

Groen, Jan J. J.; Kapetanios, George

Working Paper

Revisiting useful approaches to data-rich macroeconomic forecasting

Staff Report, No. 327

Provided in Cooperation with:
Federal Reserve Bank of New York

Suggested Citation: Groen, Jan J. J.; Kapetanios, George (2008) : Revisiting useful approaches to data-rich macroeconomic forecasting, Staff Report, No. 327, Federal Reserve Bank of New York, New York, NY

This Version is available at:

<https://hdl.handle.net/10419/60769>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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 September 2009

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in the paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

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 September 2009

JEL classification: C22, C53, E37, E47

Abstract

We compare a number of data-rich prediction methods that are widely used in macroeconomic forecasting with a lesser known alternative: partial least squares (PLS) regression. In this 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. We show theoretically that when the data have a factor structure, PLS regression can be seen as an alternative way to approximate this unobserved factor structure. In addition, we prove that when a large data set has a weak factor structure, which possibly vanishes in the limit, PLS regression still provides asymptotically the best fit for the target variable of interest. Monte Carlo experiments confirm our theoretical results that PLS regression performs at least as well as principal components regression and rivals Bayesian regression when the data have a factor structure. But when the factor structure in the data is weak, PLS regression outperforms both principal components and Bayesian regressions. Finally, we apply PLS, principal components, and Bayesian regressions to a large panel of monthly U.S. macroeconomic data to forecast key variables across different subperiods. The results indicate that PLS regression usually has the best out-of-sample performance.

Key words: macroeconomic forecasting, factor models, forecast combination, 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 (email: 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 “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 at University College 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, most traditional macroeconomic prediction approaches rarely consists of models that handle more than 10 variables, because it is either inefficient or downright impossible to incorporate a much larger number of variables in a single forecasting model and estimate it using standard econometric techniques. This failure of traditional macroeconomic forecasting methods 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. The variables that are chosen are then used in a standard forecasting model. Recent developments in this line of research has focussed on automated model selection procedures in order to be better able to select the optimal predictors from large data sets; see Krolzig and Hendry (2001). 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. These two groups of methods inevitably overlap. However, we feel that the step of variable selection is, and involves methods that are, sufficiently distinct to merit separate mention and treatment. Instead, 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. These methods are related to the strict factor models used in finance, but, starting with

Chamberlain and Rothschild (1983), they use weaker assumptions regarding the behavior of the idiosyncratic components, which allows the use of principal components in very large data sets to identify the common factors in such a data set. Stock and Watson (2002a) and Bai (2003) further formalized the underlying asymptotic theory. 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. However, the use of principal components does not always guarantee that the information extracted from a large number of predictors is useful for forecasting. Boivin and Ng (2006) make it clear 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. Under Bayesian regression one essentially estimates a multivariate regression consisting of *all* predictor variables, but with the regression coefficients shrunk 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 and the production of multiple forecasts for the target variable, which are then averaged to produce a final forecast. The distinctive feature of these two approaches is that the information in a large data set is compressed such that this has explanatory power for the target variable. Note, however, that from an econometric perspective forecast combinations are *ad hoc* in nature, whereas it has been shown in De Mol *et al.* (2008) that Bayesian regression is theoretically related to PC-based factor models.

In this paper we revisit the use of principal components (PC) and Bayesian regression for data-rich macroeconomic forecasting and compare these with the lesser known method of partial least squares (PLS) regression. We propose PLS regression as an alternative data-rich approach that can be used for macroeconomic forecasting using very large data sets, irrespective of whether such a data set involves having 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 such that the linear combinations maximize the covariance between

the target forecast variable and each of the common components constructed from the 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.

One significant contribution of our paper is that we provide theoretical results that relates PLS to PC for large data sets. When a factor structure is a dominant feature of large data sets we can show that in the limit PLS and PC regressions will be equivalent, and thus PLS regression can be seen as a valid alternative technique to uncover the relevant, underlying unobserved factor structure. However, when this underlying factor structure is very weak, or even absent, we prove that PLS regression still provides asymptotically the best fit for the target variable of interest. Next, we argue that the range of forecast combination techniques can be seen as restricted versions of PLS. PLS, therefore, has explicit theoretical links with the currently used range of data-rich macroeconomic forecasting tools. More specifically, we provide a unification of forms of PLS, PC, forecast combination and, via the work of De Mol *et al.* (2008), Bayesian regression, thereby linking all major forecasting tools. Finally, 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. Our work suggests a clear potential for PLS regression for a variety of contexts.

In the remainder of this paper we have the following structure: Section 2 discusses the most frequently used data-rich methods for macroeconomic forecasting. Then, in Section 3, we provide an overview of PLS regression, present some results on the asymptotic behavior of PLS regression, and 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 Frequently Used Methods for Data-Rich Macroeconomic Forecasting

A useful framework for studying existing methods is provided by the following general forecasting equation

$$y_t = \alpha' x_t + \epsilon_t; \quad t = 1, \dots, T, \quad (1)$$

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 $N \times 1$. It is assumed 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.

Factor Methods

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.

$$\tilde{x}_t = \Lambda' F_t + e_t \quad (2)$$

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 $r < \min(N, T)$. So, implicitly, in (1) $\alpha' = \tilde{\alpha}' \Lambda \tilde{x}_t$, where $F_t = \Lambda \tilde{x}_t$, 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 . 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:

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

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)

PC estimation of the factor structure is essentially a static exercise as no lags or leads of x_t are considered. One alternative is dynamic principal components, which, as a method of factor extraction, has been suggested in a series of papers by Forni, Hallin, Lippi and Reichlin (see, e.g., Forni *et al.* (2000, 2004) among others) is designed to address this issue. Dynamic principal components are extracted in similar fashion to static principal components but, instead of the second moment matrix, the spectral density matrices of the data at various frequencies are used. These are then used to construct estimates of the common component of the data set which is a function of the unobserved factors. This method uses leads of the data and as a result its application to forecasting has been slow for obvious reasons. Recent work by the developers of the method has addressed this issue (see, e.g., Forni *et al.* (2005)).

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

$$\hat{\alpha}_{BRR} = (X'X + vI)^{-1}X'y \quad (4)$$

where $X = (x_1, \dots, x_T)'$, $y = (y_1, \dots, y_T)'$ and v is a shrinkage scalar parameter. The shrinkage estimator (4) shrinks the OLS estimator, given by $(X'^{-1}X'y$ 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 (4) 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).

3 An Alternative: 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.¹ 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 variables, in PLS, the relationship between y_t and x_t is considered as well in constructing the factors. PLS regression does not seem to have been explicitly considered for data sets with a very large number of series, i.e., when N is assumed in the limit to converge to infinity. The

¹Herman Wold and co-workers introduced PLS regression between 1975 and 1982, see, e.g., Wold (1982). Since then it has received much attention in a variety of disciplines, especially in chemometrics, outside of economics.

latter assumption has motivated the use of PC regression for macroeconomic forecasting, as the principal components under that assumption can under certain conditions identify the common factors in the data set (see, e.g., Bai (2003)). One of our contributions is, therefore, to develop the asymptotic properties of PLS regression under similar assumptions that $N, T \rightarrow \infty$ for data sets that have a common factor structure, and we do that in Section 3.1. In addition, we also investigate in that subsection the asymptotic properties of PLS regression when large data sets have a weak factor structure that possibly vanishes asymptotically. Section 3.2 then describes how PLS regression ties up with other data-rich techniques. Finally, Section 3.3 reports on a number of Monte Carlo experiments in which we compare PLS regression with PC-based factor modeling.

3.1 PLS Regression: Methodology and Theory

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 $v_{i,t} = x_{i,t}$, $i = 1, \dots, N$. 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: $w_{ij} = \text{Cov}(u_t, v_{it})$, $i = 1, \dots, N$. 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 $v_{i,t}$, $i = 1, \dots, N$ 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 $u_t = \tilde{u}_t$, $v_{i,t} = \tilde{v}_{i,t}$ $i = 1, \dots, N$ 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 $f_{j,t}$ $j = 1, \dots, k$. Helland (1988, 1990) provide a general description of the partial least squares (PLS) regression problem. Helland (1988) shows that the estimates of the coefficients α in the regression of y_t on x_t , as in (1), obtained implicitly via PLS Algorithm 1 and a regression of y_t on $f_{j,t}$ $j = 1, \dots, k$, are mathematically equivalent to

$$\hat{\alpha}_{PLS} = V_k(V_k'X'XV_k)^{-1}V_k'X'y \quad (5)$$

with $V_{k_1} = (X'y \quad X'XX'y \quad \dots \quad (X'X)^{k-1}X'y)$, $X = (x_1 \dots x_T)'$ and $y = (y_1 \dots y_T)'$. Thus, (5) suggests that the PLS factors that result from Algorithm 1 span the Krylov subspace generated by $X'X$ and $X'y$, resulting in valid approximations of the covariance between y_t and x_t .

Next, we undertake a theoretical analysis of PLS when $T, N \rightarrow \infty$. Note that previous work focused exclusively on the case of a finite N . We consider two mutually exclusive frameworks. In the first, a factor structure exists for X . In the second there is only a very weak factor structure that can actually disappear as $N \rightarrow \infty$ but some cross-sectional dependence is allowed. We start by analysing the factor case. Here, our goal is to show how PLS regression behaves relative to PC regression, as $T, N \rightarrow \infty$. Within the general regression (1), which relates y_t to x_t , denote the PLS and PC implied regression estimates by $\hat{\alpha}_{PLS}$ and $\hat{\alpha}_{PC}$ respectively, based on (5) and

$$\hat{\alpha}_{PC} = \Lambda'(\Lambda X'X\Lambda')^{-1}\Lambda X'y \quad (6)$$

where Λ is the $r \times N$ matrix of linear combinations of the x_{it} 's that result from minimizing (3). Let the matrix norm we use be $\|A\| = \text{tr}(A'A)^{1/2}$. We follow Stoica and Söderström (1998) who assume the following :

Assumption 1 *Let*

$$X'X = S\Psi S' + C_\delta$$

where $\Psi = \text{diag}(\psi_1, \dots, \psi_r)$ for some $r < N$, $S'S = I$, C_δ denotes a term whose matrix norm is $O_p(\delta)$ and $\delta \rightarrow 0$. This assumption implies for (5) and (6) within (1):

$$\|\hat{\alpha}_{PLS} - \hat{\alpha}_{PC}\| = O_p(\delta). \quad (7)$$

An interesting aside that comes out of the above setup of Stoica and Söderström (1998) is that if the definition of k , suggested in the previous paragraph, is adopted then, for finite N at least, and assuming a reduced rank structure as in Assumption 1 we have that $k \leq r$. We will comment on that later in more detail.

For our large data set framework, with $T, N \rightarrow \infty$, we now assume the following factor structure for the explanatory variables x_t :

Assumption 2 *Let $\Sigma = \Sigma_N = [\sigma_{ij}]$ denote the $N \times N$ second moment matrix of X . Σ can be factorized as follows:*

$$\Sigma = \tilde{S}\tilde{\Psi}\tilde{S}' + R$$

where $\tilde{S}\tilde{S}' = I$, $\tilde{\Psi} = \text{diag}(\tilde{\psi}_{N1}, \dots, \tilde{\psi}_{Nr})$ is a $r \times r$ matrix, $r < N$ and $\|R\| = o_p(N)$.

Finally, we need an additional assumption regarding the second moments of x_t , i.e.,

Assumption 3 *For all $i, j = 1, \dots, N$*

$$\sum_{t=1}^T (x_{i,t}x_{j,t} - \sigma_{i,j}) = O_p(T^{-1/2})$$

Remark 1 *Assumptions 2 and 3 deserve some comment. Assumption 2 states that the variables in X are asymptotically with respect to N collinear. Under a standard factor assumption there can be only a finite number of unbounded eigenvalues, and hence unbounded singular values, for the covariance matrix of the data. In case of Assumption 2 we can have an infinity of unbounded eigenvalues as long as the sum of all but the first r eigenvalues is $o_p(N)$. In particular the remainder term, R , can, in fact, be parameterized as a neglected ‘weak’ factor model whose*

eigenvalue characterization allows for unbounded eigenvalues which, however, have to grow at a rate slower than N . Assumption 3 is a mild, high level, assumption. It is sufficient to have a central limit theorem for $(x_{i,t}x_{j,t} - \sigma_{i,j})$ for this assumption to hold.

Given the above set-up we now end-up with the following Theorem

Theorem 1 *Under Assumptions 1-3, and as $N, T \rightarrow \infty$ sequentially, we have for the PLS and PC regression-implied estimates of the individual coefficients α in the general regression framework (1):*

$$\|\hat{\alpha}_{PLS} - \hat{\alpha}_{PC}\| = o_p\left((NT)^{-1/2}\right). \quad (8)$$

Proof: In order to prove this result in our case we need to show that as $T, N \rightarrow \infty$, Assumption 1 holds in probability asymptotically; for details see Appendix A.

Theorem 1 therefore indicates that under an assumed factor structure the PLS and PC regression approaches provide an asymptotically equivalent modeling of the relationship between y_t and x_t in a general, large-dimensional, regression set-up as in (1). The intuition behind this Theorem can be described as follows. The principal components extracted from x_t in, say, (1) provide the basis vectors that span the r -dimensional space of the underlying dominant factor model for x_t . Then, the PLS factors extracted from x_t reflect an immediate rotation of this r -dimensional space that provides the best fit for y_t and hence $k \leq r$ in (5). Of course, in the limit that part of the r -dimensional space spanned by the principal components that is not relevant for the fit of y_t becomes redundant in (6). But exactly because the PLS factors yield in one step the most relevant rotation of this space one can expect a more efficient modeling of y_t in finite samples using PLS regression than using PC regression.

Next, we consider the theoretical properties of PLS when there exists a weak factor structure in X that can disappear as $N \rightarrow \infty$. We make explicit the dependence of the coefficient and variable vector on N and write the model in (1) as

$$y_t = \alpha'_N x_{N,t} + \epsilon_t \quad (9)$$

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

Assumption 4

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

$$\|\Sigma_{XX} - I\|_1 = o(N),$$

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

Part II: Further, we have

$$0 < \limsup_{N \rightarrow \infty} \sup_{i,j} \frac{\alpha_{N,i}}{\alpha_{N,j}} < \infty.$$

Assumption 5 Uniformly over $j = 1, \dots, N$

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

where $\sigma_{Xy,j} = E(x_{j,t} y_t)$.

Remark 2 Again, Assumptions 4 and 5 deserve comment. Part I of Assumption 4 states that the covariance matrix of X becomes relatively close in a particular sense to the identity matrix as $N \rightarrow \infty$. It allows, e.g., for structures such as

$$\Sigma_{XX} = \begin{pmatrix} B & 0 \\ 0 & I \end{pmatrix} \quad (10)$$

where B is a $N_1 \times N_1$ symmetric positive definite matrix ($N_1 < N$) that can have a factor structure as outlined in Assumption 2, $N_1 \rightarrow \infty$ and $N_1^2/N \rightarrow 0$. 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. Another possibility consistent with Part I of Assumption 4 is a covariance matrix where non-diagonal elements are non-zero but tending to zero as the dimension of the matrix increases. For example, denoting the representative element of Σ_{XX} by σ_{ij} , we can set $\sigma_{ij} = \frac{1}{i^\kappa j^\kappa}$. Then, the Assumption holds as long as $\kappa > 0.5$. The same result holds if we set $\sigma_{ij} = \frac{1}{N^{2\kappa}}$. Note that such a covariance matrix can be obtained by specifying a factor model where the loadings depend on and tend to zero with N . Part II of Assumption 4 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. Assumption 5 is a mild assumption of a similar nature to Assumption 3

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

Theorem 2 *Let σ^2 denote the true variance of ϵ_t in the general regression framework (1) (and thus (9)) and $\hat{\sigma}^2$ its estimate from a PLS regression, given by steps 1-3 of Algorithm 1, using the N predictor variables that correspond with (1). Under Assumptions 4-5, and as $N, T \rightarrow \infty$ sequentially, $\hat{\sigma}^2 - \sigma^2 = o_p(1)$.*

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.

3.2 Linking Alternative Data Rich Methods

In this subsection, we briefly consider links between alternative data rich methods such as PLS, PC and Bayesian regression given the results presented in the previous subsection. We first 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, i.e.,

Algorithm 2

1. Set $u_t = y_t$ and $v_{i,t} = x_{i,t}$, $i = 1, \dots, N$. Set $j = 1$.
2. Regress u_t on $v_{i,t}$, $i = 1, \dots, N$ and denote the OLS estimate of the coefficient of each regression by β_i . Construct the j -th PLS factor by taking the 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.
3. Regress u_t and $v_{i,t}$, $i = 1, \dots, N$ on $f_{j,t}$. Denote the residuals of these regressions by \tilde{u}_t and $\tilde{v}_{i,t}$ respectively.
4. If $j = k_1$ stop, else set $u_t = \tilde{u}_t$, $v_{i,t} = \tilde{v}_{i,t}$ $i = 1, \dots, N$ and $j = j + 1$ and go to step 2.

Therefore, when in this algorithm one sets $(w_{1j} \cdots w_{Nj}) = (\text{Var}(v_{1t}) \cdots \text{Var}(v_{Nt}))$ 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. In general, forecast combinations can be interpreted through Algorithm 2 as restricted approximations to 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.² By interpreting forecast combinations as a form of PLS regression, Theorems 1 and 2 provide the underpinning for the relatively good performance of forecast combinations *vis-à-vis* PC regressions within different data environments.³ 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 factor structure.⁴ Thus, given that structure, Bayesian regression should, via Theorem 1, be asymptotically equivalent to PLS regression and, under the one-factor assumption, forecast combinations.

²Granger and Ramanathan (1984) suggest regressing individual forecasts on the target variable with the resulting parameter estimates serving as combination weights. More generally, by specifying a loss function one can derive combination weights that are optimal under that specific loss function (Elliott and Timmermann (2004)). Timmermann (2006) provides a comprehensive survey of the forecast combination literature.

³For example, Faust and Wright (2007) show that forecast combination methods provide better out-of-sample performance than factor methods when applied to high-dimensional panels of U.S. macroeconomic data.

⁴It is worth noting that the assumption made by De Mol *et al.* (2008) on the idiosyncratic part of their entertained factor model bears similarities to our Assumption 2. In particular, just like the implication of our Assumption 2, discussed in Remark 1, the idiosyncratic component of their factor model can accommodate a residual ‘weak’ factor model in the sense that the eigenvalues implied by that factor model can be unbounded but have to grow at a rate slower than N .

Therefore, the introduction of the PLS regression framework provide a means to asymptotically tie together different existing data-rich forecasting methods and provides a theoretical rationale for the common empirical finding that different data-rich approaches have similar performance.

3.3 Monte Carlo Analysis

In this subsection, we explore through Monte Carlo experiments the finite sample performance of PLS regression relative to PC regression and Bayesian regression. We consider both the case when the data have a factor structure and the case where there is no factor structure.

Monte Carlo Set-up

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

$$\begin{aligned}
y_t &= \alpha'(x_{1,t} \cdots x_{N,t})' + \epsilon_t = \alpha'x_t + \epsilon_t, \quad t = 1, \dots, T, \\
x_t &= \Lambda'f_t + u_t, \\
f_t &= (f_{1,t} \cdots f_{r,t})' \sim \text{iid}N(\mathbf{0}, I_r), \quad r \ll N, \\
\epsilon_t &= \sqrt{cN}\varepsilon_t,
\end{aligned} \tag{11}$$

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 (11) uses a $N \times 1$ vector of zero-mean disturbances $u_t = (u_{1,t} \cdots u_{N,t})'$. The individual regression coefficients in (11) are determined as $\alpha_i \sim \text{iid} N(0, 1)$ and the disturbances for the N explanatory variables are determined in a similar manner: $u_{i,t} \sim \text{iid} N(0, 1)$.

We consider a number of cases for the factor loadings, which we summarize as:

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

Case II: $\lambda_{i,j} \begin{cases} = \tilde{\lambda}_{i,j} \sim \text{iid}N(0, 1) & \text{for } i = 1, \dots, N_1, N_1 = N^{\kappa_1} (\kappa_1 = 0.25, 0.75) \text{ \& } j = 1, \dots, r. \\ = 0 & \text{for } i = N_1 + 1, \dots, N \text{ \& } j = 1, \dots, r. \end{cases}$

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

Clearly, Case I represents a standard factor model where the predictor variables are driven by r common factors. Cases II and III, on the other hand, imply much weaker factor structures

in x_t than assumed under Case I and these become progressively weaker as the cross-section dimension N increases. Under Case II we assume a structure of the form (10) 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 $N \rightarrow \infty$, where κ_1 determines the speed with which this occurs. Case III 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 $N \rightarrow \infty$. 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 standard factor model, and the larger κ_2 the further away it is from such a standard factor model.

An important parameter for the Monte Carlo study is the population R^2 of the y_t regression equation in (11). We control this by controlling the variance of the y_t disturbance term ϵ_t in (11) through $c = (1 + r)\tilde{c}$, where $\epsilon_t \sim \text{iid}N(0, 1)$. 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 (Case I). For the case of weak factors (Cases II and 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 Cases II and 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 (11) for $r = 1, 6$ 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. For the case of PLS regression we argued in Section 3.1 that the number of PLS factors, 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. This is more so the case here, as both the factor loadings as well as the regression parameters α are randomly generated in our simulations and thus the factors need not be of equal strength for y_t (which indeed is also true in reality). Therefore, for PLS regression we set the number of factors k_1 to 1 and 3 to correspond to the respective k_2 PC factors. For the weak factor cases, Cases II and III, we generate data based on $r = 1$ and assume in case of PLS and PC regressions $k_1 = k_2 = 1$ in order

to focus on how a decreasing amount of collinearity within x_t affects the relative performance of these methods rather than differences in relevance for y_t across a multiple of factors. For Bayesian regression we follow De Mol *et al.* (2008) and set the shrinkage parameter proportional to the cross-section dimension N of the explanatory series: qN , where $q = 1, 5$.

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 (11) 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 $T + 1, \dots, T + 100$. 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.

Monte Carlo Results

Starting with the standard factor case in Tables 1-2, one notes that the performance of PLS and PC regressions are fairly similar. Note, though, that the relative performance of PLS regression improves when the number of entertained PC and PLS factors increases, which reflects the theoretical result of Section 3.1. The only other noteworthy feature for this case, given already available Monte Carlo work on the relative performance of PC *versus* Bayesian regressions, is the confirmation of the good predictive performance of Bayesian regression in a large number of cases.

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 both cases II and III, see Tables 3 and 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. PLS regression, on the other hand, performs in a large number of cases as well as Bayesian regression, which in Tables 3 and 4 has a slight edge *vis-à-vis* PLS regression when an appropriate shrinkage parameter is chosen. Tables 5 and 6 also report simulation results under Cases II and III for the factor loadings, but now with much more severe factor weakness than in Tables 3 and 4. What becomes clear from Tables 5 and 6 is that the relative performance of PLS regression

improves substantially to a point that it now performs at least as well, and in a lot of cases better, as Bayesian regression.⁵

To conclude we see that our Monte Carlo study suggests a great advantage for PLS regression compared to all other methods we considered in terms of forecasting performance. Especially in the case where the data do not have a standard parsimonious representation such as a (standard) common factor structure.

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

We follow standard practice in the macroeconomic forecasting literature and use as benchmark for our data-rich based forecasts an autoregressive (AR) model

$$\Delta y_{t+h,t} = \alpha^h + \sum_{i=1}^p \rho_i \Delta y_{t-i+1,t-i} + \epsilon_{t+h,t}, \quad t = 1, \dots, T \quad (12)$$

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 $i = 1, \dots, p$. The number of lagged first differences p in (12) is determined by sequentially applying the standard Schwarz (1978)'s BIC starting with a maximum lag order of $p = p_{max}$ down to $p = 1$. Next, we use as a benchmark the unconditional mean,

$$\Delta y_{t+h,t} = \alpha^h + \epsilon_{t+h,t}, \quad (13)$$

which implies a random walk (RW) forecast for the level of the forecast variable y_t . 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

⁵Indeed, if one completely turns off the factor structure in (11) one would get a similar result as in Tables 5 and 6. These simulation results are available upon request from the authors.

errors (RMSE). In Section 4.3 we will report ratios of the RMSE of the respective data-rich forecasting approaches relative to the RMSE based on either (12) or (13). Superior out-of-sample performance of a data-rich method relative to these benchmarks is, obviously, indicated by a RMSE ratio smaller than one.

Our data-rich forecasts of h period-ahead changes in y_t are generated using a model that adds the information extracted from the N explanatory variables in the $N \times 1$ vector $X_t = (x_{1,t} \cdots x_{N,t})$ to the benchmark models (12) and (13), i.e., respectively

$$\Delta y_{t+h,t} = \alpha^h + \beta^{h'} F(X_t) + \sum_{i=1}^p \rho_i \Delta y_{t-i+1,t-i} + \epsilon_{t+h,t} \quad (14)$$

and

$$\Delta y_{t+h,t} = \alpha^h + \beta^{h'} F(X_t) + \epsilon_{t+h,t}. \quad (15)$$

where β^h is $r \times 1$. In (14) and (15) $F(X_t)$ 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 $F(X_t)$ 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 $F(X_t)$ in (14) and (15).

Our forecasting models will be updated based on an expanding window of historical data, which in case of the principal components-based models 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 $t = 1, \dots, t_0 - h$.

3. Estimate either (14) or (15) with $F(X_t) = F_t$ over the sample $t = 1, \dots, t_0 - h$ for each h .
4. Extract r principal components F_t from the N indicator variables N over the sample $t = 1, \dots, t_0$.
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 $t_0 + 1, \dots, T - h$ for each h .

Bayesian Regression

When Bayesian regression is used to compress the forecast information in the N indicator variables, $F(X_t)$ in (14) and (15) simply equals X_t , whereas β^h is estimated with the shrinkage estimator (4). 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_{RR}^h . 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. In case of (14) we first regress both the demeaned $\Delta y_{t+h,t}$ and the \tilde{X} on $\Delta y_{t-i+1,t-i}$ for $i = 1, \dots, p$, and use the resulting residuals in (4) to estimate β^h . Estimates for the intercept term and the ρ_i 's in (14) can then be trivially recovered.

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

1. First forecast for all h is generated on t_0 .
2. Estimate either (15) or (14) with (4) for β^h using \tilde{X} over the sample $t = 1, \dots, t_0 - h$ 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 $t_0 + 1, \dots, T - h$ for each h .

Partial Least Squares Regression

With partial least squares (PLS) regression, $F(X_t)$ in (14) and (15) 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 3.1 can be implemented for macroeconomic forecasting as follows:

Algorithm 3

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.

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

2. The r PLS factors $F_{1,t}^{PLS}, \dots, F_{r,t}^{PLS}$ and their loadings w_1, \dots, 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 $\Delta \dot{Y}_h$:

$$F_l^{PLS} = \tilde{X}_{l|l-1} w_l; \quad w_l = \frac{1}{T-1} \tilde{X}'_{l|l-1} \Delta \dot{Y}_{h,l|l-1} \quad \text{for } l = 1, \dots, r \quad (16)$$

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 $l > 1$

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

3. Finally, we simply plug in the r PLS factors $F_t^{PLS} = (F_{1,t}^{PLS} \dots F_{r,t}^{PLS})'$ from (16) in the predictive regression (15) which we estimate in the standard way:

$$\Delta y_{t+h,t} = \alpha^h + \beta^h F_t^{PLS} + \epsilon_{t+h,t}. \quad (17)$$

When lagged predictor variables are included in the predictive regression, as in (14), one needs to control for the effect of $\Delta y_{t,t-1}, \dots, \Delta y_{t-p+1,t-p}$ on the covariances between $\Delta y_{t+h,t}$ and $x_{1,t}, \dots, x_{N,t}$. 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 $\Delta y_{t,t-1}, \dots, \Delta y_{t-p+1,t-p}$, and then using the resulting residuals in Algorithm 3 in order to be able to construct a model like

$$\Delta y_{t+h,t} = \alpha^h + \beta^{h'} F_t^{PLS} + \sum_{i=1}^p \rho_i \Delta y_{t-i+1,t-i} + \epsilon_{t+h,t}. \quad (18)$$

Finally, forecasts from (17) and (18) 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 $t = 1, \dots, t_0 - h$ for each h based on Algorithm 3.
3. Estimate either (17) or (18) over the sample $t = 1, \dots, t_0 - h$ for each h .
4. Extract r PLS factors F_t^{PLS} from the N indicator variables over the sample $t = 1, \dots, t_0$ 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 $t_0 + 1, \dots, T - h$ 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 (4). In the latter case, we are not aware of the availability of a theoretically justified approach to select the optimal shrinkage parameter v for a given data set. Instead, 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. Therefore, for BR-based estimation of (14) and (15) we use in (4) $v = qN$ with $q = 1$ and 5, which is in the range for v that De Mol *et al.* (2008) found useful in their

applications. To facilitate a fair comparison between the BR-based forecasts and both the PC-based and PLS-based forecasts, we therefore estimate the PC-based and PLS-based versions of (14) and (15) using a range of fixed numbers of factors. In case of PC-based forecasts, we will consider the predictive performance of (14) and (15) using $r = 2, 4$ and 6 PC factors, whereas in case of PLS-based forecasts we consider the performance using $r = 1, 2$ and 3 PLS factors.

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)⁶ 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.

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 C 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).⁷ 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 fore-

⁶We are very grateful to Mark Watson who provided us with the underlying raw data from Stock and Watson (2007).

⁷This particular transformation acknowledges that series like log price levels and log money aggregate levels behave as if they are $I(2)$, possibly because of mean growth shifts due to policy regime shifts, financial liberalizations and other phenomena.

cast 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 4 horizons (in months): $h = 1$, $h = 3$, $h = 12$ and $h = 24$, which are horizons commonly analyzed in the literature. The forecast evaluation spans three samples: January 1972 - December 2006, January 1972 - December 1984 and January 1985 - December 2006. The latter two sub-samples split the first sample in two around the start of the ‘Great Moderation’, 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’.⁸

4.3 Forecasting Results

As discussed in Section 4.1, we will assess the forecasting performance of our three data-rich forecast methods with two simple benchmark forecasts: those based on an autoregressive (AR) specification and those based on the unconditional mean or random walk (RW) model (respectively (12) and (13)). The first set of evaluation results can be found in Table 7 and relate to forecasting changes in annual CPI inflation. Across the three evaluation samples, the full 1972-2006 evaluation sample and the two sub-samples, PLS regression does dominates in 11 out of

⁸See, for example, D’Agostino *et al.* (2006) who compare PC-based, VAR-based and *Greenbook* forecasts for U.S. inflation and economic growth with simple benchmarks for both pre- and post-Great Moderation samples.

the 24 cases (i.e., 3 evaluation samples times 2 benchmark models times 4 forecasting horizons). In fact, only Bayesian regression can, up to a certain extent compete with PLS regression. Note, though, that when it loses out the PLS approach still provides a close second best. Of course, the overall forecasting performance of the data-rich approaches is less over the post-Great Moderation period, especially *vis-à-vis* the AR benchmark over longer horizons, which is consistent with findings elsewhere in the literature.

The Great Moderation has less of a negative effect on the predictive performance of our data-rich methods for industrial production growth and unemployment rate changes; see the results for 1985-2006 in Tables 8-9. PLS regression is clearly the overall winner in this case, as it dominates the other methods in, respectively, 13 and 20 out of the total of 24 forecasting exercises in each of these tables. In particular in case of unemployment PC regression performs poorly, and essentially only for the one-month horizon Bayesian regression is able to perform better than PLS regression.

Finally, we turn to the results for the federal funds rate in Table 10. 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. This seems certainly to be the case for PLS and Bayesian regressions, although Bayesian regression performs relatively poorly for the post-Great Moderation 1985-2006 period. PLS regression, however, performs well throughout the different evaluation samples: it outperforms the other two approaches in 18 out of the 24 evaluations.

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 minority of cases that they are outperformed by either PC or Bayesian regression approaches, the results in Tables 7-10 indicate that they are close competitors. Note also that in Tables 7-10 the performance of methods that use PLS factors and Bayesian shrinkage estimators are pretty close, with PLS usually having the edge.

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.

We provide theoretical arguments for the asymptotic similarity between principal components (PC) regression and PLS regression when the underlying data has a common factor structure. When the factor structure in a large data set is weak, and possibly vanishes as $N \rightarrow \infty$, we prove that PLS regression will continue to provide a model with the best asymptotic fit for a target variable. We also argue that forecast combinations can be considered as a specific form of PLS regression. Hence, whether or not a large panel of predictors has a clear factor structure, we would expect PLS regression, like Bayesian ridge regression, to do well in macroeconomic forecasting.

An extensive Monte Carlo analysis, which compares PC regression and Bayesian regression with PLS regression, yields a number of interesting insights. Firstly, when we assume that the predictors relate to the target variable through a standard 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, especially when the number of relevant factors for the target variable increases. When the relation between the predictors and the target variable only has a weak factor structure, PLS regression clearly has the edge in terms of out-of-sample forecasting performance.

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 regression turns out to be generally the best performing method, and even in the few cases when it is outperformed by PC or Bayesian regression, PLS regression remains a close competitor.

Table 1: 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	20	30	50	100	200	400						
PLS	20	0.92	0.96	1.00	1.02	1.04	1.05	1.01	1.02	1.05	1.06	1.07	1.07	1.06	1.07	1.08	1.09	1.09	1.10	1.06	1.07	1.08	1.09	1.09	1.10						
	30	0.90	0.94	0.98	1.01	1.03	1.03	0.99	1.01	1.02	1.03	1.05	1.05	1.03	1.04	1.05	1.06	1.06	1.07	1.03	1.04	1.05	1.06	1.06	1.07						
	50	0.88	0.93	0.96	0.99	1.01	1.02	0.96	0.98	1.00	1.02	1.03	1.03	1.01	1.02	1.03	1.03	1.04	1.04	1.01	1.02	1.03	1.03	1.04	1.04						
	100	0.86	0.87	0.91	0.95	0.96	0.96	0.91	0.91	0.98	0.96	0.96	0.97	0.97	0.96	0.99	0.98	0.99	0.98	0.97	0.96	0.99	0.98	0.99	0.98						
	200	0.83	0.87	0.91	0.94	0.96	0.97	0.91	0.92	0.94	0.95	0.96	0.96	0.94	0.95	0.96	0.97	0.96	0.97	0.94	0.95	0.96	0.97	0.96	0.97						
400	0.84	0.88	0.92	0.96	0.97	0.98	0.92	0.95	0.96	0.97	0.98	0.98	0.95	0.96	0.97	0.98	0.98	0.98	0.95	0.96	0.97	0.98	0.98	0.98							
BR(1)	20	0.71	0.78	0.86	0.93	0.96	0.98	0.94	0.95	0.97	0.98	0.99	0.99	1.04	1.03	1.02	1.01	1.00	1.00	1.04	1.03	1.02	1.01	1.00	1.00						
	30	0.62	0.70	0.80	0.89	0.94	0.97	0.90	0.94	0.96	0.98	0.99	0.99	1.04	1.04	1.03	1.02	1.00	1.00	1.04	1.04	1.03	1.02	1.00	1.00						
	50	0.53	0.59	0.70	0.83	0.90	0.95	0.86	0.90	0.93	0.96	0.98	0.99	1.01	1.04	1.04	1.03	1.01	1.01	1.01	1.04	1.04	1.03	1.01	1.01						
	100	0.38	0.38	0.41	0.48	0.52	0.57	0.67	0.67	0.69	0.63	0.63	0.61	0.86	0.80	0.81	0.71	0.72	0.68	0.86	0.80	0.81	0.71	0.72	0.68						
	200	0.33	0.31	0.29	0.28	0.27	0.26	0.62	0.59	0.54	0.45	0.36	0.31	0.75	0.72	0.64	0.53	0.40	0.33	0.75	0.72	0.64	0.53	0.40	0.33						
400	0.34	0.33	0.31	0.29	0.27	0.27	0.65	0.63	0.60	0.53	0.44	0.36	0.79	0.77	0.72	0.64	0.52	0.40	0.79	0.77	0.72	0.64	0.52	0.40							
BR(5)	20	0.86	0.90	0.94	0.97	0.99	1.00	0.93	0.95	0.96	0.98	0.99	0.99	0.97	0.97	0.98	0.99	0.99	0.99	0.97	0.97	0.98	0.99	0.99	0.99						
	30	0.79	0.85	0.90	0.95	0.98	0.99	0.91	0.93	0.96	0.98	0.99	0.99	0.96	0.97	0.98	0.99	0.99	0.99	0.96	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	0.94	0.96	0.97	0.98	0.99	0.99						
	100	0.51	0.56	0.63	0.73	0.77	0.81	0.73	0.75	0.78	0.80	0.83	0.83	0.86	0.84	0.85	0.84	0.86	0.85	0.86	0.84	0.85	0.84	0.86	0.85						
	200	0.40	0.42	0.48	0.55	0.62	0.65	0.66	0.66	0.66	0.67	0.68	0.69	0.78	0.77	0.74	0.73	0.71	0.71	0.78	0.77	0.74	0.73	0.71	0.71						
400	0.37	0.38	0.40	0.47	0.55	0.61	0.67	0.66	0.65	0.66	0.67	0.68	0.80	0.78	0.76	0.74	0.72	0.71	0.80	0.78	0.76	0.74	0.72	0.71							

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 (11) where we consider a factor loading structure as outlined under Case I assuming $r = k_2$ 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 $T + 1, \dots, T + 100$. We impose different levels of the asymptotic fit of the prediction regression, symbolized by the R^2 's. In case of PC regression k_2 factors are extracted, in case of PLS regression k_1 factors are extracted, and in case of Bayesian regression (BR) a shrinkage parameter $q \times N$ is used with $q = 1, 5$.

Table 2: Standard Common Factors Case, Relative MSE compared to PC for $(k_1, k_2) = (3, 6)$

		$R^2=0.5$												$R^2=0.33$												$R^2=0.2$											
T/N		20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400	20	30	50	100	200	400												
PLS	20	0.93	0.93	1.00	1.00	1.08	1.06	1.06	1.08	1.09	1.13	1.14	1.15	1.13	1.11	1.13	1.17	1.17	1.18	1.13	1.11	1.13	1.17	1.17	1.18												
	30	0.85	0.89	0.96	0.98	1.05	1.03	0.96	1.03	1.06	1.06	1.11	1.12	1.08	1.08	1.12	1.16	1.19	1.17	1.01	1.03	1.07	1.11	1.15	1.13												
	50	0.84	0.89	0.91	1.04	1.02	1.07	0.95	0.97	1.02	1.09	1.08	1.10	0.97	0.95	0.94	0.90	0.91	0.98	0.93	0.91	0.90	0.90	0.86	0.82												
	100	0.77	0.89	0.85	0.86	1.03	0.91	0.90	0.95	0.94	0.90	0.84	0.84	0.89	0.88	0.87	0.89	0.88	0.84	0.94	0.94	0.95	0.96	0.92	0.88												
	200	0.73	0.81	0.82	0.86	0.88	0.86	0.89	0.89	0.89	0.89	0.89	0.87	0.93	0.91	0.90	0.90	0.90	0.86	0.99	0.99	0.99	0.98	0.97	0.97												
BR(1)	20	0.75	0.81	0.88	0.93	0.96	0.97	0.93	0.95	0.95	0.96	0.97	0.97	0.99	0.99	0.99	0.99	0.98	0.97	1.02	1.02	1.01	1.00	1.00	0.99												
	30	0.67	0.74	0.82	0.91	0.95	0.97	0.93	0.94	0.96	0.97	0.98	0.98	1.00	1.04	1.04	1.02	1.00	1.00	0.85	0.82	0.83	0.77	0.69	0.66												
	50	0.56	0.62	0.72	0.83	0.92	0.96	0.89	0.90	0.94	0.97	0.98	0.99	0.81	0.75	0.67	0.54	0.41	0.33	0.83	0.80	0.75	0.65	0.52	0.40												
	100	0.44	0.46	0.45	0.49	0.57	0.59	0.80	0.71	0.71	0.66	0.65	0.67	0.90	0.91	0.93	0.92	0.92	0.93	0.94	0.94	0.95	0.96	0.96	0.96												
	200	0.40	0.35	0.32	0.29	0.28	0.27	0.70	0.64	0.56	0.46	0.37	0.31	0.94	0.94	0.95	0.97	0.97	0.98	0.94	0.94	0.95	0.96	0.96	0.96												
BR(5)	20	0.41	0.36	0.33	0.30	0.28	0.27	0.72	0.67	0.62	0.55	0.45	0.36	0.90	0.91	0.93	0.92	0.92	0.93	0.94	0.94	0.95	0.96	0.96	0.96												
	30	0.98	1.00	1.02	1.03	1.06	1.07	0.93	0.94	0.96	0.97	0.98	0.98	0.94	0.94	0.94	0.96	0.98	0.99	0.94	0.96	0.97	0.97	0.98	0.99												
	50	0.89	0.93	0.97	1.00	1.02	1.03	0.90	0.91	0.94	0.96	0.98	0.99	0.83	0.79	0.82	0.82	0.84	0.86	0.87	0.86	0.87	0.87	0.85	0.85												
	100	0.59	0.63	0.68	0.75	0.80	0.83	0.83	0.79	0.82	0.82	0.84	0.86	0.73	0.70	0.68	0.69	0.69	0.69	0.83	0.79	0.77	0.74	0.72	0.71												
	200	0.47	0.47	0.50	0.57	0.62	0.66	0.73	0.70	0.68	0.68	0.67	0.68	0.84	0.81	0.78	0.75	0.73	0.71	0.84	0.81	0.78	0.75	0.73	0.71												

Notes: See the notes for Table 1.

Table 3: Weak Factor Case II (Weak Pervasiveness $\kappa_1 = 0.75$), Relative MSE compared to PC for $(k_1, k_2) = (1, 1)$

T/N	$R^2 = 0.5$												$R^2 = 0.33$												$R^2 = 0.2$																											
	20				30				50				100				200				400				20				30				50				100				200				400							
	0.84	0.89	0.94	0.99	0.86	0.91	0.98	1.01	1.03	1.03	1.03	1.03	1.00	1.03	1.04	1.06	1.06	1.06	1.06	1.06	1.05	1.05	1.05	1.05	1.07	1.09	1.11	1.10	1.04	1.07	1.09	1.11	1.06	1.07	1.07	1.07	1.01	1.03	1.07	1.08	0.91	0.93	0.90	0.87	0.85	0.87	0.87	0.85	0.82	0.82	0.82	0.77
PLS	0.80	0.86	0.91	0.98	0.83	0.89	0.95	1.00	1.03	1.03	1.03	1.03	0.92	0.96	1.00	1.04	0.89	0.88	0.88	0.89	0.85	0.85	0.85	0.85	0.91	0.93	0.97	1.01	0.83	0.86	0.88	0.90	0.80	0.82	0.83	0.83	0.80	0.80	0.80	0.80	0.87	0.87	0.87	0.85	0.82	0.82	0.82	0.77				
BR(1)	0.65	0.69	0.75	0.78	0.72	0.76	0.82	0.84	0.86	0.86	0.86	0.86	0.82	0.84	0.87	0.88	0.88	0.88	0.88	0.88	0.87	0.87	0.87	0.87	0.89	0.90	0.90	0.91	1.02	1.02	1.02	1.01	1.02	1.02	1.02	1.01	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
BR(5)	0.71	0.78	0.86	0.92	0.78	0.80	0.89	0.94	0.98	0.98	0.98	0.98	0.94	0.95	0.97	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.02	1.03	1.04	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03

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 (11) 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 $T + 1, \dots, T + 100$. 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 $q = 1, 5$.

Table 4: Weak Factor Case III (Weak Loadings $\kappa_2 = 0.25$), 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	20	30	50	100	200	400												
PLS	20	0.81	0.87	0.91	0.96	0.98	0.99	1.00	1.02	1.05	1.05	1.04	1.03	1.09	1.12	1.12	1.12	1.09	1.06	1.04	1.09	1.12	1.12	1.12	1.09	1.06	1.04										
	30	0.76	0.82	0.88	0.94	0.97	0.99	0.96	0.99	1.03	1.05	1.05	1.03	1.06	1.09	1.10	1.11	1.11	1.08	1.06	1.06	1.09	1.10	1.11	1.08	1.06	1.04										
	50	0.72	0.76	0.83	0.90	0.96	0.98	0.92	0.95	0.99	1.04	1.06	1.06	1.01	1.05	1.09	1.11	1.11	1.08	1.06	1.01	1.05	1.09	1.11	1.11	1.08	1.06										
	100	0.61	0.63	0.64	0.67	0.68	0.65	0.79	0.83	0.82	0.77	0.71	0.67	0.90	0.90	0.88	0.85	0.85	0.82	0.73	0.90	0.90	0.88	0.85	0.85	0.82	0.73										
	200	0.57	0.60	0.60	0.60	0.55	0.48	0.76	0.76	0.74	0.69	0.61	0.49	0.84	0.83	0.80	0.73	0.62	0.49	0.84	0.83	0.80	0.73	0.62	0.49	0.49											
400	0.58	0.60	0.63	0.65	0.66	0.61	0.79	0.78	0.78	0.77	0.73	0.66	0.87	0.86	0.85	0.82	0.76	0.66	0.87	0.86	0.85	0.82	0.76	0.66	0.66												
BR(1)	20	0.71	0.79	0.86	0.93	0.96	0.98	0.94	0.95	0.97	0.98	0.99	0.99	1.03	1.03	1.02	1.02	1.00	1.00	1.03	1.03	1.02	1.02	1.00	1.00	1.00											
	30	0.63	0.71	0.80	0.89	0.94	0.97	0.91	0.93	0.96	0.98	0.99	0.99	1.03	1.04	1.04	1.02	1.01	1.00	1.03	1.04	1.04	1.02	1.01	1.00	1.00											
	50	0.52	0.59	0.70	0.82	0.91	0.95	0.87	0.90	0.93	0.97	0.98	0.99	1.01	1.04	1.04	1.04	1.02	1.01	1.01	1.04	1.04	1.04	1.02	1.01	1.01											
	100	0.37	0.40	0.40	0.48	0.55	0.58	0.68	0.69	0.66	0.66	0.62	0.61	0.84	0.84	0.79	0.70	0.70	0.66	0.84	0.84	0.79	0.70	0.70	0.70	0.66											
	200	0.33	0.31	0.30	0.28	0.27	0.26	0.62	0.59	0.54	0.45	0.36	0.31	0.75	0.71	0.64	0.52	0.41	0.33	0.75	0.71	0.64	0.52	0.41	0.33	0.33											
400	0.34	0.32	0.31	0.29	0.27	0.26	0.66	0.63	0.60	0.53	0.44	0.36	0.79	0.76	0.73	0.64	0.52	0.40	0.79	0.76	0.73	0.64	0.52	0.40	0.40												
BR(5)	20	0.86	0.91	0.94	0.97	0.99	1.00	0.94	0.95	0.97	0.98	0.99	0.99	0.96	0.97	0.98	0.98	0.99	0.99	0.96	0.97	0.98	0.98	0.99	0.99	0.99											
	30	0.80	0.86	0.91	0.96	0.98	0.99	0.91	0.93	0.96	0.97	0.99	0.99	0.95	0.97	0.98	0.98	0.99	0.99	0.95	0.97	0.98	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	0.94	0.96	0.97	0.98	0.99	0.99	0.99											
	100	0.51	0.59	0.64	0.73	0.79	0.82	0.75	0.76	0.78	0.82	0.82	0.83	0.85	0.86	0.84	0.83	0.83	0.85	0.85	0.86	0.84	0.83	0.85	0.85	0.85											
	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	0.76	0.75	0.73	0.71	0.71	0.78	0.76	0.75	0.73	0.71	0.71	0.71											
400	0.37	0.37	0.40	0.47	0.55	0.61	0.67	0.65	0.66	0.66	0.67	0.68	0.80	0.78	0.77	0.74	0.72	0.71	0.80	0.78	0.77	0.74	0.72	0.71	0.71												

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 (11) where we consider a factor loading structure as outlined under Case III 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 $T + 1, \dots, T + 100$. We impose different levels of 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 $q = 1, 5$.

Table 5: Weak Factor Case II (Weak Pervasiveness $\kappa_1 = 0.25$), 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	20	0.77	0.82	0.89	0.94	0.97	0.99	1.00	1.02	1.04	1.04	1.02	1.01	1.09	1.12	1.11	1.08	1.05	1.03
	30	0.71	0.76	0.83	0.91	0.96	0.97	0.95	0.98	1.01	1.03	1.03	1.02	1.08	1.11	1.11	1.09	1.07	1.04
	50	0.62	0.68	0.76	0.85	0.92	0.96	0.90	0.93	0.98	1.02	1.03	1.02	1.02	1.06	1.09	1.11	1.09	1.06
	100	0.47	0.48	0.54	0.51	0.51	0.49	0.68	0.73	0.72	0.71	0.66	0.66	0.89	0.82	0.83	0.80	0.71	0.64
	200	0.41	0.40	0.41	0.37	0.33	0.25	0.67	0.64	0.61	0.53	0.42	0.30	0.78	0.75	0.71	0.60	0.46	0.32
400	0.41	0.40	0.40	0.39	0.37	0.32	0.69	0.67	0.66	0.60	0.53	0.41	0.81	0.79	0.77	0.70	0.60	0.46	
BR(1)	20	0.71	0.78	0.86	0.92	0.96	0.98	0.94	0.96	0.97	0.98	0.99	1.00	1.03	1.03	1.02	1.01	1.00	1.00
	30	0.63	0.71	0.80	0.89	0.94	0.97	0.91	0.94	0.96	0.98	0.99	0.99	1.04	1.04	1.03	1.02	1.01	1.01
	50	0.53	0.61	0.70	0.82	0.91	0.95	0.87	0.90	0.94	0.96	0.98	0.99	1.01	1.04	1.03	1.03	1.01	1.01
	100	0.37	0.39	0.42	0.45	0.52	0.56	0.66	0.68	0.64	0.65	0.64	0.65	0.86	0.80	0.79	0.73	0.69	0.65
	200	0.33	0.31	0.30	0.28	0.27	0.26	0.62	0.59	0.54	0.45	0.36	0.31	0.75	0.71	0.64	0.52	0.41	0.33
400	0.34	0.33	0.31	0.29	0.27	0.27	0.65	0.63	0.60	0.53	0.44	0.36	0.79	0.77	0.73	0.64	0.52	0.40	
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	0.96	0.98	0.98	0.99	0.99	1.00
	30	0.80	0.86	0.91	0.95	0.97	0.99	0.91	0.93	0.96	0.97	0.99	0.99	0.96	0.97	0.98	0.99	0.99	1.00
	50	0.70	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	0.99
	100	0.53	0.58	0.65	0.72	0.77	0.81	0.72	0.76	0.77	0.81	0.83	0.85	0.86	0.84	0.84	0.85	0.85	0.84
	200	0.40	0.43	0.49	0.56	0.62	0.66	0.66	0.66	0.67	0.67	0.69	0.69	0.78	0.76	0.75	0.73	0.71	0.71
400	0.37	0.38	0.41	0.47	0.55	0.62	0.67	0.66	0.65	0.66	0.67	0.68	0.80	0.79	0.77	0.74	0.72	0.71	

Notes: See the notes for Table 3, but now with $N_1 = N^{0.25}$.

Table 6: Weak Factor Case III (Weak Loadings $\kappa_2 = 0.75$), 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	20	0.77	0.83	0.89	0.94	0.97	0.99	1.00	1.03	1.05	1.04	1.02	1.01	1.12	1.13	1.11	1.08	1.05	1.03
	30	0.70	0.76	0.84	0.91	0.95	0.98	0.96	0.99	1.02	1.04	1.03	1.02	1.08	1.10	1.12	1.10	1.07	1.04
	50	0.61	0.67	0.75	0.84	0.92	0.96	0.89	0.94	0.98	1.02	1.04	1.03	1.05	1.07	1.10	1.10	1.09	1.06
	100	0.46	0.45	0.51	0.51	0.52	0.51	0.70	0.72	0.74	0.71	0.67	0.67	0.81	0.82	0.81	0.75	0.69	0.64
	200	0.38	0.37	0.37	0.35	0.31	0.24	0.66	0.63	0.59	0.52	0.42	0.30	0.77	0.74	0.69	0.59	0.46	0.32
	400	0.37	0.37	0.37	0.37	0.35	0.31	0.67	0.66	0.63	0.59	0.51	0.41	0.80	0.78	0.75	0.69	0.58	0.45
BR(1)	20	0.72	0.79	0.87	0.93	0.96	0.98	0.94	0.96	0.98	0.99	0.99	1.00	1.04	1.03	1.02	1.01	1.00	1.00
	30	0.63	0.71	0.81	0.89	0.94	0.97	0.92	0.94	0.96	0.98	0.99	1.00	1.04	1.04	1.03	1.02	1.01	1.00
	50	0.53	0.60	0.70	0.83	0.91	0.95	0.87	0.91	0.93	0.96	0.98	0.99	1.03	1.04	1.04	1.02	1.02	1.01
	100	0.38	0.39	0.44	0.48	0.52	0.57	0.67	0.68	0.68	0.65	0.65	0.65	0.79	0.79	0.75	0.72	0.68	0.65
	200	0.34	0.32	0.30	0.28	0.27	0.27	0.63	0.59	0.54	0.45	0.37	0.31	0.76	0.71	0.65	0.52	0.41	0.33
	400	0.34	0.33	0.31	0.29	0.27	0.27	0.66	0.63	0.60	0.53	0.44	0.36	0.79	0.76	0.73	0.64	0.52	0.40
BR(5)	20	0.87	0.91	0.94	0.97	0.98	0.99	0.94	0.95	0.97	0.98	0.99	1.00	0.97	0.98	0.99	0.99	1.00	1.00
	30	0.80	0.86	0.91	0.95	0.97	0.99	0.91	0.94	0.96	0.98	0.99	0.99	0.96	0.97	0.98	0.99	0.99	1.00
	50	0.71	0.78	0.85	0.92	0.96	0.98	0.87	0.90	0.93	0.96	0.98	0.99	0.95	0.96	0.97	0.98	0.99	1.00
	100	0.53	0.58	0.67	0.73	0.79	0.82	0.73	0.76	0.78	0.81	0.84	0.85	0.83	0.84	0.84	0.84	0.85	0.84
	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	0.76	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	0.65	0.66	0.67	0.68	0.80	0.78	0.77	0.74	0.72	0.71

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

Table 7: Forecast evaluation for CPI inflation

h	PC Regression			PLS Regression			Bayesian Regression	
	2 PC	4 PC	6 PC	1 PLS	2 PLS	3 PLS	N	$5N$
<u>January 1972 - December 2006</u>								
<i>Benchmark: RW</i>								
1	0.9291	0.9138	0.9293	0.9114	0.9208	0.9333	0.9218	0.9017
3	0.9064	0.8951	0.8975	0.8678	0.9003	0.9213	0.9213	0.8885
12	0.8411	0.8308	0.8559	0.8101	0.8233	0.8446	0.8398	0.8185
24	0.9102	0.8659	0.8933	0.8441	0.8476	0.8636	0.8321	0.8291
<i>Benchmark: AR</i>								
1	0.9804	0.9795	0.9847	0.9643	1.0044	1.0565	0.9862	0.9640
3	0.9582	0.9692	0.9754	0.9229	0.9769	1.0059	0.9636	0.9330
12	0.8541	0.8201	0.8360	0.8032	0.8031	0.8251	0.8461	0.8244
24	0.8223	0.7694	0.7942	0.7709	0.7701	0.7884	0.8642	0.8457
<u>January 1985 - December 2006</u>								
<i>Benchmark: RW</i>								
1	0.9594	0.9395	0.9431	0.9510	0.9372	0.9508	0.9486	0.9291
3	1.0149	1.0088	0.9972	0.9928	1.0230	1.0263	1.0373	1.0008
12	1.0180	1.1165	1.1323	1.0158	1.0729	1.1061	1.1005	1.0144
24	1.0205	1.0884	1.1553	1.0127	1.0546	1.1322	1.0742	0.9287
<i>Benchmark: AR</i>								
1	0.9777	0.9587	0.9481	0.9707	0.9487	1.0340	0.9988	0.9724
3	0.9733	0.9602	0.9580	0.9696	0.9853	1.0681	1.0231	0.9852
12	0.9165	0.9798	0.9918	0.9024	0.9433	0.9859	1.0053	0.9289
24	0.9737	1.0086	1.1228	0.9446	1.0047	1.0787	1.0158	0.8738
<u>January 1972 - December 1984</u>								
<i>Benchmark: RW</i>								
1	0.8985	0.8881	0.9156	0.8710	0.9045	0.9160	0.8948	0.8742
3	0.8244	0.8082	0.8229	0.7703	0.8067	0.8432	0.8337	0.8033
12	0.7679	0.7098	0.7417	0.7242	0.7221	0.7413	0.7406	0.7418
24	0.8657	0.7984	0.8039	0.7791	0.7615	0.7570	0.7299	0.7726
<i>Benchmark: AR</i>								
1	0.9830	0.9995	1.0193	0.9579	1.0559	1.0780	0.9738	0.9558
3	0.9461	0.9757	0.9883	0.8848	0.9705	0.9552	0.9157	0.8910
12	0.8083	0.7374	0.7589	0.7504	0.7336	0.7472	0.7976	0.8104
24	0.7826	0.7134	0.7148	0.7196	0.6922	0.6959	0.7976	0.8104

Notes: The table reports the ratio of the RMSE of either a version of (14) *vis-à-vis* autoregressive model (12) or a version of (15) *vis-à-vis* the random walk model (13) for CPI inflation (see Section 4.2)) at each horizon h (in months). Versions of (14) and (15) depend on the usage of principal components (PC), partial least squares (PLS) or Bayesian regression (BR) to compress the information in the panel of predictor variables; see Section 4.1. In each case we use several sub-variants, depending either on the number of principal components, PLS factors or shrinkage parameters (BRR), where the shrinkage parameters is assumed to be proportional to the number of predictors in the panel ($N = 104$). The best performing method relative to the benchmarks are highlighted in **bold**.

Table 8: Forecast evaluation for industrial production

h	PC Regression			PLS Regression			Bayesian Regression	
	$2 PC$	$4 PC$	$6 PC$	$1 PLS$	$2 PLS$	$3 PLS$	N	$5N$
<u>January 1972 - December 2006</u>								
<i>Benchmark: RW</i>								
1	0.8600	0.8557	0.8547	0.8437	0.8589	0.8615	0.8553	0.8292
3	0.8216	0.8287	0.8412	0.7784	0.7668	0.7836	0.7832	0.7665
12	1.0414	1.0621	1.0657	0.7912	0.8558	0.8630	0.8541	0.8186
24	1.0734	1.0875	1.0929	0.9037	0.8575	0.8963	0.8749	0.8534
<i>Benchmark: AR</i>								
1	0.9388	0.9291	0.9303	0.9239	0.9246	0.9664	0.9525	0.9217
3	0.9289	0.9310	0.9411	0.8870	0.8752	0.9036	0.9279	0.9076
12	1.1922	1.2204	1.1878	0.8768	0.9598	0.9514	0.9786	0.9265
24	1.1505	1.1803	1.1840	1.0030	0.9657	0.9994	0.9956	0.9607
<u>January 1985 - December 2006</u>								
<i>Benchmark: RW</i>								
1	0.9758	0.9822	0.9751	0.9610	0.9876	0.9810	0.9854	0.9423
3	0.8486	0.8562	0.8628	0.8196	0.8737	0.9021	0.9260	0.8384
12	0.7791	0.7878	0.7980	0.7422	0.8619	0.9127	0.9125	0.8008
24	0.7044	0.7216	0.7312	0.7699	0.7432	0.8405	0.8524	0.7432
<i>Benchmark: AR</i>								
1	0.9685	0.9874	0.9831	0.9528	0.9650	1.0141	1.0120	0.9576
3	0.9784	0.9874	0.9880	0.9443	1.0287	1.0490	1.1338	1.0013
12	1.0794	1.0815	1.0794	0.9759	1.2166	1.2446	1.3238	1.1329
24	1.0725	1.0761	1.0867	1.1274	1.1642	1.2677	1.3147	1.1475
<u>January 1972 - December 1984</u>								
<i>Benchmark: RW</i>								
1	0.7950	0.7839	0.7867	0.7776	0.7858	0.7941	0.7812	0.7655
3	0.8086	0.8161	0.8307	0.7605	0.7181	0.7309	0.7181	0.7354
12	1.1660	1.1946	1.1940	0.8043	0.8338	0.8139	0.7989	0.8166
24	1.4260	1.4413	1.4439	1.0704	0.9961	0.9845	0.9297	0.9994
<i>Benchmark: AR</i>								
1	0.9186	0.8884	0.8937	0.9043	0.8969	0.9336	0.9110	0.8972
3	0.9066	0.9070	0.9206	0.8634	0.8081	0.8420	0.8373	0.8693
12	1.2216	1.2589	1.2169	0.8299	0.8448	0.8187	0.8189	0.8378
24	1.1835	1.2218	1.2184	0.9508	0.8605	0.8604	0.8269	0.8734

Notes: See the notes for Table 7.

Table 9: Forecast evaluation for unemployment

h	PC Regression			PLS Regression			Bayesian Regression	
	2 PC	4 PC	6 PC	1 PLS	2 PLS	3 PLS	N	5N
January 1972 - December 2006								
<i>Benchmark: RW</i>								
1	0.8927	0.8977	0.9073	0.8856	0.8879	0.8923	0.8866	0.8693
3	0.8095	0.8119	0.8227	0.7964	0.7705	0.8018	0.7870	0.7758
12	1.0483	1.0516	1.0424	0.8338	0.8434	0.8759	0.8872	0.8509
24	1.1398	1.1686	1.1453	0.9195	0.8471	0.8950	0.8871	0.8896
<i>Benchmark: AR</i>								
1	0.9057	0.9169	0.9228	0.9061	0.9009	0.9182	0.9507	0.9326
3	0.9031	0.9065	0.9352	0.8889	0.8616	0.8935	0.9403	0.9100
12	1.0542	1.0494	1.0438	0.8606	0.8645	0.8873	0.9182	0.8791
24	1.0761	1.1071	1.0861	0.8926	0.8221	0.8618	0.8543	0.8624
January 1985 - December 2006								
<i>Benchmark: RW</i>								
1	0.9850	0.9848	0.9867	0.9761	0.9781	0.9875	1.0125	0.9597
3	0.8959	0.8918	0.8897	0.8741	0.8926	0.9357	0.9464	0.8666
12	0.9602	0.9393	0.9332	0.8257	0.9385	0.9403	0.9749	0.8685
24	1.0090	1.0537	1.0204	1.0248	0.9305	1.0132	1.0247	0.9484
<i>Benchmark: AR</i>								
1	0.9584	0.9596	0.9592	0.9545	0.9534	0.9721	1.0410	0.9929
3	0.9484	0.9477	0.9415	0.9095	0.9483	0.9839	1.0875	0.9945
12	1.0078	0.9825	0.9543	0.8646	1.0096	0.9726	1.0629	0.9537
24	0.9999	1.0305	0.9874	1.0153	0.9224	0.9844	1.0044	0.9360
January 1972 - December 1984								
<i>Benchmark: RW</i>								
1	0.8294	0.8384	0.8536	0.8237	0.8261	0.8269	0.7972	0.8073
3	0.7757	0.7806	0.7966	0.7665	0.7211	0.7480	0.7212	0.7404
12	1.0740	1.0850	1.0765	0.8353	0.8030	0.8486	0.8530	0.8426
24	1.2186	1.2438	1.2390	0.8714	0.7931	0.8218	0.8008	0.8541
<i>Benchmark: AR</i>								
1	0.8667	0.8856	0.8963	0.8705	0.8621	0.8782	0.8816	0.8877
3	0.8833	0.8884	0.9321	0.8801	0.8222	0.8527	0.8709	0.8718
12	1.0667	1.0686	1.0701	0.8586	0.8086	0.8550	0.8662	0.8532
24	1.1536	1.1878	1.1822	0.8395	0.7597	0.7914	0.7716	0.8269

Notes: See the notes for Table 7.

Table 10: Forecast evaluation for the federal funds rate

h	PC Regression			PLS Regression			Bayesian Regression	
	2 PC	4 PC	6 PC	1 PLS	2 PLS	3 PLS	N	5N
<u>January 1972 - December 2006</u>								
<i>Benchmark: RW</i>								
1	0.9175	0.9107	0.8897	0.8541	0.8374	0.8555	0.8581	0.8314
3	0.9279	0.9560	0.9747	0.8920	0.9048	0.9196	0.9194	0.8873
12	0.9490	0.9281	0.9345	0.8939	0.9424	0.9757	1.0182	0.9398
24	1.0426	0.9947	1.0146	0.9262	1.0004	1.0154	1.0251	0.9656
<i>Benchmark: AR</i>								
1	1.0387	0.9664	0.9325	0.9713	0.8400	0.8593	0.9088	0.8928
3	1.0253	1.0525	0.9795	0.9717	0.9211	0.9214	0.9180	0.8985
12	0.8602	0.8474	0.8344	0.8136	0.8694	0.8944	0.9444	0.8794
24	0.9739	0.9288	0.9425	0.8724	0.9513	0.9665	0.9769	0.9288
<u>January 1985 - December 2006</u>								
<i>Benchmark: RW</i>								
1	0.8528	1.2544	1.2335	0.8909	1.1798	1.3448	1.2528	1.0070
3	0.7880	1.0271	1.0339	0.7546	1.1476	1.2513	1.2157	0.9454
12	0.8613	0.9935	0.9964	0.8102	1.0153	1.0762	1.0696	0.9481
24	0.9447	0.9363	0.9516	0.8836	0.9970	1.0492	1.0174	0.9477
<i>Benchmark: AR</i>								
1	1.0731	1.3156	1.2996	1.0316	1.3079	1.4715	1.3617	1.0669
3	0.9620	1.0767	1.0707	0.9482	1.1327	1.3825	1.1911	0.9495
12	0.8188	0.9445	0.9398	0.7939	0.9752	1.0251	1.0518	0.9308
24	0.8839	0.8847	0.8820	0.8404	0.9461	0.9928	0.9700	0.9000
<u>January 1972 - December 1984</u>								
<i>Benchmark: RW</i>								
1	0.9223	0.8781	0.8569	0.8511	0.8044	0.8046	0.8191	0.8160
3	0.9394	0.9492	0.9692	0.9040	0.8773	0.8819	0.8861	0.8819
12	0.9590	0.8920	0.8985	0.9000	0.9031	0.9293	0.9860	0.9226
24	1.0409	0.9758	0.9800	0.8812	0.9310	0.9179	0.9395	0.9111
<i>Benchmark: AR</i>								
1	1.0364	0.9397	0.9039	0.9673	0.8009	0.8045	0.8719	0.8805
3	1.0296	1.0504	0.9710	0.9726	0.9002	0.8704	0.8897	0.8932
12	0.8548	0.8044	0.7886	0.8003	0.8234	0.8411	0.8969	0.8511
24	0.9740	0.9104	0.9153	0.8326	0.8910	0.8789	0.9089	0.8912

Notes: See the notes for Table 7.

References

- Bai, J., 2003, Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71**, 135–173.
- Bates, J. M. and C. W. J. Granger, 1969, The Combination of Forecasts, *Operations Research Quarterly* **20**, 451–468.
- Boivin, J. and S. Ng, 2006, Are More Data Always Better for Factor Analysis?, *Journal of Econometrics* **132**, 169–194.
- Capistrán, C. and A. Timmermann, 2008, Forecast Combination With Entry and Exit of Experts, *Journal of Business & Economic Statistics* . forthcoming.
- Chamberlain, G. and M. Rothschild, 1983, Arbitrage, Factor Structure, and Mean-Variance Analysis in Large Asset Markets, *Econometrica* **51**, 1305–1324.
- D’Agostino, A. and D. Giannone, 2006, Comparing Alternative Predictors Based on Large-Panel Factor Models, *Working Paper 680*, European Central Bank.
- D’Agostino, A., D. Giannone and P. Surico, 2006, (Un)Predictability and Macroeconomic Stability, *Working Paper 605*, European Central Bank.
- De Mol, C., D. Giannone and L. Reichlin, 2008, Forecasting Using a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components?, *Journal of Econometrics* **146**, 318–328.
- Elliott, G. and A. Timmermann, 2004, Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions, *Journal of Econometrics* **122**, 47–79.
- Faust, J. and J. H. Wright, 2007, Comparing *Greenbook* and Reduced Form Forecasts using a Large Realtime Dataset, *mimeo*, Federal Reserve Board and Johns Hopkins University.
- Forni, M., M. Hallin, M. Lippi and L. Reichlin, 2000, The Generalized Dynamic Factor Model: Identification and Estimation, *Review of Economics and Statistics* **82**, 540–554.

- Forni, M., M. Hallin, M. Lippi and L. Reichlin, 2004, The Generalised Factor Model: Consistency and Rates, *Journal of Econometrics* **119**, 231–255.
- Forni, Mario, Marc Hallin, Marco Lippi and Lucrezia Reichlin, 2005, The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting, *Journal of the American Statistical Association* **100**(471), 830–840.
- Garthwaite, P. H., 1994, An Interpretation of Partial Least Squares, *Journal of the American Statistical Association* **89**, 122–127.
- Granger, C. W. J. and R. Ramanathan, 1984, Improved Methods of Combining Forecasts, *Journal of Forecasting* **3**, 197–204.
- Helland, I. S., 1988, On the Structure of Partial Least Squares Regression, *Communications in Statistics - Simulation and Computation* **17**, 581–607.
- 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.
- Krolzig, H.-M. and D. F. Hendry, 2001, Computer Automation of General-to-Specific Model Selection Procedures, *Journal of Economic Dynamics and Control* **25**, 831–866.
- McConnell, M. and G. Perez-Quiros, 2000, Output Fluctuations in the United States: What has Changed Since the Early 1980's?, *American Economic Review* **90**, 1464–1476.
- Schwarz, G., 1978, Estimating the Dimension of a Model, *Annals of Statistics* **6**, 461–464.
- Sensier, M. and D. J. C. van Dijk, 2004, Testing for Volatility Changes in U.S. Macroeconomic Time Series, *Review of Economics and Statistics* **86**, 833–839.
- Stock, J. H. and M. W. Watson, 2002a, Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association* **97**, 1167–1179.
- Stock, J. H. and M. W. Watson, 2002b, Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business & Economic Statistics* **20**, 147–162.

- Stock, J. H. and M. W. Watson, 2007, Forecasting in Dynamic Factor Models Subject to Structural Instability, *mimeo*, Harvard University and Princeton University.
- Stoica, P. and T. Söderström, 1998, Partial Least Squares: A First-Order Analysis, *Scandinavian Journal of Statistics* **25**, 17–24.
- Svensson, Lars E O, 2005, Monetary Policy with Judgment: Forecast Targeting, *International Journal of Central Banking* **1**(1), 1–54.
- Timmermann, A., 2006, Forecast Combinations, *in* G. Elliott, C. W. J. Granger and A. Timmermann (editors), *Handbook of Economic Forecasting*, North Holland, Amsterdam, pp. 135–196.
- Wold, H., 1982, Soft Modeling. The Basic Design and Some Extensions, *in* K.-G. Jöreskog and H. Wold (editors), *Systems Under Indirect Observation, Vol. 2*, North-Holland, Amsterdam.

Appendices

A Proof of Theorem 1

We first note the following obvious fact: If

$$\|\hat{\alpha}_{PLS}^n - \hat{\alpha}_{PC}^n\| = o_p(a_{N,T}), \quad (\text{A.1})$$

for some sequence $a_{N,T}$, then

$$\|\hat{\alpha}_{PLS} - \hat{\alpha}_{PC}\| = o_p(a_{N,T}/b_{N,T}), \quad (\text{A.2})$$

where $\hat{\alpha}_{PLS}^n$ and $\hat{\alpha}_{PC}^n$ are the the PLS and PC regression estimates (5) and (6) obtained when

$$X^n = \frac{1}{b_{N,T}} X$$

are used as regressors. Given this fact, we focus on $\hat{\alpha}_{PLS}^n$ and $\hat{\alpha}_{PC}^n$ where we set $b_{N,T} = (NT)^{1/2}$.

First, we have that

$$X^{n'} X^n = \frac{1}{N} \Sigma + \frac{1}{N} \left(\frac{X^{n'} X^n}{T} - \Sigma \right)$$

But, by the factor assumption,

$$\frac{1}{N} \Sigma = \frac{1}{N} \tilde{S} \tilde{\Psi} \tilde{S}' + o(1)$$

where

$$\tilde{\psi}_{Ni} = O(N), \quad i = 1, \dots, r.$$

The proof is complete if we show that

$$\left\| \frac{1}{N} \left(\frac{X^{n'} X^n}{T} - \Sigma \right) \right\| = o_p(1),$$

since, then, by Assumption 1 and (7), it follows that

$$\|\hat{\alpha}_{PLS}^n - \hat{\alpha}_{PC}^n\| = o_p(1),$$

and then, by (A.1) and (A.2), (8) follows. But,

$$\begin{aligned} & \left\| \frac{1}{N} \left(\frac{X^{n'} X^n}{T} - \Sigma \right) \right\| = \\ & \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{1}{T} \sum_{t=1}^T (x_{i,t} x_{j,t} - \sigma_{i,j}) \right)^2 \right)^{1/2} = O_p(T^{-1/2}), \end{aligned}$$

where the second equality follows by Assumption 2, proving the result.

B Proof of Theorem 2

For the proof, we first focus on population quantities and deal with the estimation problem as a last step. Although not necessary, we assume without loss of generality that the variance of y_t is finite. This implies that $\alpha'_N \Sigma_{XX} \alpha_N < \infty$. Then, Assumption 4 implies that

$$\alpha_{N,i} = \frac{\tilde{\alpha}_{N,i}}{\sqrt{N}}$$

for some $\tilde{\alpha}_{N,i}$, such that

$$0 < \sup_i |\tilde{\alpha}_{N,i}| < \infty$$

Of course, $\alpha_{N,i}$ can be equivalently written as

$$\alpha_{N,i} = \Sigma_{XX}^{-1} \sigma_{Xy}$$

Using the above we can write

$$\alpha'_N \Sigma_{XX} \alpha_N = \sigma'_{Xy} \Sigma_{XX}^{-1} \sigma_{Xy}$$

and therefore there exist $\tilde{\sigma}_{Xy,i}$ such that

$$\sigma_{Xy,i} = \frac{\tilde{\sigma}_{Xy,i}}{\sqrt{N}}$$

where

$$0 < \sup_i |\tilde{\sigma}_{Xy,i}| < \infty. \tag{B.3}$$

For PLS where for one factor we have that

$$p \lim_{N \rightarrow \infty} \hat{\alpha}_N = \sigma_{Xy}$$

it follows that

$$p \lim_{N \rightarrow \infty} \frac{1}{T} \hat{\alpha}'_N X' X \hat{\alpha} \hat{\beta}_N = \sigma'_{Xy} \Sigma_{XX} \sigma_{Xy}$$

The result we wish to prove follows if we show that

$$\sigma'_{Xy} \Sigma_{XX} \sigma_{Xy} - \sigma'_{Xy} \Sigma_{XX}^{-1} \sigma_{Xy} = o(1)$$

For a square matrix $A = [a_{ij}]$, define the following matrix norm:

$$\|A\|_{\tilde{\sigma}} = \sum_{i=1}^N \sum_{j=1}^N |\tilde{\sigma}_{Xy,1,i}| |a_{ij}| |\tilde{\sigma}_{Xy,1,j}|$$

This is easily seen to satisfy the standard conditions for it to be a valid matrix norm. Then, we wish to prove that $\|\Sigma_{XX} - \Sigma_{XX}^{-1}\|_{\tilde{\sigma}} = o(N)$. But we immediately have that

$$\|\Sigma_{XX} - \Sigma_{XX}^{-1}\|_{\tilde{\sigma}} \leq \|\Sigma_{XX} - I\|_{\tilde{\sigma}} + \|I - \Sigma_{XX}^{-1}\|_{\tilde{\sigma}} \quad (\text{B.4})$$

Using the Sherman Morrison formula, and for any $N \times N$ nonsingular matrix A , any $N \times N$ matrix M such that $(A + M)^{-1}$ exists, any matrix norm and some constant $c > 0$, it can be easily proven that

$$\left\| A^{-1} - (A + M)^{-1} \right\| \leq c \|M\|$$

Thus, it follows that as long as $\|\Sigma_{XX} - I\|_{\tilde{\sigma}} = o(N)$, (B.4) holds. But, $\|\Sigma_{XX} - I\|_{\tilde{\sigma}} = o(N)$ is equivalent to $\|\Sigma_{XX} - I\|_1 = o(N)$, for the more familiar Minkowski 1-norm $\|\cdot\|_1$ which holds by Assumption 4. Finally, we need to prove that

$$\left\| \frac{1}{T} \sum_{t=1}^T x_{j,t} y_t - \sigma_{Xy,i} \right\| = o_p(1) \quad (\text{B.5})$$

But, by assumption 5

$$\frac{1}{T} \sum_{t=1}^T x_{j,t} y_t - \sigma_{Xy,i} = O_p\left(T^{-1/2}\right), \text{ uniformly over } i. \quad (\text{B.6})$$

Hence, (B.5) follows if $N/T \rightarrow 0$.

C 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

Table C.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	$X_t = \Delta \ln Y_{t,t-12} - \Delta \ln Y_{t-1,t-13}$

use first differences of twelve-month transformations of the raw series. Table C.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:

<u>Series Y_t</u>	<u>Transformation:</u> <i>(See Table C.1)</i>
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