based on static and dynamic methods. *Communications in Applied and Industrial Mathematics*, 8(1):43–66.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Eickmeier, S. (2007). Business cycle transmission from the us to germany—a structural factor approach. *European Economic Review*, 51(3):521–551.

Engelen, S., Hubert, M., and Branden, K. V. (2005). A comparison of three procedures for robust PCA in high dimensions. *Austrian Journal of Statistics*, 34(2):117–126.

Favero, C. A., Marcellino, M., and Neglia, F. (2005). Principal components at work: the empirical analysis of monetary policy with large data sets. *Journal of Applied Econometrics*, 20(5):603–620.

Forni, M. and Gambetti, L. (2010). The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics*, 57(2):203–216.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25(5):1319–1347.

Forni, M., Giovannelli, A., Lippi, M., and Soccorsi, S. (2018). Dynamic factor model with infinite-dimensional factor space: Forecasting. *Journal of Applied Econometrics*.

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

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2004). The generalized dynamic factor model consistency and rates. *Journal of Econometrics*, 119(2):231–255.

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(471):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(2):359–371.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: asymptotic analysis. *Journal of Econometrics*, 199(1):74–92.

Forni, M. and Lippi, M. (2011). The general dynamic factor model: One-sided representation results. *Journal of Econometrics*, 163(1):23–28.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Giovannelli, A., Massacci, D., and Soccorsi, S. (2018). Forecasting stock returns with large dimensional factor models. *Working Paper available at SSRN: https://ssrn.com/abstract=2958491*.

Hallin, M. and Lippi, M. (2013). Factor models in high-dimensional time series—a time-domain approach. *Stochastic processes and their applications*, 123(7):2678–2695.

Hallin, M. and 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., Rousseeuw, P. J., and Ronchetti, E. (1981). The change-of-variance curve and optimal redescending M-estimators. *Journal of the American Statistical Association*, 76(375):643–648.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417.

Hu, Y.-P. and Tsay, R. S. (2014). Principal volatility component analysis. *Journal of Business & Economic Statistics*, 32(2):153–164.

Hubert, M., Rousseeuw, P. J., and Bossche, W. V. d. (2018). MacroPCA: An all-in-one PCA method allowing for missing values as well as cellwise and rowwise outliers. *arXiv preprint arXiv:1806.00954*.

Hubert, M., Rousseeuw, P. J., and Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics*, 47(1):64–79.

Hubert, M., Rousseeuw, P. J., and Verboven, S. (2002). A fast method for robust principal components with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 60(1-2):101–111.

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.

Li, W., Gao, J., Li, K., and Yao, Q. (2016). Modeling multivariate volatilities via latent common factors. *Journal of Business & Economic Statistics*, 34(4):564–573.

Ludvigson, S. C. and Ng, S. (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222.

Ludvigson, S. C. and Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, 22(12):5027–5067.

Ma, Y. and Genton, M. G. (2000). Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis*, 21(6):663–684.

Maronna, R. (2005). Principal components and orthogonal regression based on robust scales. *Technometrics*, 47(3):264–273.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley, Chichester, First edition.

Maronna, R. A. and Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, 44(4):307–317.

Peason, K. (1901). On lines and planes of closest fit to systems of point in space. *Philosophical Magazine*, 2(11):559–572.

Raymaekers, J. and Rousseeuw, P. J. (2018). Fast robust correlation for high dimensional data. *Technical report KU Leuven: Section of Statistcs. Available at arXiv preprint arXiv:1712.05151*.

Rousseeuw, P. J. and Croux, C. (1992). Explicit scale estimators with high breakdown point. *L1-Statistical Analysis and Related Methods*, 1:77–92.

Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical association*, 88(424):1273–1283.

Rousseeuw, P. J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics–Theory and Methods*, 22(4):965–984.

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

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for var analysis. Technical report, National Bureau of Economic Research.

Trucíos, C., Hotta, L. K., and Pereira, P. L. V. (2019a). On the robustness of the principal volatility components. *Journal of Empirical Finance*, 52:201–219.

Trucíos, C., Mazzeu, J. H. G., Hallin, M., Hotta, L. K., Valls Pereira, P. L., and Zevallos, M. (2019b). Forecasting conditional covariance matrices in high-dimensional time series: a general dynamic factor approach. *Working Paper available at SSRN: https://ssrn.com/abstract=3399782*.

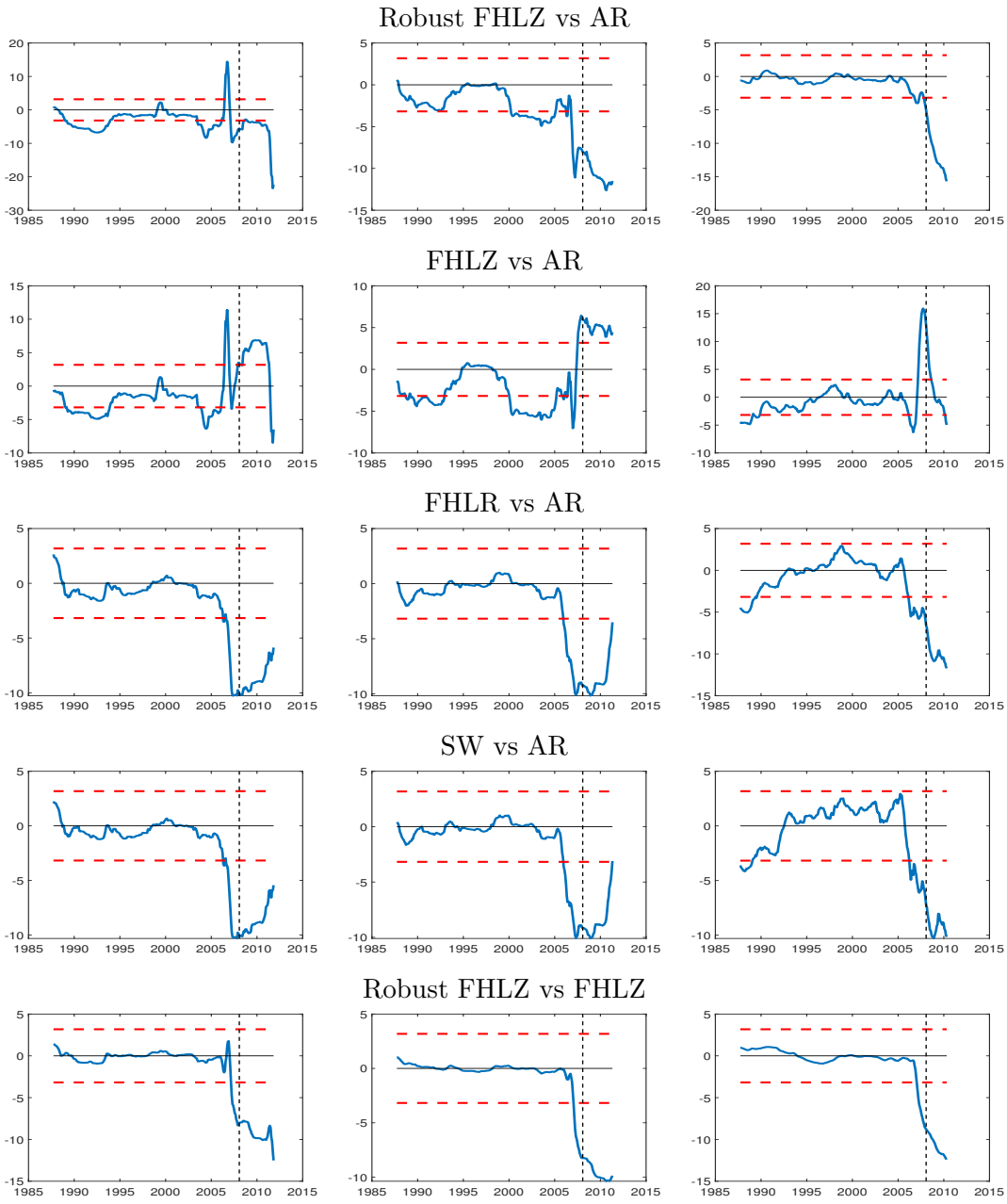

Figure 7: Equal local performance of two forecasting methods (IP). Fluctuation test statistic (solid line) and 5% two-side critical values (dotted line). If the solid is below (above) the lower (upper) critical value, the first method is significantly better (worse) than the second one.

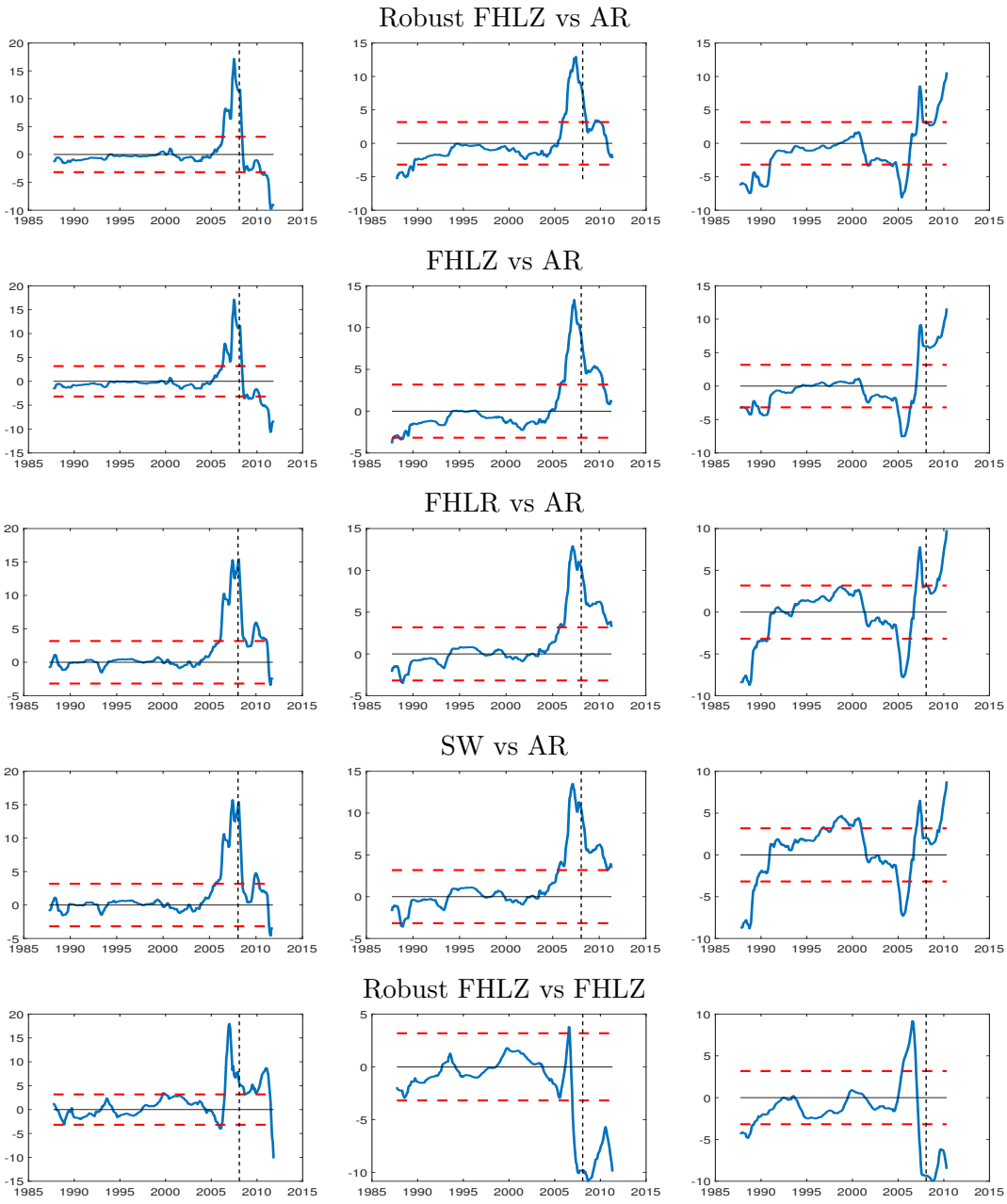

Figure 8: Equal local performance of two forecasting methods (CPI). Fluctuation test statistic (solid line) and 5% two-side critical values (dotted line). If the solid is below (above) the lower (upper) critical value, the first method is significantly better (worse) than the second one.