Federal Reserve Bank of New York
Staff Reports

Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting

Jan J. J. Groen
George Kapetanios

Staff Report No. 327
May 2008
Revised October 2015

Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting
Jan J. J. Groen and George Kapetanios
Federal Reserve Bank of New York Staff Reports, no. 327
May 2008; revised October 2015
JEL classification: C22, C53, E37, E47

Abstract

This paper analyzes the properties of a number of data-rich methods that are widely used in macroeconomic forecasting, in particular principal components (PC) and Bayesian regressions, as well as a lesser-known alternative, partial least squares (PLS) regression. In the latter method, linear, orthogonal combinations of a large number of predictor variables are constructed such that the covariance between a target variable and these common components is maximized. Existing studies have focused on modelling the target variable as a function of a finite set of unobserved common factors that underlies a large set of predictor variables, but here it is assumed that this target variable depends directly on the whole set of predictor variables. Given this setup, it is shown theoretically that under a variety of different unobserved factor structures, PLS and Bayesian regressions provide asymptotically the best fit for the target variable of interest. This includes the case of an asymptotically weak factor structure for the predictor variables, for which it is known that PC regression becomes inconsistent. Monte Carlo experiments confirm that PLS regression is close to Bayesian regression when the data has a factor structure. When the factor structure in the data becomes weak, PLS and Bayesian regressions outperform principal components. Finally, PLS, principal components, and Bayesian regressions are applied on a large panel of monthly U.S. macroeconomic data to forecast key variables across different subperiods, and PLS and Bayesian regression usually have the best out-of-sample performances.

Key words: macroeconomic forecasting, (weak) factor models, principal components, partial least squares, Bayesian ridge regression

---

Groen: Federal Reserve Bank of New York (e-mail: jan.groen@ny.frb.org). Kapetanios: Queen Mary University of London (e-mail: g.kapetanios@qmul.ac.uk). The paper has benefited from helpful comments by Alexei Onatski, James Stock, and seminar participants at Brown University, Erasmus University Rotterdam, Queen Mary University of London, the University of Tokyo, the "Time Series and Panel Modelling Workshop in honor of M. Hashem Pesaran" at the Goethe University of Frankfurt, the European Meeting of the Econometric Society in Milan, the NBER-NSF Time Series Conference in Aarhus, and the CEMMAP/UCL Conference on Unobserved Factor Models in London. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

It has been a standard assumption in theoretical macroeconomic modeling that agents are processing all the available quantities of information when forming their expectations for the future. Also, policymakers traditionally have looked at a vast array of indicator series in the run-up to major policy decisions, or in the words of Lars Svensson (Svensson (2005)) about what central bankers do in practice: '(l)arge amounts of data about the state of the economy and the rest of the world ... are collected, processed, and analyzed before each major decision.' However, generally speaking it is either inefficient or downright impossible to incorporate a large number of variables in a single forecasting model and estimate it using standard econometric techniques. This prompted a new strand of research devoted to the theory and practice of alternative macroeconomic forecasting methods that utilize large data sets.

These alternative methods can be distinguished into two main categories. As, e.g., outlined in Hendry (1995), the methods of the first category involve inherently two steps: In the first step some form of variable selection is undertaken, including automated model selection procedures as in Krolzig and Hendry (2001). The variables that are chosen are then used in a standard forecasting model. An alternative group of forecasting methods consists of estimation strategies that allow estimation of a single equation model that utilizes *all* the information in a large data set and not just an 'optimal' subset of the available predictor series. This is a diverse group of forecasting methods ranging from factor-based methods to Bayesian regression and forecast combination. We focus in this paper on the latter group of data-rich forecasting methods.

Within the group of data-rich forecasting techniques, factor methods have gained a prominent place. Building on Chamberlain and Rothschild (1983), Stock and Watson (2002a) and Bai (2003) show that under relatively weak assumptions regarding the behavior of the idiosyncratic components in a factor model, principal components can be used to identify the unobserved common factors in very large data sets. Stock and Watson (2002b) proved to be the starting point of a large empirical research output where, with mixed success, a limited number of principal components extracted from a large data set are used to forecast key macroeconomic variables.

Boivin and Ng (2006) make it clear, however, that if the forecasting power comes from a certain factor, this factor can be dominated by other factors in a large data set, as the principal components solely provide the best fit for the large data set and not for the target variable. This could explain why in some empirical applications principal components (PC) factor models are dominated by Bayesian regression and forecast combinations, as in both cases the information in a large data set is compressed such that this has explanatory power for the target variable. Under Bayesian regression one essentially estimates a multivariate regression consisting of *all* predictor variables, but with the regression coefficients shrunken to a value close to zero. Starting with Bates and Granger (1969), forecast combination involves the use of subsets of predictor variables in distinct forecasting models, which are then averaged to produce a final forecast. Note, however, that from an econometric perspective forecast combinations are *ad hoc* in nature.

Although less widely known, an alternative data-rich approach that can be used for macroeconomic forecasting using very large data sets is partial least squares (PLS) regression. We will show that PLS regression can do this irrespective of whether such a data set has a strong factor structure or not. PLS regression is implemented for large data sets through the construction of linear, orthogonal combinations of the predictor variables, which have maximize the covariance between the target forecast variable and predictor variables. Although similar in spirit to PC regression, the explicit consideration of the target forecast variable addresses a major existing criticism towards PC regression as a forecasting technique.

The main contribution of the paper rests on analysing the properties of the various data-rich methods, in particular PC, PLS and Bayesian regression, under a more general setting for the target variable. In particular, most work to date has focused on modelling the target variable as a function of a finite set of unobserved factors. We, instead, assume that the target variable depends on the whole set of available predictor variables. As the number of these variables is assumed to tend to infinity this is a more difficult problem to handle. While some work (see, e.g., Stock and Watson (2012)) allows for such a setup, there are usually strict assumptions associated with this setup such as, for example, orthogonality of the regressors which both greatly simplifies the analysis and precludes interesting models such as factor models. We consider in detail the properties of PLS, PC and Bayesian regression for forecasting using both Monte Carlo analysis and an empirical application to gauge the potential of each of these data-rich approaches.

In the remainder of this paper we have the following structure: Section 2 discusses the asymptotic behavior of PC, PLS and Bayesian regression under different factor configurations: strong factors, strong factors underlying the predictor variables but only a few of these variables are relevant for the target variable, and weak factors. Section 3 report on an extensive Monte Carlo study that focuses on the out-of-sample properties of PLS, PC and Bayesian shrinkage regression. Section 4 presents an empirical application where PLS and the other data-rich forecasting methods are used on a large monthly US macroeconomic data set. Finally, Section 5 concludes.

2 Methods for Data-Rich Macroeconomic Forecasting

A useful framework for studying data-rich based modeling methods is provided by the following general forecasting equation

$$y_t = \alpha' x_t + \epsilon_t; \quad t = 1, \ldots, T, \tag{1}$$

where y_t is the target of the forecasting exercise, $x_t = (x_{1t} \cdots x_{Nt})'$ is a vector of dimension $N \times 1$ and thus $\alpha = (\alpha_1 \cdots \alpha_N)'$ is also $N \times 1$. The error term ϵ_t in (1) is throughout the paper assumed to be a stationary, finite variance martingale difference sequence. In this paper we will focus on the case that the number of indicator variables N is too large for α to be determined by standard methods such as ordinary least squares (OLS). The literature has proposed a number of ways how one can deal with this issue of large-dimensional data sets, of which we provide a selective review below.

Before proceeding, however, we need to stress that our assumed framework given by (1) is a significant deviation from, and, we argue, generalisation of, the existing literature which analyses the case where y_t does not depend on the large, observed dataset, x_t, but a small, unobserved set of variables, referred to as factors.

We review methods that have been shown to be applicable for the data-rich case, starting with principal components (PC) regression in Section 2.1, partial least squares regression in Section 2.2 and Bayesian (shrinkage) regression in Section 2.3. In each subsection we present theoretical results and discussion on the properties of the methods under (1) while recaping the existing theory that relates to the usual factor setup utilized in the existing literature. Finally, we provide a short, qualitative comparison for the approaches in Section 2.4.

2.1 Principal Components Regression

The most widely used class of data-rich forecasting methods are factor methods. Factor methods have been at the forefront of developments in forecasting with large data sets and in fact started this literature with the influential work of Stock and Watson (2002a). The defining characteristic of most factor methods is that relatively few summaries of the large data sets are used in forecasting equations which thereby becomes a standard forecasting equation as they only involve a few variables. The assumption is that the co-movements across the indicator variables can be captured by a $r \times 1$ vector of unobserved factors $F_t = (F_{1t} \cdots F_{rt})'$, i.e.

$$\tilde{x}_t = \Lambda' F_t + e_t \tag{2}$$

where $\tilde{x}_t$ may be equal to x_t or may involve other variables such as, e.g., lags and leads of x_t and Λ is a $r \times N$ matrix of parameters describing how the individual indicator variables relate to each of the r factors, which we denote with the terms 'loadings'. In (2) e_t represents a zero-mean $I(0)$ vector of errors that represent for each indicator variable the fraction of dynamics unexplained by F_t, the 'idiosyncratic components'. The number of factors is assumed to be small, meaning $r < \min(N,\ T)$. The main difference between different factor methods relate to how Λ is estimated.

The use of principal components (PC) for the estimation of factor models is, by far, the most popular factor extraction method. It has been popularised by Stock and Watson (2002a,b), in the context of large data sets, although the idea had been well established in the traditional multivariate statistical literature. The method of principal components (PC) is simple. Estimates of Λ and the factors F_t are obtained by solving:

$$V(r) = \min_{\Lambda, F} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tilde{x}_{it} - \lambda_i' F_t)^2, \tag{3}$$

where λ_i is a $r \times 1$ vector of loadings that represent the N columns of $\Lambda = (\lambda_1 \cdots \lambda_N)$. One, non-unique, solution of (3) can be found by taking the eigenvectors corresponding to the r largest eigenvalues of the second moment matrix $X'X$, which then are assumed to represent the rows in Λ, and the resulting estimate of Λ provides the forecaster with an estimate of the r factors $\hat{F}_t = \hat{\Lambda} \tilde{x}_t$. To identify the factors up to a rotation, the data are usually normalized to have zero mean and unit variance prior to the application of principal components; see Stock and Watson (2002a) and Bai (2003).

The main assumptions that underpin the validity of PC estimation of the unobserved common factors F_t in (2) can be found in Stock and Watson (2002a,b), Bai and Ng (2002) and Bai (2003), which can be summarized as follows:

Assumption 1 **(a)** *$E\|F_t\|^4 \leq M < \infty$ for some positive constant M, and $T^{-1} \sum_{t=1}^{T} F_t' F_t \to \Sigma_F$ for some positive definite $r \times r$ matrix Σ_F.*

(b) $\|\lambda_i\| \leq \bar{\lambda} < \infty$ *and* $\|\Lambda'\Lambda/N - \Sigma_N\| \to 0$, *as* $N \to \infty$, *for some positive definite* $r \times r$ *matrix* Σ_N.

(c) $E(e_{i,t}) = 0$, $E|e_{i,t}|^8 \leq M$ *where* $e_t = (e_{1,t} \cdots e_{N,t})'$. *The variance-covariance matrix of* e_t *equals* Ω_e, *and* F_s *and* e_t *are independent for all* $s,\ t$.

(d) *For* $\tau_{i,j,t,s} \equiv E(e_{j,s}e_{i,t})$:

- $(NT)^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} |\sum_{i+1} N\tau_{i,i,t,s}| \leq M$
- $1/N \sum_{i=1}^{N} |\tau_{i,i,s,s}| \leq M$ *for all* s
- $N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} |\tau_{i,j,t,s}| \leq M$
- $(NT)^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} |\tau_{i,j,t,s}| \leq M$
- *For every* $(t,\ s)$, $E|(N)^{-1/2} \sum_{i=1}^{N}(e_{i,s}e_{i,t} - \tau_{i,i,t,s})|^4 \leq M$

Assumption 1(a) and 1(b) are necessary conditions for the common factors to dominate the dynamics of the predictor variables $\tilde{x}_t$ and Assumption 1(c) and 1(d) means that the idiosyncratic components e_t are at the most weakly correlated and heteroskedastic. This implies for the second moment matrix $X'X$, under the aforementioned normalization,

$$E(X'X) = \Lambda'\Lambda + \Sigma_e, \tag{4}$$

where $E(e'e) = \Sigma_e$, and Assumption 1 states that all eigenvalues of $\Lambda'\Lambda$ in (4) are $O(N)$ whereas those on Σ_e are bounded as $N \to \infty$. Thus the eigenvalues of $X'X$ are $O(N)$, which means that PC will provide consistent estimates of the unobserved common factors F_t in (2).

The setup whereby y_t asymptotically depends on F_t solely, rather than directly on the observed x_t, has been dominant in the literature:

$$y_t = \zeta' F_t + \epsilon_t, \tag{5}$$

with $\zeta = (\zeta_1 \cdots \zeta_r)'$ and r equals the number of unobserved factors that underlie the N predictor variables x_t. To the best of our knowledge, the assumption in (5) that for *any* potential target variable y_t one cannot find a better set of predictors than F_t has been made in all existing theoretical econometric work involving forecasting or modelling using large datasets, such as Bai and Ng (2006), De Mol et al. (2008) and Kelly and Pruitt (2012). However, it is reasonable to cast doubt on such a setup from a variety of directions. To see this one needs to realize that the unobserved factor structure (2) implies for (1)

$$y_t = \zeta' F_t + (\alpha' e_t + \epsilon_t), \tag{6}$$

where $\zeta' = \alpha'\Lambda'$ and $F_t = \Lambda\tilde{x}_t$, which means that a small, r, number of linear combinations of $\tilde{x}_t$ represent the factors and act as the predictors for y_t.

To see why making the assumption that the target variable can solely depend on the unobserved common factors is not warranted even within the confines of (2) where the eigenvalues of $\Lambda'\Lambda$ are $O(N)$, it is instructive to consider allowing $y_t = x_{it}$ for some i. Denote by x_{-it} the subset of x_t that does not include x_{it}. Then, one can ask whether x_{-it} may contain information useful in forecasting x_{it}, that is not contained in F_t. To see that this is indeed the case we note that the standard assumptions for strong factor models in (2) allows for some correlation between elements of e_t (see Assumption 1). If that is the case, then there exist elements of x_{-it} whose idiosyncratic components are correlated with e_{it}. It then immediately follows that including these elements of x_{-it} as predictors in a predictive regression for y_t can result in a better specified predictive regression, with a smaller forecasting error, than the one obtained by simply using F_t as predictors, at least for a finite cross-section dimension N. This casts doubt on the assumption that the generating model for y_t should be one that involves only F_t.

Having provided a justification for the use of (1), we note some of the properties of PC regression under (1). Via (6) using PC to estimate the unobserved common factors yields for (1)

$$y_t = \hat{\zeta}'\hat{F}_t + \left((\zeta' F_t - \hat{\zeta}'\hat{F}_t) + \alpha' e_t + \epsilon_t \right), \tag{7}$$

where $\hat{F}_t$ is the PC estimate of the r unobserved factors F_t and $\hat{\zeta}$ is the OLS estimator of ζ in the regression of y_t on $\hat{F}_t$. The resulting forecast error clearly has larger variance than ϵ_t for finite N. As $N,\ T \to \infty$, there are a number of observations one can make regarding the asymptotic behavior of the PC regression based forecast error. Under forecasting model (1) with (2) and if Assumption 1 holds, principal components regression can only achieve the asymptotically best feasible forecast if $\|\alpha\| = O(N^{-1/2})$. To see why this holds it is obvious from Assumption 1 and Bai (2003, Theorem 2) that for $N \to \infty$, we have in (7), $\zeta' F_t - \hat{\zeta}'\hat{F}_t = o_p(1)$. However, when the parameter vector α in (1) is sparse, in the sense that $\|\alpha\| = O_p(1)$, which implies that there exists a finite subset of the predictor variables that has a non-negligible effect on y_t as $N \to \infty$, then in (7) $\|\alpha' e_t\| = O_p(1)$. PC regression now can clearly never achieve the asymptotically best feasible forecast for y_t. Conversely, if $\|\alpha\| = O(N^{-1/2})$ then $\|\alpha' e_t\| = O_p(N^{-1/2})$.

But even when we have a non-sparse α parameter vector in (1), the traditionally utilized strong factor assumptions might be invalid themselves. Recently, Bailey et al. (2013) have analyzed extensively weak factor models and propose both a measure of the strength of factor models and an estimator for this measure. They also find that some datasets which have been used in the past to illustrate the use of strong factor models may in fact not follow such models. It is important to stress that this possibility has not been analysed extensively within the context of forecasting since the dominant assumption is a strong factor one, where all eigenvalues of $\Lambda'\Lambda$ in (4) are $O(N)$.

It is then worth noting that the validity of PC as an estimator of the unobserved factors rests on assuming that the underlying factor model is relatively strong. To get a more concrete view on what "relatively strong" means, one can follow the set up entertained in Kapetanios and Marcellino (2010) where the factor loadings matrix Λ in (2) is replaced by a factor loadings matrix Λ_N that depends on the size of N by means of a local-to-zero structure, i.e.,

$$\Lambda_N = N^{-\kappa}\tilde{\Lambda}, \tag{8}$$

where $\tilde{\Lambda}$ is a $r \times N$ matrix of fixed loadings. Within such a context it is proven in Kapetanios and Marcellino (2010, Theorem 3) that PC estimation of the unobserved common factors is inconsistent if in (8) $\kappa \geq 0.5$, as the eigenvalues of $\Lambda_N'\Lambda_N$ are now $O(N^{(1-\kappa)})$ for $0 \leq \kappa < 1$, which can be shown to result in a covariance matrix of $\tilde{x}_t$ in (2) that has bounded eigenvalues for $N \to \infty$ at each t when $\kappa \geq 0.5$. In this case it becomes hard to distinguish common and idiosyncratic components of the predictor variables in $\tilde{x}_t$ in (2) and consequently, even when the $N \times 1$ parameter vector α is non-sparse, as $N \to \infty$, in (7) $\zeta' F_t - \hat{\zeta}'\hat{F}_t = O_p(1)$. But even for $0 < \alpha < 0.5$, which is considered by Chudik et al. (2011) as a case of a 'semi-strong' factor model underlying the data, it is not necessarily the case that PC regression would be able to achieve the asymptotically best feasible forecast for y_t in (1), as is witnessed by the following Theorem:

Theorem 1 *Let Assumption 1 hold, except for Assumption 1(b). Let* (8) *apply to* (2) *with* $0 \leq \kappa < 1/4$. *Then, under forecasting model* (1) *with* (2), $\|\alpha\| = O_p(N^{-1/2})$ *and as* $N,\ T \to \infty$ *we have for principal components regression*

$$\left\|\hat{\zeta}'\hat{F}_t - \alpha' x_t\right\| = O_p\left(C_{NT}^{-1/2}\right) + O_p(N^{-1/2}), \tag{9}$$

where

$$C_{NT} \equiv \min\left(N^{1/2-3\kappa}T^{1/2}, N^{1-3\kappa}, N^{-2\kappa}\min(N,T)\right),$$

as long as $0 \leq \kappa < 1/4$ *and* $N = o(T^{1/4\kappa})$.

Proof: The proof is given in Appendix A.

Thus, only under very minor deviations from the strong factor model Assumption 1, and then only when the number of predictor variables N is restricted to grow at a *much* lower rate than T, PC regression can be guaranteed to provide a consistent forecast for y_t. Hence, given the results in Bailey et al. (2013), PC regression might in practice not be the most useful data-rich approach to macroeconomic forecasting.

2.2 Partial Least Squares Regression

Partial least squares (PLS) is a relatively new method for estimating regression equations, introduced in order to facilitate the estimation of multiple regressions when there is a large, but finite, amount of regressors (see, e.g., Wold (1982)). The basic idea is similar to principal component analysis in that factors or components, which are linear combinations of the original regression variables, are used, instead of the original variables, as regressors. A major difference between PC and PLS is that, whereas in PC regressions the factors are constructed taking into account only the values of the x_t predictor variables, in PLS, the relationship between y_t and x_t is considered as well in constructing the factors. There is only a limited literature that has considered PLS regression for data sets with a very large number of series, i.e., when N is assumed in the limit to converge to infinity, which is an assumption that has motivated the use of PC regression for macroeconomic forecasting. In this subsection we discuss the asymptotic properties of PLS regression under similar assumptions that $N,\ T \to \infty$ for data sets that have a common factor structure. In addition, we also investigate the asymptotic properties of PLS regression when large data sets have a weak factor structure that possibly vanishes asymptotically.

There are a variety of definitions for PLS and accompanying specific PLS algorithms that inevitably have much in common. A conceptually powerful way of defining PLS is to note that the PLS factors are those linear combinations of x_t, denoted by Υx_t, that give maximum covariance between y_t and Υx_t while being orthogonal to each other. Of course, in analogy to PC factors, an identification assumption is needed, to construct PLS factors, in the usual form of a normalization.

A simple algorithm to construct k PLS factors is discussed among others, in detail, in Helland (1990). Assuming for simplicity that y_t has been demeaned and x_t have been normalized to have zero mean and unit variance, a simplified version of the algorithm is given below

Algorithm 1 *1. Set $u_t = y_t$ and $v_{i,t} = x_{i,t}$, $i = 1, ...N$. Set $j = 1$.*

2. *Determine the $N \times 1$ vector of indicator variable weights or loadings $w_j = (w_{1j} \cdots w_{Nj})'$ by computing individual covariances: $w_{ij} = Cov(u_t,\ v_{it})$, $i = 1, ..., N$. Construct the j-th PLS factor by taking the linear combination given by $w_j' v_t$ and denote this factor by $f_{j,t}$.*

3. *Regress u_t and $v_{i,t}$, $i = 1, ..., N$ on $f_{j,t}$. Denote the residuals of these regressions by $\tilde{u}_t$ and $\tilde{v}_{i,t}$ respectively.*

4. *If $j = k$ stop, else set $u_t = \tilde{u}_t$, $v_{i,t} = \tilde{v}_{i,t}$ $i = 1, .., N$ and $j = j + 1$ and go to step 2.*

This algorithm makes clear that PLS is computationally tractable for very large data sets. Once PLS factors are constructed y_t can be modeled or forecast by regressing y_t on $f_{j,t}$ $j = 1, ..., k$.

Next, we undertake a theoretical analysis of PLS when $T,\ N \to \infty$. We consider two mutually exclusive frameworks. In the first, a (dominant) factor structure exists for X. In the second there is only a weak factor structure as $N \to \infty$. We start by analysing the (dominant) factor case. Here, work by Kelly and Pruitt (2012) has shown that, under (5) and Assumption 1, PLS behaves equivalently to PC. However, we consider (1) and it is important to discuss what happens then. Given the representation (6) it can be easily seen using the results of Kelly and Pruitt (2012) that under the forecasting model (1) with (2) and if Assumption 1 holds, that if the number of PLS factors $k = r$, partial least squares regression is asymptotically equivalent to principal components regression and it can achieve the asymptotically best feasible forecast for y_t as long as $\|\alpha\| = O(N^{-1/2})$.

The following example makes the relevance of the above discussion clear. Assume that the number of PLS factors k is equal to the number of factors r in the strong factor model, which we set equal to 1. The model is

$$y_t = x_{1t} + \epsilon_t \tag{10}$$

with

$$x_t = f_t + e_t \tag{11}$$

where $x_t = (x'_{1t}, x'_{2t})'$. The vector x'_{1t} is of finite dimension q and $\alpha = (\alpha_1, ..., \alpha_q)'$. Partition e_t conformably so that $e_t = (e'_{1t}, e'_{2t})'$, resulting in

$$y_t = f_t + \iota'_q e_{1t} + \epsilon_t$$

where ι_q is a q dimensional vector of ones. Let $Ef_t = 0$, $Ef_t^2 = 1$, $Ee_{it} = 0$, $Ee_{it}e_{jt} = 0$, $Ee_{it}f_t = 0$, $E\epsilon_t f_t = 0$, $Ee_{1t}^2 = 1$. $E\epsilon_t = 0$, $E\epsilon_t^2 = 1$. Then, $Ex_{it} = 0$, $Ex_{it}^2 = 2$ and $Ey_t = 0$, $Ey_t^2 = q + 2$, Denote by $f_t^{(PLS)}$ the unfeasible PLS factor assuming that (10) and (11) were fully known and a single PLS factor was used. We assume that $f_t^{(PLS)}$ is normalised to have variance 1. Note that use of the standard Bai and Ng (2002) information criteria on x_t would suggest to use a single factor. In addition it is assumed that $cov(y_t, x_{it}) = 2$, $i = 1, ..., q$ and $cov(y_t, x_{it}) = 1$, $i = q + 1, ..., N$. So $f_t^{(PLS)} = \left((N+q)\, f_t + 2\iota'_q e_{1t} + \iota'_{N-q} e_{2t}\right) / \sqrt{(N+q)^2 + 2q + N - q}$. As $N \to \infty$, $\left((N+q)\, f_t\right) / \sqrt{(N+q)^2 + 2q + N - q} \to f_t$ and $\left(2\iota'_q e_{1t} + \iota'_{N-q} e_{2t}\right) / \sqrt{(N+q)^2 + 2q + N - q} = O_p\left(N^{-1/2}\right)$ and so $f_t^{(PLS)} - f_t = O_p\left(N^{-1/2}\right)$. Therefore, PLS is as efficient a forecasting model as PC, but PC regression is not efficient here, in the sense that $\|\alpha\| = O(1)$. However, it is important to note that the above result depends crucially on the number of PLS factors used. We conjecture that if more PLS factors are used then PLS can do better than PC. The intuition for this is that, once the true factors have been extracted, PLS will construct a new 'pseudo' factor that attempts to approximate $\iota'_q e_{1t}$.

This counterexample illustrates that the strategy of Kelly and Pruitt (2012), who suggest the use of standard information criteria, such as those proposed by Bai and Ng (2002), may be suboptimal. It is possible that information criteria such as those proposed by Groen and Kapetanios (2013) may be of use instead as they focus on the forecasting equation rather than on the number of factors in (2). Another suggestion that may be of use is to adapt the strategy of Giraitis et al. (2013) who use cross-validation to determine from the data various aspects of a given forecasting strategy. In particular, in our case we may consider the forecasting performance of models with various numbers of PLS factors and choose the one that minimise the sum of the pseudo out-of-sample squared forecasting errors using observed data. We explore this possibility in the Monte Carlo section.

Next, we consider some theoretical properties of PLS when there exist weaker correlation structures than implied by a strong factor model for x_t when Assumption 1 holds. Such weaker correlation structures can arise from weak factor models discussed, e.g., in Bailey et al. (2013) but we choose to express them through more primitive assumptions relating to the actual covariance of x_t. We make explicit the dependence of the coefficient and variable vector on N, in the forecasting regression, and write the model in (1) as

$$y_t = \alpha_N' x_{N,t} + \epsilon_t \tag{12}$$

where $x_{N,t} = (x_{1,N,t}, ..., x_{N,N,t})'$. We make the following assumptions

Assumption 2 ***Part I:*** *Let $\Sigma = \Sigma_N = [\sigma_{ij}]$ denote the $N \times N$ second moment matrix of X. Then,*

$$\|\Sigma - I\|_1 = o(N),$$

where $\|.\|_1$ denotes the Minkowski 1-norm.

Part II: *Further, we have*

$$0 < \lim_{N\to\infty} \sup_{i,j} \frac{\alpha_{N,i}}{\alpha_{N,j}} < \infty.$$

Assumption 3 *Uniformly over $j = 1, ..., N$*

$$\frac{1}{T}\sum_{t=1}^{T} x_{j,t}y_t - \sigma_{Xy,j} = O_p\left(T^{-1/2}\right)$$

where $\sigma_{Xy,j} = E(x_{j,t}y_t)$. Further, assume that $\max_{i,t} x_{it} < \infty$.

Assumptions 2 and 3 deserve a number of comments.

Remark 1 **(a)** *Part I of Assumption 2 states that the covariance matrix of X becomes relatively close in a particular sense to the identity matrix as $N \to \infty$, suggesting a covariance matrix where non-diagonal elements are non-zero but tending to zero as the dimension of the matrix increases. Such a covariance matrix is obtained when in factor model (2) for the predictor variables x_t the factor loadings are local-to-zero in N as in (8), which implies for the representative element of Σ, denoted by σ_{ij}, $\sigma_{ij} = \frac{1}{N^{2\kappa}}$. Then, Part I of Assumption 2 holds when $\kappa > 0.5$, for which all eigenvalues of $X'X$ are bounded and it will be hard to distinguish in that case between common and idiosyncratic components for the predictor variables x_t (see Kapetanios and Marcellino (2010, Theorem 3)).*

(b) *Part I of Assumption 2 also allows for structures such as*

$$\Sigma = \begin{pmatrix} B & 0 \\ 0 & I \end{pmatrix} \tag{13}$$

where B is a $N_1 \times N_1$ symmetric positive definite matrix ($N_1 < N$) that can have a factor structure, $N_1 \to \infty$ and $N_1^2/N \to 0$. In such a set-up relatively more variables are added to X, as N grows, that are unrelated to the factor structure implied by B, which therefore gets more and more diluted. Note that this covariance matrix structure can be nested in that of Remark 1(a), along the lines of Kapetanios and Marcellino (2010, Section 2.3).

(c) *Part II of Assumption 2 states that no variable in $x_{N,t}$ has a dominant effect in the forecasting regression. This assumption can certainly be relaxed to allow, e.g., for a small subset of variables to be redundant. Further extensions are also possible but we choose to have this assumption so as to provide a relatively simple theoretical analysis while illustrating the desirable properties of PLS in a lower collinearity setting than assumed under a standard factor structure.*

(d) *Assumption 3 is a mild uniformity assumption.*

Based on such a weak factor structure we now can formulate the following theorem:

Theorem 2 *Let σ^2 denote the variance of ϵ_t in the regression framework* (1) *(and thus* (12)*) and $\hat{\sigma}^2$ its estimate from a one factor PLS regression, using the N predictor variables that correspond with* (1)*. Under Assumptions 2-3, and as $N, T \to \infty$ with $N/T^{1/2} = o(1)$, $\hat{\sigma}^2 - \sigma^2 = o_p(1)$.*

Proof: The proof is given in Appendix B.

Theorem 2 suggests that under a weak factor (or 'near factor') structure PLS regression is still able to estimate a model that has asymptotically the best fit, as the theorem implies that the PLS regression R^2 will converge to the population R^2 of the general regression (1). Hence, PLS regression should continue to do well in modeling a target variable even if the collinearity amongst the predictor variables is not as strong as it is assumed to be for conventional factor models. Finally, note the relative rate condition $N/T^{1/2} = o(1)$ which is reminiscent, but stronger, to that of Theorem 1 of Chun and Keleş (2010), which is a result for strong factor models.

2.3 Bayesian (Shrinkage) Regression

Bayesian regression is a standard tool for providing inference for α in (1) and there exist a large variety of approaches for implementing Bayesian regression. We will provide a brief exposition of this method. A starting point is the specification of a prior distribution for α. Once this is in place standard Bayesian analysis proceeds by incorporating the likelihood from the observed data to obtain a posterior distribution for α which can then be used for a variety of inferential purposes, including, of course, forecasting.

A popular and simple implementation of Bayesian regression results in a shrinkage estimator for α in (1) given by

$$\hat{\alpha}_{BRR} = (X'X + vI)^{-1}X'y \tag{14}$$

where $X = (x_1, ..., x_T)'$, $y = (y_1, .., y_T)'$ and v is a shrinkage scalar parameter. The shrinkage estimator (14) shrinks the OLS estimator, given by $(X'X)^{-1}X'y$, towards zero, thus enabling a reduction in the variance of the resulting estimator. This is a major feature of Bayesian regression that makes it useful in forecasting when large data sets are available. This particular implementation of Bayesian regression implies that elements of α are small but different from zero ensuring that all variables in x_t are used for forecasting. In this sense, Bayesian regression can be linked to other data-rich approaches. When a certain factor structure is assumed in the data, Bayesian regression through (14) will forecast y_t by projecting it on a weighted sum of all N principal components of X, with decaying weights, instead of projecting it on a limited number of r principal components with equal weights as in PC regression; see De Mol et al. (2008). De Mol et al. (2008) produce a consistency result forecasts produced using the above implementation of Bayesian regression. However, they assume (5) rather than (1). The Theorem below extends their result to our setting

Theorem 3 *Under the forecasting model* (1) *with* (2)*, Assumption 3 and as* $N, T \to \infty$*, we have*

$$||\hat{\alpha}'_{BRR}x_t - \alpha' x_t|| = O_p\left(\frac{N^{1/2}v\,\|\alpha\|}{\mu_{\min}(\Lambda\Lambda')}\right) + O_p\left(\frac{N}{vT^{1/2}}\left[1 + N^{1/2}\|\alpha\|\right]\right). \tag{15}$$

where $\mu_{\min}(\Lambda\Lambda')$ *is the minimum eigenvalue of* $\Lambda\Lambda'$.

Proof: The proof is given in Appendix C.

It is important to comment on this result.

Remark 2 (a) *Under a strong factor model,* $\mu_{\min}(\Lambda\Lambda') = O(N)$*. Then Theorem 3 implies*

$$||\hat{\alpha}'_{BRR}x_t - \alpha' x_t|| = O_p\left(N^{-1/2}v\,\|\alpha\|\right) + O_p\left(\frac{N}{vT^{1/2}}\left[1 + N^{1/2}\|\alpha\|\right]\right).$$

We note the crucial role played by $\|\alpha\|$. Assuming that $||\alpha' x_t|| = O(1)$, there are two distinct extreme scenarios with different implications. Firstly, if every element of α is non zero then $\|\alpha\| = O\left(N^{-1/2}\right)$. On the other hand, if a finite number of elements of α are non zero then $\|\alpha\| = O(1)$. In the first case, $||\hat{\alpha}'_{BRR} x_t - \alpha x_t|| = O_p\left(N^{-1}v\right) + O_p\left(\frac{N}{vT^{1/2}}\right)$, meaning shrinkage is consistent and yields the asymptotically best feasible forecast for y_t. For the second case $||\hat{\alpha}'_{BRR} x_t - \alpha x_t|| = O_p\left(N^{-1/2}v\right) + O_p\left(\frac{N^{3/2}}{vT^{1/2}}\right)$ and this is clearly the more problematic than the first case, which is intuitive as the shrinkage method shrinks all coefficients uniformly and this is less appropriate when there are a finite number of important variables in the forecasting regression. In this case, shrinkage is consistent if $N = o\left(T^{1/2}\right)$.

(b) *The theorem does not depend on assuming that the dynamics of the N predictor variables x_t are driven by a strong factor model: Under a factor model where the factor loadings depend on the number of predictor variables N as in (8), we have for the second moment matrix $X'X$ that $\mu_{\min}(\Lambda_N \Lambda'_N) = O\left(N^{(1-\kappa)}\right)$ for $0 \leq \kappa < 1$. Then, as long as $\|\alpha\| = O\left(N^{-1/2}\right)$ Theorem 3 implies that shrinkage remains consistent and that Bayesian regression continues to yield the asymptotically best feasible forecast for y_t.*

2.4 A Comparison of the Methods

Regarding PLS regression, we note that Garthwaite (1994) provides a rationale to cast (*ad hoc*) forecast combinations in terms of the above described PLS framework. Essentially what Garthwaite (1994) shows is that a general PLS algorithm like Algorithm 1 can be expressed in terms of sequences of univariate regressions, where we regress u_t on $v_{i,t}$, $i = 1, ..., N$ and denote the OLS estimate of the coefficient of each regression by β_i. The j-th PLS factor then equals weighted average of $\beta_i v_{it}$: $f_{j,t} = \tilde{w}'_j v_t$ with $\tilde{w}_j = ((\beta_1 w_{1j}) \cdots (\beta_N w_{Nj}))'$ where $(w_{1j} \cdots w_{Nj})$ are given. Therefore, if one sets $(w_{1j} \cdots w_{Nj}) = (Var(v_{1t}) \cdots Var(v_{Nt}))$ Algorithm 1 follows, but if one assumes $(w_{1j} \cdots w_{Nj}) = (\frac{1}{N} \cdots \frac{1}{N})$ than in the *one-factor* case the Capistrán and Timmermann (2008) projection on equal-weighted mean (PEW) forecast combination approach follows. Thus, forecast combinations can be interpreted as restricted approximations to a one-factor PLS regression, with alternative specifications for $(w_{1j} \cdots w_{Nj})$ and often with *zero* intercept and slope coefficients in the final forecast regression. Timmermann (2006) provides a comprehensive survey of the forecast combination literature.

By interpreting forecast combinations as a form of PLS regression, we obtain the underpinning for the relatively good performance of forecast combinations *vis-à-vis* PC regression within different data environments; see, for example, Faust and Wright (2009). Note, though, that PLS is much more general and it allows for several factors to be included in the forecast regression. In addition, De Mol et al. (2008) prove the existence of a form of asymptotic equivalence between PC regression and Bayesian regression when the underlying data comply with a dominant factor structure in (5). Therefore, given such a dominant factor structure, Bayesian regression should be asymptotically equivalent to PLS regression and, under the one-factor assumption, forecast combinations. Thus, the PLS regression framework provide a means to asymptotically tie together different existing data-rich forecasting methods when a dominant factor structure is assumed to be present in the data.

However, when we have a much weaker factor assumption, these asymptotic links no longer seem to hold, except for the link between PLS and forecast combinations, but in this case PLS regression nonetheless retains its desirable properties. PC regression, however, is only guaranteed to produce a consistent forecasting model when deviations from the strong factor model assumption underlying the predictor variables are very minor in nature (Theorem 1). In contrast, Theorems 2 and 3 make clear that both PLS and Bayesian regression preserve their attractive forecasting features under much less stringent factor model assumptions for the predictor variables and much less strict rate conditions on the number of predictor variables.

3 Monte Carlo Analysis

In this section, we explore through Monte Carlo experiments the finite sample performance of PLS regression, PC regression and Bayesian regression. Here, as well as in the Appendix, we consider several variations on a data set that is driven by common factors.

3.1 Monte Carlo Set-up

For our Monte Carlo experiments we consider the following data generating processes (DGPs):

$$\begin{aligned} y_t &= \alpha'(x_{1,t} \cdots x_{N,t})' + \epsilon_t = \alpha' x_t + \epsilon_t, \quad t = 1, ..., T, \\ x_t &= \Lambda' f_t + c_2 u_t, \\ f_t &= (f_{1,t} \cdots f_{r,t})' \sim \mathrm{iid} N(\mathbf{0}, I_r), \quad r << N, \\ \epsilon_t &= \sqrt{cN} \varepsilon_t, \end{aligned} \tag{16}$$

with the $N \times 1$ vector of regression parameters $\alpha = (\alpha_1 \cdots \alpha_N)'$, ϵ_t is a zero-mean disturbance term that we discuss in more detail later, and $\Lambda = (\lambda_1 \cdots \lambda_N)$ is a $r \times N$ matrix of factor loadings that corresponds with the $r \times 1$ vector of factors f_t with $\lambda_i = (\lambda_{i,1} \cdots \lambda_{i,r})'$. The DGP for x_t in (16) uses a $N \times 1$ vector of zero-mean disturbances $u_t = (u_{1,t} \cdots u_{N,t})'$. The disturbances for the N explanatory variables are determined in a similar manner: $u_{i,t} \sim$ iid $N(0,1)$.

Here, we consider a number of cases for the factor model:

Case I: $\lambda_{i,j} \sim \text{iid}N(0,1)$ for $i = 1, \ldots, N$ & $j = 1, \ldots, r$, $c_2 = 1$. $\alpha_i \sim$ iid $N(0,1)$.

Case II: $\lambda_{i,j} \sim \text{iid}N(0,1)$ for $i = 1, \ldots, N$ & $j = 1, \ldots, r$, $c_2 = 1$. $\alpha_i \sim$ iid $N(0,1)$, for $i = 1, ..., 5$; $\alpha_i = 0$, for $i = 6, ..., N$.

Case III: $\lambda_{i,j} = \frac{\tilde{\lambda}_{i,j}}{N^{\kappa_2}}$ and $\tilde{\lambda}_{i,j}$ ∼iid $N(0,1)$ for $i = 1, \ldots, N$, $j = 1, \ldots, r$ & $\kappa_2 = 0.25,\ 0.75$, $c_2 = 1$, $\alpha_i \sim$ iid $N(0,1)$.

Note that apart from Case II above, the individual regression coefficients in (16) are determined as $\alpha_i \sim$ iid $N(0,1)$.

Clearly, Case I represents a standard, dominant factor model where the predictor variables are driven by r common factors. Case II is a strong factor model where α is sparse. In contrast, Case III implies much weaker factor structures in x_t than assumed under Cases I and II, and these become progressively weaker as the cross-section dimension N increases. It assumes that the representative non-diagonal element of the covariance matrix of x_t is given by $\frac{1}{N^{2\kappa_2}}$. We obtain this by using factor loadings that tend to zero as $N \to \infty$. As such Case III represents a case where we have a 'near-factor model' in x_t in which κ_2 determines how close this structure is to a dominant factor model, where an increase in κ_2 signifies a move away from such a factor model.

An important parameter for the Monte Carlo study is the population R^2 of the y_t regression equation in (16). We control this by controlling the variance of the y_t disturbance term ϵ_t in (16) through $c = (1+r)\tilde{c}$, where $\varepsilon_t \sim \text{iid}N(0,1)$. Setting $\tilde{c} = 1,\ 4,\ 9$ gives a population R^2 equal to $0.5,\ 0.2$ and 0.1 in case of the standard factor model and the no factor model (Cases I and V). For the case of weak factors (Case III) we assume that $r = 0$ for the purposes of setting c. Therefore, in this case, the calibrated population R^2 is slightly lower than for the standard factor case but this is a minor deviation since the factor loadings under Case III are small. These values for R^2 provide, in our view, reasonable representations of empirically relevant situations.

When we assume a standard factor structure, we generate data through (16) for $r = 1, 3$ and we set that the assumed number of PC factors, k_2, is equal to the true number of factors, r, when carrying out PC regression. In the case of PLS and Bayesian regression, we are not aware of the availability of a theoretically justified approach to select either the optimal number of PLS factors r or the optimal shrinkage parameter v for a given data set. For the case of PLS regression we can argue, based on Section 2.2, that the number of PLS factors under strong factors and non-sparse α's, k_1 is at the most equal and very likely smaller than the number of PC factors, as not all factors have to be relevant for the target variable. For PLS regression we therefore could suffice by setting the number of factors k_1 to correspond to the respective k_2 PC factors. De Mol et al. (2008) suggest that an appropriate Bayesian regression for a large data set under a factor structure should be based on a shrinkage parameter that is proportional to the cross-section dimension of the data set: $v = qN$ with q ranging from 0.5 to 10. The latter range is based on the best in-sample fit De Mol et al. (2008) found in their empirical applications using BR models. In the Monte Carlo experiments we will consider $q = 5,\ 10$.

For the weak factor case, we generate data based on $r = 1$ and assume in case of PLS and PC regressions $k_1 = k_2 = 1$ and focus on how a decreasing amount of collinearity within x_t affects the relative performance of these methods.

We evaluate the competing methods using the relative out-of-sample mean squared prediction error (MSE) compared to PC regression. To construct these relative MSEs, we generate $T+100$ data points according to (16) and estimate all models over the first T observations. We then use the implied regression weights to do one-step ahead forecasting and get forecast errors for $T+1, \ldots, T+100$. The results across all variants are computed for $N,\ T = 20,\ 30,\ 50,\ 100,\ 200,\ 400$ and are each based on 1000 Monte Carlo replications.

Finally, as an alternative to the above for selecting the k_1 factors for PLS regression and the q shrinkage parameter in Bayesian regression, we follow the suggestion made in Section 2.2 and also use the following out-of-sample cross-validation algorithm:

Algorithm 2 *1. We produce one-step-ahead forecasts over the last 20 (10 if the sample size is 20) observations of the estimation period T for a grid of possible values for k_1 ($k_1 = 1, .., 8$) as well as q ($q = 0.5, 1, 2, 5, 10$) for, respectively, PLS and Bayesian regression.*

2. Then we use the values of k_1 and q that minimise the MSE for the forecast errors of the first step, and use these to produce PLS and Bayesian regression based forecasts for the *evaluation period* $T+1,\ldots,T+100$.

The resulting performances are denoted by $PLS(A)$ and $BR(A)$ in the next subsection.

3.2 Monte Carlo Results

Starting with the standard factor case in Tables 1-2, one notes that the performance of PLS regression for a fixed number of factors is better than that of PC for finite samples, especially for $N < T$, but becomes comparable for both large N and T in case of $r = 1$ as we would expect from our theoretical analysis. However, for the $r = 3$ case in Table 2 PLS regression with fixed factors outperforms PC regression also in large samples, as the random drawings of coefficients make different subset combinations of the factors relevant for predicting y_t across the simulations. The other noteworthy feature for the strong factor case is the confirmation of the very good predictive performance of Bayesian regression in a large number of cases.

When we allow the number of PLS factors and shrinkage parameter to be chosen in a data-dependent way it is clear that both methods perform even better relative to PC regression in Tables 1-2. The advantage is much more pronounced for PLS again in accordance with our theoretical results. It is clear that a data-dependent method for determining tuning parameters, that relates explicitly to the forecasting regression, is crucial for both methods but especially for PLS, casting doubt on the standard factor number selection methods. The above conclusions are confirmed if we also consider a strong factor case with a higher variance for the idiosyncratic components of the x_t variables; see Table D.1 in Section D of the Appendix.

For Case II where there is a strong factor structure but a sparse α, we note in Table 3 that PLS is equivalent to PC when the number of PLS factors is set equal to those used by PC regression, but PLS regression outperforms PC regression when out-of-sample cross-validation is employed to determine the number of PLS factors. In this case, although certainly worse in absolute terms, PC regression does relatively well compared to PLS and Bayesian regression, which is to be expected as the PC factors are good proxies for the few x_t variables that are included in the true model.

For the intermediate cases of weak factor models a dichotomy in performance emerges between PC regression on one side and both PLS and Bayesian regression on the other. Already with moderate degrees of factor weakness under case III, see Table 4, it becomes clear that PC regression performs poorly when the dimensions of the underlying panel of predictor variables become large, as it is almost always outperformed by the two other methods. This breakdown of the performance of PC regression relates to the inability of this method to uncover sensible factor estimates when one has a weak factor structure, and this is also confirmed both theoretically and through Monte Carlo experiments elsewhere in the literature; see Uhlig (2009) and Onatski (2009). PLS regression, on the other hand, performs in a large number of cases as well as Bayesian regression. Table 5 report simulation results under Case III for the factor loadings, but now with much more severe factor weakness than in Table 4. What becomes clear from Table 5 is that the relative performance of PLS regression improves substantially to a point that it now performs better in many cases than Bayesian regression. In Appendix D we also consider cases with a different weak factor structure than discussed here, based on (13), and one where we completely turn off the factor structure; see Tables D.2-D.3 and Table D.4, respectively. These results point to a qualitatively similar result as in Tables 4-5. In all the above cases the methods based on the data-dependent tuning parameter perform extremely well suggesting that they should be considered very carefully in future work.

To conclude we see that our Monte Carlo study suggests a great advantage for PLS and Bayesian regression compared to PC regression, in terms of forecasting performance, when (1) holds. This is especially the case where the data do not have a standard parsimonious representation such as the standard common factor structure suggested by Assumption 1. The Monte Carlo simulation results therefore confirm the theoretical results from Section 2, which suggested that, in contrast to PC regression, PLS and Bayesian regression retain desirable properties when modeling a target variable y_t irrespective of the assumed strength of a common factor structure or multi-collinearity within a large dataset of predictor variables.

4 Empirical Applications

In this section we further analyze the properties of PLS, PC and Bayesian regressions within an empirical context. We describe in Section 4.1 how we implement the different methods on the data. In Section 4.2 we provide details on the utilized large data set and on how we construct our predictor and explanatory variables. Finally, the results of the different forecast exercises are reported in Section 4.3.

4.1 Implementation of the Data-Rich Methods and the Forecast Comparison

Our data-rich forecasts of h period-ahead changes in y_t are generated using a model that either combines the information extracted from the N explanatory variables in the $N \times 1$ vector

$X_t = (x_{1,t} \cdots x_{N,t})$ with or without lagged changes in y_t, i.e.,

$$\Delta y_{t+h,t} = \alpha^h + \beta^{h\prime} \digamma(X_t) + \sum_{i=1}^{p} \rho_i \Delta y_{t-i+1,t-i} + \epsilon_{t+h,t}, \tag{17}$$

where β^h is $r \times 1$. In (17) $\digamma(X_t)$ represents a $r \times 1$ function of X_t that compresses the information in the N indicator variables, i.e., through principal components (PC), partial least squares (PLS) or by estimating the β^h's through Bayesian regression (BR, where $r = N$).

We operationalize the construction of $\digamma(X_t)$ on our data sets as follows:

Principal Components Regression

Following Stock and Watson (2002b) we take our $T \times N$ matrix of N indicator variables $X = (X_1' \cdots X_T')'$ and normalize this such that the variables are in zero-mean and unity variance space, which results in the $T \times N$ matrix $\tilde{X}$. We then compute the r eigenvectors of the $N \times N$ matrix $\tilde{X}'\tilde{X}$ that correspond to the first r largest eigenvalues of that matrix, which we assemble in the $N \times r$ matrix Λ^r. These eigenvectors are then used to approximate the common factors F that determine the series in X, i.e., $F = \tilde{X}\Lambda^r$, which gives us $\digamma(X_t)$ in (17).

In case of the principal components-based models updating based on an expanding window of historical data evolves as follows:

1. First forecast for all h is generated on t_0.

2. Extract r principal components F_t from the N indicator variables over the sample $t = 1, \ldots, t_0 - h$.

3. Estimate (17) with $\digamma(X_t) = F_t$ over the sample $t = 1, \ldots, t_0 - h$ for each h.

4. Extract r principal components F_t from the N indicator variables N over the sample $t = 1, \ldots, t_0$.

5. Generate for h the forecast $\Delta\hat{y}_{t+h,t}$ using the parameter estimates from step 3 and F_t from step 4.

6. Repeat for $t_0 + 1, \ldots, T - h$ for each h.

Bayesian Regression

When Bayesian regression is used to compress the forecast information in the N indicator variables, $\mathcal{F}(X_t)$ in (17) simply equals X_t, whereas β^h is estimated with the shrinkage estimator (14). As in the case of principal components-based regressions we use a normalized version of the $T \times N$ matrix of explanatory variables $X = (X_1' \cdots X_T')'$, indicated with $\tilde{X}$, and we also demean $\Delta y_{t+h,t}$ first before we estimate β^h_{BRR}. By doing this we follow De Mol et al. (2008), as the regression can then be interpreted as a Bayesian regression with a Gaussian prior. If we include lagged Δy_t's in (17) we first regress both the demeaned $\Delta y_{t+h,t}$ and the $\tilde{X}$ on $\Delta y_{t-i+1,t-i}$ for $i = 1, \ldots, p$, and use the resulting residuals in (14) to estimate β^h. Estimates for the intercept term and the ρ_i's in (17) can then be trivially recovered.

The Bayesian regression forecasts are updated using an expanding window of data like this:

1. First forecast for all h is generated on t_0.
2. Estimate (17) with (14) for β^h using $\tilde{X}$ over the sample $t = 1, \ldots, t_0 - h$ for each h.
3. Generate for h the forecast $\Delta\hat{y}_{t+h,t}$ using the parameter estimates from step 2.
4. Repeat for $t_0 + 1, \ldots, T - h$ for each h.

Partial Least Squares Regression

With partial least squares (PLS) regression, $\mathcal{F}(X_t)$ in (17) is constructed by computing r orthogonal combinations from the N indicator variables, where the weights of the individual indicator variables in the respective combinations are chosen such that the covariance with $\Delta y_{t+h,t}$ is maximized. The general PLS algorithm from Section 2.2 can be implemented for macroeconomic forecasting as follows:

Algorithm 3 *Suppose we do not include lagged ΔY_t's in* (17)*. Then:*

1. *Denote, as before, the $T \times N$ matrix of indicator variables, each normalized to have a zero mean and unit variance, as $\tilde{X}$ and demean the predictor variable, i.e.*

$$\Delta\dot{Y}_h = \left(I_T - \iota(\iota'^{-1}\iota')\right) \begin{pmatrix} \Delta y_{h+1,1} \\ \vdots \\ \Delta y_{T,T-h} \end{pmatrix}.$$

2. *The r PLS factors $F_{1,t}^{PLS}, \ldots, F_{r,t}^{PLS}$ and their loadings $w_1, \ldots, w_r$ are iteratively build up through projections on lower order PLS factors followed by computing the covariances between the resulting residuals of the columns from $\tilde{X}$ and those of $\Delta\dot{Y}_h$:*

$$F_l^{PLS} = \tilde{X}_{l|l-1} w_l; \quad w_l = \frac{1}{T-1} \tilde{X}'_{l|l-1} \Delta\dot{Y}_{h,l|l-1} \quad \text{for} \quad l = 1, \ldots, r \tag{18}$$

where for $l = 1$ (the first PLS factor)

$$\tilde{X}_{1|0} = \tilde{X}, \quad \Delta\dot{Y}_{h,1|0} = \Delta\dot{Y}_h,$$

and for $l > 1$

$$\tilde{X}_{l|l-1} = \left(I_T - F_{l-1}^{PLS} (F_{l-1}^{PLS'} F_{l-1}^{PLS})^{-1} F_{l-1}^{PLS'} \right) \tilde{X}_{l-1|l-2} \quad \text{and}$$
$$\Delta\dot{Y}_{h,l|l-1} = \left(I_T - F_{l-1}^{PLS} (F_{l-1}^{PLS'} F_{l-1}^{PLS})^{-1} F_{l-1}^{PLS'} \right) \Delta\dot{Y}_{h,l-1|l-2}.$$

3. *Finally, we simply plug in the r PLS factors $F_t^{PLS} = (F_{1,t}^{PLS} \cdots F_{r,t}^{PLS})'$ from (18) in the predictive regression (17) without lagged Δy_t's, which we estimate in the standard way:*

$$\Delta y_{t+h,t} = \alpha^h + \beta^{h'} F_t^{PLS} + \epsilon_{t+h,t}. \tag{19}$$

When lagged predictor variables are included in the predictive regression (17), one needs to control for the effect of $\Delta y_{t,t-1}, \ldots, \Delta y_{t-p+1,t-p}$ on the covariances between $\Delta y_{t+h,t}$ and $x_{1,t}, \ldots, x_{N,t}$. Like in the BR case we do that by projecting the demeaned $\Delta y_{t+h,t}$ as well as the columns of $\tilde{X}$ on $\Delta y_{t,t-1}, \ldots, \Delta y_{t-p+1,t-p}$, and then using the resulting residuals in Algorithm 3 in order to be able to construct a model like

$$\Delta y_{t+h,t} = \alpha^h + \beta^{h'} F_t^{PLS} + \sum_{i=1}^{p} \rho_i \Delta y_{t-i+1,t-i} + \epsilon_{t+h,t}. \tag{20}$$

Finally, forecasts from (19) and (20) are generated as follows, again using an expanding window of historical data:

1. First forecast for all h is generated on t_0.
2. Extract r PLS factors F_t^{PLS} from the N indicator variables over the sample $t = 1, \ldots, t_0 - h$ for each h based on Algorithm 3.
3. Estimate either (19) or (20) over the sample $t = 1, \ldots, t_0 - h$ for each h.

4. Extract r PLS factors F_t^{PLS} from the N indicator variables over the sample $t = 1, \ldots, t_0$ for each h using the corresponding loadings w_r from step 2 based on Algorithm 3.

5. Generate for h the forecast $\Delta\hat{y}_{t+h,t}$ using the parameter estimates from step 3 and F_t^{PLS} from step 4.

6. Repeat for $t_0 + 1, \ldots, T - h$ for each h.

This leaves us with one more issue: either the appropriate number of factors r for PC or PLS regression, or in case of Bayesian regression the appropriate value for the shrinkage parameter v in (14). For PC regression, we assumed in the Monte Carlo exercises that the number of common factors underlying our panel of predictor variables was known and we used this to inform how many PCs to include in the corresponding predictive regression. Obviously, in reality one cannot know for certain the true number of unobserved common factors. However, Bai and Ng (2002) proposed information criteria to determine the number of principal components that consistently estimates the unobserved common factor space underlying a panel of N predictor variables, and this has been the dominant approach in macroeconomic forecasting for selecting the appropriate number of PCs for predictive regressions. We will therefore also utilize this approach in our empirical exercise, in particular by selecting the number of dominant PCs that minimizes Bai and Ng (2002)'s IC_2 metric:

$$IC_2 = \ln(\sigma_\epsilon^2) + r\left(\frac{(N+T)\ln(\min(T,N))}{NT}\right), \tag{21}$$

where σ_ϵ^2 is the average idiosyncratic error variance in the panel of N predictor variables. In (17), conditional on the number of selected principal components by means of minimizing (21), we then select the optimal lag order by minimizing BIC across $p = 0, 1, \ldots, p^{max}$. As a final note, we reiterate here the point made in Groen and Kapetanios (2013) that the principal components selected in the aforementioned described manner does not necessarily result in the best performing predictive model.

In the Monte Carlo section we found that out-of-sample cross-validation, following Giraitis et al. (2013) who have shown that such an approach can consistently uncover from the data various characteristics of a forecasting model, could work well for both PLS and Bayesian regression under a variety of specifications of the underlying unobserved factor model. We therefore adopt Algorithm 2 used in the Monte Carlo experiments to our data and use it to select *both* the optimal lag order p and either the optimal number of factors or optimal degree of shrinkage for (17):

Algorithm 4 1. *The first forecast for all horizons h is generated at $t = t_0$.*

2. *For **each** of $t_0 - 35, t_0 - 34, \ldots, t_0 - h$ we re-estimate for:*

   - *Bayesian regression: we consider up to 21 different values for the shrinkage parameter $v = qN$ (see (14)) with $q = 0.5, 1, 1.5, \ldots, 10$ in (17), where we re-estimate for **each** of $t_0 - 35, t_0 - 34, \ldots, t_0 - h$ $(p^{max} + 1) \times 21$ versions of (17) using at each lag order $p = 0, \ldots, p^{max}$ one of $v = 0.5N, N, 1.5N, \ldots, 10N$ in (14).*
   - *PLS regression: we consider up to 6 PLS factors in (17), where we re-estimate for **each** of $t_0 - 35, t_0 - 34, \ldots, t_0 - h$ $(p^{max} + 1) \times 6$ versions of (20) using at each lag order $p = 0, \ldots, p^{max}$ one of $r = 1, \ldots, 6$ PLS factors extracted from X_t.*

3. *From the previous step we have for both Bayesian and PLS regression a set of $36 - h$ h-period ahead forecast errors across the corresponding range of potential specifications of (17). These are used to compute h-period ahead MSE measures, and the specification of (17) that has the minimum MSE for horizon h is chosen as the optimal specification for the data-rich method of interest at this horizon in t_0, where we estimate this optimal specification for $t = 1, \ldots, t_0 - h$ and generate a h-period ahead forecast with it at t_0.*

4. *Repeat for $t_0 + 1, \ldots, T - h$ for each h.*

We follow the set up of the Monte Carlo experiments in the previous section and use PC regression as the benchmark for our alternative data-rich based forecasts. In addition, we assess the forecasting performance of naive times series relative to this benchmark, as it is common practice to evaluate in the macroeconomic forecasting literature new approaches relative to hard-to-beat models like the random walk and autoregressive models. As additional naive time series models we use an autoregressive (AR) model

$$\Delta y_{t+h,t} = \alpha^h + \sum_{i=1}^{p} \rho_i \Delta y_{t-i+1,t-i} + \epsilon_{t+h,t}, \quad t = 1, \ldots, T \tag{22}$$

as well as the unconditional mean,

$$\Delta y_{t+h,t} = \alpha^h + \epsilon_{t+h,t}, \tag{23}$$

with $\Delta y_{t+h,t} = y_{t+h} - y_t$ for $h > 0$ and $\Delta y_{t-i+1,t-i} = y_{t-i+1} - y_{t-i}$ for $i = 1, \ldots, p$. The number of lagged first differences p in (22) is determined by sequentially applying the standard Schwarz (1978)'s BIC starting with a maximum lag order of $p = p_{max}$ down to $p = 1$. Unconditional mean forecast (23) implies a random walk (RW) forecast for the level of the forecast variable y_t. As mentioned earlier, all models in the forecast evaluation exercise are updated recursively for each forecast based on an expanding window of historical data.

Our assessment of the forecasting performance of the data-rich methods relative to pure AR-based and random walk-based forecasts is based on the square root of the mean of the squared forecast errors (RMSE). In Section 4.3 we will report ratios of the RMSE of the Bayesian and PLS regression approaches as well as (22) and (23) relative to the RMSE of PC regression. Superior out-of-sample performance of a method relative to PC regression is, obviously, indicated by a RMSE ratio smaller than one. While these RMSE ratios would give the reader an initial sense of the relative performance of the different data-rich based methods in terms of our benchmark models, it is not clear whether we find statistically significant differences for the different RMSEs. The usual approach to relative forecast performance inference is to construct a t-statistic for the null hypothesis that the MSE of the alternative model equals that of the benchmark model, as in, e.g., Diebold and Mariano (1995). However, the possibility of our models being nested in the PC regression model combined with an expanding data window for parameter updating affects the limiting distribution of such a t-statistic under the null; see Clark and McCracken (2013) for an overview of these effects.

To deal with these issues, we build on Monte Carlo results in Clark and McCracken (2013) who find that in case of RMSE-based evaluation comparing the Harvey et al. (1997) small sample variance correction of the Diebold and Mariano (1995) t-statistic (HLN-DM statistic hereafter) to standard normal critical values results in a good sized test of the null hypothesis of equal finite sample forecast MSE precision for both nested and non-nested models, including cases with expanded window-based model updating. It is not clear whether the assumptions underlying the Clark and McCracken (2013) results continue to hold when utilizing data-rich forecasting methods, but investigating this is beyond the scope of this paper. We, like Faust and Wright (2012) and Groen et al. (2013), make a judgement call that from a practitioner's point of view the Clark and McCracken (2013) approach provides a useful forecast assessment tool when data-rich methods are part of the set of models being evaluated, as the final predictive regression very often is still estimated by means of OLS. Thus, we follow a similar procedure, where we compare the value of the HLN-DM statistic to one-sided critical values from the standard normal distribution, and thus we test the null of equal finite sample MSE forecast accuracy *versus* the alternative that a model outperformed PC regression. As in Clark and McCracken (2013), this statistic is based on a rectangular heteroscedasticity and autocorrelation robust variance estimator with a $h-1$ truncation lag.

4.2 The Data Set and Variable Construction

Stock and Watson (2007) reorganize the large panel of macroeconomic, financial and survey-based predictor variables for the United States from Stock and Watson (2002b) and update the span of the data to the end of 2006. Both our forecast variables and our panel of indicator variables are extracted from Stock and Watson (2007) and we focus on the 109 monthly series from this U.S. data set, which before transformation span a sample starting in January 1959 and ending in December 2006. We do not use any later data as our aim is to evaluate the relative performance of the methods we consider, within a stationary framework. This aim would be adversely affected by the presence of potential nonstationarities due to the 2007-2008 financial crisis.

The panel of predictor variables consist of 105 series spanning real variables, labor market data, data on price indices (and subcomponents) and wages, money and credit series, asset prices and surveys. The predictor variables are transformed such that they are $I(0)$, which in general means that the real variables are expressed in log first differences and we use simply first differences of series expressed in rates, such as interest rates; see Appendix E for more details. With respect to nominal series we transform these into first differences of annual growth rates in order to guarantee that the dynamic properties of these transformed series are comparable to those of the rest of the panel, as for example motivated in D'Agostino and Giannone (2006, Appendix B). Hence, after transforming the indicator variables we end up with an effective span of the data that starts in February 1960 (i.e. 1960.2) and ends in December 2006 (i.e. 2006.12).

This predictor variables panel will be used to forecast appropriate transformations of CPI inflation, industrial production, the unemployment rate and the federal funds rate. These forecast variables are not part of the panel of predictors and are transformed such to guarantee stationarity:

Y_t	$\Delta y_{t,t-1}$	$\Delta y_{t+h,t}$
CPI index	$\Delta \ln Y_{t,t-12} - \Delta \ln Y_{t-1,t-13}$	$\Delta \ln Y_{t+h,t+h-12} - \Delta \ln Y_{t,t-12}$
Industrial Production index	$\Delta \ln Y_{t,t-1}$	$\Delta \ln Y_{t+h,t}$
Unemployment rate	$\Delta Y_{t,t-1}$	$\Delta Y_{t+h,t}$
Federal Funds rate	$\Delta Y_{t,t-1}$	$\Delta Y_{t+h,t}$

As described in the previous subsection, the forecasting models are updated based on an expanding window of data and all forecasts are direct forecasts for 3 horizons (in months): $h = 1$, $h = 3$ and $h = 12$, which are horizons commonly analyzed in the literature. The forecast evaluation spans two samples: January 1972 - December 2006 and January 1985 - December 2006. The latter sample commences around the start of the 'Great Moderation', as, e.g., McConnell and Perez-Quiros (2000) and Sensier and van Dijk (2004) find evidence for a downward, exogenous, shift in the volatility of a large number of U.S. macroeconomic time series around 1985. This sample split is of particular importance for forecasting U.S. economic time series, as it has been shown that it is difficult for a lot of data-rich, approaches (including *Greenbook* projections from the Federal Reserve Board) to beat simple, non-structural benchmarks like RW and AR models after the occurrence of the 'Great Moderation'; see, e.g., D'Agostino et al. (2006).

4.3 Forecasting Results

As discussed in Section 4.1, we will assess the forecasting performance of the Bayesian and PLS regression approaches as well as the two simple time series models (respectively (22) and (23)) relative to forecasts from PC regression. The evaluation results for our four variables can be found in Table 6.

The first panel of Table 6 relates to forecasting changes in annual CPI inflation. Across the two evaluation samples, the full 1972-2006 evaluation sample and the Great Moderation subsample, both PLS and Bayesian regression dominate over PC regression. This is not only the case when PC regression performs as well as or better than the RW and AR models, but also when it performs worse than those naive time series models. Between PLS and Bayesian regression, the latter seems to perform slightly better but the PLS approach still provides a close second best, in particular for one-year ahead forecasts. Of course, the overall forecasting performance of the data-rich approaches is less over the post-Great Moderation period, especially *vis-à-vis* the RW benchmark beyond the one-month ahead horizon, which is consistent with findings for inflation prediction elsewhere in the literature.

For our two real activity measures, the unemployment rate and industrial production (see the second and third panels of Table 6, respectively), there is no significant difference in forecasting performance across PC, PLS and Bayesian regressions at the one-month ahead horizon. Beyond that horizon, PLS and Bayesian regressions generally perform better than PC regression (often significantly so), which upholds even when PC regression itself predicts worse than one of the naive time series models. When we compare PLS and Bayesian regression, the former clearly performs best at the three-month ahead horizon as well as the *full sample* one-year ahead horizon. The Great Moderation has an impact in so far forecasts from an AR model become pretty much impossible to beat by any of the data-rich approaches at the one-year ahead horizon.

Finally, the bottom panel in Table 6 summarizes the forecast evaluation results for the federal funds rate. As the federal funds rate is determined by the Federal Reserve Board, which sets the target for the federal funds rate by taking into account both nominal and real developments, data-rich methods, which feed off of both nominal and real series, are expected to perform well in predicting fed funds rate changes. Although less dominating than in case of industrial production growth and unemployment changes, Bayesian regression and PLS regression perform well especially well at the one-month ahead horizon throughout the different evaluation samples, with PLS regression dominating at the three-month horizon for the full evaluation sample and Bayesian regression dominating at that horizon for the Great Moderation sample. One-year ahead there is no difference in forecasting performance across the data-rich approaches across both evaluation samples, nor relative to the naive time series models in the Great Moderation sample.

The empirical forecast evaluations in this subsection lead to a number of general observations. First, it is clear that the PLS-based forecast models are, generally speaking, amongst the best performing models. And even in the cases that they are outperformed by Bayesian regression approaches, the results in Table 6 indicate that they are close competitors. Note also that in Table 6 the performance of methods that use PLS factors and Bayesian regression are generally better than those based on PC regression, particularly at longer horizons.

5 Conclusions

In this paper we have revisited a number of approaches for prediction with many predictors that are widely used in macroeconomic forecasting and we compare these with a less widely known alternative approach: partial least squares (PLS) regression. Under PLS regression, one constructs a number of linear combinations of the predictor variables such that the covariances between the target variable and each of these linear combinations are maximized.

Based on the work of Kelly and Pruitt (2012), principal components (PC) regression and PLS regression are asymptotically similar when the underlying data has a dominant common factor structure. When the factor structure in a large data set is weak as $N \to \infty$, we prove that both PLS and Bayesian regression will continue to provide a model with the best asymptotic fit for a target variable. In contrast, PC regression is only guaranteed to provide the best asymptotic fit for the target variable under very minor deviations of a strong factor structure in a large panel of predictor variables with very stringent rate conditions. Hence, whether or not a large panel of predictors has a clear factor structure, we would expect PLS and Bayesian regression to do well in macroeconomic forecasting, in contrast to PC regression.

An extensive Monte Carlo analysis, which compares PLS and Bayesian regressions with PC regression, yields a number of interesting insights. Firstly, when we assume that the predictors relate to the target variable through a standard, dominant, common factor structure, PLS regression is shown to have an out-of-sample performance that is at least comparable to, and often better than PC regression. PLS regression also compares well to Bayesian regression under this data specification. When the relation between the predictors and the target variable has a weak factor structure, PLS and Bayesian regression clearly have the edge in terms of out-of-sample forecasting performance relative to PC regression. Also, in this setting PLS regression is at least performing as well as Bayesian regression, and actually outperforms it the weaker the factor structure becomes.

Finally, we apply PC, PLS and Bayesian regression on a panel of 105 U.S. monthly macroeconomic and financial variables to forecast CPI inflation, industrial production, unemployment and the federal funds rate, where these forecasts are evaluated across several sub-samples. PLS and Bayesian regression turn out to be generally the best performing methods, . Hence, if a forecaster is unsure how strong the commonality is amongst a large dataset of predictor variables, he or she is therefore best off modeling a target variable through PLS or Bayesian regression, as our paper shows that these approaches will retain their desirable properties under different factor structure assumptions.

Table 1: Case I, Standard Common Factor Case, Relative MSE compared to PC for $(k_1, k_2) = (1, 1)$

		R^2=0.5						R^2=0.33						R^2=0.2					
	T/N	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400
PLS(1)	20	0.84	0.89	0.95	1.00	1.03	1.04	0.98	1.01	1.03	1.05	1.07	1.07	1.04	1.07	1.09	1.11	1.09	1.09
	30	0.80	0.86	0.93	0.99	1.01	1.05	0.95	0.98	1.00	1.04	1.06	1.06	1.01	1.04	1.05	1.10	1.08	1.09
	50	0.75	0.82	0.90	0.97	1.00	1.02	0.90	0.95	0.98	1.03	1.04	1.05	0.98	1.00	1.02	1.06	1.07	1.07
	100	0.64	0.75	0.88	0.88	0.92	0.92	0.91	0.80	0.91	0.91	0.91	0.93	0.93	0.95	0.93	0.98	0.95	0.92
	200	0.67	0.74	0.80	0.88	0.91	0.92	0.80	0.84	0.87	0.91	0.92	0.92	0.88	0.89	0.90	0.92	0.93	0.92
	400	0.67	0.74	0.83	0.89	0.94	0.96	0.83	0.86	0.90	0.93	0.95	0.96	0.89	0.91	0.93	0.95	0.96	0.96
BR(5)	20	0.70	0.77	0.85	0.92	0.96	0.97	0.87	0.93	0.95	0.96	0.97	0.98	0.98	0.99	0.99	0.98	0.98	0.99
	30	0.60	0.68	0.79	0.88	0.94	0.97	0.85	0.90	0.92	0.95	0.97	0.98	0.98	0.99	0.99	0.99	0.99	0.99
	50	0.48	0.56	0.69	0.82	0.90	0.95	0.81	0.84	0.88	0.93	0.96	0.98	0.95	0.96	0.98	0.99	1.00	1.00
	100	0.33	0.35	0.36	0.49	0.55	0.63	0.67	0.56	0.63	0.63	0.68	0.65	0.88	0.77	0.73	0.71	0.67	0.63
	200	0.29	0.28	0.27	0.28	0.30	0.30	0.59	0.55	0.50	0.44	0.39	0.35	0.73	0.68	0.62	0.53	0.44	0.38
	400	0.30	0.27	0.26	0.26	0.27	0.29	0.61	0.58	0.55	0.49	0.43	0.38	0.76	0.74	0.69	0.61	0.52	0.43
BR(10)	20	0.77	0.83	0.89	0.94	0.97	0.98	0.88	0.92	0.95	0.96	0.97	0.98	0.95	0.96	0.97	0.97	0.98	0.98
	30	0.68	0.76	0.85	0.92	0.96	0.98	0.86	0.90	0.92	0.96	0.98	0.98	0.94	0.95	0.97	0.97	0.98	0.99
	50	0.56	0.66	0.77	0.86	0.93	0.97	0.81	0.85	0.89	0.93	0.96	0.98	0.92	0.93	0.95	0.97	0.99	0.99
	100	0.37	0.42	0.46	0.61	0.66	0.73	0.69	0.61	0.67	0.71	0.76	0.74	0.86	0.78	0.77	0.77	0.74	0.73
	200	0.31	0.31	0.33	0.39	0.44	0.47	0.60	0.57	0.55	0.54	0.53	0.52	0.74	0.70	0.66	0.61	0.57	0.54
	400	0.31	0.29	0.29	0.32	0.38	0.44	0.61	0.59	0.57	0.54	0.53	0.52	0.77	0.74	0.70	0.65	0.61	0.56
PLS(A)	20	0.83	0.86	0.91	0.96	0.99	1.00	1.09	1.14	1.09	1.05	1.03	1.04	1.22	1.21	1.17	1.12	1.06	1.05
	30	0.68	0.77	0.87	0.93	0.98	0.99	1.00	1.15	1.09	1.07	1.05	1.03	1.07	1.20	1.22	1.13	1.08	1.07
	50	0.55	0.66	0.79	0.88	0.95	0.98	0.90	0.97	1.04	1.06	1.06	1.04	1.00	1.05	1.17	1.17	1.13	1.07
	100	0.32	0.32	0.37	0.48	0.50	0.58	0.77	0.60	0.79	0.71	0.80	0.66	0.92	0.92	0.83	0.90	0.81	0.76
	200	0.28	0.26	0.23	0.20	0.20	0.20	0.60	0.57	0.54	0.57	0.56	0.51	0.77	0.74	0.73	0.73	0.76	0.67
	400	0.30	0.27	0.26	0.22	0.19	0.18	0.62	0.60	0.57	0.55	0.54	0.50	0.78	0.77	0.75	0.74	0.74	0.77
BR(A)	20	1.28	0.84	0.86	0.94	0.97	0.98	1.25	1.07	1.03	0.99	0.99	0.99	1.40	1.07	1.05	1.01	1.00	1.00
	30	0.64	1.35	0.85	0.90	0.95	0.97	0.94	1.21	1.04	1.01	0.99	1.00	1.00	1.21	1.09	1.02	1.01	1.01
	50	0.47	0.59	1.26	0.85	0.91	0.96	0.83	0.92	1.35	1.03	1.00	1.00	0.94	0.97	1.26	1.05	1.03	1.01
	100	0.30	0.30	0.29	0.41	0.47	0.58	0.69	0.58	0.65	0.66	0.72	0.65	0.86	0.78	0.76	0.76	0.69	0.69
	200	0.28	0.26	0.23	0.18	0.16	0.14	0.59	0.55	0.51	0.48	0.41	0.36	0.73	0.69	0.65	0.58	0.54	0.48
	400	0.30	0.27	0.25	0.22	0.18	0.15	0.61	0.58	0.55	0.51	0.45	0.40	0.76	0.74	0.70	0.64	0.58	0.53

Notes: The entries are average one-period ahead, out-of-sample MSE ratios relative to principal components-based one-period ahead forecasts across 1,000 Monte Carlo replications. The target and indicator variables are generated through the DGPs in (16) where we consider a factor loading structure as outlined under Case I assuming $r = 1$ factors; we generate $T + 100$ observations on the predictor and explanatory variables, use the first T observations to estimate the models and use the resulting parameter estimates to forecast the predictor variable over $T + 1, \ldots, T + 100$. We impose different levels of for the asymptotic fit of the prediction regression, symbolized by the R^2's. In case of PC regression 1 factor is extracted, in case of PLS(1) regression 1 factor is extracted, and in case of Bayesian regression (BR) a shrinkage parameter $q \times N$ is used with either $q = 5, 10$ or in case of $BR(A)$ with a data determined q as discussed in Algorithm 2. Similarly in the case of $PLS(A)$ the number of factors is data-determined as discussed in Algorithm 2.

Table 2: Case I, Standard Common Factor Case, Relative MSE compared to PC for $(k_1, k_2) = (3, 3)$

		R^2=0.5						R^2=0.33						R^2=0.2					
	T/N	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400
	20	1.05	1.09	1.08	1.09	1.08	1.08	1.26	1.25	1.21	1.15	1.11	1.10	1.31	1.29	1.23	1.16	1.13	1.09
	30	1.00	1.04	1.08	1.09	1.10	1.08	1.19	1.22	1.22	1.17	1.14	1.10	1.25	1.27	1.26	1.21	1.15	1.11
PLS(3)	50	0.94	0.98	1.03	1.07	1.10	1.09	1.11	1.16	1.17	1.19	1.16	1.12	1.17	1.22	1.24	1.23	1.18	1.13
	100	0.81	0.80	0.80	0.80	0.70	0.72	0.87	0.92	0.95	0.83	0.79	0.76	0.99	0.96	0.95	0.89	0.81	0.67
	200	0.72	0.70	0.67	0.61	0.51	0.41	0.86	0.83	0.78	0.67	0.54	0.39	0.90	0.86	0.80	0.69	0.54	0.38
	400	0.74	0.73	0.71	0.68	0.62	0.52	0.89	0.88	0.85	0.78	0.68	0.54	0.93	0.91	0.88	0.80	0.69	0.54
	20	0.93	0.95	0.96	0.97	0.98	0.98	0.95	0.95	0.95	0.96	0.96	0.96	0.96	0.95	0.94	0.95	0.95	0.95
	30	0.91	0.94	0.96	0.98	0.98	0.99	0.96	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
BR(5)	50	0.88	0.90	0.94	0.96	0.98	0.99	0.98	0.98	0.98	0.99	0.99	0.99	1.00	1.00	1.00	1.00	0.99	0.99
	100	0.79	0.79	0.81	0.83	0.82	0.84	0.87	0.89	0.89	0.86	0.86	0.86	0.95	0.90	0.90	0.88	0.86	0.85
	200	0.69	0.67	0.67	0.68	0.69	0.69	0.85	0.82	0.79	0.75	0.73	0.71	0.89	0.86	0.82	0.77	0.74	0.72
	400	0.69	0.68	0.66	0.66	0.67	0.68	0.87	0.85	0.82	0.78	0.75	0.73	0.92	0.90	0.87	0.81	0.77	0.74
	20	0.97	0.99	1.00	1.00	1.01	1.01	0.94	0.94	0.95	0.95	0.95	0.95	0.94	0.93	0.92	0.93	0.93	0.93
	30	0.95	0.98	0.99	1.00	1.01	1.01	0.95	0.97	0.97	0.97	0.97	0.98	0.96	0.96	0.97	0.97	0.96	0.97
BR(10)	50	0.93	0.94	0.97	0.98	1.00	1.00	0.97	0.97	0.98	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.98	0.98
	100	0.84	0.85	0.87	0.90	0.90	0.91	0.90	0.92	0.92	0.91	0.91	0.92	0.95	0.93	0.93	0.93	0.92	0.91
	200	0.74	0.74	0.76	0.79	0.81	0.82	0.88	0.86	0.85	0.84	0.83	0.83	0.91	0.89	0.87	0.85	0.84	0.84
	400	0.72	0.71	0.72	0.76	0.78	0.80	0.88	0.87	0.86	0.84	0.84	0.83	0.93	0.91	0.89	0.87	0.85	0.84
	20	1.10	1.11	1.06	1.06	1.06	1.06	1.17	1.15	1.13	1.09	1.07	1.06	1.18	1.18	1.14	1.10	1.08	1.06
	30	1.03	1.09	1.09	1.06	1.06	1.05	1.08	1.16	1.14	1.10	1.08	1.06	1.11	1.14	1.15	1.12	1.09	1.07
PLS(A)	50	0.99	1.03	1.06	1.06	1.06	1.04	1.04	1.08	1.12	1.12	1.10	1.07	1.05	1.09	1.15	1.14	1.11	1.08
	100	0.82	0.83	0.87	0.91	0.86	0.84	0.96	0.98	0.99	0.96	0.95	0.95	0.98	1.00	0.98	0.98	0.96	0.90
	200	0.67	0.66	0.65	0.68	0.70	0.70	0.89	0.89	0.89	0.91	0.90	0.88	0.94	0.93	0.93	0.94	0.93	0.90
	400	0.71	0.68	0.65	0.64	0.64	0.68	0.90	0.90	0.90	0.89	0.92	0.91	0.95	0.95	0.94	0.94	0.94	0.95
	20	1.18	1.09	1.02	1.01	1.00	1.00	1.17	1.06	1.03	1.00	0.99	0.98	1.13	1.06	1.00	0.98	0.97	0.96
	30	0.97	1.26	1.08	1.02	1.01	1.00	1.02	1.19	1.06	1.02	1.00	0.99	1.02	1.07	1.04	1.01	1.00	0.99
BR(A)	50	0.91	0.98	1.37	1.05	1.02	1.01	1.00	1.03	1.18	1.04	1.03	1.01	1.01	1.03	1.11	1.04	1.01	1.00
	100	0.76	0.75	0.75	0.78	0.74	0.78	0.87	0.90	0.90	0.86	0.84	0.84	0.95	0.92	0.91	0.90	0.87	0.84
	200	0.66	0.63	0.59	0.56	0.53	0.52	0.86	0.83	0.80	0.77	0.73	0.70	0.90	0.88	0.85	0.81	0.78	0.74
	400	0.69	0.66	0.63	0.59	0.54	0.51	0.87	0.86	0.83	0.80	0.77	0.74	0.92	0.90	0.88	0.84	0.81	0.78

Notes: See the notes for Table 1, but now with $r = 3$ factors.

Table 3: Case II, Standard Common Factor Case with sparse α, Relative MSE compared to PC for $(k_1, k_2) = (1, 1)$

		R^2=0.5						R^2=0.33						R^2=0.2					
	T/N	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400
PLS(1)	20	0.96	1.00	1.03	1.04	1.05	1.06	1.05	1.07	1.08	1.08	1.09	1.10	1.10	1.12	1.11	1.11	1.11	1.11
	30	0.94	0.97	1.01	1.02	1.04	1.04	1.02	1.05	1.05	1.07	1.07	1.07	1.06	1.07	1.09	1.09	1.09	1.08
	50	0.92	0.96	0.98	1.01	1.02	1.02	1.00	1.02	1.03	1.04	1.04	1.04	1.03	1.04	1.05	1.06	1.06	1.06
	100	0.90	0.90	0.94	0.95	0.97	0.97	0.95	0.96	0.96	0.99	0.95	0.96	0.95	0.96	0.98	0.96	0.97	0.97
	200	0.87	0.90	0.93	0.95	0.96	0.97	0.94	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	400	0.88	0.91	0.94	0.96	0.98	0.98	0.95	0.96	0.97	0.98	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98
BR(5)	20	0.90	0.93	0.96	0.98	0.99	0.99	0.97	0.97	0.98	0.98	0.99	0.98	0.98	0.99	0.98	0.99	0.99	0.99
	30	0.87	0.90	0.94	0.97	0.98	0.99	0.95	0.97	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	50	0.80	0.85	0.90	0.94	0.97	0.99	0.94	0.96	0.97	0.98	0.99	0.99	0.98	0.99	0.99	0.99	1.00	1.00
	100	0.66	0.69	0.73	0.77	0.81	0.83	0.83	0.83	0.84	0.84	0.85	0.84	0.88	0.89	0.87	0.87	0.86	0.85
	200	0.56	0.56	0.60	0.63	0.66	0.67	0.77	0.76	0.74	0.72	0.71	0.71	0.84	0.82	0.79	0.75	0.73	0.72
	400	0.55	0.55	0.56	0.59	0.63	0.66	0.79	0.78	0.76	0.74	0.72	0.71	0.87	0.86	0.83	0.79	0.75	0.73
BR(10)	20	0.95	0.97	0.99	1.00	1.00	1.01	0.97	0.98	0.98	0.99	0.99	0.98	0.97	0.98	0.98	0.98	0.98	0.98
	30	0.93	0.95	0.97	0.99	1.00	1.00	0.97	0.98	0.98	0.99	0.99	0.99	0.98	0.98	0.98	0.99	0.99	0.99
	50	0.86	0.90	0.94	0.97	0.98	1.00	0.95	0.96	0.98	0.99	0.99	1.00	0.98	0.98	0.99	0.99	0.99	1.00
	100	0.74	0.78	0.82	0.86	0.89	0.90	0.87	0.88	0.89	0.90	0.91	0.91	0.90	0.92	0.90	0.91	0.92	0.91
	200	0.62	0.65	0.70	0.75	0.79	0.81	0.81	0.81	0.81	0.82	0.82	0.83	0.87	0.86	0.85	0.84	0.83	0.83
	400	0.58	0.60	0.64	0.70	0.75	0.79	0.81	0.80	0.80	0.81	0.81	0.82	0.88	0.87	0.86	0.85	0.84	0.83
PLS(A)	20	1.07	1.05	1.04	1.03	1.01	1.02	1.25	1.24	1.17	1.10	1.06	1.04	1.34	1.30	1.20	1.13	1.08	1.05
	30	0.93	1.02	1.05	1.02	1.02	1.01	1.12	1.21	1.19	1.13	1.07	1.05	1.19	1.23	1.25	1.17	1.10	1.06
	50	0.84	0.91	1.04	1.02	1.02	1.01	1.03	1.11	1.21	1.20	1.11	1.06	1.08	1.14	1.27	1.24	1.15	1.09
	100	0.62	0.64	0.69	0.66	0.76	0.71	0.86	0.89	0.92	0.91	0.85	0.79	0.94	0.97	0.95	0.95	0.90	0.82
	200	0.51	0.50	0.49	0.47	0.47	0.46	0.79	0.78	0.79	0.81	0.81	0.76	0.88	0.88	0.88	0.89	0.88	0.83
	400	0.54	0.52	0.51	0.49	0.46	0.46	0.81	0.79	0.79	0.78	0.80	0.83	0.90	0.89	0.88	0.90	0.91	0.90
BR(A)	20	1.20	1.05	1.00	1.00	0.99	1.00	1.18	1.11	1.05	1.03	1.01	1.00	1.22	1.15	1.06	1.03	1.01	1.00
	30	0.93	1.18	1.03	0.99	0.99	0.99	1.01	1.21	1.09	1.05	1.02	1.01	1.07	1.17	1.11	1.05	1.02	1.01
	50	0.76	0.87	1.25	1.00	0.99	0.99	0.97	1.01	1.22	1.07	1.03	1.02	1.01	1.02	1.19	1.06	1.04	1.02
	100	0.58	0.60	0.61	0.64	0.69	0.67	0.82	0.81	0.84	0.82	0.81	0.77	0.89	0.91	0.87	0.87	0.84	0.81
	200	0.51	0.49	0.48	0.43	0.41	0.39	0.77	0.75	0.72	0.69	0.66	0.63	0.85	0.83	0.80	0.77	0.74	0.71
	400	0.54	0.52	0.50	0.47	0.44	0.42	0.79	0.78	0.75	0.72	0.69	0.67	0.87	0.86	0.84	0.81	0.78	0.74

Notes: See the notes for Table 1, but now with $\alpha_i \sim$ iid $N(0,1)$, for $i = 1, ..., 5$; $\alpha_i = 0$, for $i = 6, ..., N$.

Table 4: Case III, Weak Loadings, $\kappa_2 = 0.25$, Relative MSE compared to PC for $(k_1, k_2) = (1, 1)$

		R^2=.5						R^2=.33						R^2=.2					
	T/N	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400
	20	0.81	0.86	0.91	0.96	0.98	0.99	1.00	1.02	1.05	1.05	1.04	1.03	1.09	1.11	1.11	1.10	1.06	1.04
	30	0.76	0.82	0.88	0.94	0.98	0.99	0.96	1.00	1.04	1.05	1.05	1.04	1.06	1.09	1.10	1.10	1.09	1.05
PLS(1)	50	0.72	0.76	0.82	0.91	0.96	0.99	0.92	0.95	1.00	1.04	1.06	1.05	1.01	1.05	1.09	1.12	1.11	1.08
	100	0.64	0.64	0.67	0.70	0.69	0.63	0.79	0.81	0.81	0.78	0.76	0.72	0.84	0.90	0.90	0.79	0.78	0.73
	200	0.57	0.59	0.60	0.60	0.57	0.49	0.76	0.75	0.74	0.69	0.62	0.50	0.84	0.83	0.79	0.74	0.63	0.49
	400	0.58	0.60	0.63	0.65	0.65	0.62	0.78	0.78	0.79	0.77	0.72	0.65	0.87	0.86	0.85	0.82	0.75	0.66
	20	0.86	0.90	0.94	0.97	0.99	0.99	0.93	0.95	0.97	0.98	0.99	0.99	0.96	0.97	0.98	0.98	0.99	0.99
	30	0.80	0.86	0.91	0.95	0.98	0.99	0.91	0.93	0.95	0.97	0.99	0.99	0.95	0.97	0.97	0.98	0.99	0.99
BR(5)	50	0.71	0.77	0.85	0.92	0.96	0.98	0.87	0.90	0.93	0.96	0.98	0.99	0.94	0.96	0.97	0.98	0.99	0.99
	100	0.51	0.59	0.65	0.74	0.80	0.82	0.74	0.77	0.78	0.80	0.84	0.83	0.81	0.85	0.84	0.84	0.84	0.86
	200	0.40	0.43	0.48	0.56	0.62	0.66	0.66	0.66	0.66	0.67	0.68	0.69	0.78	0.77	0.74	0.73	0.72	0.71
	400	0.37	0.38	0.41	0.47	0.55	0.61	0.66	0.66	0.65	0.66	0.67	0.68	0.80	0.78	0.77	0.74	0.72	0.71
	20	0.94	0.96	0.98	1.00	1.00	1.00	0.96	0.97	0.98	0.99	0.99	0.99	0.97	0.97	0.98	0.98	0.99	0.99
	30	0.89	0.93	0.96	0.99	1.00	1.00	0.94	0.96	0.97	0.99	0.99	0.99	0.96	0.97	0.98	0.99	0.99	0.99
BR(10)	50	0.81	0.86	0.92	0.96	0.98	0.99	0.91	0.94	0.96	0.98	0.99	0.99	0.95	0.97	0.97	0.98	0.99	0.99
	100	0.64	0.72	0.77	0.85	0.89	0.91	0.80	0.84	0.86	0.88	0.91	0.90	0.86	0.89	0.89	0.91	0.91	0.92
	200	0.50	0.55	0.62	0.71	0.77	0.80	0.71	0.73	0.75	0.78	0.81	0.82	0.81	0.82	0.82	0.82	0.83	0.83
	400	0.41	0.45	0.51	0.61	0.70	0.76	0.69	0.70	0.72	0.75	0.78	0.80	0.81	0.81	0.81	0.81	0.82	0.82
	20	0.87	0.90	0.93	0.96	0.98	0.99	1.17	1.18	1.13	1.08	1.04	1.02	1.28	1.30	1.20	1.13	1.07	1.04
	30	0.75	0.86	0.90	0.94	0.97	0.98	1.04	1.16	1.18	1.10	1.06	1.03	1.17	1.27	1.27	1.17	1.10	1.05
PLS(A)	50	0.63	0.72	0.85	0.91	0.95	0.98	0.93	1.02	1.17	1.15	1.09	1.05	1.04	1.12	1.26	1.27	1.16	1.09
	100	0.40	0.42	0.46	0.54	0.60	0.52	0.72	0.77	0.76	0.72	0.75	0.64	0.80	0.89	0.91	0.76	0.76	0.69
	200	0.34	0.31	0.28	0.25	0.24	0.20	0.64	0.62	0.59	0.56	0.48	0.34	0.78	0.77	0.74	0.68	0.57	0.40
	400	0.34	0.33	0.31	0.27	0.24	0.22	0.67	0.65	0.63	0.59	0.55	0.48	0.81	0.79	0.77	0.74	0.68	0.57
	20	1.06	0.92	0.93	0.96	0.98	0.99	1.15	1.09	1.04	1.02	1.01	1.00	1.18	1.13	1.06	1.03	1.02	1.01
	30	0.76	1.10	0.93	0.93	0.97	0.98	0.97	1.24	1.09	1.03	1.01	1.01	1.04	1.17	1.11	1.04	1.02	1.01
BR(A)	50	0.58	0.71	1.14	0.93	0.94	0.97	0.88	0.96	1.26	1.06	1.02	1.01	0.97	1.01	1.17	1.07	1.03	1.03
	100	0.37	0.41	0.43	0.58	0.59	0.59	0.70	0.72	0.71	0.71	0.77	0.72	0.80	0.84	0.84	0.81	0.81	0.80
	200	0.33	0.31	0.28	0.28	0.28	0.24	0.63	0.61	0.58	0.55	0.52	0.51	0.78	0.76	0.73	0.71	0.68	0.65
	400	0.34	0.33	0.31	0.28	0.27	0.27	0.66	0.64	0.62	0.58	0.54	0.53	0.80	0.78	0.76	0.73	0.70	0.67

Notes: The entries are average one-period ahead, out-of-sample MSE ratios relative to principal components-based one-period ahead forecasts across 1,000 Monte Carlo replications. The target and indicator variables are generated through the DGPs in (16) where we consider a factor loading structure as outlined under Case II assuming $\lambda_{i,j} = \tilde{\lambda}_{i,j}/(N^{0.25})$ and $r = 1$ factors; we generate $T + 100$ observations on the predictor and explanatory variables, use the first T observations to estimate the models and use the resulting parameter estimates to forecast the predictor variable over $T+1, \ldots, T+100$. We impose different levels of for the asymptotic fit of the prediction regression, symbolized by the R^2's. In case of both PC regression and PLS regression 1 factor is extracted, and in case of Bayesian regression (BR) a shrinkage parameter $q \times N$ is used with $q = 5, 10$ or in case of $BR(A)$ with a data determined q as discussed in Algorithm 2. Similarly in the case of $PLS(A)$ the number of factors is data-determined as discussed in Algorithm 2.

Table 5: Case III, Weak Loadings, $\kappa_2 = 0.75$, Relative MSE compared to PC for $(k_1, k_2) = (1, 1)$

		R^2=.5						R^2=.33						R^2=.2					
	T/N	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400
	20	0.77	0.83	0.89	0.94	0.98	0.98	1.00	1.03	1.05	1.04	1.03	1.01	1.12	1.12	1.11	1.08	1.05	1.03
	30	0.70	0.76	0.83	0.91	0.95	0.98	0.97	1.00	1.02	1.04	1.03	1.02	1.09	1.11	1.12	1.10	1.06	1.04
PLS(1)	50	0.61	0.67	0.76	0.85	0.92	0.96	0.89	0.93	0.98	1.02	1.03	1.03	1.03	1.07	1.10	1.11	1.09	1.06
	100	0.49	0.45	0.51	0.54	0.57	0.52	0.68	0.72	0.72	0.73	0.65	0.57	0.84	0.87	0.87	0.78	0.76	0.62
	200	0.38	0.38	0.37	0.36	0.31	0.24	0.65	0.64	0.59	0.52	0.42	0.29	0.77	0.74	0.69	0.59	0.45	0.32
	400	0.37	0.37	0.37	0.37	0.35	0.31	0.67	0.65	0.63	0.59	0.51	0.41	0.80	0.78	0.75	0.68	0.58	0.45
	20	0.87	0.91	0.94	0.97	0.98	0.99	0.94	0.96	0.97	0.98	0.99	1.00	0.97	0.98	0.98	0.99	1.00	1.00
	30	0.81	0.86	0.91	0.95	0.97	0.99	0.92	0.94	0.96	0.98	0.99	0.99	0.96	0.97	0.98	0.99	0.99	1.00
BR(5)	50	0.71	0.78	0.85	0.92	0.96	0.98	0.87	0.90	0.94	0.96	0.98	0.99	0.94	0.96	0.97	0.98	0.99	1.00
	100	0.55	0.57	0.67	0.74	0.80	0.82	0.72	0.76	0.79	0.83	0.84	0.84	0.84	0.85	0.86	0.83	0.85	0.85
	200	0.41	0.44	0.49	0.56	0.62	0.66	0.67	0.67	0.67	0.68	0.69	0.69	0.78	0.77	0.75	0.73	0.72	0.71
	400	0.37	0.38	0.41	0.48	0.56	0.62	0.67	0.66	0.66	0.66	0.67	0.68	0.80	0.79	0.77	0.74	0.72	0.71
	20	0.93	0.95	0.97	0.98	0.99	1.00	0.96	0.97	0.98	0.99	1.00	1.00	0.98	0.98	0.99	0.99	1.00	1.00
	30	0.89	0.92	0.95	0.97	0.99	0.99	0.94	0.96	0.97	0.99	0.99	1.00	0.97	0.98	0.99	0.99	0.99	1.00
BR(10)	50	0.81	0.86	0.91	0.95	0.98	0.99	0.91	0.93	0.96	0.98	0.99	0.99	0.95	0.97	0.98	0.99	0.99	1.00
	100	0.67	0.70	0.79	0.84	0.89	0.90	0.80	0.83	0.85	0.90	0.91	0.91	0.88	0.89	0.90	0.89	0.91	0.91
	200	0.51	0.56	0.63	0.71	0.77	0.80	0.72	0.74	0.76	0.79	0.81	0.82	0.82	0.82	0.82	0.82	0.83	0.83
	400	0.42	0.46	0.52	0.62	0.70	0.76	0.70	0.70	0.72	0.75	0.79	0.80	0.81	0.81	0.81	0.81	0.82	0.82
	20	0.89	0.92	0.93	0.95	0.98	0.99	1.22	1.20	1.13	1.07	1.04	1.02	1.35	1.32	1.21	1.12	1.06	1.03
	30	0.74	0.85	0.90	0.93	0.96	0.98	1.07	1.20	1.17	1.10	1.05	1.03	1.19	1.32	1.31	1.18	1.09	1.05
PLS(A)	50	0.59	0.68	0.85	0.91	0.93	0.96	0.92	1.02	1.19	1.16	1.09	1.04	1.06	1.16	1.32	1.27	1.15	1.08
	100	0.44	0.39	0.45	0.49	0.55	0.50	0.68	0.72	0.72	0.72	0.65	0.57	0.85	0.88	0.87	0.77	0.76	0.62
	200	0.33	0.31	0.28	0.25	0.22	0.16	0.63	0.61	0.56	0.50	0.40	0.27	0.76	0.73	0.68	0.58	0.45	0.31
	400	0.35	0.33	0.31	0.27	0.24	0.20	0.66	0.64	0.61	0.56	0.50	0.40	0.79	0.78	0.74	0.67	0.58	0.45
	20	1.06	0.95	0.93	0.95	0.98	0.99	1.22	1.12	1.04	1.02	1.01	1.00	1.22	1.14	1.06	1.04	1.02	1.01
	30	0.75	1.07	0.94	0.94	0.97	0.98	1.00	1.20	1.08	1.03	1.02	1.01	1.03	1.21	1.12	1.05	1.03	1.02
BR(A)	50	0.57	0.70	1.17	0.93	0.94	0.97	0.89	0.97	1.27	1.06	1.03	1.01	0.97	1.03	1.23	1.08	1.04	1.02
	100	0.42	0.37	0.44	0.54	0.60	0.58	0.69	0.71	0.74	0.73	0.74	0.72	0.85	0.84	0.85	0.82	0.83	0.80
	200	0.33	0.31	0.29	0.28	0.28	0.24	0.64	0.62	0.59	0.55	0.54	0.49	0.78	0.76	0.73	0.70	0.67	0.64
	400	0.35	0.33	0.31	0.28	0.27	0.27	0.66	0.65	0.62	0.59	0.55	0.52	0.80	0.79	0.77	0.73	0.70	0.68

Notes: See the notes for Table 4, but now with $\lambda_{i,j} = \tilde{\lambda}_{i,j}/(N^{0.75})$.

Table 6: Forecast evaluation results

h	*RW*	*AR*	*PLS*	*BR*	*RW*	*AR*	*PLS*	*BR*
	January 1972 - December 2006				January 1985 - December 2006			
				CPI Inflation				
1	1.083	1.010	0.973	0.949***	1.030	1.026	0.986	0.951*
	(2.455)	(0.452)	(−1.101)	(−2.020)	(0.774)	(1.052)	(−0.555)	(−1.553)
3	1.063	1.062	0.993	0.961*	0.937*	1.036	0.977	0.938***
	(1.349)	(1.625)	(−0.251)	(−1.377)	(−1.647)	(1.101)	(−1.171)	(−2.148)
12	1.008	1.037	0.965*	0.942**	0.849***	0.981	0.954*	0.924***
	(0.147)	(0.680)	(−1.380)	(−1.850)	(−2.436)	(−0.313)	(−1.287)	(−2.032)
				Unemployment				
1	1.148	1.102	1.002	1.012	1.059	1.047	1.003	1.015
	(3.479)	(3.413)	(0.112)	(0.580)	(1.807)	(1.608)	(0.200)	(0.765)
3	1.219	1.105	0.953*	0.965	1.142	1.085	0.953**	1.001
	(2.425)	(2.518)	(−1.496)	(−1.027)	(2.364)	(2.118)	(−1.763)	(0.041)
12	1.006	0.991	0.885**	0.915	1.024	0.985	0.960	0.951
	(0.082)	(−0.136)	(−1.641)	(−1.226)	(0.517)	(−0.356)	(−0.766)	(−1.137)
				Industrial Production				
1	1.172	1.077	1.012	0.984	1.006	1.001	0.976	0.939***
	(3.402)	(2.702)	(0.571)	(−0.787)	(0.183)	(0.051)	(−1.127)	(−2.597)
3	1.200	1.073	0.930**	0.963	1.130	1.007	0.998	0.966
	(2.553)	(1.131)	(−1.920)	(−0.879)	(1.836)	(0.135)	(−0.066)	(−1.133)
12	0.982	0.902*	0.798***	0.847***	1.171	0.890***	0.972	0.933**
	(−0.208)	(−1.603)	(−2.759)	(−2.354)	(1.427)	(−2.748)	(−0.671)	(−1.957)
				Federal Funds Rate				
1	1.060	1.068	0.925*	0.871**	0.792**	0.719***	0.960	0.710***
	(0.862)	(0.677)	(−1.503)	(−2.014)	(−2.013)	(−2.728)	(−0.482)	(−2.996)
3	0.984	1.044	0.897*	0.901	0.937	0.942	0.964	0.849***
	(−0.219)	(0.485)	(−1.549)	(−1.216)	(−0.863)	(−0.854)	(−0.712)	(−3.391)
12	1.073	1.157	1.004	1.031	1.045	1.077	1.030	0.955
	(1.485)	(2.158)	(0.110)	(0.728)	(0.639)	(1.032)	(0.476)	(−0.863)

Notes: The table reports the ratio of the RMSE of the random walk (RW) model (23), the autoregressive (AR) model (22) and versions of (17) based on PLS and Bayesian regression (BR) *vis-à-vis* the RMSE of (17) based on principal components (PC) regression; see Section 4.1. The optimal number of PLS factors or shrinkage parameter (BR) as well as the optimal lag order p are for each forecast determined by means of out-of-sample cross-validation using Algorithm 4 with $p^{max} = 12$. In case of PC regression, we use for each forecast the Bai and Ng (2002) information criterion IC_2 (21) based on a upper bound of 6 principal components, where conditional on the selected number of principal components that minimizes (21) we use BIC with $p^{max} = 12$ to select the optimal lag order. In parentheses we report the Diebold and Mariano (1995) statistic with the Harvey et al. (1997) variance estimation correction utilizing a rectangular HAC estimator with lag truncation $h-1$ for the null hypothesis of equal finite sample prediction accuracy *versus* the alternative hypothesis that a model outperforms PC regression, where *, ** and *** indicates rejection of this null at the 10%, 5% and 1% level, respectively, based on one-sided standard normal critical values.

References

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–173.

Bai, J. and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

— (2006), "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74, 1133–1150.

Bailey, N., Kapetanios, G., and Pesaran, M. H. (2013), "Exponent of Cross-sectional Dependence: Estimation and Inference," *Mimeo, University of Cambridge.*

Bates, J. M. and Granger, C. W. J. (1969), "The Combination of Forecasts," *Operations Research Quarterly*, 20, 451–468.

Boivin, J. and Ng, S. (2006), "Are More Data Always Better for Factor Analysis?" *Journal of Econometrics*, 132, 169–194.

Capistrán, C. and Timmermann, A. (2008), "Forecast Combination With Entry and Exit of Experts," *Journal of Business & Economic Statistics*, forthcoming.

Chamberlain, G. and Rothschild, M. (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1305–1324.

Chudik, A., Pesaran, M. H., and Tosetti, E. (2011), "Weak and Strong Cross-Section Dependence and Estimation of Large Panels," *Econometrics Journal*, 14, C45–C90.

Chun, H. and Keleş, S. (2010), "Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Variable Selection," *Journal of the Royal Statistical Society B*, 72, 3–25.

Clark, T. E. and McCracken, M. W. (2013), "Advances in Forecast Evaluation," in *Handbook of Economic Forecasting, Vol. 2*, eds. Timmermann, A. and Elliott, G., Amsterdam: Elsevier, pp. 1107–1201.

D'Agostino, A. and Giannone, D. (2006), "Comparing Alternative Predictors Based on Large-Panel Factor Models," Working Paper 680, European Central Bank.

D'Agostino, A., Giannone, D., and Surico, P. (2006), "(Un)Predictability and Macroeconomic Stability," Working Paper 605, European Central Bank.

De Mol, C., Giannone, D., and Reichlin, L. (2008), "Forecasting Using a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components?" *Journal of Econometrics*, 146, 318–328.

Diebold, F. X. and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

Faust, J. and Wright, J. H. (2009), "Comparing Greenbook and Reduced Form Forecasts Using a Large Real-Time Dataset," *Journal of Business & Economic Statistics*, 27, 468–479.

— (2012), "Forecasting Inflation," in *Handbook of Economic Forecasting*, Amsterdam: Elsevier.

Garthwaite, P. H. (1994), "An Interpretation of Partial Least Squares," *Journal of the American Statistical Association*, 89, 122–127.

Giraitis, L., Kapetanios, G., and Price, S. (2013), "Adaptive forecasting in the presence of recent and ongoing structural change," *Forthcoming in the Journal of Econometrics.*

Groen, J. J. J. and Kapetanios, G. (2013), "Model Selection Criteria for Factor-Augmented Regressions," *Oxford Bulletin of Economics and Statistics*, 75, 37–63.

Groen, J. J. J., Paap, R., and Ravazzolo, F. (2013), "Real-Time Inflation Forecasting in a Changing World," *Journal of Business & Economic Statistics*, 21, 29–44.

Harvey, D., Leybourne, S., and Newbold, P. (1997), "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13, 281–291.

Helland, I. S. (1990), "Partial Least Squares Regression and Statistical Models," *Scandinavian Journal of Statistics*, 17, 97–114.

Hendry, D. F. (1995), *Dynamic Econometrics*, Oxford University Press.

Kapetanios, G. and Marcellino, M. (2010), "Factor-GMM estimation with large sets of possibly weak instruments," *Computational Statistics and Data Analysis*, 54, 2655–2675.

Kelly, B. and Pruitt, S. (2012), "The Three-Pass Regression Filter: A New Approach to Forecasting Using Many Predictors," *Mimeo, University of Chicago.*

Krolzig, H.-M. and Hendry, D. F. (2001), "Computer Automation of General-to-Specific Model Selection Procedures," *Journal of Economic Dynamics and Control*, 25, 831–866.

McConnell, M. and Perez-Quiros, G. (2000), "Output Fluctuations in the United States: What has Changed Since the Early 1980's?" *American Economic Review*, 90, 1464–1476.

Onatski, A. (2009), "Asymptotics of the Principal Components Estimator of Large Factor Models with Weak Factors," mimeo, Columbia University.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Sensier, M. and van Dijk, D. J. C. (2004), "Testing for Volatility Changes in U.S. Macroeconomic Time Series," *Review of Economics and Statistics*, 86, 833–839.

Stock, J. H. and Watson, M. W. (2002a), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

— (2002b), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–162.

— (2007), "Forecasting in Dynamic Factor Models Subject to Structural Instability," mimeo, Harvard University and Princeton University.

— (2012), "Generalized Shrinkage Methods for Forecasting Using Many Predictors," *Journal of Business and Economic Statistics*, 30, 481–493.

Svensson, L. E. O. (2005), "Monetary Policy with Judgment: Forecast Targeting," *International Journal of Central Banking*, 1, 1–54.

Timmermann, A. (2006), "Forecast Combinations," in *Handbook of Economic Forecasting*, eds. Elliott, G., Granger, C. W. J., and Timmermann, A., Amsterdam: North Holland, pp. 135–196.

Uhlig, H. (2009), "Macroecnomic Dynamics in the Euro Area. Discussion by Harald Uhlig," in *NBER Macroeconomics Annual 2008*, eds. Acemoglu, D., Rogoff, K., and Woodford, W., Chicago: University of Chicago Press, vol. 23.

Wold, H. (1982), "Soft Modeling. The Basic Design and Some Extensions," in *Systems Under Indirect Observation, Vol. 2*, eds. Jöreskog, K.-G. and Wold, H., Amsterdam: North-Holland.

Appendix
for
"Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting"

Jan J. J. Groen
Federal Reserve Bank of New York

George Kapetanios
Queen Mary University of London

Outline

This Appendix provides proofs of our theoretical analysis, some additional simulation results and details on data used in the article. In Section A we provide the proof underlying Theorem 1, whereas in Sections B and C we do that for Theorems 2 and 3, respectively. Section D reports the results for a number of Monte Carlo experiments that complement those in Section 3 of the main article. Finally, in Section E describes the data and the necessary data transformations as used in Section 4 of the main article. Note that when we refer in this Appendix numerically to equations, tables, sections and so on, these pertain to the ones in the main article.

A Proof of Theorem 1

The local nature of the factor loadings in (8) we need to use a different normalization than $\Lambda'\Lambda/N = I$, which is implied by Assumption 1(b), and instead we us $\Lambda'^{1-2\kappa} = I$. This results in $\hat{F} = N^{-1+2\kappa}\tilde{X}\tilde{\Lambda}$ and $\tilde{\Lambda} = T^{-1}\tilde{X}'\tilde{F}$, where $\tilde{F}$ is the solution to the optimization problem of maximizing $tr(F'(\tilde{X}'\tilde{X})F)$ subject to $F'F/T = I$. Let $H = \left((\tilde{F}'F/T)(\Lambda_N'\Lambda_N/N^{1-2\kappa})\right)'$. Then, (7) becomes

$$(\hat{\zeta}'\hat{F}_t - \alpha'x_t) = \left(\zeta'HF_t - \hat{\zeta}\hat{F}_t\right) + \alpha'e_t. \tag{A.1}$$

Rearranging (A.1) using the definition of ζ in (6) and assuming that for forecasting model (1) we have $\|\alpha\| = O(N^{-1/2})$, the result directly follows from Corollary 1 of Bai and Ng (2006) and Lemma 1 of Kapetanios and Marcellino (2010).

B Proof of Theorem 2

We wish to show that

$$E\left(y_T^2|x_T\right) - E\left(\hat{y}_T^2|x_T\right) = \sigma_\epsilon^2,$$

$$\begin{aligned} E\left(y_T^2|x_T\right) - E\left(\hat{y}_T^2|x_T\right) &= E\left(\alpha x_T x_T'\alpha|x_T\right) - E\left(\alpha_N' x_T x_T'\alpha_N|x_T\right) + E\left(\alpha_N' x_T x_T'\alpha_N|x_T\right) - E\left(\hat{\alpha}_N' x_T x_T'\hat{\alpha}_N|x_T\right) \\ &= \left\|x_T'\alpha - x_T'\alpha_N + x_T'\alpha_N - x_T'\hat{\alpha}_N\right\| \leq \left\|x_T'\alpha - x_T'\alpha_N\right\| + \left\|x_T'\alpha_N - x_T'\hat{\alpha}_N\right\|, \end{aligned}$$

and

$$\left\|E\left(\alpha x_T x_T'\alpha|x_T\right) - E\left(\alpha_N' x_T x_T'\alpha_N|x_T\right)\right\| \leq C\left(E\left(\alpha x_T x_T'\alpha\right) - E\left(\alpha_N' x_T x_T'\alpha_N\right)\right),$$

where $\alpha_N = cov(y_t, x_t) = E\left(y_t x_t\right) = \Sigma\alpha$.

We can write

$$E\left(\alpha x_T x_T'\alpha\right) - E\left(\alpha_N' x_T x_T'\alpha_N\right) \leq \left\|\alpha'\left(\Sigma - I\right)\alpha\right\|$$

and

$$E\left(\alpha_N' x_T x_T'\alpha_N | x_T\right) - E\left(\hat{\alpha}_N' x_T x_T'\hat{\alpha}_N | x_T\right) \leq C\left\|E\left(\hat{\alpha}_N - \alpha_N\right)'\left(\hat{\alpha}_N - \alpha_N\right)\right\| \leq$$

$$\left(\sum_{j=1}^{N} E\left(\left(\frac{1}{T}\sum_{t=1}^{T} x_{j,t}y_t - \sigma_{Xy,j}\right)^2\right) + \sum_{j=1}^{N}\sum_{i=1}^{N} E\left(\left(\frac{1}{T}\sum_{t=1}^{T} x_{j,t}y_t - \sigma_{Xy,j}\right)\left(\frac{1}{T}\sum_{t=1}^{T} x_{i,t}y_t - \sigma_{Xy,i}\right)\right)\right)^{1/2}.$$

But

$$E\left(\left(\frac{1}{T}\sum_{t=1}^{T} x_{j,t}y_t - \sigma_{Xy,j}\right)^2\right) = O\left(T^{-1}\right), \text{ uniformly over } i,$$

and

$$E\left(\left(\frac{1}{T}\sum_{t=1}^{T} x_{j,t}y_t - \sigma_{Xy,j}\right)\left(\frac{1}{T}\sum_{t=1}^{T} x_{i,t}y_t - \sigma_{Xy,i}\right)\right) = O\left(T^{-1}\right), \text{ uniformly over } i, j.$$

Therefore,

$$\left(\sum_{j=1}^{N} E\left(\left(\frac{1}{T}\sum_{t=1}^{T} x_{j,t}y_t - \sigma_{Xy,j}\right)^2\right) + \sum_{j=1}^{N}\sum_{i=1}^{N} E\left(\left(\frac{1}{T}\sum_{t=1}^{T} x_{j,t}y_t - \sigma_{Xy,j}\right)\left(\frac{1}{T}\sum_{t=1}^{T} x_{i,t}y_t - \sigma_{Xy,i}\right)\right)\right)^{1/2}$$
$$= O\left(T^{-1/2}N\right).$$

C Proof of Theorem 3

Let α_{BRR} denote the population counterpart of $\hat{\alpha}_{BRR}$, which via (4) equals

$$\alpha_{BRR} = \left(\Lambda'\Lambda + \Sigma_e + vI\right)^{-1}\left(\Lambda'\Lambda + \Sigma_e\right)\alpha,$$

where $\Sigma_e = E(e_t e_t')$. Given Lemma 4 of De Mol et al. (2008), we only need to show that

$$\|\alpha - \alpha_{BRR}\| = O\left(\frac{v\|\alpha\|}{\mu_{\min}\left(\Lambda\Lambda'\right)}\right), \tag{C.1}$$

where $\mu_{\min}\left(\Lambda\Lambda'\right)$ is the minimum eigenvalue of $\Lambda\Lambda'$, and

$$\|\alpha_{BRR} - \hat{\alpha}_{BRR}\| = O_p\left(\frac{N^{1/2}}{vT^{1/2}}\left[1 + N^{1/2}\|\alpha\|\right]\right). \tag{C.2}$$

Then the result follows. For (C.1), we have that

$$\alpha - \alpha_{BRR} = \left[\left(\Lambda'\Lambda + \Sigma_e + vI\right)^{-1} - \left(\Lambda'\Lambda + \Sigma_e\right)^{-1}\right]\left(\Lambda'\Lambda + \Sigma_e\right)\alpha =$$

$$\left(\Lambda'\Lambda + \Sigma_e\right)^{-1} v \left(\Lambda'\Lambda + \Sigma_e + vI\right)^{-1} \left(\Lambda'\Lambda + \Sigma_e\right) \alpha$$

So,

$$\|\alpha - \alpha_{BRR}\| \leq \left\| \left(\Lambda'\Lambda + \Sigma_e\right)^{-1} v \left(\Lambda'\Lambda + \Sigma_e + vI\right)^{-1} \left(\Lambda'\Lambda + \Sigma_e\right) \alpha \right\| \leq v \left\| \left(\Lambda'\Lambda + \Sigma_e\right)^{-1} \alpha \right\| \leq \frac{v \|\alpha\|}{\mu_{\min}\left(\Lambda\Lambda'\right)}.$$

For (C.2), we simply note that $\|\alpha_{BRR}\| \leq \|\alpha\|$ and the result follows by Lemma 2 and the proof of Lemma 3 of De Mol et al. (2008).

D Additional Monte Carlo Experiments

In this section we report the simulation results for a number of additional variations of the Monte Carlo DGP (16) in 3:

Case IV: $\lambda_{i,j} \sim \text{iid}N(0,1)$ for $i = 1, \ldots, N$ & $j = 1, \ldots, r$, $c_2 = 2$. $\alpha_i \sim \text{iid } N(0,1)$

Case V: $\lambda_{i,j} \left\{ \begin{array}{ll} = \tilde{\lambda}_{i,j} \sim \text{iid}N(0,1) & \text{for } i = 1, \ldots, N_1,\ N_1 = N^{\kappa_1}\ (\kappa_1 = 0.25,\ 0.75)\ \&\ j = 1, \ldots, r. \\ = 0 & \text{for } i = N_1 + 1, \ldots, N\ \&\ j = 1, \ldots, r. \end{array} \right\}$, , $c_2 = 1$.

Case VI: $\lambda_{i,j} = 0$, $c_2 = 1$. $\alpha_i \sim \text{iid } N(0,1)$

Case IV (Table D.1) is one where while the factor is pervasive the idiosyncratic noise term is more dominant implying that PLS and Baeysian regression should have a larger advantage over PC. Under Case V (Tables D.2-D.3) we assume a structure of the form (13) where we set $N_1 = N^{\kappa_1}$ and $\kappa_1 = 0.25,\ 0.75$. This is operationalized by using a factor model where the factors are not pervasive but affect a subset N_1 of the variables in x_t. The subset of predictor variables $(N_1 - N)$ are non-informative for the factors and this subset will dominate the panel of N predictor variables as $N \to \infty$, where κ_1 determines the speed with which this occurs. Finally, Case VI (Table D.4) is a case where there are no common factors driving the dynamics of the x_t predictor variables.

Table D.1: Case IV, Strong Factor, Relative MSE compared to PC for $(k_1, k_2) = (1, 1)$

		R^2=0.5						R^2=0.33						R^2=0.2					
	T/N	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400
PLS(1)	20	0.75	0.83	0.90	0.97	1.02	1.04	0.91	0.95	0.99	1.03	1.05	1.06	1.01	1.04	1.07	1.08	1.09	1.08
	30	0.70	0.78	0.87	0.96	1.01	1.04	0.87	0.92	0.98	1.02	1.06	1.07	0.98	1.01	1.05	1.07	1.09	1.10
	50	0.64	0.74	0.83	0.94	1.00	1.04	0.83	0.87	0.94	1.00	1.04	1.06	0.94	0.97	1.01	1.05	1.07	1.09
	100	0.52	0.58	0.68	0.82	0.88	0.89	0.70	0.77	0.82	0.82	0.92	0.91	0.81	0.86	0.85	0.86	0.89	0.90
	200	0.51	0.59	0.68	0.77	0.84	0.86	0.69	0.72	0.77	0.81	0.86	0.86	0.78	0.80	0.83	0.85	0.85	0.85
	400	0.52	0.60	0.71	0.81	0.87	0.92	0.70	0.74	0.80	0.86	0.90	0.92	0.80	0.83	0.86	0.89	0.91	0.92
BR(5)	20	0.64	0.74	0.83	0.91	0.96	0.98	0.83	0.87	0.91	0.95	0.97	0.99	0.94	0.96	0.97	0.98	0.99	0.99
	30	0.53	0.63	0.75	0.87	0.93	0.96	0.77	0.82	0.88	0.93	0.96	0.98	0.92	0.94	0.96	0.97	0.99	0.99
	50	0.40	0.50	0.63	0.79	0.88	0.94	0.71	0.75	0.82	0.89	0.94	0.97	0.88	0.91	0.94	0.96	0.98	0.99
	100	0.24	0.28	0.34	0.44	0.53	0.59	0.52	0.55	0.59	0.55	0.64	0.65	0.69	0.74	0.74	0.66	0.68	0.67
	200	0.21	0.20	0.21	0.24	0.27	0.29	0.49	0.46	0.43	0.39	0.36	0.34	0.65	0.62	0.56	0.48	0.41	0.36
	400	0.21	0.20	0.20	0.20	0.23	0.27	0.50	0.48	0.46	0.42	0.38	0.35	0.68	0.66	0.62	0.56	0.48	0.41
BR(10)	20	0.73	0.81	0.88	0.94	0.97	0.99	0.85	0.89	0.93	0.96	0.98	0.99	0.92	0.94	0.96	0.98	0.98	0.99
	30	0.64	0.72	0.82	0.91	0.95	0.98	0.80	0.85	0.90	0.94	0.97	0.98	0.90	0.92	0.95	0.97	0.98	0.99
	50	0.50	0.60	0.72	0.84	0.92	0.96	0.73	0.78	0.84	0.91	0.95	0.98	0.87	0.90	0.93	0.96	0.98	0.99
	100	0.30	0.36	0.45	0.56	0.64	0.70	0.55	0.59	0.65	0.65	0.73	0.74	0.71	0.76	0.76	0.73	0.76	0.76
	200	0.23	0.24	0.28	0.35	0.42	0.46	0.50	0.49	0.48	0.49	0.50	0.51	0.66	0.64	0.60	0.57	0.55	0.53
	400	0.22	0.22	0.23	0.28	0.35	0.42	0.50	0.49	0.48	0.48	0.48	0.49	0.68	0.66	0.64	0.60	0.57	0.54
PLS(A)	20	0.69	0.78	0.85	0.93	0.97	1.00	1.02	1.05	1.02	1.02	1.02	1.02	1.19	1.18	1.13	1.09	1.05	1.04
	30	0.57	0.69	0.80	0.89	0.95	0.98	0.92	1.02	1.05	1.02	1.02	1.02	1.07	1.18	1.17	1.10	1.07	1.04
	50	0.45	0.54	0.70	0.83	0.91	0.96	0.80	0.89	1.02	1.03	1.01	1.02	0.96	1.04	1.17	1.15	1.09	1.05
	100	0.23	0.25	0.29	0.39	0.49	0.54	0.57	0.62	0.69	0.61	0.75	0.67	0.77	0.82	0.80	0.77	0.78	0.75
	200	0.20	0.19	0.17	0.14	0.11	0.13	0.49	0.47	0.43	0.42	0.41	0.41	0.67	0.65	0.63	0.64	0.62	0.56
	400	0.21	0.20	0.18	0.16	0.12	0.10	0.51	0.49	0.46	0.43	0.41	0.41	0.69	0.68	0.65	0.63	0.65	0.65
BR(A)	20	1.00	0.78	0.83	0.91	0.96	0.98	1.19	1.02	0.97	0.98	0.99	0.99	1.29	1.08	1.04	1.02	1.01	1.00
	30	0.53	1.13	0.78	0.87	0.93	0.96	0.87	1.41	0.99	0.98	0.98	0.99	0.98	1.29	1.05	1.02	1.01	1.00
	50	0.35	0.47	1.13	0.80	0.89	0.94	0.74	0.82	1.48	0.97	0.97	0.99	0.90	0.95	1.32	1.03	1.02	1.01
	100	0.21	0.23	0.25	0.40	0.42	0.50	0.53	0.56	0.60	0.56	0.65	0.62	0.70	0.75	0.75	0.71	0.71	0.71
	200	0.20	0.18	0.16	0.13	0.10	0.10	0.48	0.46	0.42	0.38	0.33	0.31	0.65	0.62	0.58	0.53	0.48	0.43
	400	0.21	0.20	0.18	0.16	0.13	0.10	0.50	0.48	0.46	0.42	0.37	0.33	0.68	0.66	0.63	0.58	0.53	0.48

Notes: See the notes for Table 1, but now with $c_2 = 2$.

Table D.2: Case II, Weak Pervasiveness, $\kappa_1 = 0.75$, Relative MSE compared to PC for $(k_1, k_2) = (1, 1)$

		R^2=.5						R^2=.33						R^2=.2					
	T/N	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400
PLS(1)	20	0.84	0.90	0.95	1.00	1.01	1.01	1.00	1.02	1.05	1.06	1.06	1.05	1.07	1.10	1.10	1.11	1.09	1.07
	30	0.80	0.86	0.92	0.97	1.01	1.03	0.96	1.00	1.02	1.06	1.07	1.07	1.05	1.07	1.09	1.10	1.10	1.09
	50	0.75	0.83	0.88	0.95	1.00	1.03	0.92	0.96	1.00	1.04	1.06	1.08	1.01	1.03	1.06	1.10	1.11	1.11
	100	0.69	0.76	0.81	0.84	0.87	0.85	0.84	0.86	0.87	0.88	0.86	0.86	0.90	0.92	0.93	0.91	0.87	0.83
	200	0.64	0.71	0.74	0.78	0.80	0.79	0.81	0.82	0.83	0.84	0.82	0.79	0.87	0.88	0.87	0.85	0.82	0.78
	400	0.65	0.71	0.77	0.82	0.84	0.86	0.82	0.84	0.87	0.88	0.89	0.88	0.89	0.90	0.91	0.90	0.89	0.87
BR(5)	20	0.86	0.91	0.94	0.97	0.99	1.00	0.93	0.95	0.96	0.98	0.99	0.99	0.96	0.97	0.98	0.98	0.98	0.99
	30	0.80	0.85	0.91	0.95	0.98	1.00	0.91	0.93	0.96	0.98	0.99	0.99	0.95	0.97	0.98	0.99	0.99	0.99
	50	0.70	0.77	0.85	0.92	0.96	0.98	0.87	0.90	0.93	0.96	0.98	0.99	0.94	0.96	0.97	0.98	0.99	0.99
	100	0.54	0.57	0.65	0.74	0.79	0.81	0.75	0.78	0.78	0.82	0.80	0.83	0.83	0.83	0.84	0.85	0.84	0.85
	200	0.40	0.43	0.48	0.55	0.62	0.65	0.67	0.66	0.66	0.67	0.68	0.69	0.78	0.77	0.74	0.73	0.71	0.71
	400	0.36	0.37	0.40	0.47	0.55	0.61	0.67	0.66	0.65	0.66	0.67	0.68	0.80	0.78	0.77	0.74	0.72	0.71
BR(10)	20	0.94	0.97	0.99	1.00	1.02	1.02	0.96	0.97	0.98	0.99	1.00	1.00	0.97	0.98	0.98	0.98	0.98	0.99
	30	0.89	0.92	0.97	0.99	1.01	1.02	0.94	0.96	0.98	0.99	1.00	1.00	0.96	0.97	0.98	0.99	0.99	0.99
	50	0.81	0.86	0.92	0.97	0.99	1.00	0.91	0.93	0.96	0.98	0.99	1.00	0.95	0.97	0.98	0.99	0.99	0.99
	100	0.65	0.70	0.77	0.84	0.87	0.90	0.81	0.84	0.85	0.89	0.88	0.90	0.87	0.88	0.89	0.90	0.91	0.92
	200	0.49	0.55	0.62	0.70	0.76	0.79	0.72	0.73	0.75	0.78	0.80	0.82	0.81	0.82	0.81	0.82	0.82	0.83
	400	0.41	0.45	0.51	0.61	0.70	0.76	0.69	0.70	0.72	0.75	0.78	0.80	0.81	0.81	0.81	0.81	0.82	0.82
PLS(A)	20	0.89	0.92	0.93	0.98	0.99	1.00	1.15	1.16	1.12	1.07	1.04	1.03	1.28	1.27	1.19	1.12	1.07	1.05
	30	0.76	0.87	0.92	0.95	0.98	1.00	1.04	1.14	1.13	1.10	1.06	1.04	1.14	1.24	1.24	1.16	1.10	1.06
	50	0.63	0.74	0.85	0.92	0.96	1.00	0.94	1.01	1.13	1.13	1.08	1.05	1.04	1.11	1.24	1.23	1.14	1.08
	100	0.42	0.41	0.43	0.54	0.59	0.61	0.78	0.78	0.80	0.80	0.72	0.69	0.84	0.90	0.90	0.90	0.79	0.71
	200	0.32	0.31	0.28	0.24	0.24	0.25	0.65	0.63	0.61	0.61	0.59	0.52	0.79	0.78	0.76	0.75	0.70	0.60
	400	0.34	0.33	0.31	0.27	0.24	0.22	0.67	0.65	0.63	0.61	0.61	0.59	0.81	0.80	0.79	0.78	0.77	0.75
BR(A)	20	1.10	0.94	0.92	0.96	0.98	0.99	1.17	1.09	1.04	1.01	1.01	1.00	1.19	1.13	1.06	1.03	1.01	1.01
	30	0.77	1.11	0.93	0.94	0.97	0.99	0.99	1.18	1.06	1.03	1.01	1.01	1.03	1.26	1.10	1.04	1.02	1.01
	50	0.56	0.71	1.16	0.92	0.94	0.97	0.89	0.95	1.25	1.03	1.02	1.01	0.98	1.01	1.19	1.07	1.04	1.02
	100	0.42	0.39	0.42	0.58	0.56	0.59	0.72	0.72	0.71	0.75	0.71	0.74	0.83	0.83	0.82	0.84	0.81	0.81
	200	0.32	0.31	0.28	0.27	0.28	0.25	0.64	0.62	0.58	0.55	0.53	0.51	0.77	0.76	0.73	0.70	0.67	0.63
	400	0.34	0.33	0.31	0.28	0.27	0.27	0.66	0.64	0.62	0.58	0.54	0.52	0.80	0.78	0.77	0.73	0.70	0.67

Notes: The entries are average one-period ahead, out-of-sample MSE ratios relative to principal components-based one-period ahead forecasts across 1,000 Monte Carlo replications. The target and indicator variables are generated through the DGPs in (16) where we consider a factor loading structure as outlined under Case II assuming $N_1 = N^{0.75}$ and $r = 1$ factors; we generate $T + 100$ observations on the predictor and explanatory variables, use the first T observations to estimate the models and use the resulting parameter estimates to forecast the predictor variable over $T + 1, \ldots, T + 100$. We impose different levels of for the asymptotic fit of the prediction regression, symbolized by the R^2's. In case of both PC regression and PLS regression 1 factor is extracted, and in case of Bayesian regression (BR) a shrinkage parameter $q \times N$ is used with $q = 5, 10$ or in case of $BR(A)$ with a data determined q as discussed in Algorithm 2. Similarly in the case of $PLS(A)$ the number of factors is data-determined as discussed in Algorithm 2.

Table D.3: Case II, Weak Pervasiveness, $\kappa_1 = 0.25$, Relative MSE compared to PC for $(k_1, k_2) = (1, 1)$

		R^2=.5						R^2=.33						R^2=.2					
	T/N	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400
PLS(1)	20	0.77	0.83	0.89	0.94	0.97	0.99	0.99	1.02	1.03	1.04	1.03	1.01	1.11	1.11	1.11	1.08	1.05	1.03
	30	0.71	0.77	0.84	0.91	0.95	0.98	0.96	0.99	1.01	1.03	1.03	1.02	1.07	1.10	1.11	1.10	1.06	1.04
	50	0.63	0.67	0.75	0.85	0.92	0.96	0.89	0.93	0.99	1.02	1.03	1.03	1.03	1.06	1.09	1.11	1.09	1.06
	100	0.45	0.48	0.57	0.55	0.55	0.52	0.73	0.71	0.72	0.75	0.66	0.62	0.85	0.85	0.87	0.72	0.71	0.64
	200	0.41	0.39	0.40	0.37	0.32	0.25	0.67	0.64	0.62	0.53	0.42	0.30	0.78	0.75	0.70	0.60	0.47	0.32
	400	0.41	0.40	0.40	0.39	0.37	0.31	0.69	0.67	0.66	0.60	0.53	0.42	0.81	0.79	0.76	0.70	0.59	0.46
BR(5)	20	0.86	0.90	0.94	0.97	0.98	0.99	0.93	0.95	0.97	0.98	0.99	1.00	0.97	0.97	0.98	0.99	0.99	1.00
	30	0.80	0.86	0.91	0.95	0.97	0.99	0.92	0.93	0.95	0.98	0.99	0.99	0.96	0.97	0.98	0.99	0.99	1.00
	50	0.71	0.77	0.85	0.92	0.96	0.98	0.87	0.90	0.93	0.96	0.98	0.99	0.94	0.96	0.97	0.98	0.99	0.99
	100	0.52	0.58	0.67	0.74	0.79	0.81	0.75	0.76	0.78	0.83	0.83	0.85	0.84	0.83	0.86	0.83	0.84	0.84
	200	0.41	0.43	0.48	0.56	0.62	0.66	0.67	0.66	0.67	0.68	0.69	0.69	0.78	0.77	0.75	0.73	0.72	0.71
	400	0.37	0.38	0.40	0.48	0.55	0.61	0.67	0.66	0.65	0.66	0.67	0.68	0.80	0.78	0.77	0.74	0.72	0.71
BR(10)	20	0.93	0.95	0.97	0.98	0.99	1.00	0.96	0.97	0.98	0.99	0.99	1.00	0.97	0.98	0.98	0.99	0.99	1.00
	30	0.89	0.92	0.95	0.97	0.99	0.99	0.94	0.96	0.97	0.99	0.99	1.00	0.97	0.98	0.98	0.99	0.99	1.00
	50	0.82	0.86	0.91	0.95	0.98	0.99	0.91	0.93	0.96	0.98	0.99	0.99	0.95	0.96	0.98	0.98	0.99	1.00
	100	0.65	0.71	0.78	0.84	0.88	0.90	0.81	0.83	0.85	0.89	0.90	0.92	0.88	0.88	0.91	0.89	0.91	0.90
	200	0.50	0.55	0.62	0.71	0.77	0.80	0.72	0.74	0.76	0.79	0.81	0.82	0.82	0.82	0.82	0.82	0.83	0.83
	400	0.42	0.45	0.51	0.62	0.70	0.76	0.70	0.70	0.72	0.75	0.78	0.80	0.81	0.81	0.81	0.81	0.82	0.82
PLS(A)	20	0.89	0.91	0.93	0.95	0.97	0.99	1.19	1.18	1.11	1.07	1.04	1.02	1.31	1.30	1.21	1.12	1.06	1.03
	30	0.73	0.86	0.90	0.93	0.96	0.98	1.05	1.18	1.19	1.10	1.05	1.03	1.17	1.27	1.29	1.17	1.09	1.05
	50	0.60	0.68	0.83	0.91	0.94	0.96	0.92	1.02	1.19	1.16	1.08	1.05	1.07	1.15	1.31	1.26	1.16	1.08
	100	0.38	0.41	0.47	0.53	0.51	0.49	0.71	0.70	0.71	0.75	0.65	0.62	0.85	0.85	0.87	0.71	0.72	0.64
	200	0.33	0.31	0.28	0.25	0.22	0.17	0.63	0.61	0.58	0.50	0.40	0.28	0.77	0.73	0.69	0.59	0.46	0.31
	400	0.35	0.33	0.31	0.27	0.23	0.21	0.66	0.64	0.62	0.56	0.50	0.40	0.80	0.78	0.74	0.68	0.58	0.45
BR(A)	20	1.11	0.95	0.93	0.96	0.98	0.99	1.23	1.10	1.04	1.02	1.01	1.00	1.18	1.13	1.06	1.04	1.02	1.01
	30	0.77	1.13	0.94	0.94	0.97	0.98	0.99	1.19	1.08	1.03	1.02	1.01	1.02	1.21	1.12	1.05	1.02	1.01
	50	0.57	0.70	1.19	0.92	0.94	0.97	0.88	0.96	1.22	1.07	1.02	1.01	0.98	1.01	1.21	1.07	1.04	1.02
	100	0.37	0.40	0.45	0.58	0.53	0.57	0.71	0.71	0.72	0.75	0.72	0.77	0.84	0.82	0.86	0.80	0.80	0.79
	200	0.33	0.30	0.29	0.28	0.28	0.25	0.64	0.62	0.59	0.55	0.53	0.51	0.78	0.76	0.73	0.71	0.67	0.63
	400	0.35	0.33	0.31	0.28	0.27	0.27	0.67	0.65	0.62	0.58	0.55	0.53	0.80	0.79	0.76	0.74	0.70	0.68

Notes: See the notes for Table D.2, but now with $N_1 = N^{0.25}$.

Table D.4: No Factor Case, Relative MSE compared to PC for $(k_1, k_2) = (1, 1)$

		R^2=.5						R^2=.33						R^2=.2					
	T/N	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400
PLS(1)	20	0.77	0.84	0.89	0.94	0.97	0.99	1.00	1.03	1.04	1.04	1.03	1.01	1.11	1.12	1.11	1.08	1.05	1.03
	30	0.70	0.77	0.83	0.92	0.95	0.98	0.95	0.98	1.02	1.04	1.03	1.02	1.08	1.12	1.12	1.10	1.07	1.04
	50	0.61	0.67	0.75	0.85	0.92	0.96	0.90	0.94	0.98	1.02	1.04	1.03	1.03	1.06	1.11	1.10	1.09	1.06
	100	0.44	0.45	0.50	0.51	0.58	0.62	0.69	0.73	0.76	0.73	0.68	0.61	0.80	0.82	0.76	0.78	0.72	0.67
	200	0.38	0.38	0.37	0.35	0.31	0.25	0.66	0.63	0.59	0.52	0.41	0.30	0.77	0.74	0.69	0.59	0.46	0.32
	400	0.37	0.37	0.37	0.36	0.35	0.31	0.67	0.65	0.63	0.59	0.51	0.41	0.80	0.78	0.75	0.69	0.58	0.45
BR(5)	20	0.87	0.90	0.94	0.97	0.98	0.99	0.93	0.96	0.97	0.98	0.99	1.00	0.97	0.98	0.98	0.99	1.00	1.00
	30	0.81	0.86	0.91	0.95	0.97	0.99	0.91	0.93	0.96	0.98	0.99	0.99	0.96	0.97	0.98	0.99	1.00	1.00
	50	0.72	0.78	0.85	0.92	0.96	0.98	0.87	0.90	0.93	0.96	0.98	0.99	0.94	0.96	0.97	0.98	0.99	1.00
	100	0.52	0.58	0.66	0.73	0.81	0.82	0.73	0.76	0.80	0.82	0.83	0.84	0.82	0.85	0.82	0.85	0.85	0.86
	200	0.41	0.44	0.49	0.56	0.62	0.66	0.67	0.67	0.67	0.68	0.69	0.70	0.78	0.77	0.75	0.73	0.72	0.71
	400	0.37	0.38	0.41	0.48	0.56	0.62	0.67	0.66	0.66	0.66	0.67	0.68	0.80	0.79	0.77	0.74	0.72	0.71
BR(10)	20	0.93	0.95	0.97	0.98	0.99	1.00	0.96	0.97	0.98	0.99	0.99	1.00	0.98	0.98	0.99	0.99	1.00	1.00
	30	0.89	0.92	0.95	0.97	0.99	0.99	0.94	0.96	0.97	0.99	0.99	1.00	0.97	0.98	0.98	0.99	1.00	1.00
	50	0.82	0.86	0.91	0.95	0.97	0.99	0.91	0.94	0.96	0.98	0.99	0.99	0.95	0.97	0.98	0.99	0.99	1.00
	100	0.64	0.72	0.78	0.84	0.89	0.90	0.80	0.83	0.87	0.89	0.90	0.91	0.86	0.90	0.88	0.91	0.92	0.92
	200	0.51	0.56	0.63	0.71	0.77	0.80	0.73	0.74	0.76	0.79	0.81	0.82	0.82	0.82	0.82	0.82	0.83	0.83
	400	0.42	0.46	0.52	0.62	0.71	0.76	0.70	0.70	0.72	0.75	0.78	0.80	0.81	0.81	0.81	0.82	0.82	0.82
PLS(A)	20	0.90	0.92	0.92	0.95	0.98	0.99	1.20	1.20	1.13	1.07	1.04	1.02	1.31	1.32	1.22	1.11	1.06	1.03
	30	0.74	0.86	0.90	0.94	0.96	0.98	1.05	1.16	1.18	1.11	1.05	1.03	1.19	1.31	1.29	1.18	1.10	1.04
	50	0.59	0.69	0.84	0.89	0.93	0.97	0.93	1.03	1.17	1.17	1.09	1.04	1.06	1.15	1.34	1.26	1.15	1.08
	100	0.40	0.39	0.45	0.48	0.56	0.58	0.68	0.72	0.76	0.73	0.67	0.59	0.80	0.82	0.76	0.77	0.74	0.68
	200	0.34	0.31	0.29	0.24	0.22	0.16	0.64	0.60	0.57	0.50	0.40	0.27	0.76	0.73	0.68	0.58	0.45	0.31
	400	0.35	0.33	0.31	0.27	0.24	0.20	0.66	0.64	0.61	0.56	0.49	0.40	0.79	0.78	0.74	0.68	0.58	0.45
BR(A)	20	1.09	0.97	0.93	0.96	0.98	0.99	1.18	1.10	1.06	1.02	1.01	1.01	1.18	1.13	1.08	1.04	1.02	1.01
	30	0.76	1.08	0.94	0.94	0.97	0.98	0.99	1.20	1.10	1.03	1.02	1.01	1.03	1.21	1.11	1.05	1.03	1.01
	50	0.57	0.71	1.15	0.91	0.94	0.97	0.89	0.95	1.25	1.06	1.03	1.01	0.98	1.01	1.20	1.08	1.04	1.03
	100	0.39	0.38	0.44	0.57	0.59	0.63	0.68	0.71	0.75	0.73	0.76	0.74	0.82	0.84	0.81	0.82	0.82	0.82
	200	0.34	0.31	0.29	0.28	0.28	0.25	0.64	0.62	0.59	0.55	0.53	0.53	0.78	0.76	0.74	0.70	0.67	0.64
	400	0.35	0.33	0.31	0.28	0.27	0.27	0.67	0.65	0.62	0.59	0.55	0.52	0.80	0.79	0.77	0.74	0.70	0.67

Notes: The entries are average one-period ahead, out-of-sample MSE ratios relative to principal components-based one-period ahead forecasts across 1,000 Monte Carlo replications. The target and indicator variables are generated through the DGPs in (16) where we consider the absence of common factors in the predictor variables, as outlined under Case IV assuming $\lambda_{i,j} = 0$ and thus $r = 0$; we generate $T + 100$ observations on the predictor and explanatory variables, use the first T observations to estimate the models and use the resulting parameter estimates to forecast the predictor variable over $T + 1, \ldots, T + 100$. We impose different levels of for the asymptotic fit of the prediction regression, symbolized by the R^2's. In case of both PC regression and PLS regression 1 factor is extracted, and in case of Bayesian regression (BR) a shrinkage parameter $q \times N$ is used with $q = 5, 10$ or in case of $BR(A)$ with a data determined q as discussed in Algorithm 2. Similarly in the case of $PLS(A)$ the number of factors is data-determined as discussed in Algorithm 2.

Table E.1: Transformation of the predictor variables

Transformation code	Transformation X_t of raw series Y_t
1	$X_t = Y_t$
2	$X_t = \Delta Y_{t,t-1}$
3	$X_t = \Delta Y_{t,t-12} - \Delta Y_{t-1,t-13}$
4	$X_t = \ln Y_t$
5	$X_t = \Delta \ln Y_{t,t-1}$
6	$X_t = \Delta \ln Y_{t,t-12} - \Delta \ln Y_{t-1,t-13}$

E Data Set

The data set used for forecasting are the monthly series from the panel of U.S. predictor series as employed in Stock and Watson (2007), but excluding our four forecast variables: CPI inflation, (aggregate) industrial production, (aggregate) unemployment rate and the (effective) federal funds rate. In order to have $I(0)$ predictor variables, the underlying raw series need to be appropriately transformed; generally we employ the same transformation as Stock and Watson (2007), except for the nominal series where we follow, e.g., D'Agostino and Giannone (2006) and use first differences of twelve-month transformations of the raw series. Table E.1 summarizes our potential transformations for the raw series.

Hence, we are using as predictor variables the following 105 series, which span before transformation the sample January 1959 - December 2006 and we refer to Stock and Watson (2007) for more details regarding data construction and sources:

Series Y_t	**Transformation:** ***(See Table E.1)***
INDUSTRIAL PRODUCTION INDEX - FINAL PRODUCTS	5
INDUSTRIAL PRODUCTION INDEX - CONSUMER GOODS	5
INDUSTRIAL PRODUCTION INDEX - DURABLE CONSUMER GOODS	5
INDUSTRIAL PRODUCTION INDEX - NONDURABLE CONSUMER GOODS	5
INDUSTRIAL PRODUCTION INDEX - BUSINESS EQUIPMENT	5
INDUSTRIAL PRODUCTION INDEX - MATERIALS	5
INDUSTRIAL PRODUCTION INDEX - DURABLE GOODS MATERIALS	5
INDUSTRIAL PRODUCTION INDEX - NONDURABLE GOODS MATERIALS	5
INDUSTRIAL PRODUCTION INDEX - MANUFACTURING (SIC)	5
INDUSTRIAL PRODUCTION INDEX - RESIDENTIAL UTILITIES	5
INDUSTRIAL PRODUCTION INDEX - FUELS	5
NAPM PRODUCTION INDEX (PERCENT)	1
CAPACITY UTILIZATION - MANUFACTURING (SIC)	1
AVG HRLY EARNINGS, PROD WRKRS, NONFARM - GOODS-PRODUCING	6
AVG HRLY EARNINGS, PROD WRKRS, NONFARM - CONSTRUCTION	6
AVG HRLY EARNINGS, PROD WRKRS, NONFARM - MFG	6
REAL AVG HRLY EARNINGS, PROD WRKRS, NONFARM - GOODS-PRODUCING	5

REAL AVG HRLY EARNINGS, PROD WRKRS, NONFARM - CONSTRUCTION	5
REAL AVG HRLY EARNINGS, PROD WRKRS, NONFARM - MFG	5
EMPLOYEES, NONFARM - TOTAL PRIVATE	5
EMPLOYEES, NONFARM - GOODS-PRODUCING	5
EMPLOYEES, NONFARM - MINING EMPLOYEES, NONFARM - CONSTRUCTION	5
EMPLOYEES, NONFARM - MFG	5
EMPLOYEES, NONFARM - DURABLE GOODS	5
EMPLOYEES, NONFARM - NONDURABLE GOODS	5
EMPLOYEES, NONFARM - SERVICE-PROVIDING	5
EMPLOYEES, NONFARM - TRADE, TRANSPORT, UTILITIES	5
EMPLOYEES, NONFARM - WHOLESALE TRADE	5
EMPLOYEES, NONFARM - RETAIL TRADE	5
EMPLOYEES, NONFARM - FINANCIAL ACTIVITIES	5
EMPLOYEES, NONFARM - GOVERNMENT	5
INDEX OF HELP-WANTED ADVERTISING IN NEWSPAPERS (1967=100;SA)	2
EMPLOYMENT: RATIO; HELP-WANTED ADS:NO. UNEMPLOYED CLF	2
CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS.,SA)	5
CIVILIAN LABOR FORCE: EMPLOYED, NONAGRIC.INDUSTRIES (THOUS.,SA)	5
UNEMPLOY.BY DURATION: PERSONS UNEMPL.LESS THAN 5 WKS (THOUS.,SA)	5
UNEMPLOY.BY DURATION: PERSONS UNEMPL.5 TO 14 WKS (THOUS.,SA)	5
UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 WKS + (THOUS.,SA)	5
UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 TO 26 WKS (THOUS.,SA)	5
UNEMPLOY.BY DURATION: PERSONS UNEMPL.27 WKS + (THOUS,SA)	5
AVG WKLY HOURS, PROD WRKRS, NONFARM - GOODS-PRODUCING	1
AVG WKLY OVERTIME HOURS, PROD WRKRS, NONFARM - MFG	2
HOUSING AUTHORIZED: TOTAL NEW PRIV HOUSING UNITS (THOUS.,SAAR)	4
HOUSING STARTS:NONFARM(1947-58);TOTAL FARM&NONFARM(1959-)(THOUS.,U)SA	4
HOUSING STARTS:NORTHEAST (THOUS.U.)S.A.	4
HOUSING STARTS:MIDWEST(THOUS.U.)S.A.	4
HOUSING STARTS:SOUTH (THOUS.U.)S.A.	4
HOUSING STARTS:WEST (THOUS.U.)S.A.	4
INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,3-MO.(% PER ANN,NSA)	2
INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,6-MO.(% PER ANN,NSA)	2
INTEREST RATE: U.S.TREASURY CONST MATURITIES,1-YR.(% PER ANN,NSA)	2
INTEREST RATE: U.S.TREASURY CONST MATURITIES,5-YR.(% PER ANN,NSA)	2
INTEREST RATE: U.S.TREASURY CONST MATURITIES,10-YR.(% PER ANN,NSA)	2
BOND YIELD: MOODY'S AAA CORPORATE (% PER ANNUM)	2
BOND YIELD: MOODY'S BAA CORPORATE (% PER ANNUM)	2
INTEREST RATE SPREAD: 6-MO. TREASURY BILLS MINUS 3-MO. TREASURY BILLS	1
INTEREST RATE SPREAD: 1-YR. TREASURY BONDS MINUS 3-MO. TREASURY BILLS	1
INTEREST RATE SPREAD: 10-YR. TREASURY BONDS MINUS 3-MO. TREASURY BILLS	1
INTEREST RATE SPREAD: AAA CORPORATE MINUS 10-YR. TREASURY BONDS	1
INTEREST RATE SPREAD: BAA CORPORATE MINUS 10-YR. TREASURY BONDS	1
MONEY STOCK: M1(CURR,TRAV.CKS,DEM DEP,OTHER CK'ABLE DEP)(BIL$,SA)	6
MZM (SA) FRB St. Louis	6
MONEY STOCK:M2(M1+O'NITE RPS,EURO$,G/P&B/D MMMFS&SAV&SM TIME DEP)(BIL$,SA)	6
MONETARY BASE, ADJ FOR RESERVE REQUIREMENT CHANGES(MIL$,SA)	6
DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS(MIL$,SA)	6
DEPOSITORY INST RESERVES:NONBORROWED,ADJ RES REQ CHGS(MIL$,SA)	6
Commercial and Industrial Loans at All Commercial Banks (FRED) Billions $ (SA)	6
CONSUMER CREDIT OUTSTANDING - NONREVOLVING(G19)	6
Personal Consumption Expenditures, Price Index (2000=100) , SAAR	6
Personal Consumption Expenditures - Durable Goods, Price Index (2000=100), SAAR	6
Personal Consumption Expenditures - Nondurable Goods, Price Index (2000=100), SAAR	6
Personal Consumption Expenditures - Services, Price Index (2000=100) , SAAR	6
PCE Price Index Less Food and Energy (SA) Fred	6
PRODUCER PRICE INDEX: FINISHED GOODS (82=100,SA)	6
PRODUCER PRICE INDEX:FINISHED CONSUMER GOODS (82=100,SA)	6
PRODUCER PRICE INDEX:INTERMED MAT.SUPPLIES & COMPONENTS(82=100,SA)	6
PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA)	6
Real PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA)	5
SPOT MARKET PRICE INDEX:BLS & CRB: ALL COMMODITIES(1967=100)	6
Real SPOT MARKET PRICE INDEX:BLS & CRB: ALL COMMODITIES(1967=100)	5
PRODUCER PRICE INDEX: CRUDE PETROLEUM (82=100,NSA)	6
PPI Crude (Relative to Core PCE)	5

NAPM COMMODITY PRICES INDEX (PERCENT)	1
UNITED STATES;EFFECTIVE EXCHANGE RATE(MERM)(INDEX NO.)	5
FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER U.S.$)	5
FOREIGN EXCHANGE RATE: JAPAN (YEN PER U.S.$)	5
FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)	5
FOREIGN EXCHANGE RATE: CANADA (CANADIAN $ PER U.S.$)	5
S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)	5
S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS (1941-43=10)	5
S&P'S COMPOSITE COMMON STOCK: DIVIDEND YIELD (% PER ANNUM)	2
S&P'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO (%,NSA)	2
COMMON STOCK PRICES: DOW JONES INDUSTRIAL AVERAGE	5
S&P'S COMPOSITE COMMON STOCK: DIVIDEND YIELD (% PER ANNUM)	2
U. OF MICH. INDEX OF CONSUMER EXPECTATIONS(BCD-83)	2
PURCHASING MANAGERS' INDEX (SA)	1
NAPM NEW ORDERS INDEX (PERCENT)	1
NAPM VENDOR DELIVERIES INDEX (PERCENT)	1
NAPM INVENTORIES INDEX (PERCENT)	1
NEW ORDERS (NET) - CONSUMER GOODS & MATERIALS, 1996 DOLLARS (BCI)	5
NEW ORDERS, NONDEFENSE CAPITAL GOODS, IN 1996 DOLLARS (BCI)	5