

Johannes Tang Kristensen*

Factor-based forecasting in the presence of outliers: Are factors better selected and estimated by the median than by the mean?

Abstract: Macroeconomic forecasting using factor models estimated by principal components has become a popular research topic with many both theoretical and applied contributions in the literature. In this paper we attempt to address an often neglected issue in these models: The problem of outliers in the data. Most papers take an ad-hoc approach to this problem and simply screen datasets prior to estimation and remove anomalous observations. We investigate whether forecasting performance can be improved by using the original unscreened dataset and replacing principal components with a robust alternative. We propose to use an estimator based on least absolute deviations (LAD) as this alternative and establish a tractable method for computing the estimator. In addition to this we demonstrate the robustness features of the estimator through a number of Monte Carlo simulation studies. Finally, we apply the estimator in a simulated real-time forecasting exercise to test its merits. We use a newly compiled dataset of US macroeconomic series spanning the period 1971:2–2012:10. Our findings suggest that the chosen treatment of outliers does affect forecasting performance and that in many cases improvements can be made using a robust estimator such as the proposed LAD estimator.

Keywords: factors models; forecasting; least absolute deviations; principal components analysis; robust estimation.

JEL classification codes: C38; C53; E37.

*Corresponding author: Johannes Tang Kristensen, CREATES and Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark, e-mail: jtang@creates.au.dk

1 Introduction

As time goes by we accumulate information at an ever increasing rate. This coupled with vast improvements in computation power has driven an intensive research interest into the area of econometric analysis of large-dimensional datasets. In macroeconomic forecasting in particular we often have many hundreds if not thousands of time series at our disposal for forecasting. These are, however, of no interest without the right set of tools for analysing them. The class of dynamic factor models has found its way into many econometricians' toolboxes and especially developments in this area over the last decade has made these models increasingly popular for modelling large-dimensional data. A recent survey by Stock and Watson (2011) provides a thorough overview of the state of the literature, and is an important addition to previous surveys (e.g., Stock and Watson, 2006; Bai and Ng, 2008).

Although a number of estimation methods for factor models have been proposed in the literature, this paper will focus on probably the most popular method, namely estimation by principal components (PC). One of the main advantages of PC estimation is its ease of use. The estimation is quick and easily done even for very large datasets with implementations only requiring a few lines of code. The literature on macroeconomic forecasting using factor models estimated by PC was popularized by two papers by Stock and Watson (2002a,b). In their papers they coined the term diffusion indexes which refers to the estimated factors. The term was chosen by Stock and Watson because they interpret estimated factors in terms of the diffusion indexes developed by NBER business cycle analysts. So the name only relates to the interpretation, the actual estimates are simply factors from a factor model.

Stock and Watson (2002b) demonstrated the superiority of diffusion index forecasting when compared to many traditional models. Since their paper numerous papers have applied factor models in forecasting settings. However, not all arrive at similarly impressive results. In fact several papers have been pointing towards a possible break-down of factor model forecasts in recent years, see e.g., Schumacher (2007), Schumacher and Dreger (2004), Banerjee, Marcellino, and Masten (2006), Gosselin and Tkacz (2001), Angelini, Henry, and Mestre (2001). In an attempt to investigate the forecasting performance in more detail Eickmeier and Ziegler (2008) conduct a meta-analysis of 52 forecasting applications of dynamic factor models, and they also find mixed results regarding the performance of these models.

It is hence not clear why factor-based forecasts in some cases perform very poorly. However, the literature is starting to see attempts to investigate this and further develop the ideas behind the PC factor estimator. In this paper we will try to address an often neglected issue in factor models, namely that of outliers in the data. We believe that the presence of outliers might be one possible explanation of the forecasting performance issues seen in the empirical literature.

The problem of outliers is generally acknowledged in the literature albeit in a rather indirect fashion. Often people will screen data for anomalous observations and either remove these if the method used is capable of handling missing observations or replace them with more “normal” values. In the classical paper by Stock and Watson (2002b) outliers were defined as observations exceeding the median of the series by more than 10 times the interquartile range. It, however, appears that the consensus in the empirical literature is leaning towards defining outliers as observations exceeding the median of the series by more than six times the interquartile range. Examples of this include Banerjee, Marcellino, and Masten (2008), Breitung and Eickmeier (2011), Stock and Watson (2009),¹ and Artis, Banerjee, and Marcellino (2005).

In this paper we wish to challenge the common practice of screening for outliers. By screening one is implicitly assuming the outliers to be errors. If this is indeed not the case, and they are in fact relevant features of the data, it should be possible to improve forecasting performance by replacing PC with a robust alternative better adapted to such a setting. Our goal is therefore to find a tractable robust alternative to PC that is able to capture such features of the data and investigate whether using such a method does lead to improvements in forecasting performance.

The contributions of the paper are threefold: (i) Working within the diffusion index forecasting framework we will propose to replace the traditional PC factor estimator with a robust estimator based on least absolute deviations (LAD). An idea that, to the best of our knowledge, has not been considered in macroeconomic forecasting before.² We detail how such an estimator can be estimated in a tractable manner. The proposed approaches are based on existing methods. We give recommendations on which to use and provide new results on convergence of the methods. (ii) Through a series of Monte Carlo studies we will demonstrate the features of the estimator and uncover similarities with the common screening approach. Particular attention is given to the problem of determining the true number of factors, as this is also greatly affected by outliers. (iii) Using a new dataset we will conduct a forecasting experiment where we illustrate the importance of taking outliers into account.

The main findings are: (i) Based on simulation results we conclude that the LAD factor estimator is indeed robust. Screening alleviates the problem of outliers for the PC factor estimator but is outperformed by the LAD factor estimator. (ii) A major part of factor modelling is determining the number of factors. We find that the commonly used information criteria of Bai and Ng (2002) is severely inflated by outliers when based on the PC factor estimates. This is not the case for the LAD factor estimates. Again screening lessens the problem, but not to the same extent as LAD. (iii) The forecasting experiment shows that using the LAD factor estimates does increase forecasting performance in a number of cases. Care must, however, be taken

¹ Although not discussed in the paper, the details can be found in their replication files.

² Since the first version of this paper was written a similar idea has been investigated by Croux and Exterkate (2011). In their paper a number of alternatives to PC are examined including LAD-based approaches. They find mixed results as to which approach is preferred when judging by forecasting performance. This is somewhat in contrast to the results presented in this paper. A possible explanation of this could be that they use BIC to select the number of factors. The results in this paper suggest that this may be an inappropriate choice.

when scaling the data. Our results suggest that using the mean absolute deviation as a robust alternative to the standard deviation is recommendable. The exercise also confirms the results from the simulations in that LAD generally chooses fewer factors than PC.

The paper is organized as follows: First we setup the general factor model in Section 2. We define the LAD factor estimator and give results on estimation methods. In Section 3 we provide a number of simulation results on both the precision of the factor estimates and the effect outliers have on determining the true number of factors. In Section 4 we conduct a typical macroeconomic forecasting exercise, however, using a newly constructed dataset. Finally, we conclude in Section 5.

2 A large-dimensional factor model

Our point of departure is the classical dynamic factor model. Let X_t be n macroeconomic variables we observe for $t=1, \dots, T$ and wish to use for forecasting some variable of interest y_t . We will assume that X_t is stationary, centered at zero, and that it follows a dynamic factor model of the form:

$$X_{it} = \bar{\lambda}_i(L) f_t + e_{it} \quad (1)$$

for $i=1, \dots, n$. Assuming that $\bar{\lambda}_i(L)$ is of finite order of at most q , such that $\bar{\lambda}_i(L) = \sum_{j=0}^q \bar{\lambda}_{i,j} L^j$, we can rewrite the model in static form as:

$$X_{it} = \lambda_i F_t + e_{it} \quad (2)$$

where $F_t = (f_t', \dots, f_{t-q}')'$ is $r \times 1$, and $\lambda_i = (\bar{\lambda}_{i,0}, \dots, \bar{\lambda}_{i,q})$ is $1 \times r$ with $r \leq (q+1)\bar{r}$. We will exclusively consider this static representation as is also common in the literature. Often it will be more convenient to work with the model in vector or matrix form in which cases we will write it as:

$$X_t = \Lambda F_t + e_t \quad \text{or} \quad X = F \Lambda' + e \quad (3)$$

where $\Lambda = (\lambda_1', \dots, \lambda_n')'$ is $n \times r$, $F = (F_1', \dots, F_T')'$ is $T \times r$, and $X = (X_1', \dots, X_T')'$ is $T \times n$.

Estimation of factor models entails a number of difficulties. Clearly, if either the factors or loadings were known estimation would be easily accomplished since the model would simplify to a multivariate linear regression. However, since only X_t is observed we must estimate both loadings Λ and factors F . Due to this we are faced with an inherent identification problem because $\Lambda F_t = \Lambda R R^{-1} F_t$ for any non-singular $r \times r$ matrix R . We therefore generally need to impose identifying restrictions in order to make estimation feasible. In the present paper we are interested in the factors solely for forecasting purposes, and because of this need not worry too much about identification. Any rotation of the estimated factors will be captured by the estimated parameters in the forecasting model and should not affect forecasting performance.

2.1 Least squares estimation

The theory underlying PC estimation of factor models has been extensively studied and is well understood. We will, however, spend a few moments defining the estimator and discussing its details. This is meant to serve as a prelude to the introduction of the LAD factor estimator below.

Consistency of the PC factor estimator was first shown by Connor and Korajczyk (1986) for the exact static factor model assuming T fixed and $n \rightarrow \infty$. Stock and Watson (2002a) extended their results to the case of the approximate static factor model, approximate in the sense of Chamberlain and Rothschild (1983), i.e., allowing for weak correlation across time and series. Their results were derived assuming that both $T \rightarrow \infty$ and $n \rightarrow \infty$. These results have since been supplemented by e.g., Bai and Ng (2002) who provided improved rates, Bai (2003) who derived the asymptotic distribution of the factors and loadings, and Bai and Ng (2006) who provided results for constructing confidence intervals for common components estimated using the factors.

Thus it would seem that the asymptotic theory of the PC factor estimator has been very well covered. However, common for all these results is that moment assumptions are made on the error terms, and hence they do not allow for typical outlier distributions such as the Student- t distribution with a low degree of freedom which is the premise we will explore in this paper.

One very appealing characteristic of the PC estimator is that it is in fact simply a least squares (LS) estimator. Consider the nonlinear LS objective function:

$$V^{\text{LS}}(F, \Lambda; X) = (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (X_{it} - \lambda_i F_t)^2 \quad (4)$$

Imposing the identifying restriction that $\Lambda' \Lambda = I_r$, where I_r is the identity matrix of dimension r , this problem yields the usual PC factor estimator. This can be solved easily as an eigenvalue problem, and it turns out that the estimated factors are given as linear combinations of the X -variables, i.e., $\hat{F} = X \hat{\Lambda}$, where $\hat{\Lambda}$ are the estimated loadings, see e.g., Stock and Watson (2002a) for details.

Alternatively, we could view the above as a problem of maximizing the variance of the factors, i.e., the first PC is defined as the linear combination of the variables with maximal variance, and subsequent PCs are similarly defined but with the restriction that their loadings must be orthogonal the loadings of all preceding PCs. Hence PCs have a very nice interpretation where the first PC tries to explain as much of the variation in the data as possible, the second PC then explains as much of the variation left over after extracting the first PC as possible, and so on, see e.g., Johnson and Wichern (2007) for details.

This sequential interpretation of the PC estimation can also be seen in an LS perspective and is the way we choose to define the estimator in Definition 1 below. Note that this sequential approach gives exactly the same estimates as the usual eigenvalue approach to solving the problem, but serves as a more natural starting point for the introduction of the LAD factor estimator.

Definition 1. *PC Factor Estimator: The PC estimates of the first factor and associated loadings are defined as:*

$$(\hat{F}_1, \hat{\Lambda}_1) = \underset{F, \Lambda}{\operatorname{argmin}} V^{\text{LS}}(F, \Lambda; X) \text{ s.t. } \Lambda' \Lambda = 1 \quad (5)$$

Let the residuals from the estimation of the k th factor be defined as e_k , then the subsequent estimates are given as:

$$(\hat{F}_k, \hat{\Lambda}_k) = \underset{F, \Lambda}{\operatorname{argmin}} V^{\text{LS}}(F, \Lambda; e_{k-1}) \text{ s.t. } \Lambda' \Lambda = 1 \quad (6)$$

Hence the PC factor estimates of r factors and associated loadings are given as $\hat{F} = (\hat{F}_1, \dots, \hat{F}_r)$ and $\hat{\Lambda} = (\hat{\Lambda}_1, \dots, \hat{\Lambda}_r)$. Note that the estimated loadings will be orthogonal, i.e., $\hat{\Lambda}' \hat{\Lambda} = I_r$.

2.2 Least absolute deviations

LAD has a long-standing position in the literature as a robust alternative to LS. However, robustness is a somewhat vague and often misunderstood concept. In econometrics we tend to define it in quite broad terms, e.g., Amemiya (1985, 71):

In general, we call an estimator, such as the median, that performs relatively well under distributions heavier-tailed than normal a “robust” estimator.

Since this definition is stated in relative terms we must choose what to compare to. With the obvious choice being LS, we can then for example compute the asymptotic relative efficiency (ARE). Hence under this definition LAD is a robust estimator when compared to LS for distributions sufficiently heavy-tailed.

In statistics much effort has been put into quantifying robustness and thus a number of measures have been developed. One of the most prominent measures is undoubtedly the breakdown point. A treatment of this subject can be found in Huber and Ronchetti (2009) where the breakdown point is defined (p. 8):

The breakdown point is the smallest fraction of bad observations that may cause an estimator to take on arbitrarily large aberrant values.

Naturally the breakdown point can take values between 0 and 0.5. In the case of the linear regression model the breakdown point for an LAD fit is 0.5 when the contamination occurs in the error term (Huber and Ronchetti, 2009, Sec. 11.2.3). This can be compared to the breakdown point for an LS fit which is $1/n$ for a sample size of n . The important thing to remark is that LAD is not robust to contamination or outliers in the explanatory variables. Note that the breakdown point is typically considered a finite sample measure as opposed to ARE.

In the case of our factor model we will assume that outliers occur in the error term, and the presence of outliers will be defined as meaning that the distribution of the error term is heavier-tailed than a normal distribution. In this setting we would expect the LAD estimator to be robust in terms of the definitions above. In cases where both LAD and PC are consistent we would expect LAD to be more efficient since PC is an LS estimator. Furthermore, due to the high breakdown point of LAD we would expect it to perform well even in extreme cases where PC would fail.

The theory of LAD in the regression case is quite well developed. Bassett and Koenker (1978) proved asymptotic normality of the LAD estimator for linear regressions in the i.i.d. case. Alternative proofs have since been given by Pollard (1991) and Phillips (1991). The theory has also been extended to more intricate settings such as correlated errors (Weiss, 1990) and nonlinearity (Weiss, 1991). See Dielman (2005) for a general review of the LAD literature. The robustness of LAD has furthermore been demonstrated to be beneficial in forecasting by Dielman (1986, 1989) and Dielman and Rose (1994).

Considering the LS formulation of the PC estimation problem it seems natural to attempt to estimate the components using LAD in order to gain robustness. The basic idea is to replace the objective function in (4) with

$$V^{\text{LAD}}(F, \Lambda; X) = (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T |X_{it} - \lambda_i F_t| \quad (7)$$

and estimate the model in the same sequential manner as Definition 1. Thus we propose the following LAD factor estimator:

Definition 2. *LAD Factor Estimator: The LAD estimates of the first factor and associated loadings are defined as:*

$$(\hat{F}_1, \hat{\Lambda}_1) = \underset{F, \Lambda}{\operatorname{argmin}} V^{\text{LAD}}(F, \Lambda; X) \text{ s.t. } \Lambda' \Lambda = 1 \quad (8)$$

Let the residuals from the estimation of the k th factor be defined as e_k , then the subsequent estimates are given as:

$$(\hat{F}_k, \hat{\Lambda}_k) = \underset{F, \Lambda}{\operatorname{argmin}} V^{\text{LAD}}(F, \Lambda; e_{k-1}) \text{ s.t. } \Lambda' \Lambda = 1 \quad (9)$$

Hence the LAD factor estimates of r factors and associated loadings are given as $\hat{F} = (\hat{F}_1, \dots, \hat{F}_r)$ and $\hat{\Lambda} = (\hat{\Lambda}_1, \dots, \hat{\Lambda}_r)$.

It is important to note that in contrast to PC this estimator will not produce a solution with orthogonal loadings. This should, however, not be considered a problem since the orthogonality restriction is simply an identification device. One might, however, worry about the fact that the factors are not identified in

this approach. Since we do not impose any identifying restrictions, such as orthogonality, the estimated factors may be rotated in an arbitrary way. In some settings this could be a serious problem, but since we are interested in forecasting it is not as argued above. Furthermore, this approach also reflects our interpretation of the model, i.e., the notion that factors should explain “what is left over” after extracting previous factors, and not be defined by arbitrary identification restrictions. In the case of PC, however, these two cases coincide.

Additionally, a potential pitfall of the sequential approach is that estimation error is accumulated as we estimate more factors. This is not a problem for PC as sequential and simultaneous estimation give numerically the same results. For LAD it could, however, be a problem, but as we shall see in the simulation results below it appears not to be the case.

2.3 Estimation in practice

Even though we have defined the estimator we still need to find a tractable way of performing the estimation. The literature does contain methods for estimating nonlinear LAD (e.g., Koenker and Park, 1996). However, these do not easily generalize to a factor model setting. A simple approach to the problem could instead be to approximate the objective function based on the smoothed LAD estimator of Hitomi and Kagihara (2001). We then define the smoothed objective function as:

$$V^{\text{SLAD}}(F, \Lambda; X) = (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T \sqrt{(X_{it} - \lambda_i F_t)^2 + d^2} \quad (10)$$

for some $d > 0$. Obviously, for $d = 0$ the problem is equivalent to (7) which is not tractable. But by introducing the smoothing parameter d we have a differentiable approximation which we can solve using standard optimization methods. The smoothing parameter naturally controls the approximation error as summarized in the following proposition which follows directly from sub-additivity of the square root function:

Proposition 1. *The approximation error of using (10) will at most be d , i.e.,*

$$0 \leq V^{\text{SLAD}}(F, \Lambda; X) - V^{\text{LAD}}(F, \Lambda; X) \leq d \quad (11)$$

Alternatively, we could exploit the fact that the model is greatly simplified if either the factors or loadings are known. If the loadings are known the factors can be estimated by performing T separate linear LAD regressions, and if the factors are known we can estimate the loadings by n separate linear LAD regressions, all of which can be done using standard methods. This idea gives rise to the following algorithm for solving the estimation problem:

Algorithm 1. *Iterative approach to solving (8). In order to start the algorithm a starting value for the factor is needed. Let this be denoted $\hat{F}^{(0)}$, and in general let superscripts denote the iteration, then iteration v of the algorithm is given as:*

- (i) Calculate the loadings as $\hat{\Lambda}^{(v)} = \operatorname{argmin}_{\Lambda} V^{\text{LAD}}(\hat{F}^{(v-1)}, \Lambda; X)$.
- (ii) Normalize the loadings to have length 1.
- (iii) Calculate the factor as $\hat{F}^{(v)} = \operatorname{argmin}_F V^{\text{LAD}}(F, \hat{\Lambda}^{(v)}; X)$.
- (iv) Check for convergence, e.g., squared difference of the factor estimates: Stop algorithm if $(\hat{F}^{(v)} - \hat{F}^{(v-1)})'(\hat{F}^{(v)} - \hat{F}^{(v-1)})$ is sufficiently small.

Hence steps (i) and (iii) are simply linear regressions. Note that steps (i) and (ii) combined are equivalent to $\hat{\Lambda}^{(i)} = \operatorname{argmin}_{\Lambda} V^{\text{LAD}}(\hat{F}^{(i)}, \Lambda; X)$ s.t. $\hat{\Lambda}^{(i)'} \hat{\Lambda}^{(i)} = 1$. However, since unconstrained estimation is more easily implemented, this step is split in two. Convergence of the algorithm is given in the following proposition, proof of which can be found in the appendix.

Proposition 2. Consider the sequence $\{(\hat{F}^{(\nu)}, \hat{\Lambda}^{(\nu)})\}$ defined by Algorithm 1 and let $\lim_{\nu \rightarrow \infty} (\hat{F}^{(\nu)}, \hat{\Lambda}^{(\nu)}) = (\tilde{\Lambda}, \tilde{F})$ be an accumulation point of the sequence, then:

- (i) $\{(\hat{F}^{(\nu)}, \hat{\Lambda}^{(\nu)})\}$ has at least one accumulation point.
- (ii) If $(\tilde{F}, \tilde{\Lambda})$ and $(\check{F}, \check{\Lambda})$ are two accumulation points then $V^{\text{LAD}}(\tilde{F}, \tilde{\Lambda}) = V^{\text{LAD}}(\check{F}, \check{\Lambda})$.
- (iii) For every accumulation point $(\tilde{F}, \tilde{\Lambda})$

$$\max_F V^{\text{LAD}}(F, \tilde{\Lambda}) = \max_{\Lambda, \Lambda' \Lambda = 1} V^{\text{LAD}}(\tilde{F}, \Lambda) = V^{\text{LAD}}(\tilde{F}, \tilde{\Lambda})$$

Using such a simple iterative scheme is of course not a new idea. In the PC literature a similar approach is used for computing PCs in the Nonlinear Iterated Partial Least Squares (NIPALS) algorithm (see e.g., Esbensen, Geladi, and Wold 1987), and for the LAD case the idea was investigated by Croux et al. (2003).

We thus have two quite different possible approaches for solving the problem at hand. Both are based on approximations. In the first case this is directly evident from the introduction of the smoothing parameter d , and in the second case the approximation arises since in practice we do not have $\nu \rightarrow \infty$. It is of course desirable to minimize the approximation error. In the first case this requires d to be small. However, as d becomes smaller we increase the computational burden and numerical issues arise as we approach the infeasible case of $d=0$. In the second case we reduce the approximation error with each step of the algorithm and hence we should be able to reduce the approximation error to any desired level with only computation time as a constraint.

Based on this it would appear that estimation based on (10) is inferior to Algorithm 1. However, Algorithm 1 is not without pitfalls. Although Proposition 2 guarantees convergence of the method, it still hinges on the assumption that the minima of steps (i) and (iii) of the algorithm can be found. In order to do this we must employ numerical routines and they might still fail. It is our experience that the algorithm is very sensitive to the choice of starting values, and that especially for small sample sizes the method might not converge or falsely report convergence without having reached the optimum. On the other hand, we generally found estimation based on (10) to be quite robust to the choice of starting values.

We have not provided any results on these stability issues as they rarely show up in simulation settings. However, when applying the methods to real data these complications do show up, and the experiences described above are based on the empirical application given later in the paper. It is thus not immediately clear which approach is preferred. However, our recommendation is the following two-step approach: First, use standard numerical routines, e.g., BFGS, to minimize (10) subject to $\Lambda' \Lambda = 1$.³ This requires a set of starting values, one could for example use arbitrary constants or random values. As the method is quite robust to starting values the choice is less important.⁴ Furthermore, we must set d . In this first step some approximation error is acceptable, and therefore the choice is again less important. However, it seems plausible that the approximation error is more easily reduced when more data is available. We therefore set d based on the sample size, namely $d = (nT)^{-1}$. This can be scaled up if convergence problems are encountered. This gives us an initial estimate of the factor and loadings which are then used as starting values in Algorithm 1 in a second step. By doing so we combine the best parts of both approaches, the robustness to starting values of the first, and the potentially smaller approximation error of the second. In the remainder of this paper we will thus refer to estimates obtained from Algorithm 1 using starting values from (10) as LAD factor estimates.

3 Monte Carlo results

The main aim of this paper is to obtain better macroeconomic forecasts, and even though we have not set out to provide an asymptotic analysis of the LAD factor estimator, we will nonetheless still examine the statistical properties of it through a series of Monte Carlo simulations.

³ Note that this can be carried out as an unconstrained minimization problem as long as we normalize the estimated loadings, just as in Algorithm 1, and scale the estimated factors appropriately.

⁴ It could be tempting to use the PC estimates as starting values. However, as we shall see in the simulation results, PC can produce very poor results in the presence of outliers and hence may produce very poor starting values. It is therefore not recommendable.

As already discussed our view on outliers in this paper is that these occur in the error term and not in the factors nor the loadings. Although the latter could also be a problem, in light of what we know about the LAD estimator, it would not seem plausible that it would be able to handle this (as neither would ordinary PC), and hence we refrain from investigating this possibility. Considering the problem at hand we hypothesize the following three possible effects of outliers:

- (i) The factor estimates will become inconsistent or less efficient.
- (ii) Determining the true number of factors will become difficult or impossible.
- (iii) The ordering of the estimated factors will change.

Clearly, the first two points are potentially very critical to forecasting performance, and hence we will investigate both below. The third point is more subtle. In forecasting the ordering is not important if all relevant factors are included in the forecasting model. However, if the number of factors is misspecified it could have a large impact on the performance. We will not address this point here, but postpone it until the empirical application.

Setting up the Monte Carlo framework will be done rather straightforwardly. We want to keep the setup as simple as possible while still encompassing the relevant cases. We will follow Stock and Watson (2002b) and define the data-generating process as:

$$X_{it} = \lambda_i F_t + e_{it} \quad (12)$$

$$(1-aL)e_{it} = (1+b^2)v_{it} + bv_{i+1,t} + bv_{i-1,t} \quad (13)$$

As discussed in Stock and Watson (2002b) it is important to take into account the possibility of correlation across both time and variables since the application we have in mind is macroeconomic forecasting. We therefore include the possibility that the error term e_{it} is correlated, i.e., it will be serially correlated with an AR(1) coefficient of a and cross-series correlated with a (spatial) MA(1) coefficient b . The error is driven by the random variable v_{it} and this is where we will introduce the outliers. We will be using the prototypical outlier distribution for v_{it} , namely a Student- t distribution with a low degree of freedom. In addition to this we will also include results for the limiting case of the Student- t , the standard normal distribution, as the reference of no outliers. Finally, both the factors F_t and loadings λ_i will be generated as independent standard normal variables, and the dimension of F_t , i.e., the number of factors, will be either one or four.⁵

Owing to our ultimate goal of using the estimated factors for forecasting we will also examine how estimation difficulties propagate to the actual forecasts. We therefore generate a uni-variate time series to be forecast

$$y_{t+1} = \iota' F_t + \varepsilon_{t+1} \quad (14)$$

where ι is a vector of ones and ε_{t+1} is an independent standard normal error term.

Hence the different simulation scenarios are defined in terms of the parameters a and b , the distribution of v_{it} , the number of factors r , the number of variables n , and the sample size T . Results will be presented both for the traditional PC factor estimates and the proposed LAD factor estimates. In the latter case the estimation is carried out as described in the previous section. In addition to this we will also be comparing to the effects of screening the data. As mentioned in the introduction screening is common practice in the literature and hence PC estimation based on screened data is the natural alternative to the LAD estimator. We will use one of the most typical screening rules, i.e., define outliers as observations exceeding the median of the series by more than six times the interquartile range, and since our estimation method cannot handle missing observations we will replace any outlier with the median of the five observations preceding it. This is the method used in e.g., Breitung and Eickmeier (2011), and Stock and Watson (2009). PC estimates obtained using data screened according to this rule will be labelled PC-S.

⁵ Note that this choice is merely in the interest of space. The simulation study has also been conducted for other values of r and the results were found to be in line with the ones presented here.

3.1 Precision of the factor estimates

Before examining the Monte Carlo results let us briefly review what we know about the LAD estimator in the classical i.i.d. linear regression case. Assuming the regression model has an error term with a symmetric distribution $f(\cdot)$ centered at zero and finite variance σ^2 , and that the explanatory variables are well-behaved, then both LAD and LS provide estimates that are asymptotically normal. Hence in these cases we can directly compare the two using e.g., their ARE. The ARE of LAD with respect to LS is $ARE = \sigma^2 4f(0)^2$, which for $t(3)$ $ARE \approx 1.62$, for $t(4)$ $ARE \approx 1.13$, for $t(5)$ $ARE \approx 0.96$, and for $N(0, 1) \approx 0.64$. In addition to this it is important to also recall that even if no moments of the error terms exist, e.g., $t(1)$ errors, the LAD estimator will still be consistent and asymptotically normal. In contrast if the degrees of freedom is less than three the distribution will not have a finite variance and hence LS cannot be expected to work well. Therefore, if the error terms are Student- t distributed, then in the classical model we expect LAD to outperform LS when the degrees of freedom in the distribution is less than five. Furthermore, in the extreme cases with degrees of freedom less than three we expect LS to work very poorly or outright break down. In the case of the LAD factor model, however, things are clearly more complicated. But we should remember that the model is still basically a regression model, so we do expect it to share some traits with the classical regression model.

Assessing the precision of the factor estimates is done, as is common in the literature, by computing the trace R^2 of a multivariate regression of the factor estimates on the true factors

$$R^2 = \text{tr}[\hat{F}'F(F'F)^{-1}F'\hat{F}] / \text{tr}[\hat{F}'\hat{F}] \quad (15)$$

and averaging this across Monte Carlo replications. Hence we obtain a statistic that measure how well the estimated factors span the space of the true factors, with values as close to 1 as possible being the desired goal.

For each Monte Carlo run we also perform a 1-step out-of-sample forecast based on (14). Hence we fit the forecasting equation (14) by OLS using data until $T-1$ and obtain the forecast \hat{y}_{T+1} by fitting the equation using the data for period T . On the basis of this we can compute the mean square forecast error (MSFE) where the mean is taken across Monte Carlo replications. The statistic presented in the tables is the infeasible MSFE obtained by estimating (14) using the true factors relative to the MSFE obtained using the estimated factors. Thus a relative MSFE close to 1 indicates that we are close to the infeasible forecasts we could have made if the factors were known, whereas a smaller relative MSFE suggests that the forecasting performance is adversely affected by the poor estimation of the factors.

We start by considering the simplest case where the error terms are i.i.d., hence $a=b=0$. In Table 1 results are given for both Student- t and normal errors with either one or four factors. We quite clearly see the resemblance to the usual linear regression case. Judging by the R^2 we see that for the $t(1)$ errors PC is inconsistent and LAD performs very well. Attempting to mitigate the outlier problem by screening the data does have some effect, but PC-S is still clearly inferior to LAD. The precision (or lack thereof) of the factor estimates carry over to the forecasting performance as measured by the relative MSFE. Comparing the case of $r=1$ to $r=4$ we see a slight drop in performance. This is not surprising since, in the latter case, we are estimating four times as many parameters with an unchanged sample size. However, the comparable performance does illustrate the validity of the sequential estimation approach.

Moving to the $t(2)$ distribution we notice an increase in the PC performance, although still clearly below LAD, and in spite of the increase the forecasting performance of PC is still very low. Interestingly, the performance of PC-S is now only slightly below that of LAD. In the case of the $t(3)$ and $t(5)$ distributions all three methods perform very similarly, but especially in the $t(5)$ case we see that LAD is losing ground to PC. The distance between the estimators is further increased in the case of normality where PC is the best performing method (although not by a large margin). Note that in the case of normality PC and PC-S are identical since there of course are no outliers to screen out.

Table 1 Precision of the factor estimates in the case of i.i.d. errors.

<i>n</i>	<i>T</i>	Dist.	<i>r=1</i>						<i>r=4</i>					
			Factor R^2			Rel. MSFE			Factor R^2			Rel. MSFE		
			PC	PC-S	LAD	PC	PC-S	LAD	PC	PC-S	LAD	PC	PC-S	LAD
25	50	<i>t</i> (1)	0.02	0.32	0.77	0.01	0.60	0.83	0.08	0.45	0.47	0.00	0.32	0.32
25	100	<i>t</i> (1)	0.01	0.41	0.83	0.02	0.59	0.85	0.04	0.49	0.58	0.00	0.33	0.41
50	100	<i>t</i> (1)	0.01	0.43	0.92	0.03	0.61	0.92	0.04	0.56	0.88	0.00	0.43	0.63
50	200	<i>t</i> (1)	0.01	0.53	0.93	0.00	0.66	0.93	0.02	0.62	0.89	0.00	0.48	0.70
100	200	<i>t</i> (1)	0.01	0.49	0.97	0.00	0.66	0.96	0.02	0.63	0.94	0.00	0.56	0.85
150	200	<i>t</i> (1)	0.01	0.42	0.98	0.00	0.60	0.97	0.02	0.62	0.96	0.00	0.61	0.88
25	50	<i>t</i> (2)	0.31	0.81	0.88	0.41	0.85	0.90	0.53	0.82	0.84	0.20	0.54	0.56
25	100	<i>t</i> (2)	0.29	0.83	0.89	0.11	0.87	0.90	0.50	0.84	0.86	0.12	0.59	0.60
50	100	<i>t</i> (2)	0.31	0.90	0.94	0.11	0.92	0.95	0.55	0.91	0.92	0.22	0.75	0.78
50	200	<i>t</i> (2)	0.32	0.91	0.95	0.11	0.90	0.94	0.55	0.91	0.93	0.11	0.72	0.77
100	200	<i>t</i> (2)	0.36	0.95	0.97	0.10	0.96	0.97	0.58	0.95	0.96	0.05	0.83	0.90
150	200	<i>t</i> (2)	0.36	0.96	0.98	0.12	0.96	0.98	0.59	0.97	0.97	0.06	0.88	0.91
25	50	<i>t</i> (3)	0.79	0.88	0.89	0.77	0.88	0.90	0.84	0.89	0.87	0.52	0.67	0.65
25	100	<i>t</i> (3)	0.83	0.89	0.91	0.86	0.87	0.87	0.86	0.90	0.88	0.65	0.69	0.67
50	100	<i>t</i> (3)	0.90	0.94	0.95	0.92	0.93	0.93	0.92	0.94	0.93	0.69	0.83	0.82
50	200	<i>t</i> (3)	0.91	0.95	0.95	0.86	0.96	0.97	0.93	0.95	0.94	0.38	0.81	0.83
100	200	<i>t</i> (3)	0.95	0.97	0.98	0.96	0.98	0.99	0.96	0.97	0.97	0.33	0.92	0.93
150	200	<i>t</i> (3)	0.96	0.98	0.98	0.97	0.99	1.00	0.97	0.98	0.98	0.72	0.95	0.96
25	50	<i>t</i> (5)	0.91	0.91	0.90	0.94	0.94	0.92	0.92	0.92	0.88	0.75	0.76	0.66
25	100	<i>t</i> (5)	0.92	0.92	0.91	0.94	0.95	0.96	0.93	0.93	0.90	0.77	0.77	0.72
50	100	<i>t</i> (5)	0.96	0.96	0.95	0.98	0.98	0.97	0.96	0.96	0.94	0.86	0.86	0.80
50	200	<i>t</i> (5)	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.86	0.86	0.84
100	200	<i>t</i> (5)	0.98	0.98	0.98	0.99	0.99	0.99	0.98	0.98	0.97	0.94	0.94	0.91
150	200	<i>t</i> (5)	0.98	0.98	0.98	0.99	0.99	1.00	0.98	0.98	0.98	0.96	0.96	0.93
25	50	N	0.94	0.94	0.91	0.91	0.91	0.88	0.94	0.94	0.90	0.87	0.87	0.74
25	100	N	0.95	0.95	0.92	0.92	0.92	0.90	0.95	0.95	0.91	0.85	0.85	0.73
50	100	N	0.97	0.97	0.95	0.94	0.94	0.94	0.97	0.97	0.95	0.94	0.93	0.84
50	200	N	0.97	0.97	0.96	0.94	0.94	0.93	0.98	0.98	0.95	0.91	0.91	0.83
100	200	N	0.98	0.98	0.98	0.96	0.96	0.95	0.99	0.99	0.97	0.95	0.95	0.93
150	200	N	0.99	0.99	0.98	0.96	0.96	0.95	0.99	0.99	0.98	0.97	0.97	0.93

The results are based on 1,000 Monte Carlo replications. The Factor R^2 column is the (trace) R^2 from a regression of the estimated factor(s) on the true factor(s). The Rel. MSFE column is the mean squared forecast error from the infeasible forecast using the true factors relative to the MSFE using the estimated factors.

Abandoning the i.i.d. premise of these results we now turn to Table 2 where we consider cases of both serial and cross-sectional correlation. We choose to focus on the case of four factors and compare normality to $t(3)$ errors. Overall we see a decline in performance compared to the i.i.d. case. This can, however, partly be explained by the fact that the unconditional variance of the error term is larger than in the i.i.d. case whereas the variance of the factors is unchanged, and hence we would expect lower precision in the estimates. That being said, in the first scenario where we have only a modest degree of persistence in the errors ($a=0.5$, $b=0$), we see that the performance is quite close to the i.i.d. case and the results are very similar. LAD performs best under $t(3)$ and PC is best under normality. In the second scenario we increase the persistence to $a=0.9$ and overall the performance drops as expected. It is, however, interesting to notice that for the $t(3)$ errors the precision (as measured by R^2) of PC and PC-S does not increase with the sample size. This is in contrast to LAD and illustrates that in cases of high persistence PC is further hampered by outliers. In the final two scenarios we also include correlation across series ($b \neq 0$). Besides the general level of performance, including this only gives rise to minor differences.

Table 2 Precision of the factor estimates in the case of non-i.i.d. errors and four factors.

<i>n</i>	<i>T</i>	<i>a</i>	<i>b</i>	<i>t</i> (3) errors						Normal errors					
				Factor <i>R</i> ²			Rel. MSFE			Factor <i>R</i> ²			Rel. MSFE		
				PC	PC-S	LAD	PC	PC-S	LAD	PC	PC-S	LAD	PC	PC-S	LAD
25	50	0.5	0	0.79	0.83	0.81	0.45	0.52	0.48	0.93	0.93	0.89	0.81	0.81	0.69
25	100	0.5	0	0.81	0.86	0.83	0.58	0.62	0.59	0.94	0.94	0.89	0.81	0.81	0.70
50	100	0.5	0	0.89	0.92	0.91	0.73	0.77	0.75	0.97	0.97	0.94	0.91	0.91	0.86
50	200	0.5	0	0.90	0.93	0.91	0.71	0.79	0.74	0.97	0.97	0.94	0.90	0.91	0.82
100	200	0.5	0	0.95	0.96	0.95	0.82	0.88	0.86	0.98	0.98	0.97	0.95	0.95	0.91
150	200	0.5	0	0.96	0.97	0.97	0.87	0.91	0.90	0.99	0.99	0.98	0.96	0.96	0.93
25	50	0.9	0	0.35	0.36	0.35	0.28	0.28	0.28	0.67	0.67	0.62	0.38	0.38	0.35
25	100	0.9	0	0.30	0.30	0.30	0.27	0.27	0.26	0.67	0.67	0.61	0.38	0.38	0.34
50	100	0.9	0	0.29	0.30	0.32	0.28	0.29	0.29	0.74	0.74	0.69	0.46	0.46	0.40
50	200	0.9	0	0.33	0.34	0.38	0.29	0.30	0.30	0.83	0.83	0.77	0.65	0.65	0.55
100	200	0.9	0	0.35	0.37	0.44	0.30	0.30	0.32	0.90	0.90	0.86	0.78	0.78	0.71
150	200	0.9	0	0.36	0.38	0.47	0.31	0.33	0.35	0.93	0.93	0.90	0.83	0.83	0.79
25	50	0	1	0.31	0.37	0.34	0.22	0.24	0.23	0.67	0.67	0.58	0.43	0.43	0.34
25	100	0	1	0.28	0.35	0.33	0.27	0.28	0.27	0.69	0.69	0.59	0.46	0.46	0.38
50	100	0	1	0.42	0.55	0.56	0.30	0.34	0.35	0.85	0.85	0.77	0.67	0.67	0.55
50	200	0	1	0.44	0.59	0.60	0.31	0.38	0.38	0.86	0.86	0.79	0.63	0.63	0.52
100	200	0	1	0.65	0.80	0.81	0.41	0.57	0.61	0.93	0.93	0.89	0.77	0.77	0.66
150	200	0	1	0.75	0.87	0.87	0.47	0.65	0.67	0.95	0.95	0.92	0.84	0.85	0.77
25	50	0.5	1	0.23	0.26	0.24	0.21	0.25	0.24	0.54	0.54	0.48	0.35	0.35	0.29
25	100	0.5	1	0.20	0.23	0.21	0.23	0.23	0.23	0.56	0.56	0.48	0.37	0.37	0.31
50	100	0.5	1	0.27	0.34	0.34	0.25	0.28	0.28	0.74	0.74	0.65	0.50	0.50	0.41
50	200	0.5	1	0.29	0.37	0.38	0.25	0.28	0.28	0.78	0.78	0.70	0.57	0.57	0.48
100	200	0.5	1	0.46	0.61	0.63	0.32	0.40	0.39	0.90	0.90	0.85	0.69	0.69	0.61
150	200	0.5	1	0.58	0.74	0.76	0.38	0.49	0.51	0.93	0.93	0.89	0.76	0.76	0.68

The results are based on 1,000 Monte Carlo replications. The Factor *R*² column is the (trace) *R*² from a regression of the estimated factor(s) on the true factor(s). The Rel. MSFE column is the mean squared forecast error from the infeasible forecast using the true factors relative to the MSFE using the estimated factors.

3.2 Determining the number of factors

We now turn to the problem of determining the correct number of factors in the true model. This is often tackled as a model selection problem, and e.g., Stock and Watson (2002b) used the Bayesian Information Criterion (BIC) applied to the forecasting equation. The intuition behind this choice is described in an earlier version of their paper, where they proposed a modified version of the BIC and showed that it would asymptotically select the correct number of factors. However, their Monte Carlo results showed that the ordinary BIC outperformed their modified version, and hence the ordinary BIC is often used. These results can be found in Stock and Watson (1998). Their work has, however, since been surpassed by Bai and Ng (2002), and the information criteria suggested in their work has become the preferred approach in the literature. We will therefore focus on their information criterion which is defined as:

$$IC_j(k) = \log(V^{LS}(\hat{F}, \hat{\Lambda}; X)) + kg_j(n, T) \tag{16}$$

with three possible penalty terms:

$$g_1(n, T) = \left(\frac{n+T}{nT}\right) \log\left(\frac{nT}{n+T}\right) \tag{17}$$

$$g_2(n, T) = \left(\frac{n+T}{nT} \right) \log(C_{nT}^2) \quad (18)$$

$$g_3(n, T) = \left(\frac{\log(C_{nT}^2)}{C_{nT}^2} \right) \quad (19)$$

where k is the number of factors in the estimated model and $C_{nT}^2 = \min(n, T)$. We can then estimate the true number of factors as $\hat{r} = \operatorname{argmin}_{1 \leq k \leq k_{\max}} IC_j(k)$.

Although their results were derived in relation to PC, it is important to note that according to Bai and Ng (2002, Corollary 2) their results hold for any consistent estimator regardless of the estimation method used as long as the convergence rate C_{nT} is correctly specified. So even though we have not provided a formal proof of the consistency of the LAD estimator, and hence do not know what the rate of convergence might be, we would still expect IC_j to perform well for the LAD estimator in light of the Monte Carlo results provided above.⁶

The Monte Carlo setup is the same as outlined above. We set the true number of factors to four, i.e., $r=4$ and the maximum number of estimated factors to 12, i.e., $k_{\max}=12$. The number of factors is estimated using the three variants of IC_j and the PC, PC-S and LAD methods.

In Table 3 results are given for the i.i.d. case, i.e., where $a=b=0$. Again, just as the precision of the PC estimates is poor in the case of $t(1)$ errors, the ability to correctly determine the number of factors is also adversely affected. The estimated number of factors diverge to the maximum of $k_{\max}=12$ as the sample size increases. Both LAD and PC-S appear to have the opposite problem. PC-S sets the true number of factors to one in all cases, and LAD seems to converge to one as the sample size increases. However, for small sample sizes LAD is closer to the true value of four. In the case of $t(2)$ errors we see both LAD and PC-S increasing their performance whereas PC still provides more or less divergent results. In the remaining cases all methods are quite close to the true value. One small difference that should be noted is that for the smallest sample size ($n=25, T=50$) IC_3 tends to overestimate the number of factors for the PC and PC-S estimates. This is, however, not a problem for the LAD estimates.

In Table 4 we compare $t(3)$ errors to normality in three different non-i.i.d. scenarios. The results are much in line with what was observed in Table 2 regarding precision. In the first scenario with only moderate persistence ($a=0.5, b=0$) PC-S and LAD perform best with the PC estimates being inflated due to the outliers in the $t(3)$ distribution. It should, however, again be noted that for PC and PC-S, IC_3 overestimates the number of factors. In the second scenario we only include cross-sectional correlation ($a=0, b=1$). What is interesting about this case is that LAD outperforms PC (and PC-S) under normality where especially IC_3 tends to diverge to the maximum of 12 factors. In the case of $t(3)$ errors we do, nonetheless, notice that LAD has a tendency to underestimate the number of factors. Finally, in the last scenario we have both time and cross-sectional correlation ($a=0.5, b=1$). The results here are quite close to the second scenario indicating that correlation across series is more crucial to the performance than correlation across time (at least in the case of only moderate persistence).

Clearly, in real-world forecasting applications we cannot separate the issues of determining the number of factors and estimating the factors precisely. Hence we conclude the simulation study by examining the combined effect outliers have on precision and selecting the number of factors. We look at the relative MSFEs, as in Tables 1 and 2, but where we now select the number of factors using the IC_j criterion. The results are much as one would expect, and for this reason the tables have been included in the appendix. The i.i.d. case is covered in Table 7. From Table 3 we know that in cases of Student- t errors with a low degree of freedom it is difficult to determine the true number of factors and this shines through in the forecasting

⁶ It might seem more intuitive to use V^{LAD} in (16) for the LAD estimator. However, in addition to having no theoretical justification for using it, our (non-reported) Monte Carlo results showed comparable, but slightly lower performance, when compared to using V^{LS} , hence we use the traditional IC_j .

Table 3 Estimated number of factors in the case of i.i.d. errors and four factors.

<i>n</i>	<i>T</i>	Dist.	PC			PC-S			LAD		
			IC ₁	IC ₂	IC ₃	IC ₁	IC ₂	IC ₃	IC ₁	IC ₂	IC ₃
25	50	<i>t</i> (1)	5.46	4.68	7.02	1.00	1.00	1.00	2.67	2.49	2.93
25	100	<i>t</i> (1)	7.17	6.65	8.05	1.00	1.00	1.00	2.59	2.46	2.69
50	100	<i>t</i> (1)	8.17	7.49	9.40	1.00	1.00	1.00	1.49	1.46	1.59
50	200	<i>t</i> (1)	9.55	9.21	10.13	1.00	1.00	1.00	1.43	1.40	1.54
100	200	<i>t</i> (1)	10.35	10.10	10.86	1.00	1.00	1.00	1.12	1.11	1.14
150	200	<i>t</i> (1)	10.71	10.52	11.10	1.00	1.00	1.00	1.04	1.04	1.07
25	50	<i>t</i> (2)	5.39	3.50	11.28	1.53	1.25	2.21	1.52	1.28	2.24
25	100	<i>t</i> (2)	5.36	4.41	8.67	1.50	1.36	1.87	1.45	1.32	1.87
50	100	<i>t</i> (2)	5.78	4.97	9.06	2.07	1.72	2.88	1.86	1.52	2.73
50	200	<i>t</i> (2)	5.92	5.56	7.08	2.28	2.09	2.81	2.14	1.92	2.67
100	200	<i>t</i> (2)	6.68	6.25	9.13	3.22	2.99	3.65	3.13	2.89	3.59
150	200	<i>t</i> (2)	7.13	6.56	10.93	3.50	3.31	3.83	3.46	3.25	3.81
25	50	<i>t</i> (3)	4.43	4.04	9.58	3.66	3.39	4.55	3.11	2.63	3.72
25	100	<i>t</i> (3)	4.33	4.21	4.78	3.83	3.75	3.93	3.51	3.36	3.76
50	100	<i>t</i> (3)	4.36	4.26	4.95	3.99	3.98	4.00	3.98	3.96	4.00
50	200	<i>t</i> (3)	4.27	4.24	4.44	4.00	4.00	4.00	4.00	4.00	4.00
100	200	<i>t</i> (3)	4.32	4.26	4.68	4.00	4.00	4.00	4.00	4.00	4.00
150	200	<i>t</i> (3)	4.32	4.25	5.00	4.00	4.00	4.00	4.00	4.00	4.00
25	50	<i>t</i> (5)	4.03	3.98	6.82	4.00	3.97	5.90	3.81	3.68	3.94
25	100	<i>t</i> (5)	4.02	4.01	4.03	4.00	4.00	4.01	3.96	3.94	3.99
50	100	<i>t</i> (5)	4.01	4.00	4.04	4.00	4.00	4.00	4.00	4.00	4.00
50	200	<i>t</i> (5)	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
100	200	<i>t</i> (5)	4.01	4.00	4.01	4.00	4.00	4.00	4.00	4.00	4.00
150	200	<i>t</i> (5)	4.00	4.00	4.02	4.00	4.00	4.00	4.00	4.00	4.00
25	50	N	4.00	4.00	4.83	4.00	4.00	4.83	3.98	3.96	3.99
25	100	N	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.99	4.00
50	100	N	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
50	200	N	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
100	200	N	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
150	200	N	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00

The results are based on 1,000 Monte Carlo replications.

results as the performance is lower compared to Table 1. There is one interesting exception, however. Consider PC with *t*(2) errors, here the forecasting performance has gone slightly up. This could very well be due to IC₃ overestimating the number of factors and hence including more information in the forecasting equation. The case of LAD with *t*(1) errors for the smallest sample size is also interesting as here there is no forecasting power. This is in contrast to PC-S which does quite well. We know from the previous results that the precision of LAD and PC-S are close and that LAD for this sample size is closer (on average) to selecting the true number of factors. However, it turns out that occasionally LAD includes too many factors and that some of these contain very large values which drives up the MSFE. However, this appears to only be a problem for this small sample size.

Table 8 gives similar results for the non-i.i.d. case. Again we see clear effects of incorrectly selecting the number of factors. Consider for example the case of *t*(3) errors and *a*=0, *b*=1. Here we know from Table 2 that PC underperforms in terms of MSFE. This is no longer as evident in Table 8 as the number of factors is often overestimated for PC whereas it is often underestimated in the other cases. This decreases the difference and in some cases causes PC to perform better than PC-S and LAD. Thus overestimating *r* can be preferable to underestimating it.

Summing up the results in this section we in general see many of the same characteristics of the LAD estimator in the factor setting as we would in a standard linear regression setting. In the i.i.d. case LAD is

Table 4 Estimated number of factors in the case of non-i.i.d. errors and four factors.

n	T	a	b	Dist.	PC			PC-S			LAD		
					IC ₁	IC ₂	IC ₃	IC ₁	IC ₂	IC ₃	IC ₁	IC ₂	IC ₃
25	50	0.5	0	$t(3)$	4.61	3.70	9.75	3.57	2.99	6.18	2.58	2.00	3.55
25	100	0.5	0	$t(3)$	4.37	4.13	5.87	3.68	3.50	4.00	3.05	2.78	3.52
50	100	0.5	0	$t(3)$	4.43	4.28	7.20	3.97	3.93	4.30	3.94	3.84	4.07
50	200	0.5	0	$t(3)$	4.29	4.24	4.52	3.99	3.99	4.00	3.99	3.98	4.00
100	200	0.5	0	$t(3)$	4.33	4.26	5.18	4.00	4.00	4.01	4.00	4.00	4.00
150	200	0.5	0	$t(3)$	4.38	4.28	7.74	4.00	4.00	4.07	4.00	4.00	4.00
25	50	0.5	0	N	4.09	4.00	7.49	4.09	4.00	7.47	3.92	3.85	3.99
25	100	0.5	0	N	4.00	4.00	4.11	4.00	4.00	4.11	3.99	3.99	4.00
50	100	0.5	0	N	4.00	4.00	4.45	4.00	4.00	4.45	4.00	4.00	4.00
50	200	0.5	0	N	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
100	200	0.5	0	N	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
150	200	0.5	0	N	4.00	4.00	4.29	4.00	4.00	4.29	4.00	4.00	4.00
25	50	0	1	$t(3)$	11.89	8.21	12.00	4.03	1.55	8.37	1.33	1.14	2.88
25	100	0	1	$t(3)$	12.00	11.99	12.00	4.22	2.52	7.68	1.26	1.18	1.74
50	100	0	1	$t(3)$	2.93	1.41	12.00	1.22	1.04	6.90	1.05	1.02	1.63
50	200	0	1	$t(3)$	3.25	1.98	12.00	1.31	1.14	2.76	1.05	1.03	1.31
100	200	0	1	$t(3)$	3.73	2.68	11.90	2.34	1.67	3.98	1.49	1.16	3.39
150	200	0	1	$t(3)$	4.29	3.46	12.00	3.36	2.44	4.21	2.50	1.67	3.98
25	50	0	1	N	11.80	8.15	12.00	11.80	8.13	12.00	1.65	1.30	3.21
25	100	0	1	N	12.00	12.00	12.00	12.00	12.00	12.00	2.09	1.73	2.96
50	100	0	1	N	4.05	3.85	12.00	4.05	3.85	12.00	3.43	3.00	3.97
50	200	0	1	N	4.01	3.98	11.99	4.01	3.98	11.99	3.84	3.74	3.97
100	200	0	1	N	4.00	4.00	8.91	4.00	4.00	8.90	4.00	4.00	4.00
150	200	0	1	N	4.00	4.00	11.55	4.00	4.00	11.55	4.00	4.00	4.00
25	50	0.5	1	$t(3)$	8.72	3.86	11.57	4.47	1.83	8.56	1.51	1.25	3.44
25	100	0.5	1	$t(3)$	11.91	11.61	11.97	6.61	4.34	9.20	1.40	1.22	2.37
50	100	0.5	1	$t(3)$	4.98	1.48	11.99	1.53	1.06	9.97	1.15	1.09	2.33
50	200	0.5	1	$t(3)$	5.31	1.74	11.99	1.23	1.06	6.83	1.07	1.04	1.29
100	200	0.5	1	$t(3)$	2.97	1.68	12.00	1.52	1.08	9.29	1.07	1.04	2.63
150	200	0.5	1	$t(3)$	3.73	2.05	12.00	2.24	1.28	11.26	1.28	1.06	4.09
25	50	0.5	1	N	9.56	4.61	11.83	9.55	4.60	11.83	1.49	1.16	3.39
25	100	0.5	1	N	11.99	11.86	12.00	11.99	11.86	12.00	1.61	1.34	2.63
50	100	0.5	1	N	6.66	3.65	12.00	6.65	3.65	12.00	2.56	1.89	4.54
50	200	0.5	1	N	6.02	4.10	12.00	6.00	4.10	12.00	3.21	2.90	3.80
100	200	0.5	1	N	4.08	4.00	12.00	4.08	4.00	12.00	3.99	3.97	4.21
150	200	0.5	1	N	4.02	4.00	12.00	4.02	4.00	12.00	4.00	4.00	5.09

The results are based on 1,000 Monte Carlo replications.

capable of estimating the factors even in extreme case of $t(1)$ errors, whereas PC require at least 3 degrees of freedom in the Student- t distribution for acceptable results. The closer we get to normality, the better PC performs with a turning point around 5 degrees of freedom. Adding correlation to the error term lowers performance, and in particular the combination of high persistence and heavy tails hampers performance of the PC estimator. Determining the number of factors is also affected by outliers. In general LAD also outperforms PC here, though LAD has difficulties giving correct estimates in the extreme cases. Furthermore, there is a tendency for IC₃ to overestimate the number of factors (in some cases quite severely) when applied to the PC and PC-S estimates. When considering the joint effects of precision and selecting the number of factors on forecasting performance the consequences of incorrectly selecting r is further emphasized. Screening before applying PC lessens the problem in all cases but is not as effective as using the LAD estimator.

4 Empirical application

Forecasting in the factor model was already alluded to in the previous section. We shall now consider our empirical application and hence return to it in more detail. The forecasting framework used is very similar to what is often applied in the literature, e.g., Stock and Watson (2002b). The aim is to obtain h -step-ahead forecasts of a number of macroeconomic variables. In order to avoid having to specify a process for the factors we will only consider direct forecasting. In general we will consider the following forecasting model:

$$y_{t+h}^h = \alpha_h + \beta_h(L)' F_t + \gamma_h(L) y_t + \epsilon_{t+h}^h \quad (20)$$

where $\beta_h(L)$ and $\gamma_h(L)$ are lag polynomials. The explicit dependence on the forecast horizon should be noted as the model is specific to the horizon.

The variables to be forecast are assumed to be either I(1) or I(2) in logarithms. Let z_t be the original variable of interest recorded at a monthly frequency, then in the case of I(1) we define the h -step-ahead variable as:

$$y_{t+h}^h = (1200/h) \log(z_{t+h}/z_t) \quad (21)$$

i.e., annualized growth over the horizon in percent. In the case of I(2) we define it as:

$$y_{t+h}^h = (1200/h) \log(z_{t+h}/z_t) - 1200 \log(z_t/z_{t-1}) \quad (22)$$

i.e., the difference between annualized growth over the horizon in percent and annualized growth over the last month.

Since the forecasting model (20) contains the unobserved factors F_t , estimation is done in a two-step approach. First the factors are estimated using either PC or the proposed LAD procedure. Then in the second step (20) is estimated by OLS with the estimated factors in place of the unobserved true ones. Hence, for a dataset ending at time T we obtain the forecast of $T+h$ by fitting the forecasting equation using the OLS estimates:

$$\hat{y}_{T+h|T}^h = \hat{\alpha}_h + \sum_{j=1}^k \hat{\beta}_{h,j} \hat{F}_{T,j} + \sum_{j=1}^p \hat{\gamma}_{h,j} y_{T-j+1} \quad (23)$$

Clearly, in order to fully specify the forecasting model, k and p need to be specified, i.e., the number of factors to include and how many lags of y_t to include. Note that in the following we will also allow for the possibility of including no lags of y_t , this case will be referred to as $p=0$. Either BIC or the IC_j criterion by Bai and Ng (2002) as described earlier will be used to select the number of factors, and BIC to select the number of lags. In addition to this the possibility of simply fixing k and p will also be considered.

Before estimation can be carried out the data need to be prepared. Recall that the factor model assumes the variables to be stationary and centered at zero. We will need to ensure that the data reflect this. Therefore all series are transformed to be stationary, the details on which can be found in the data appendix. For the PC estimation we further center all variables such that they have mean zero. The intuition behind this is that the factor estimation is basically a regression without an intercept, centering corrects for this. In some cases the mean of the series is even referred to as the zeroth PC. The LAD estimation will also require centering. However, since LAD estimates the median, centering will be done at the median.

Finally, the data have to be scaled. The need for this is not immediately apparent from the general model setup, but is in fact quite critical. Recall that the solution to the PC estimation problem is the loadings that maximize the variances of the individual factors. Therefore it is quite clear that if some variables have a high variance and others a low variance the latter will be crowded out by the former. To avoid this problem it is common to scale all variables to have unit variances. By doing this we ensure that the “choice” of variables in the factors is not driven by differences in the variances.

Although scaling the variables to have unit variances is the obvious choice, we must remember that the variance is not a robust measure of dispersion. In the case of the LAD estimator one could imagine more appropriate scalings. A popular robust alternative to the variance (or rather the standard deviation) is the median absolute deviation (MAD):⁷

$$\text{MAD}(x) = \text{med}[|x_i - \text{med}[x]|] \quad (24)$$

Consequently, we will consider the following methods for estimating the factors. PC: Estimation by PC using data centered at the mean and scaled to have unit variance; LAD: Estimation by LAD using data centered at the median and scaled to have unit variance; LAD-MAD: Estimation by LAD using data centered at the median and scaled to have unit MAD. We will also compare to the effects of screening the data, hence we also have a PC-S model, i.e., estimation by PC using data centered at the mean, scaled to have unit variance and screened for outliers. The screening procedure used is the same as in the previous section.⁸

In addition to these factor models we will also consider a number of benchmark models. Firstly, our main benchmark will be an AR model, i.e., (20) without factors, all results will be relative to this model. Secondly, we include a naïve scenario where the forecast is computed as the unconditional mean (denoted U.Mean) or median (denoted U.Median) of either the entire data series or a window containing the last h observations (abbreviated W). Lastly, we also include what could be referred to as a low-dimensional factor model based on the economic indexes developed by Stock and Watson (1989). The model is the same basic factor model as before, i.e., (2), however, we will now specify the dynamics of the factors and error terms. Written in general terms, we have

$$X_{it} = \lambda_i F_t + e_{it} \quad (25)$$

$$\Gamma(L)F_t = \eta_t \quad (26)$$

$$\delta_i(L)e_{it} = v_{it} \quad (27)$$

where $\Gamma(L)$ and $\delta_i(L)$ are lag polynomials and $(v_{1t}, \dots, v_{nt}, \eta_{1t}, \dots, \eta_{nt})$ are i.i.d. normal and mutually independent error terms.

Stock and Watson (1989) only considered the case of a single factor, we will also limit ourselves to this case. Furthermore, (26) and (27) will be specified as AR(1) processes.⁹ Identification will be achieved by setting the top block of the loading matrix equal to the identity matrix, i.e., $\Lambda = (I_r, \Lambda_2)'$. Even though this identification scheme is different to the one used in e.g., PC estimation, it is merely a different rotation and for forecasting purposes not important. Under these assumptions we can use the Kalman filter to compute the likelihood and thus obtain the maximum likelihood estimates of the model. See Stock and Watson (2006) for further details. The interest of Stock and Watson (1989) was to build coincident and leading economic indexes by using the model (25)–(27). Our interest is to forecast key macroeconomic variables, and to ensure comparability we will again compute these as direct forecasts. Hence we use the Kalman smoother to extract the factors and use these in (23), the same forecasting relationship used for the other factor models.

Such a model, where we fully specify the dynamics of the factors and error terms, is an example of confirmatory factor analysis, as opposed to exploratory factor analysis often conducted using PC. As pointed out by a referee the LAD approach can also be seen in a confirmatory factor analysis context as we are implicitly assuming that the errors have fat tails (or equivalently that they contain outliers) when choosing to use this approach.

⁷ Often MAD is scaled by a constant (≈ 1.4826) to make it consistent for the standard deviation in the case of Gaussian data. See Huber and Ronchetti (2009) for details. In our case this is of course irrelevant.

⁸ I could perhaps be tempting to use MAD scaling in conjunction with the PC estimator as a shortcut to robustness. However, as argued this is an illogical choice and indeed it produces very poor results. We therefore do not include it in the paper.

⁹ Other specifications, two factors and higher order (V)AR processes, were also considered but were not supported by the data.

4.1 The dataset

Many of the papers written on the subject of factor model forecasting use the same few datasets available online. More specifically data from Stock and Watson (2002b, 2005), or an updated version of the latter from Ludvigson and Ng (2010). For this paper, however, a more up-to-date dataset shall be considered, and hence we have collected a new dataset. All variables used are either directly from the Federal Reserve Economic Data (FRED) database made available by the Federal Reserve of St. Louis or have been computed on the basis of these. We have attempted to keep the composition of the dataset close to the original Stock and Watson datasets. A total of 111 monthly US macroeconomic variables are included in our dataset, and hence it is slightly smaller than the typical datasets. The data have been taken in seasonally adjusted form from FRED and hence we have not performed any adjustments ourselves.

We will be forecasting six variables; three variables measuring real economic activity: Industrial production (INDPRO), real personal income excluding current transfer receipts (W875RX1), and the number of employees on nonagricultural payrolls (PAYEMS); as well as three price indexes: The consumer price index less food and energy (CPILFESL), the personal consumption expenditures price index (PCEPI), and the producer price index for finished goods (PPIFGS). We assume that the first three are $I(1)$ in logarithms and that the last three are $I(2)$ in logarithms.

A balanced dataset is needed for the estimation procedures and hence the span of the dataset will be determined by the variable with the shortest span after possibly being differenced to obtain stationarity. We will therefore be considering data spanning the period 1971:2–2012:10. Data are available for the variables we wish to forecast prior to 1971:2 and hence the span is not reduced when lags are included in the forecasting model.

The forecasting exercise will be carried out as a pseudo-real-time forecasting experiment; real-time in the sense that both the factors and the forecasting model are reestimated in each period using only data available up until that period; and pseudo in the sense that we are not using actual real-time data, but rather final vintage data which may have undergone revision. Forecasting will be done for the period 1981:1–2012:10.

4.2 Results

Before examining the forecasting results we start with an inspection of the factor estimates for the different models. In Figure 1 the first four factors estimated using the four main models and the entire dataset are plotted. All four methods give very similar estimates of the first factor, with one difference being the LAD-MAD estimate which appears to be scaled differently. Clearly, this is due to the MAD scaling of the dataset. For the next three factors the methods agree less. Although it might not be obvious from these plots, it does appear that some of the factors have interchanged. To illustrate this further we have provided the correlation between the different factor estimates in Table 5.

The correlations confirm that the methods agree on the estimation of the first factor and that the difference for the LAD-MAD estimate is indeed simply a scale effect. For the second factor, we see that the two LAD methods agree on the estimate. For the PC estimates the screening has caused factors two and three to switch, and these factors are different from the second factor estimated by the LAD methods. For the third factor the LAD methods agree on the estimate, and finally, the PC and PC-S methods agree on the fourth factor.

We thus see that the choice of estimation method can give very different estimates of the factors. As this is based on real data it is of course impossible to determine the reason for this. However, returning to our previous conjecture that outliers affect the ordering of estimated factors, we do see that in the case of the PC method controlling for outliers though screening does affect the ordering as factors 2 and 3 change places. The ordering is of course not important for the forecasting performance if we indeed include all factors in the forecasting model. However, if the number of factors is not correctly determined it may be crucial for the performance.

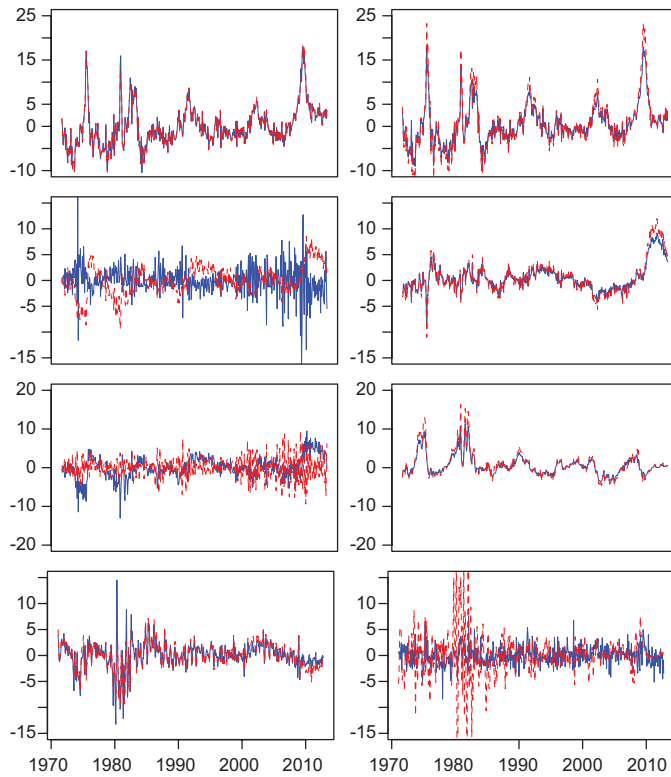


Figure 1 Plots of the estimated factors. The left column of the figure depicts factors 1 through 4 (top to bottom) for the PC estimates (blue solid line) and the PC-S estimates (red dashed line). Likewise for the right column with the LAD estimates (blue solid line) and the LAD-MAD estimates (red dashed line).

Table 5 Correlation matrix for the first four factors estimated using the four methods.

	Factor 1				Factor 2				Factor 3				Factor 4			
	PC	PC-S	LAD	L-M	PC	PC-S	LAD	L-M	PC	PC-S	LAD	L-M	PC	PC-S	LAD	L-M
Factor 1																
PC	1.00															
PC-S	1.00	1.00														
LAD	0.97	0.98	1.00													
L-M	0.97	0.98	1.00	1.00												
Factor 2																
PC					1.00											
PC-S					1.00											
LAD						1.00										
L-M						1.00	1.00									
Factor 3																
PC						0.92		1.00								
PC-S					0.83			1.00	1.00							
LAD										1.00						
L-M										0.99	1.00					
Factor 4																
PC													1.00			
PC-S													0.88	1.00		
LAD															1.00	
L-M																1.00

The coefficients are shown in absolute values and only entries larger than 0.8 in the lower triangular part of the matrix are included. LAD-MAD has been abbreviated L-M.

In Table 6 we present the main forecasting results for the 12-month horizon. The results are divided into seven main scenarios. In the first scenario we fix the number of factors at four and include no AR terms. In the next three scenarios we determine the number of factors using IC_1 and include no AR terms, six AR terms, or let BIC determine the number of AR terms. The fifth and sixth scenarios use BIC to determine the number of AR terms and either IC_2 or IC_3 for the number of factors. Finally, the seventh scenario uses BIC for both the number of AR terms and factors. In addition we include the benchmark models mentioned above, the naïve scenario and the Kalman-filter-based model of Stock and Watson (1989). For the latter we consider two different specifications. First, we will mimic their coincident index which modelled four variables: Industrial production; personal income, total less transfer payments; manufacturing and trade sales; employee-hours in non-agricultural establishments. We do not have data on manufacturing and trade sales, however, production, personal income, and employment are the first three variables we wish to forecast. We therefore

Table 6 Forecasting results for $h = 12$ reported as relative MSFEs.

Model				Variables					
Factor method	p	k	\bar{k}	IP	PI	Emp.	CPI	PCE	PPI
PC	0	4	4	0.978	0.914 [‡]	1.005 [‡]	1.908 ^{‡‡}	2.188 ^{‡‡}	<u>2.073^{‡‡}</u>
PC-S	0	4	4	0.947	0.910 [*]	<u>0.993</u>	<u>1.900^{‡‡}</u>	<u>2.108^{‡‡}</u>	2.088 ^{‡‡}
LAD	0	4	4	1.046 [*]	0.967 [‡]	1.071 [*]	2.074 ^{‡‡}	2.399 ^{‡‡}	2.889 ^{‡‡}
LAD-MAD	0	4	4	0.988	<u>0.871</u>	1.002	2.165 ^{‡‡}	2.383 ^{‡‡}	2.621 ^{‡‡}
PC	0	IC_1	9.44	1.057	0.958 [‡]	1.043	2.241 ^{‡‡}	<u>2.116^{‡‡}</u>	<u>2.027^{‡‡}</u>
PC-S	0	IC_1	4.34	1.044 [‡]	0.949 [‡]	1.146 [‡]	2.085 ^{‡‡}	2.118 ^{‡‡}	2.175 ^{‡‡}
LAD	0	IC_1	4.99	1.079 [*]	0.992 [‡]	1.071	2.082 ^{‡‡}	2.284 ^{‡‡}	2.145 ^{‡‡}
LAD-MAD	0	IC_1	2.42	<u>0.955</u>	<u>0.833[‡]</u>	<u>0.880</u>	<u>1.806^{‡‡}</u>	2.356 ^{‡‡}	2.893 ^{‡‡}
PC	6	IC_1	9.44	1.061	0.969 [‡]	0.995	1.468 [*]	1.247	1.212
PC-S	6	IC_1	4.34	1.068 [‡]	0.976 [‡]	1.136 [‡]	1.528	1.271	1.208
LAD	6	IC_1	4.99	1.091 [*]	1.015 [‡]	1.093	1.592 [*]	1.236	1.185
LAD-MAD	6	IC_1	2.42	<u>0.955</u>	<u>0.858[*]</u>	<u>0.889</u>	1.052	<u>1.094</u>	<u>1.151</u>
PC	BIC	IC_1	9.44	1.057	0.958 [‡]	0.990	1.468 [*]	1.256	1.212
PC-S	BIC	IC_1	4.34	1.047 [‡]	0.949 [‡]	1.133 [‡]	1.528	1.272	1.208
LAD	BIC	IC_1	4.99	1.080 [*]	0.992 [‡]	1.080	1.592 [*]	1.246	1.185
LAD-MAD	BIC	IC_1	2.42	<u>0.955</u>	<u>0.833[‡]</u>	<u>0.889</u>	<u>1.053</u>	<u>1.115</u>	<u>1.151</u>
PC	BIC	IC_2	8.27	1.036	0.944 [‡]	0.995	1.413	1.240	1.189
PC-S	BIC	IC_2	4.05	1.046 [‡]	0.944 [‡]	1.133 [‡]	1.515	1.268	1.206
LAD	BIC	IC_2	4.65	1.062 [*]	0.976 [‡]	1.080 [*]	1.585 [*]	1.263	1.196
LAD-MAD	BIC	IC_2	1.93	<u>0.965</u>	0.801[‡]	0.862	<u>1.103</u>	<u>1.116</u>	<u>1.148</u>
PC	BIC	IC_3	12	1.081	0.952 [‡]	<u>0.974</u>	1.576 [‡]	1.238	1.212
PC-S	BIC	IC_3	6.98	<u>1.054[*]</u>	0.970 [‡]	1.097	1.566 [*]	1.266	1.215
LAD	BIC	IC_3	7.76	1.103 [*]	1.017 [‡]	1.091	1.638 [‡]	1.244	1.189
LAD-MAD	BIC	IC_3	4.41	1.067	<u>0.910[‡]</u>	1.018 [*]	<u>1.129</u>	<u>1.109</u>	<u>1.157</u>
PC	BIC	BIC	–	1.111 [*]	<u>0.974[‡]</u>	<u>0.997</u>	<u>1.393</u>	<u>1.229</u>	<u>1.185</u>
PC-S	BIC	BIC	–	<u>1.080</u>	0.995 [‡]	1.012	1.403	1.280	1.207
LAD	BIC	BIC	–	1.195 [‡]	1.023 [‡]	1.002	1.640 [*]	1.272	1.200
LAD-MAD	BIC	BIC	–	1.190 [‡]	0.989 [‡]	1.068	1.426 [*]	1.324	1.265
KF(3)	BIC	1	1	<u>1.012</u>	<u>0.829[‡]</u>	<u>1.022</u>	1.098	1.093	1.121
KF(6)	BIC	1	1	1.012	0.830 [‡]	1.024	<u>1.097</u>	1.092	1.121
U.Mean	–	–	–	1.106 [*]	1.047	1.769 ^{‡‡}	1.829 [‡]	2.281 ^{‡‡}	2.835 ^{‡‡}
U.Mean W	–	–	–	2.036 ^{‡‡}	1.673 [*]	2.155 ^{‡‡}	2.124 [‡]	2.477 ^{‡‡}	3.229 ^{‡‡}
U.Median	–	–	–	1.215 [*]	1.078	2.100 ^{‡‡}	1.821 [‡]	2.287 ^{‡‡}	2.838 ^{‡‡}
U.Median W	–	–	–	2.203 ^{‡‡}	1.824 [*]	2.270 ^{‡‡}	2.011 [‡]	2.463 ^{‡‡}	3.209 ^{‡‡}
RMSFE, AR Model	–	–	–	4.325	2.924	1.552	1.155	1.554	4.187

Bold indicates lowest value in a column, underlined indicates lowest value within each of the seven main scenarios and the benchmark models. Two sets of Diebold-Mariano tests have been performed: 1) Equal predictive accuracy compared to bold case; significant at 5%(*), 10%(†). 2) Equal predictive accuracy compared to benchmark AR model; significant at 5%(*), 10%(†). The average number of included factors is \bar{k} . The last row gives the RMSFE of the benchmark AR model.

consider the case where X_t consists of these three variables, a case we denote KF(3). Second, we will also consider a model where X_t consists of all six variables we wish to forecast, this case will be denoted KF(6).

For each variable of interest we report the mean squared forecast error (MSFE) relative to the MSFE of an AR(p) forecast with $0 \leq p \leq 6$ chosen by BIC, i.e., forecasts based on model (20) without factors. The root mean squared forecast error (RMSFE) of this model is included in the final row of the table. For each scenario the lowest MSFE is underlined and the overall lowest MSFE is bold. Two sets of Diebold and Mariano (1995) tests have been performed. First, we test for equal predictive accuracy comparing the best performing model, i.e., the one in bold, to all other models. Significance is indicated as: 5%(*), 10%(*). Second, we test for equal predictive accuracy comparing the benchmark AR model to to all other models. Significance is indicated as: 5%(*), 10%(*).

First, notice that the best performing model within a scenario, i.e., the underlined number, is very often associated with the LAD-MAD method, especially if we focus on the cases using the IC_j criterion. Furthermore, for three of the six variables LAD-MAD provides the overall best results. To a large extent it could appear that this is driven by the number of factors included in the forecasting equation. Looking at the average number of factors across the experiment horizon \bar{k} we see that the number of factors is consistently lowest for the LAD-MAD method. PC-S and LAD on the other hand roughly choose the same number of factors, which is perhaps to be expected since our Monte Carlo results did show very similar performance for these two methods in many cases. It is, however, surprising that LAD-MAD chooses fewer factors than LAD. This suggests that scaling using a non-robust dispersion measure may incorrectly increase the number of factors and thereby hamper forecasting performance. Furthermore, if the true number of factors is as low as 2–3 as suggested by the LAD-MAD method the ordering of the factors becomes very important. Even if the other methods do estimate these factors but ordered differently, we may need to include many more estimated factors to ensure we include the few truly relevant ones. This may then in turn lead to degradation of the performance since we include irrelevant factors.

Looking at the individual variables, we first have for IP that PC-S is the preferred method. Its performance is, however, close to that of the best performing LAD-MAD case and the Diebold-Mariano test shows no significant difference in predictive accuracy. For PI the LAD-MAD method is preferred with significantly lower MSFE than PC, PC-S and LAD in all the scenarios and the benchmark AR model. Interestingly, we cannot reject equal predictive accuracy when comparing to the KF models. For Employment LAD-MAD is again preferred, but is in many cases not significantly different from the alternatives. The last three variables are in some sense less interesting since we in none of the cases outperform the benchmark AR model. However, we do recognize that if we for a moment exclude the benchmark models then the LAD-MAD method is the best performing method for all three variables. However, for PCE and PPI we have that KF(6) slightly outperforms LAD-MAD. Hence, in this application, we generally see that LAD-MAD is either on par with the alternatives or it outperforms them. Often it is difficult to show that these differences are significant according to the Diebold-Mariano tests. However, in the case of PI where we do see significant differences the results also point to LAD-MAD as the preferred method.

The results for the PC model are very interesting since they highlight the effects of screening when compared to PC-S. In general PC tends to choose a quite high number of factors. The drop in the number of factors between PC and PC-S suggests that we do have outliers in the data. However, it is not always the case that this translates into higher forecasting performance. This could again be explained by the ordering of the factors. If the ordering is such that not all true factors are included in PC-S and the severity of not including all true factors is greater than the performance decrease due to possibly including irrelevant factors, then we should include more estimated factors as in the PC model.

In the final two scenarios either IC_3 or BIC is used to choose k . None of these scenarios provide results of comparable performance. For IC_3 the Monte Carlo results demonstrated a tendency to overestimate the number of factors. In these empirical results we also find that IC_3 in general includes more factors than IC_1 and IC_2 . This may very well be the reason for the poor performance. Further, as we mentioned earlier, the use of BIC is no longer common, and hence the poor performance is expected. Note, that one crucial difference between the IC_j criterion and BIC is that BIC is applied to the forecasting equation and hence depends on

the variable being forecast. Although this could give rise to better performance than IC_j , it is not in line with the general theory of the factor model which assumes one model with r factors regardless of what is being forecast. This is also why \bar{k} is not reported in the tables for the BIC case. Finally, it is clear that the naïve alternatives are inferior to the models considered as they underperform in all cases.

Similar forecasting experiments for 6- and 24-month horizons have also been conducted. These results can be found in the appendix in Tables 9 and 10. For the 6-month horizon the results are quite similar to the 12-month forecasts. In general LAD-MAD still performs very well. For IP it is now the preferred method, although not significantly different from PC. For PI LAD-MAD is no longer preferred, instead KF(3) is associated with the lowest MSFE, albeit the MSFEs are very close and not significantly different. Interestingly, we are now able to outperform the AR forecast for CPI, i.e., the relative MSFE is less than one. In the 12-month case the factor-based forecasts did not improve on the AR forecasts for any of the price indexes. For the 24-month horizon it is generally harder to show significant differences between the different methods. Putting that aside we do, however, have that LAD-MAD is associated with the lowest MSFE for the first four variables. For the last two, PCE and CPI, we again see that KF(6) is preferred among the factor methods. However, as it was also the case for the 6- and 12-month horizons, the benchmark AR model is the best performing model for these two variables. Note, however, that in general the sizes of the results are difficult to compare across horizons since the definition of the variable being forecast changes with the horizon. Specifically this means that the variable being forecast becomes smoother as the horizon increases. This is also evident from the RMSFE of the AR model which actually decreases with the horizon.

5 Concluding remarks

In this paper we set out to challenge the common perception that outliers can easily be dealt with in factor models by simply screening the data. Building on the virtues of the LAD estimator we have established a tractable LAD factor estimator and demonstrated that in the presence of outliers a robust estimation method is preferable to ad-hoc screening.

Throughout the paper we have attempted to cling firmly to an ultimate goal of applicability. After all, what good is a model that does not perform when faced with real data? Because of this we have especially focussed on two issues: Establishing not only a tractable, but also fairly quick estimation method; and demonstrating its use on a relevant dataset. We particularly find the last point important since a model's ability to forecast an old dataset says nothing about its applicability today.

In our Monte Carlo simulation study we have demonstrated that outliers not only affect the precision of the PC factor estimates adversely, but also tend to inflate the estimated number of factors. The proposed LAD factor estimator was shown to share many traits with its regression counterpart and was in general not negatively affected by the presence of the outliers.

Taking the LAD factor estimator to our newly collected dataset covering 111 US macroeconomic variables illustrated the importance of taking outliers into account in factor-based forecasting. When applying the LAD estimator to a dataset that had been robustly scaled using MAD we were able to achieve a gain in forecasting performance compared to the traditional PC approach using screened data.

Since our focus has been the applicability of the model we have not delved into the theoretical aspects of the model. However, considering the encouraging forecasting results this would undoubtedly be a very interesting thing to do as part of the future research into this area.

Appendix A: Proof of Proposition 2

Proof. The proof makes use of Oberhofer and Kmenta (1974, Lemma 1) and follows the same line of reasoning as the proof of Oberhofer and Kmenta (1974, Theorem 1). Let $a=(F, \Lambda)$ be the parameter of interest, and write the objective function as

$$f(a)=-nT \cdot V^{\text{LAD}}(F, \Lambda; X)=-\sum_{i=1}^n \sum_{t=1}^T |X_{it}-\lambda_i F_t| \quad (28)$$

where $a \in U$, $U=U_F \times U_\Lambda$. Since we minimize the objective function subject to $\Lambda' \Lambda=1$ we have that $U_\Lambda=\{\Lambda | \Lambda' \Lambda=1\}$. We further take U_F to be \mathbb{R}^T . The proposition then follows directly from Oberhofer and Kmenta (1974, Lemma 1) given their three assumptions hold:

- (i) There exists an s such that the set $S=\{a | a \in U, f(a) \geq s\}$ is nonempty and bounded.
- (ii) $f(a)$ is continuous in S .
- (iii) U_Λ is closed and $U_F=\mathbb{R}^T$.

Starting from the bottom, (iii) is satisfied by the definition of the parameters spaces. (ii) is satisfied by continuity of the absolute value function. Hence the crucial assumption to verify is (i). The choice of s is arbitrary but can e.g., be set to $f(a^{(0)})$ where $a^{(0)}$ is the starting value for the algorithm. To show S is bounded, assume the opposite is the case to arrive at a contradiction. Then there must be a sequence $a^{(\nu)}$ in S such that

$$\lim_{\nu \rightarrow \infty} [a^{(\nu)'} a^{(\nu)}] = \infty \quad (29)$$

Since $\Lambda \in U_\Lambda$ which is closed this implies that

$$\lim_{\nu \rightarrow \infty} [F^{(\nu)'} F^{(\nu)}] = \infty \quad (30)$$

which in turn implies that

$$\lim_{\nu \rightarrow \infty} (X - F^{(\nu)} \Lambda^{(\nu)'})'(X - F^{(\nu)} \Lambda^{(\nu)'}) = nT \cdot V^{\text{LS}}(F^{(\nu)}, \Lambda^{(\nu)}; X) = \infty \quad (31)$$

However, by subadditivity of the square root function we have that

$$0 \leq \sqrt{nT \cdot V^{\text{LS}}(F^{(\nu)}, \Lambda^{(\nu)}; X)} \leq nT \cdot V^{\text{LAD}}(F^{(\nu)}, \Lambda^{(\nu)}; X) \quad (32)$$

It therefore must follow that $\lim_{\nu \rightarrow \infty} f(a^{(\nu)}) = -\infty$. However, this contradicts the fact that $a^{(\nu)} \in S$ and thus S must be bounded. \square

Appendix B: Additional results

Table 7 Precision of the factor-based forecasts in the case of i.i.d. errors and four factors when the number of factors is estimated.

<i>n</i>	<i>T</i>	Dist.	PC			PC-S			LAD		
			IC ₁	IC ₂	IC ₃	IC ₁	IC ₂	IC ₃	IC ₁	IC ₂	IC ₃
25	50	<i>t</i> (1)	0.00	0.00	0.00	0.23	0.23	0.23	0.00	0.00	0.00
25	100	<i>t</i> (1)	0.03	0.03	0.03	0.25	0.25	0.25	0.27	0.26	0.27
50	100	<i>t</i> (1)	0.01	0.01	0.01	0.27	0.27	0.27	0.29	0.29	0.29
50	200	<i>t</i> (1)	0.01	0.01	0.01	0.26	0.26	0.26	0.30	0.30	0.30
100	200	<i>t</i> (1)	0.00	0.00	0.00	0.27	0.27	0.27	0.30	0.30	0.30
150	200	<i>t</i> (1)	0.00	0.00	0.00	0.27	0.27	0.27	0.30	0.30	0.30
25	50	<i>t</i> (2)	0.23	0.18	0.27	0.27	0.26	0.34	0.28	0.26	0.32
25	100	<i>t</i> (2)	0.14	0.13	0.25	0.29	0.28	0.32	0.28	0.28	0.30
50	100	<i>t</i> (2)	0.34	0.32	0.38	0.33	0.31	0.43	0.32	0.30	0.41
50	200	<i>t</i> (2)	0.24	0.20	0.21	0.36	0.34	0.41	0.34	0.32	0.40
100	200	<i>t</i> (2)	0.27	0.27	0.28	0.51	0.45	0.66	0.51	0.44	0.67
150	200	<i>t</i> (2)	0.15	0.15	0.19	0.64	0.58	0.78	0.63	0.55	0.79
25	50	<i>t</i> (3)	0.61	0.58	0.54	0.58	0.52	0.64	0.44	0.38	0.55
25	100	<i>t</i> (3)	0.64	0.64	0.64	0.64	0.62	0.68	0.55	0.52	0.61
50	100	<i>t</i> (3)	0.78	0.79	0.78	0.82	0.81	0.82	0.81	0.80	0.81
50	200	<i>t</i> (3)	0.81	0.81	0.81	0.81	0.81	0.81	0.83	0.83	0.83
100	200	<i>t</i> (3)	0.91	0.91	0.91	0.92	0.92	0.92	0.92	0.92	0.92
150	200	<i>t</i> (3)	0.93	0.93	0.93	0.95	0.94	0.95	0.95	0.95	0.95
25	50	<i>t</i> (5)	0.75	0.73	0.70	0.75	0.74	0.73	0.66	0.62	0.66
25	100	<i>t</i> (5)	0.76	0.77	0.76	0.76	0.76	0.76	0.72	0.72	0.73
50	100	<i>t</i> (5)	0.86	0.86	0.85	0.86	0.86	0.86	0.81	0.81	0.81
50	200	<i>t</i> (5)	0.86	0.86	0.86	0.86	0.86	0.86	0.83	0.83	0.83
100	200	<i>t</i> (5)	0.94	0.94	0.94	0.94	0.94	0.94	0.91	0.91	0.91
150	200	<i>t</i> (5)	0.96	0.96	0.96	0.96	0.96	0.96	0.93	0.93	0.93
25	50	N	0.87	0.87	0.85	0.87	0.87	0.85	0.75	0.74	0.75
25	100	N	0.86	0.86	0.86	0.85	0.85	0.85	0.76	0.76	0.76
50	100	N	0.94	0.94	0.94	0.93	0.93	0.93	0.86	0.86	0.86
50	200	N	0.91	0.91	0.91	0.91	0.91	0.91	0.84	0.84	0.84
100	200	N	0.95	0.95	0.95	0.95	0.95	0.95	0.93	0.93	0.93
150	200	N	0.97	0.97	0.97	0.97	0.97	0.97	0.94	0.94	0.94

The results are based on 1,000 Monte Carlo replications. The reported numbers are the mean squared forecast error from the infeasible forecast using the true factors, and hence the true number of factors, relative to the MSFE using the estimated factors where the number of factors is estimated using the IC_{*j*} criterion.

Table 8 Precision of the factor-based forecasts in the case of non-i.i.d. errors and four factors when the number of factors is estimated.

n	T	a	b	Dist.	PC			PC-S			LAD		
					IC ₁	IC ₂	IC ₃	IC ₁	IC ₂	IC ₃	IC ₁	IC ₂	IC ₃
25	50	0.5	0	t(3)	0.50	0.43	0.47	0.47	0.40	0.51	0.34	0.31	0.43
25	100	0.5	0	t(3)	0.61	0.58	0.62	0.57	0.54	0.60	0.44	0.41	0.50
50	100	0.5	0	t(3)	0.75	0.75	0.74	0.76	0.75	0.76	0.73	0.72	0.74
50	200	0.5	0	t(3)	0.76	0.76	0.76	0.79	0.78	0.78	0.73	0.71	0.73
100	200	0.5	0	t(3)	0.86	0.85	0.86	0.88	0.88	0.88	0.85	0.85	0.85
150	200	0.5	0	t(3)	0.90	0.90	0.89	0.91	0.91	0.91	0.89	0.89	0.89
25	50	0.5	0	N	0.81	0.81	0.75	0.81	0.81	0.76	0.68	0.68	0.70
25	100	0.5	0	N	0.81	0.81	0.81	0.81	0.81	0.81	0.71	0.71	0.71
50	100	0.5	0	N	0.91	0.91	0.91	0.91	0.91	0.91	0.83	0.83	0.83
50	200	0.5	0	N	0.90	0.90	0.90	0.90	0.90	0.90	0.82	0.82	0.82
100	200	0.5	0	N	0.95	0.95	0.95	0.95	0.95	0.95	0.91	0.91	0.91
150	200	0.5	0	N	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.94	0.94
25	50	0	1	t(3)	0.27	0.25	0.27	0.23	0.22	0.26	0.20	0.22	0.21
25	100	0	1	t(3)	0.33	0.33	0.33	0.28	0.26	0.31	0.24	0.24	0.24
50	100	0	1	t(3)	0.28	0.25	0.37	0.27	0.26	0.35	0.26	0.26	0.28
50	200	0	1	t(3)	0.31	0.28	0.39	0.27	0.26	0.33	0.26	0.26	0.27
100	200	0	1	t(3)	0.39	0.33	0.52	0.36	0.31	0.54	0.28	0.27	0.49
150	200	0	1	t(3)	0.55	0.42	0.61	0.52	0.39	0.65	0.40	0.32	0.68
25	50	0	1	N	0.48	0.42	0.48	0.48	0.42	0.48	0.30	0.28	0.33
25	100	0	1	N	0.55	0.55	0.55	0.55	0.55	0.55	0.31	0.29	0.35
50	100	0	1	N	0.67	0.64	0.67	0.67	0.64	0.67	0.51	0.46	0.55
50	200	0	1	N	0.62	0.62	0.64	0.62	0.62	0.64	0.50	0.50	0.52
100	200	0	1	N	0.77	0.77	0.76	0.77	0.77	0.76	0.67	0.67	0.67
150	200	0	1	N	0.84	0.84	0.82	0.84	0.84	0.82	0.78	0.78	0.78
25	50	0.5	1	t(3)	0.23	0.20	0.25	0.25	0.23	0.27	0.23	0.22	0.22
25	100	0.5	1	t(3)	0.26	0.26	0.26	0.24	0.23	0.26	0.21	0.21	0.22
50	100	0.5	1	t(3)	0.25	0.23	0.29	0.24	0.23	0.31	0.23	0.23	0.25
50	200	0.5	1	t(3)	0.26	0.22	0.32	0.23	0.23	0.29	0.23	0.22	0.23
100	200	0.5	1	t(3)	0.29	0.25	0.43	0.25	0.24	0.46	0.24	0.24	0.32
150	200	0.5	1	t(3)	0.36	0.28	0.50	0.31	0.25	0.54	0.26	0.24	0.50
25	50	0.5	1	N	0.39	0.35	0.43	0.39	0.35	0.43	0.26	0.25	0.29
25	100	0.5	1	N	0.46	0.45	0.46	0.46	0.45	0.46	0.26	0.25	0.28
50	100	0.5	1	N	0.53	0.46	0.57	0.53	0.46	0.57	0.34	0.31	0.42
50	200	0.5	1	N	0.57	0.56	0.60	0.57	0.56	0.61	0.42	0.39	0.47
100	200	0.5	1	N	0.69	0.69	0.70	0.69	0.69	0.71	0.63	0.62	0.63
150	200	0.5	1	N	0.76	0.76	0.75	0.76	0.76	0.75	0.69	0.69	0.68

The results are based on 1,000 Monte Carlo replications. The reported numbers are the mean squared forecast error from the infeasible forecast using the true factors, and hence the true number of factors, relative to the MSFE using the estimated factors where the number of factors is estimated using the IC _{j} criterion.

Table 9 Forecasting results for $h=6$ reported as relative MSFEs.

Model				Variables					
Factor method	p	k	\bar{k}	IP	PI	Emp.	CPI	PCE	PPI
PC	0	4	4	0.990	0.857 [‡]	1.021 [‡]	2.095 ^{‡*}	2.045 ^{‡*}	1.822 ^{‡*}
PC-S	0	4	4	<u>0.945</u>	0.856 [‡]	0.991 [‡]	<u>2.066</u> ^{‡*}	<u>1.956</u> ^{‡*}	<u>1.801</u> ^{‡*}
LAD	0	4	4	1.057	0.897 [‡]	1.023 [‡]	2.180 ^{‡*}	2.196 ^{‡*}	2.472 ^{‡*}
LAD-MAD	0	4	4	0.977	<u>0.811</u>	<u>0.966</u> [‡]	2.226 ^{‡*}	2.212 ^{‡*}	2.282 ^{‡*}
PC	0	IC ₁	9.46	1.046	0.926 [‡]	1.026 [‡]	2.254 ^{‡*}	2.080 ^{‡*}	<u>1.819</u> ^{‡*}
PC-S	0	IC ₁	4.35	1.060	0.906 [‡]	1.166 [‡]	2.151 ^{‡*}	<u>1.954</u> ^{‡*}	1.860 ^{‡*}
LAD	0	IC ₁	4.99	1.063	0.900 [*]	1.001 [*]	2.167 ^{‡*}	2.117 ^{‡*}	1.883 ^{‡*}
LAD-MAD	0	IC ₁	2.45	<u>0.977</u>	<u>0.788</u>	<u>0.856</u>	<u>2.111</u> ^{‡*}	2.181 ^{‡*}	2.453 ^{‡*}
PC	6	IC ₁	9.46	1.087	0.941 [‡]	0.914	1.130 [‡]	1.233	1.195
PC-S	6	IC ₁	4.35	1.108	0.931 [‡]	1.048 [‡]	1.211 [‡]	1.266	1.225
LAD	6	IC ₁	4.99	1.118	0.945 [‡]	0.998 [‡]	1.153 [‡]	1.238	1.205
LAD-MAD	6	IC ₁	2.45	<u>1.032</u>	<u>0.826</u>	<u>0.847</u>	0.867	<u>1.146</u>	<u>1.159</u>
PC	BIC	IC ₁	9.46	1.046	0.931 [‡]	0.952	1.167 [‡]	1.237	1.195
PC-S	BIC	IC ₁	4.35	1.077	0.910 [‡]	1.092 [‡]	1.243 [‡]	1.270	1.225
LAD	BIC	IC ₁	4.99	1.052	0.904 [*]	1.005 [*]	1.162 [‡]	1.240	1.205
LAD-MAD	BIC	IC ₁	2.45	<u>0.977</u>	<u>0.792</u>	<u>0.867</u>	<u>0.898</u>	<u>1.158</u>	<u>1.159</u>
PC	BIC	IC ₂	8.31	1.010	0.917 [‡]	0.944	1.147 [‡]	1.228	1.181
PC-S	BIC	IC ₂	4.06	1.082	0.903 [‡]	1.106 [‡]	1.236 [‡]	1.269	1.224
LAD	BIC	IC ₂	4.66	1.046	0.898 [*]	1.008 [*]	1.184 [‡]	1.244	1.212
LAD-MAD	BIC	IC ₂	1.97	0.929	<u>0.734</u> [‡]	0.844	<u>0.946</u>	<u>1.140</u>	<u>1.151</u>
PC	BIC	IC ₃	12	1.076	0.930 [*]	<u>0.930</u>	1.219 [‡]	1.264	1.205
PC-S	BIC	IC ₃	6.97	1.079	0.961 [‡]	1.074 [‡]	1.213 [‡]	1.274	1.230
LAD	BIC	IC ₃	7.73	1.085	0.927 [‡]	1.028 [‡]	1.168 [‡]	1.249	1.212
LAD-MAD	BIC	IC ₃	4.41	<u>1.043</u>	<u>0.836</u>	0.957 [‡]	<u>0.951</u>	<u>1.153</u>	<u>1.160</u>
PC	BIC	BIC	–	<u>1.029</u>	0.916 [‡]	<u>0.948</u>	1.151 [‡]	<u>1.137</u>	<u>1.134</u>
PC-S	BIC	BIC	–	1.045	0.920 [‡]	0.995 [*]	1.182 [‡]	1.205	1.150
LAD	BIC	BIC	–	1.114 [*]	0.906 [*]	0.989	1.181 [‡]	1.204	1.149
LAD-MAD	BIC	BIC	–	1.090 [*]	<u>0.898</u> [‡]	1.024 [*]	<u>1.101</u> [*]	1.208	1.148
KF(3)	BIC	1	1	<u>1.005</u>	0.732 [‡]	1.051	<u>1.048</u>	1.100	1.110
KF(6)	BIC	1	1	1.005	0.733 ^{‡*}	<u>1.048</u>	1.049	1.098	1.110
U.Mean	–	–	–	1.243	1.055	2.604 ^{‡*}	2.075 ^{‡*}	2.130 ^{‡*}	2.435 ^{‡*}
U.Mean W	–	–	–	1.612 [‡]	1.234 [*]	1.691 ^{‡*}	2.161 ^{‡*}	2.591 ^{‡*}	2.903 ^{‡*}
U.Median	–	–	–	1.313 [‡]	1.081	3.061 ^{‡*}	2.067 ^{‡*}	2.151 ^{‡*}	2.442 ^{‡*}
U.Median W	–	–	–	1.665 [‡]	1.273 [*]	1.730 ^{‡*}	2.465 ^{‡*}	2.643 ^{‡*}	3.107 ^{‡*}
RMSFE, AR Model				4.680	3.406	1.330	1.103	1.634	4.648

Bold indicates lowest value in a column, underlined indicates lowest value within each of the seven main scenarios and the benchmark models. Two sets of Diebold-Mariano tests have been performed: 1) Equal predictive accuracy compared to bold case; significant at 5%([†]), 10%([°]). 2) Equal predictive accuracy compared to benchmark AR model; significant at 5%([†]), 10%([°]). The average number of included factors is \bar{k} . The last row gives the RMSFE of the benchmark AR model.

Table 10 Forecasting results for $h=24$ reported as relative MSFEs.

Model				Variables						
	Factor method	p	k	\bar{k}	IP	PI	Emp.	CPI	PCE	PPI
PC		0	4	4	0.991	1.049	0.997 [‡]	<u>1.702[†]</u>	2.340 ^{‡‡}	<u>2.381^{†‡}</u>
PC-S		0	4	4	0.993	1.057	1.019 [‡]	1.765 ^{‡†}	<u>2.286^{†‡}</u>	2.476 ^{†‡}
LAD		0	4	4	1.022	1.080	1.083 [*]	2.051 ^{†‡}	2.680 ^{‡‡}	3.481 ^{†‡}
LAD-MAD		0	4	4	<u>0.927</u>	<u>0.999</u>	<u>0.995[*]</u>	2.205 ^{†‡}	2.745 ^{†‡}	3.207 ^{†‡}
PC		0	IC ₁	9.39	0.956	1.027	0.946	2.383 ^{††}	2.513 ^{†‡}	<u>2.416^{†‡}</u>
PC-S		0	IC ₁	4.32	1.040 [*]	1.072	1.089 [‡]	2.062 ^{†‡}	<u>2.339^{†‡}</u>	2.619 ^{†‡}
LAD		0	IC ₁	4.96	1.037	1.083 [*]	1.063	2.178 ^{†‡}	2.552 ^{†‡}	2.519 ^{†‡}
LAD-MAD		0	IC ₁	2.3	<u>0.923</u>	0.929	0.892	<u>1.604[‡]</u>	2.561 ^{†‡}	3.465 ^{†‡}
PC		6	IC ₁	9.39	0.976	1.050	<u>0.913</u>	1.661	1.567	1.425
PC-S		6	IC ₁	4.32	1.036 [*]	1.100 [*]	1.135 [‡]	1.678	1.475	<u>1.328</u>
LAD		6	IC ₁	4.96	1.055 [*]	1.133 [‡]	1.105 [*]	1.658	1.474	1.343
LAD-MAD		6	IC ₁	2.3	0.921	<u>0.972[‡]</u>	0.919	<u>1.121</u>	<u>1.253</u>	1.328
PC		BIC	IC ₁	9.39	0.956	1.027	<u>0.896</u>	1.664	1.567	1.425
PC-S		BIC	IC ₁	4.32	1.029	1.072	1.102 [‡]	1.678	1.479	<u>1.324</u>
LAD		BIC	IC ₁	4.96	1.037	1.083 [*]	1.069	1.658	1.483	1.343
LAD-MAD		BIC	IC ₁	2.3	<u>0.924</u>	0.929	0.896	1.120	<u>1.274</u>	1.327
PC		BIC	IC ₂	8.18	<u>0.948</u>	1.021	0.919	1.681	1.554	1.385
PC-S		BIC	IC ₂	4.02	1.024	1.069	1.099 [‡]	1.666	1.473	<u>1.325</u>
LAD		BIC	IC ₂	4.61	1.018	1.068	1.060	1.672 [*]	1.485	1.347
LAD-MAD		BIC	IC ₂	1.79	1.028	<u>0.960</u>	<u>0.906</u>	<u>1.227</u>	<u>1.314</u>	1.344
PC		BIC	IC ₃	12	0.994	1.030	<u>0.899</u>	1.634	1.587	1.465
PC-S		BIC	IC ₃	7.01	1.038	1.081 [*]	1.064 [‡]	1.659	1.483	1.339
LAD		BIC	IC ₃	7.81	1.071 [*]	1.109 [‡]	1.087 [*]	1.652	1.488	1.355
LAD-MAD		BIC	IC ₃	4.36	<u>0.981</u>	<u>1.000</u>	0.978 [*]	<u>1.205</u>	<u>1.296</u>	<u>1.332</u>
PC		BIC	BIC	–	1.000	1.037	<u>0.910</u>	<u>1.553</u>	<u>1.427</u>	<u>1.340</u>
PC-S		BIC	BIC	–	<u>0.966</u>	<u>1.022</u>	0.925	1.681	1.497	1.361
LAD		BIC	BIC	–	1.063 [*]	1.135 [‡]	0.938	1.692 [*]	1.454	1.361
LAD-MAD		BIC	BIC	–	1.001	1.088 [‡]	0.980	1.615	1.501	1.373
KF(3)		BIC	1	1	1.038 [‡]	0.967	0.966 [‡]	1.206	1.235	1.276
KF(6)		BIC	1	1	1.038 [‡]	<u>0.967</u>	<u>0.958[‡]</u>	<u>1.192</u>	1.233	1.274
U.Mean		–	–	–	<u>1.000[‡]</u>	1.000	1.207	1.517 [‡]	2.291 ^{†‡}	3.247 ^{†‡}
U.Mean W		–	–	–	1.758 ^{†‡}	1.615 ^{†‡}	1.863 ^{†‡}	1.987 [*]	2.555 ^{†‡}	3.420 ^{†‡}
U.Median		–	–	–	1.110	1.051 [†]	1.496 ^{†‡}	1.643 [‡]	2.321 ^{†‡}	3.284 ^{†‡}
U.Median W		–	–	–	1.829 ^{†‡}	1.706 ^{†‡}	1.995 ^{†‡}	2.078 [‡]	2.590 ^{†‡}	3.513 ^{†‡}
RMSFE, AR Model					3.357	2.401	1.626	1.414	1.570	3.896

Bold indicates lowest value in a column, underlined indicates lowest value within each of the seven main scenarios and the benchmark models. Two sets of Diebold-Mariano tests have been performed: 1) Equal predictive accuracy compared to bold case; significant at 5%(*), 10%(†). 2) Equal predictive accuracy compared to benchmark AR model; significant at 5%(*), 10%(†). The average number of included factors is \bar{k} . The last row gives the RMSFE of the benchmark AR model.

Appendix C: Data

The dataset has been collected from the FRED database at the Federal Reserve Bank of St. Louis (<http://research.stlouisfed.org/fred2/>). Transforming the variables to be stationary is done according to the transformation codes (TC): 1, no transformation; 2, first difference; 3, second difference; 4, logarithms; 5, first difference of logarithms; 6, second difference of logarithms. In addition to this, some variables have been seasonally adjusted according to the S.Adj. column: SA, seasonally adjusted; SAAR, seasonally adjusted at an annual rate; NA, not applicable.

Table 11 Variable list for the FRED dataset.

Mnemonic	Description	TC	Units	S.Adj.
DSPIC96	Real Disposable Personal Income	5	Bil. of Chn. 2005 \$	SAAR
W875RX1	Personal Income Excluding Current Transfer Receipts	5	Bil. of Chn. 2005 \$	SAAR
INDPRO	Industrial Production Index	5	Index 2007=100	SA
IPBUSEQ	Industrial Production: Business Equipment	5	Index 2007=100	SA
IPCONGD	Industrial Production: Consumer Goods	5	Index 2007=100	SA
IPDCONGD	Industrial Production: Durable Consumer Goods	5	Index 2007=100	SA
IPFINAL	Industrial Production: Final Products (Market Group)	5	Index 2007=100	SA
IPMAT	Industrial Production: Materials	5	Index 2007=100	SA
IPNCONGD	Industrial Production: Nondurable Consumer Goods	5	Index 2007=100	SA
TCU	Capacity Utilization: Total Industry	2	% of Capacity	SA
PAYEMS	All Employees: Total Nonfarm	5	Thous.	SA
MANEMP	All Employees: Manufacturing	5	Thous.	SA
USCONS	All Employees: Construction	5	Thous.	SA
DMANEMP	All Employees: Durable Goods	5	Thous.	SA
USEHS	All Employees: Education & Health Services	5	Thous.	SA
USFIRE	All Employees: Financial Activities	5	Thous.	SA
USGOOD	All Employees: Goods-Producing Industries	5	Thous.	SA
USGOVT	All Employees: Government	5	Thous.	SA
USINFO	All Employees: Information Services	5	Thous.	SA
USLAH	All Employees: Leisure & Hospitality	5	Thous.	SA
USMINE	All Employees: Mining and Logging	5	Thous.	SA
NDMANEMP	All Employees: Nondurable Goods	5	Thous.	SA
USSERV	All Employees: Other Services	5	Thous.	SA
USPBS	All Employees: Professional & Business Services	5	Thous.	SA
USTRADE	All Employees: Retail Trade	5	Thous.	SA
SRVPRD	All Employees: Service-Providing Industries	5	Thous.	SA
USPRIV	All Employees: Total Private Industries	5	Thous.	SA
USTPU	All Employees: Trade, Transportation & Utilities	5	Thous.	SA
USWTRADE	All Employees: Wholesale Trade	5	Thous.	SA
UEMP15OV	Civilians Unemployed – 15 Weeks & Over	5	Thous. of Persons	SA
UEMP15T26	Civilians Unemployed for 15–26 Weeks	5	Thous. of Persons	SA
UEMP27OV	Civilians Unemployed for 27 Weeks and Over	5	Thous. of Persons	SA
UEMP5T014	Civilians Unemployed for 5–14 Weeks	5	Thous. of Persons	SA
UEMPLT5	Civilians Unemployed – <5 Weeks	5	Thous. of Persons	SA
UNRATE	Civilian Unemployment Rate	2	%	SA
UEMPMEAN	Average (Mean) Duration of Unemployment	2	Weeks	SA
ICSA	Initial Claims	5	Number	SA
AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	1	Hours	SA
AWOTMAN	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing	2	Hours	SA
AWHNONAG	Average Weekly Hours Of Production And Nonsupervisory Employees: Total private	1	Hours	SA
HOUSTMW	Housing Starts in Midwest Census Region	4	Thous. of Units	SAAR

(Table 11 Continued)

Mnemonic	Description	TC	Units	S.Adj.
HOUSTNE	Housing Starts in Northeast Census Region	4	Thous. of Units	SAAR
HOUSTS	Housing Starts in South Census Region	4	Thous. of Units	SAAR
HOUSTW	Housing Starts in West Census Region	4	Thous. of Units	SAAR
HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	4	Thous. of Units	SAAR
PERMIT	New Private Housing Units Authorized by Building Permits	4	Thous. of Units	SAAR
PERMITMW	New Private Housing Units Authorized by Building Permits in the Midwest Census Region	4	Thous. of Units	SAAR
PERMITNE	New Private Housing Units Authorized by Building Permits in the Northeast Census Region	4	Thous. of Units	SAAR
PERMITS	New Private Housing Units Authorized by Building Permits in the South Census Region	4	Thous. of Units	SAAR
PERMITW	New Private Housing Units Authorized by Building Permits in the West Census Region	4	Thous. of Units	SAAR
NAPMII	ISM Manufacturing: Inventories Index	1	Index	SA
NAPMNOI	ISM Manufacturing: New Orders Index	1	Index	SA
NAPMSDI	ISM Manufacturing: Supplier Deliveries Index	1	Index	SA
M1SL	M1 Money Stock	6	Bil. of \$	SA
M2SL	M2 Money Stock	6	Bil. of \$	SA
CURRSL	Currency Component of M1	6	Bil. of \$	SA
BOGAMBSL	Board of Governors Monetary Base, Adjusted for Changes in Reserve Requirements	3	Bil. of \$	SA
TRARR	Board of Governors Total Reserves, Adjusted for Changes in Reserve Requirements	6	Bil. of \$	SA
BUSLOANS	Commercial and Industrial Loans at All Commercial Banks	6	Bil. of \$	SA
BOGNONBR	Non-Borrowed Reserves of Depository Institutions	6	Bil. of \$	SA
NONREVSL	Total Nonrevolving Credit Outstanding	6	Bil. of \$	SA
SP500	S&P 500 Index	5	Index	NA
FEDFUNDS	Effective Federal Funds Rate	2	%	NA
TB3MS	3-Month Treasury Bill: Secondary Market Rate	2	%	NA
TB6MS	6-Month Treasury Bill: Secondary Market Rate	2	%	NA
GS1	1-Year Treasury Constant Maturity Rate	2	%	NA
GS5	5-Year Treasury Constant Maturity Rate	2	%	NA
GS10	10-Year Treasury Constant Maturity Rate	2	%	NA
AAA	Moody's Seasoned Aaa Corporate Bond Yield	2	%	NA
BAA	Moody's Seasoned Baa Corporate Bond Yield	2	%	NA
STB3MS	Spread: TB3MS–FEDFUNDS	1	%	NA
STB6MS	Spread: TB6MS–FEDFUNDS	1	%	NA
SGS1	Spread: SGS1–FEDFUNDS	1	%	NA
SGS5	Spread: SGS5–FEDFUNDS	1	%	NA
SGS10	Spread: SGS10–FEDFUNDS	1	%	NA
SAAA	Spread: AAA–FEDFUNDS	1	%	NA
SBAA	Spread: BAA–FEDFUNDS	1	%	NA
EXCAUS	Canada/US Foreign Exchange Rate	5	Canadian \$ to 1 US \$	NA
EXDNUS	Denmark/US Foreign Exchange Rate	5	Danish Kroner to 1 US \$	NA
EXUSAL	US/Australia Foreign Exchange Rate	5	US \$ to 1 Australian \$	NA
EXUSUK	US/UK Foreign Exchange Rate	5	US \$ to 1 British Pound	NA
EXSZUS	Switzerland/US Foreign Exchange Rate	5	Swiss Francs to 1 US \$	NA
EXJPUS	Japan/US Foreign Exchange Rate	5	Japanese Yen to 1 US \$	NA
PPICFF	Producer Price Index: Crude Food-stuffs & Feedstuffs	6	Index 1982=100	SA
PPICRM	Producer Price Index: Crude Materials for Further Processing	6	Index 1982=100	SA
PPIFCF	Producer Price Index: Finished Consumer Foods	6	Index 1982=100	SA
PPIFCG	Producer Price Index: Finished Consumer Goods	6	Index 1982=100	SA
PFCGEF	Producer Price Index: Finished Consumer Goods Excluding Foods	6	Index 1982=100	SA
PPIFGS	Producer Price Index: Finished Goods	6	Index 1982=100	SA
PPICPE	Producer Price Index: Finished Goods: Capital Equipment	6	Index 1982=100	SA
PPIFLF	Producer Price Index: Finished Goods Excluding Foods	6	Index 1982=100	SA

(Table 11 Continued)

Mnemonic	Description	TC	Units	S.Adj.
PPIIFF	Producer Price Index: Intermediate Foods & Feeds	6	Index 1982=100	SA
PPIITM	Producer Price Index: Intermediate Materials: Supplies & Components	6	Index 1982=100	SA
CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items	6	Index 1982–1984=100	SA
CPILEGSL	Consumer Price Index for All Urban Consumers: All Items Less Energy	6	Index 1982–1984=100	SA
CPIULFSL	Consumer Price Index for All Urban Consumers: All Items Less Food	6	Index 1982–1984=100	SA
CPILFESL	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	6	Index 1982–1984=100	SA
CPIAPPSL	Consumer Price Index for All Urban Consumers: Apparel	6	Index 1982–1984=100	SA
CPIENGSL	Consumer Price Index for All Urban Consumers: Energy	6	Index 1982–1984=100	SA
CPIUFDSL	Consumer Price Index for All Urban Consumers: Food	6	Index 1982–1984=100	SA
CPIFABSL	Consumer Price Index for All Urban Consumers: Food and Beverages	6	Index 1982–1984=100	SA
CPIHOSSL	Consumer Price Index for All Urban Consumers: Housing	6	Index 1982–1984=100	SA
CPIMEDSL	Consumer Price Index for All Urban Consumers: Medical Care	6	Index 1982–1984=100	SA
CPIOGSSL	Consumer Price Index for All Urban Consumers: Other Goods and Services	6	Index 1982–1984=100	SA
CPITRNSL	Consumer Price Index for All Urban Consumers: Transportation	6	Index 1982–1984=100	SA
PCEPI	Personal Consumption Expenditures: Chain-type Price Index	6	Index 2005=100	SA
PCEPILFE	Personal Consumption Expenditures: Chain-Type Price Index Less Food and Energy	6	Index 2005=100	SA
PCE	Personal Consumption Expenditures	5	Bil. of \$	SAAR
PCEDG	Personal Consumption Expenditures: Durable Goods	5	Bil. of \$	SAAR
PCEND	Personal Consumption Expenditures: Nondurable Goods	5	Bil. of \$	SAAR
PCES	Personal Consumption Expenditures: Services	5	Bil. of \$	SAAR

Acknowledgement: This paper formed part of my PhD dissertation and I am grateful to the members of my assessment committee, Hans Christian Kongsted, Alessandra Luati and Timo Teräsvirta, for their helpful comments and suggestions. I would also like to thank Christian M. Dahl, Allan Würtz, Siem Jan Koopman and two anonymous referees for their comments and suggestions. Finally, support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation is gratefully acknowledged.

References

- Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.
- Angelini, E., J. Henry, and R. Mestre. 2001. “Diffusion Index-Based Inflation Forecasts for the Euro Area,” Working Paper 61, European Central Bank.
- Artis, M., A. Banerjee, and M. Marcellino. 2005. “Factor Forecasts for the UK.” *Journal of Forecasting* 24, 279–298.
- Bai, J. 2003. “Inferential Theory for Factor Models of Large Dimensions.” *Econometrica* 71, 135–171.
- Bai, J., and S. Ng. 2002. “Determining the Number of Factors in Approximate Factor Models.” *Econometrica* 70, 191–221.
- Bai, J., and S. Ng. 2006. “Confidence Intervals for Diffusion index Forecasts and Inference for Factor-Augmented Regressions.” *Econometrica* 74, 1133–1150.
- Banerjee, A., M. Marcellino, and I. Masten. 2006 “Forecasting Macroeconomic Variables for the New Member States.” In *The Central and Eastern European Countries and the European Union*, edited by M. Artis, A. Banerjee, and M. Marcellino, 108–134. Cambridge, UK: Cambridge University Press.

- Banerjee, A., M. Marcellino, and I. Masten. 2008. "Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change." In *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, *Frontiers of Economics and Globalization*, vol. 3, edited by D. E. Rapach and M. E. Wohar, 149–194. Emerald Group Publishing Limited.
- Bassett, G., and R. Koenker. 1978. "Asymptotic Theory of Least Absolute Error Regression." *Journal of the American Statistical Association* 73, 618–622.
- Breitung, J., and S. Eickmeier. 2011. "Testing for Structural Breaks in Dynamic Factor Models." *Journal of Econometrics* 163, 71–84.
- Chamberlain, G., and M. Rothschild. 1983. "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets." *Econometrica* 51, 1281–1304.
- Connor, G., and R. Korajczyk. 1986. "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis." *Journal of Financial Economics* 15, 373–394.
- Croux, C., and P. Exterkate. 2011. "Sparse and Robust Factor Modelling." Discussion Paper TI 2011-122/4, Tinbergen Institute.
- Croux, C., P. Filzmoser, G. Pison, and P. Rousseeuw. 2003. "Fitting Multiplicative Models by Robust Alternating Regressions." *Statistics and Computing*, 13, 23–36.
- Diebold, F. X., and R. S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13, 253–265.
- Dielman, T. 1986. "A Comparison of Forecasts from Least Absolute Value and Least Squares Regression." *Journal of Forecasting* 5, 189–195.
- Dielman, T. 1989. "Corrections to a Comparison of Forecasts from Least Absolute Value and Least Squares Regression." *Journal of Forecasting* 8, 419–420.
- Dielman, T. 2005. "Least Absolute Value Regression: Recent Contributions." *Journal of Statistical Computation and Simulation* 75, 263–286.
- Dielman, T., and E. Rose. 1994. "Forecasting in Least Absolute Value Regression with Autocorrelated Errors: A Small-Sample Study." *International Journal of Forecasting* 10, 539–547.
- Eickmeier, S., and C. Ziegler. 2008. "How Successful are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach." *Journal of Forecasting* 27, 237–265.
- Gosselin, M., and G. Tkacz. 2001. "Evaluating Factor Models: An Application to Forecasting Inflation in Canada." Working Paper 2001-18, Bank of Canada.
- Hitomi, K., and M. Kagihara. 2001. "Calculation Method for Nonlinear Dynamic Least-Absolute Deviations Estimator." *Journal of the Japan Statistical Society* 31, 39–51.
- Huber, P., and E. Ronchetti. 2009. *Robust Statistics*. 2nd ed. New York: John Wiley and Sons.
- Johnson, R. A., and D. W. Wichern. 2007. *Applied Multivariate Statistical Analysis*, 6th ed. Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Ludvigson, S., and S. Ng. 2010. "A Factor Analysis of Bond Risk Premia." In *Handbook of Empirical Economics and Finance, Statistics: A Series of Textbooks and Monographs*, edited by A. Ulah and D. Giles, 313–372. Boca Raton, FL: Chapman and Hall.
- Phillips, P. 1991. "A Shortcut to LAD Estimator Asymptotics." *Econometric Theory* 7, 450–463.
- Pollard, D. 1991. "Asymptotics for Least Absolute Deviation Regression Estimators." *Econometric Theory* 7, 186–199.
- Schumacher, C. 2007. "Forecasting German GDP using Alternative Factor Models Based on Large Datasets." *Journal of Forecasting* 26, 271–302.
- Schumacher, C., and C. Dreger. 2004. "Estimating Large-Scale Factor Models for Economic Activity in Germany: Do They Outperform Simpler Models." *Jahrbücher für Nationalökonomie und Statistik*, 224, 731–750.
- Stock, J., and M. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators." In *NBER Macroeconomics Annual 1989*, Vol. 4, edited by O. J. Blanchard and S. Fischer, 351–409. MIT Press.
- Stock, J., and M. Watson. 1998. "Diffusion Indexes." NBER Working Paper 6702, National Bureau of Economic Research.
- Stock, J., and M. Watson. 2002a. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97, 1167–1179.
- Stock, J., and M. Watson. 2002b. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics* 20, 147–162.
- Stock, J., and M. Watson. 2005. "Implications of Dynamic Factor Models for VAR Analysis." NBER Working Paper 11467, National Bureau of Economic Research.
- Stock, J., and M. Watson. 2006. "Forecasting with Many Predictors." In *Handbook of Economic Forecasting*, vol. 1, edited by G. Elliott, C. W. Granger, and A. Timmermann, 515–554. Elsevier.
- Stock, J., and M. Watson. 2009. "Forecasting in Dynamic Factor Models Subject to Structural Instability." In *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, edited by J. Castle and N. Shephard, 173–205. USA: Oxford University Press.
- Stock, J., and M. Watson. 2011. "Dynamic Factor Models." In *Oxford Handbook of Economic Forecasting*, edited by M. P. Clements and D. F. Hendry, 35–59. USA: Oxford University Press.
- Weiss, A. 1990. "Least Absolute Error Estimation in the Presence of Serial Correlation." *Journal of Econometrics* 44, 127–158.
- Weiss, A. 1991. "Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation." *Econometric Theory*, 7, 46–68.