Federal Reserve Bank of New York
Staff Reports

Model Selection Criteria for Factor-Augmented Regressions

Jan J. J. Groen
George Kapetanios

Staff Report no. 363
February 2009

Model Selection Criteria for Factor-Augmented Regressions
Jan J. J. Groen and George Kapetanios
Federal Reserve Bank of New York Staff Reports, no. 363
February 2009
JEL classification: C22, C52, E37

Abstract

In a factor-augmented regression, the forecast of a variable depends on a few factors estimated from a large number of predictors. But how does one determine the appropriate number of factors relevant for such a regression? Existing work has focused on criteria that can consistently estimate the appropriate number of factors in a large-dimensional panel of explanatory variables. However, not all of these factors are necessarily relevant for modeling a specific dependent variable within a factor-augmented regression. This paper develops a number of theoretical conditions that selection criteria must fulfill in order to provide a consistent estimate of the factor dimension relevant for a factor-augmented regression. Our framework takes into account factor estimation error and does not depend on a specific factor estimation methodology. It also provides, as a by-product, a template for developing selection criteria for regressions that include standard generated regressors. The conditions make it clear that standard model selection criteria do not provide a consistent estimate of the factor dimension in a factor-augmented regression. We propose alternative criteria that do fulfill our conditions. These criteria essentially modify standard information criteria so that the corresponding penalty function for dimensionality also penalizes factor estimation error. We show through Monte Carlo and empirical applications that these modified information criteria are useful in determining the appropriate dimensions of factor-augmented regressions.

Key words: factor models, information criteria, macroeconomic forecasting

---

Groen: Federal Reserve Bank of New York (e-mail: jan.groen@ny.frb.org). Kapetanios: Queen Mary University of London (e-mail: g.kapetanios@qmul.ac.uk). The authors thank Craig Kennedy for excellent research assistance. 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

When forecasting an economic variable, it is often necessary to incorporate information from a large set of potential explanatory variables into the forecasting model. Most traditional macroeconomic prediction approaches, however, are unable to deal with this, either because it is inefficient or downright impossible to incorporate a large number of variables in a single forecasting model and estimate it using standard econometric techniques. As an alternative approach to this problem factor-augmented regressions have gained a prominent place. A seminal application is Stock and Watson (2002b), where a limited number of principal components extracted from a large data set are added to a standard linear regression model which then is used to forecast key macroeconomic variables. Stock and Watson (2002a) and Bai (2003) formalized the underlying asymptotic theory, which allows the use of principal components in very large data sets to identify the common factors in such a data set.

Dynamic factor research in econometrics has spend substantial effort on developing tests and selection criteria aimed at determining that number of factors that describes best the dynamics in a large data set of explanatory variables. A well-known contribution is Bai and Ng (2002), who derive a range of consistent information criteria that can be used to identify the common factor space underlying a large panel of predictor series. While the number of factors selected in such a way provides an upper bound for the number of factors that should enter the forecasting regression for a particular variable, there is no *a priori* reason to suppose that all factors should enter this regression. Therefore, it is of importance that a form of factor selection is carried out that is tailored at determining a factor-based forecasting model for a specific variable. This problem has received far less attention in the literature than the aforementioned issue of determining the number of factors that best explains the dynamics in large data sets of explanatory variables.

Intuitively, since the aim is to specify a regression model for a single variable, standard information criteria may be considered useful in selecting the optimal number of factors for a particular forecasting regression. However, factor variables are not observed and as a result this estimation error may matter and make standard information criteria invalid. Stock and Watson (1998) propose a selection criterion that takes into account this estimation error - note, though, that their criterion depends on an unknown parameter which needs to be calibrated before hand. Building on Bai and Ng (2006), Bai and Ng (2008) propose a final prediction error (FPE) criterion in which an extra penalty term is added to proxy for the effect of factor estimation error on the forecasting regression. Optimizing this FPE will yield the number of factors that asymptotically minimizes the prediction error, but it does not necessarily provide an asymptotically consistent estimate of the number of factors present in the regression of interest. Also, the finite sample performance of this FPE criterion depends on the choice of a consistent estimator of the factor estimation error variance.

In this paper, we propose a number of novel insights with respect to this issue of determining the relevant factors for a specific factor-augmented regression. We show that standard information criteria are inconsistent estimators of the true dimension of the relevant factor space, in particular when the time series dimension of the underlying panel of predictor variables grows faster than its cross-section dimension. As an alternative we suggest new criteria that are consistent estimators in all cases - essentially we build on existing consistent information criteria for time series analysis and modify them to take into account the effects of factor estimation error. Further, we generalize our analysis to factor estimation methods other than principal components. In general, our analysis is valid for setups where variable selection has to be carried out in the presence of generated regressors. Both Monte Carlo and empirical exercises show the relevance and added value of our proposed framework.

The paper is structured as follows. In Section 2, we present our setup and theoretical results. Section 3 reports on a detailed Monte Carlo study of our new selection criteria in comparison with existing ones. Section 4 presents an empirical forecasting application and Section 5 concludes.

2 Theory

We focus on a single variable y_t that we wish to model using an N-dimensional set of variables x_t and the latter is assumed to have a factor structure. In particular, we posit the following model for x_t:

$$x_t = \Lambda f_t + u_t, \quad t = 1, \ldots, T \tag{1}$$

where f_t is an $r \times 1$ vector of factor variables such that $r << N$ and u_t is an $N \times 1$ vector of zero-mean errors. The factors f_t are not observed and need to be estimated from the $N \times 1$ data vector, x_t. Let the forecasting equation for y_t, be specified as

$$y_t = f_t^{0'} \beta + e_t \tag{2}$$

where f_t^0 is $r_0 \times 1$ vector of factor variables that is possibly a subset of f_t, i.e., $1 \leq r_0 \leq r$. The aim of our work is to provide information criteria for selecting the appropriate set of factors that should be entered in (2). There has been a considerable amount of work on determining r, which is the true number of factors needed to explain x_t (see, e.g., Bai and Ng (2002)). Our focus is different in the sense that not all factors underlying x_t maybe relevant for modeling y_t. It is clear that standard information criteria may be of use in specifying (2), but care needs to be taken given that f_t are not observed and must be estimated from x_t.

Now let us consider the class of information criteria (IC) given by

$$IC = \frac{T}{2} \ln\{\hat{\sigma}_{\hat{e}}^2\} + C_{T,N} \tag{3}$$

where $\hat{\sigma}_{\hat{e}}^2$ denotes the estimated residual variance from the regression

$$y_t = \hat{f}_t' \beta + \hat{e}_t \tag{4}$$

and $\hat{f}_t$ denotes some subset of the estimated factor set obtained by applying principal components (PC) to (1). We further specify that $C_{T,N} = i\tilde{C}_{T,N}$ where i denotes the dimension of the candidate set of factors to be entered in (2) and $\tilde{C}_{T,N}$ denotes a penalty term that depends solely on T and N. This class includes all popular IC such as the Akaike (1974) IC (AIC), the Bayesian IC (BIC - see Schwarz (1978)) and the IC proposed by Hannan and Quinn (1979) (HQIC). We have the following theorem concerning the consistency of factor selection using IC of the above form.

Theorem 1 *Let Assumptions A-E of Bai and Ng (2006) hold. Let*

$$\hat{\mathcal{F}} = \left\{ \left\{ \hat{f}_t^{(1)} \right\}_{t=1}^T, \left\{ \hat{f}_t^{(2)} \right\}_{t=1}^T \ldots, \left\{ \hat{f}_t^{(r)} \right\}_{t=1}^T \right\}$$

denote the set of estimated factor vectors over which the information criterion search is carried over and let

$$\mathcal{F} = \left\{ \left\{ f_t^{(1)} \right\}_{t=1}^\infty, \left\{ f_t^{(2)} \right\}_{t=1}^\infty \ldots, \left\{ f_t^{(r)} \right\}_{t=1}^\infty \right\}$$

denote the probability limit of $\hat{\mathcal{F}}$ *as* $N, T \to \infty$. *Further, denote by* $\left\{f_t^0\right\}_{t=1}^{\infty}$ *the set of true factors entering (2). Assume that there exists a unique* i, $1 \leq i \leq r$, *such that* $\left\{f_t^{(i)}\right\}_{t=1}^{\infty}$ *is both spanned by and spans* $\left\{f_t^0\right\}_{t=1}^{\infty}$. *Let the penalty term of the IC, denoted by* $C_{T,N}$ *satisfy the following: (i) Suppose that* $\left\{f_t^{(i)}\right\}_{t=1}^{\infty}$ *does not span* $\left\{f_t^0\right\}_{t=1}^{\infty}$ *but* $\left\{f_t^0\right\}_{t=1}^{\infty}$ *spans* $\left\{f_t^{(i)}\right\}_{t=1}^{\infty}$. *Then,* $C_{T,N}^0 - C_{T,N} = o(T)$. *(ii) Suppose that* $\left\{f_t^0\right\}_{t=1}^{\infty}$ *does not span* $\left\{f_t^{(i)}\right\}_{t=1}^{\infty}$ *but* $\left\{f_t^{(i)}\right\}_{t=1}^{\infty}$ *spans* $\left\{f_t^0\right\}_{t=1}^{\infty}$. *Then,* $\frac{C_{T,N}^0 - C_{T,N}}{T \min(N,T)^{-1}} \to -\infty$. *Under the above conditions, the IC search will choose the unique* i *such that* $\left\{f_t^{(i)}\right\}_{t=1}^{\infty}$ *is both spanned by and spans* $\left\{f_t^0\right\}_{t=1}^{\infty}$, *with probability approaching 1.*

Proof: *See Appendix A for details on the proof of this theorem.*

Remark 1 *Theorem 1 is a factor selection consistency result. Note that we assume that there exists an* $\left\{f_t^{(i)}\right\}_{t=1}^{\infty}$ *such that* $\left\{f_t^{(i)}\right\}_{t=1}^{\infty}$ *is both spanned by and spans* $\left\{f_t^0\right\}_{t=1}^{\infty}$. *Given that only the space of the true factors is consistently estimated by PC, this assumption is not guaranteed to hold. Then, our result can easily be seen to hold for the set* $\left\{f_t^{(i)}\right\}_{t=1}^{\infty}$ *that spans* $\left\{f_t^0\right\}_{t=1}^{\infty}$ *and is of the minimum dimension among all such sets that are considered in the IC search.*

The intuition behind Theorem 1 can be summarized as follows. Condition (i) of Theorem 1 implies that the penalty term of a standard consistent IC is sufficient to guarantee that any $i < r^0$ number of factors is not selected with probability approaching one. Now to avoid selecting $i > r^0$ in the limit, Condition (ii) of Theorem 1 implies that if $T/N \to c$ for $0 \leq c < \infty$ the penalty term of an IC needs to be of larger order of magnitude than T^{-1}, whereas if $N/T \to 0$, it needs to be of larger order of magnitude than N^{-1}. Therefore, standard consistent IC can only guarantee consistency if in (2) $T/N \to c$. Hence, in practice such a standard consistent IC, like BIC, is only useful for determining the factor dimension in a factor-augmented regression if in the underlying panel $N > T$, if not it will overestimate the number of factors entering such a regression.

Theorem 1 relates explicitly to factor estimates obtained by static PC and although this is the most widely used method for estimating factors, there exist a variety of other estimation methods. For example, we have dynamic principal components as suggested in, e.g., Forni *et al.* (2000), there are methods based on maximum likelihood estimation of state space factor models (Doz *et al.* (2006)) or one can follow Groen and Kapetanios (2008) and use partial least squares to directly estimate the factors relevant for a specific dependent variable. These methods may have different consistency rates both for the factor estimates and the coefficients entering (2). It is therefore useful to generalize our consistency result to cover cases where factors are estimated by some other method. As we wish our result to be general we make the following high level assumption where $\hat{\sigma}_e^2$ denotes the residual variance from (2):

Assumption 1 $\hat{\sigma}_e^2 - \hat{\sigma}_{\hat{e}}^2 = O_p(q_{NT})$ *where* $q_{NT} \to 0$.

Theorem 2 spells out the generalization of the consistency result in Theorem 1.

Theorem 2 *Let Assumption 1 and Assumption A of Bai and Ng (2006) hold. Further, assume that* e_t *in* (2) *has finite variance and satisfies a law of large numbers. Let*

$$\hat{\mathcal{F}} = \left\{ \left\{\hat{f}_t^{(1)}\right\}_{t=1}^{T}, \left\{\hat{f}_t^{(2)}\right\}_{t=1}^{T} \ldots, \left\{\hat{f}_t^{(r)}\right\}_{t=1}^{T} \right\}$$

denote the set of estimated factor vectors over which the information criterion search is carried over and let

$$\mathcal{F} = \left\{ \left\{ f_t^{(1)} \right\}_{t=1}^{\infty}, \left\{ f_t^{(2)} \right\}_{t=1}^{\infty} \dots, \left\{ f_t^{(r)} \right\}_{t=1}^{\infty} \right\}$$

denote the probability limit of $\hat{\mathcal{F}}$ as $N, T \to \infty$. Further, denote by $\left\{ f_t^0 \right\}_{t=1}^{\infty}$ the set of true factors entering (2). Assume that there exists a unique i, $1 \leq i \leq r$, such that $\left\{ f_t^{(i)} \right\}_{t=1}^{\infty}$ is both spanned by and spans $\left\{ f_t^0 \right\}_{t=1}^{\infty}$. Let the penalty term of the IC, denoted by $C_{T,N}$ satisfy the following: (i) Suppose that $\left\{ f_t^{(i)} \right\}_{t=1}^{\infty}$ does not span $\left\{ f_t^0 \right\}_{t=1}^{\infty}$ but $\left\{ f_t^0 \right\}_{t=1}^{\infty}$ spans $\left\{ f_t^{(i)} \right\}_{t=1}^{\infty}$. Then, $C_{T,N}^0 - C_{T,N} = o(T)$. (ii) Suppose that $\left\{ f_t^0 \right\}_{t=1}^{\infty}$ does not span $\left\{ f_t^{(i)} \right\}_{t=1}^{\infty}$ but $\left\{ f_t^{(i)} \right\}_{t=1}^{\infty}$ spans $\left\{ f_t^0 \right\}_{t=1}^{\infty}$. Then, $C_{T,N}^0 - C_{T,N} \to -\infty$ and $\frac{C_{T,N}^0 - C_{T,N}}{T q_{NT}} \to -\infty$. Under the above conditions, the information criterion search will choose the unique i such that $\left\{ f_t^{(i)} \right\}_{t=1}^{\infty}$ is both spanned by and spans $\left\{ f_t^0 \right\}_{t=1}^{\infty}$, with probability approaching 1.

Proof: *The proof of Theorem 2 is straightforward and given in Appendix A.*

Remark 1 holds for Theorem 2 as well. Note further, that Theorem 2 need not be applied just to factor models. It provides a general template for determining penalty terms for consistency of information criteria when generated regressors are considered.

In the context of regression (2) it is easy to see that modified versions of the BIC and HQIC given by

$$\begin{aligned} BICM &= \frac{T}{2} \ln\{\hat{\sigma}_{\hat{e}}^2\} + i \ln(T) \left(1 + \frac{T}{N} \right), \\ HQICM &= \frac{T}{2} \ln\{\hat{\sigma}_{\hat{e}}^2\} + 2i \ln\ln(T) \left(1 + \frac{T}{N} \right), \end{aligned} \tag{5}$$

with $1 \leq i \leq r$, fulfill the conditions of Theorems 1 and 2. Of course, other variables can enter the regression and one can also envisage other types of selection. The most obvious one is lag selection where lags of y_t or possibly other variables enter the regression and the number of lags needs to be selected. Given that the conditions of the Theorems 1 and 2 imply that the relevant IC will be consistent also for lag selection, it is clear that such joint searches are feasible. Therefore, we can modify regression (2) such that

$$y_t = z_t' \gamma + f_t^{0'} \beta + e_t \tag{6}$$

with z_t is a $k \times 1$ vector of non-generated regressors and γ is the corresponding parameter vector; z_t can contain an intercept, lags of y_t and so on. The following versions of the modified criteria in (5) are valid for regression (6) under the framework spelled out in Theorems 1 and 2:

$$\begin{aligned} BICM &= \frac{T}{2} \ln\{\hat{\sigma}_{\hat{e}}^2\} + k \ln(T) + i \ln(T) \left(1 + \frac{T}{N} \right), \\ HQICM &= \frac{T}{2} \ln\{\hat{\sigma}_{\hat{e}}^2\} + 2k \ln\ln(T) + 2i \ln\ln(T) \left(1 + \frac{T}{N} \right). \end{aligned} \tag{7}$$

Hence, searching for the optimal values of the modified ICs in (7) will provide the econometrician with a consistent, simultaneous, estimate of the optimal values of k and i in regression (6).

3 Monte Carlo Analysis

In this section we carry out a Monte Carlo study of the new selection criteria for factor-augmented regressions suggested in Section 2. The set-up of the Monte Carlo experiments are spelled out in Section 3.1. In the experiments we compare our suggested criteria with existing ones and the results of this comparison are reported in Section 3.2.

3.1 Set-Up

Our Monte Carlo experiments are based on the following data generating processes (DGPs):

$$\begin{aligned} y_t &= \alpha' x_t + \epsilon_t, \quad t = 1, ..., T, \\ x_t &= \Lambda' f_t + u_t, \\ \epsilon_t &= \sqrt{cN}\varepsilon_t \end{aligned} \tag{8}$$

In (8), $x_t = (x_{1,t} \cdots x_{N,t})'$ is the $N \times 1$ vector of explanatory variables with corresponding $N \times 1$ vector of regression parameters $\alpha = (\alpha_1 \cdots \alpha_N)'$. The term ϵ_t is a zero-mean disturbance term, which we discuss in more detail below. The explanatory variables x_t in (8) are generated by r factors with a $N \times 1$ vector of zero-mean disturbances $u_t = (u_{1,t} \cdots u_{N,t})'$ and a $r \times N$ matrix of factor loadings $\Lambda = (\lambda_1 \cdots \lambda_N)$ that corresponds with the $r \times 1$ vector of factors $f_t = (f_{1,t} \cdots f_{r,t})'$ with $\lambda_i = (\lambda_{i,1} \cdots \lambda_{i,r})'$. The individual regression coefficients in (8) are drawn from a standard normal distribution: $\alpha_i \sim \text{iid}N(0, 1)$ and the disturbances for the N explanatory variables are generated in an equal manner: $u_{i,t} \sim \text{iid}N(0, 1)$. Similarly, the individual factor loadings in Λ are determined as $\lambda_{i,j} \sim \text{iid}N(0, 1)$ and the r factors are each generated as $f_{j,t} \sim \text{iid}N(0, 1)$.

The regression model for y_t in (8) has the alternative representation given by

$$y_t = \alpha'\Lambda' f_t + \left(\alpha' u_t + \epsilon_t\right), \quad t = 1, ..., T,$$

which warrants the use of a factor-augmented type of regression model to explain the dynamics in y_t in a parsimonious manner. In each artificial sample generated from (8), we therefore select the optimal number of factors in a regression of y_t on factors that are estimated by applying PC on x_t. We do this for different selection criteria, both the ones we considered in Section 2, in particular (5) given the set-up in (8), as well as standard information criteria and the FPE criterion suggested in Bai and Ng (2008). Crucial for the Monte Carlo study is the population R^2 of the y_t regression equation in (8). This is controlled by controlling the variance of the y_t disturbance term ϵ_t in (8) through c, where $\varepsilon_t \sim \text{iid}N(0, 1)$. Setting $c = 0.5,\ 1,\ 4,\ 9$ gives a population R^2 equal to $0.66,\ 0.50,\ 0.20$ and 0.10, respectively. We generate data through (8) for $r = 1,\ 2,\ 4,\ 6$ and for $N,\ T = 20,\ 30,\ 50,\ 100,\ 200,\ 400$. The respective Monte Carlo experiments are based on 1000 replications and we report for each selection criterion under consideration the average number of factors selected across the replications. The best performing criterion should on average select a number of factors that is close to the assumed factor order r in a particular Monte Carlo experiment.

3.2 Results

Apart from the modified information criteria (5) in Section 2, which are relevant given DGP (8) for the Monte Carlo experiments, we analyze in our study also the performance of standard information criteria. As AIC is known to be inconsistent, we will in particular focus on the performance of the BIC and HQIC criteria. Given that the extra penalty term in our criteria will be most relevant for cases were $T \geq N$, we expect specially for those cases to observe large differences for the criteria in (5) *vis-à-vis* BIC and HQIC.

Our modified information criteria are not the first set of selection criteria that are specifically developed to determine the dimensions of factor-augmented regressions, albeit that ours are the first *consistent* criteria to be proposed for this purpose. Stock and Watson (1998) derive a selection criterion that penalizes factor estimation error variance. However, this criterion depends on a nuisance parameter that needs to be calibrated before it can be applied to a factor-augmented regression, and this calibration can differ significantly across regression models. Bai and Ng (2008) suggest a forecast prediction error (FPE) criterion that in the limit minimizes the mean squared prediction error of a factor-augmented regression. This FPE criterion essentially entails adding a cross-sectional penalty factor, which depends on an estimate of the factor covariance matrix, to a standard information criterion. So using BIC the FPE for regression model (6) becomes

$$FPE = \ln\{\hat{\sigma}^2_{\hat{e}}\} + (k+i)\left(\frac{\ln(T)}{T}\right) + c_i\left(\frac{\ln(N)}{N}\right) \tag{9}$$

with

$$c_r = \frac{\hat{\beta}'\Sigma_i\hat{\beta}}{\hat{\sigma}^2_{\hat{e}}}$$

where Σ_i is a consistent estimator of the covariance matrix of the i factors included in (6).[1] We use this FPE criterion as a third alternative, next to BIC and HQIC, for our BICM and HQICM criteria, which we implement by setting $k = 0$ in (9).

The results of these experiments are reported in Tables 1-5. When we first focus on the results for the standard information criteria in Tables 1 and 2, it becomes quite apparent that these criteria overestimate the number of factors by a considerable margin. This is particularly the case when the time series dimension T is larger than the cross-section dimension N of the regressor variable vector x_t, indicating the potential severity of the impact of factor estimation error variance in that case. Also, the results for both large T and N in Tables 1 and 2 suggest that BIC and HQIC are not able to provide consistent estimates of the optimal number of factors that underlie a factor-augmented regression. On the other hand our modified criteria, Tables 3 and 4, and to a lesser extent the FPE criterion, Table 5, seem to be behaving consistently and outperform the standard criteria across most experiments. This shows the relevance of our framework and it would be of interest to see how our criteria behave in a real world data setting - something that we will explore in the next section.

4 Empirical Application

The purpose of this section is to assess the performance of our proposed framework when applied on real world data, in particular by assessing its impact on the out-of-sample forecasting performance of factor-augmented regressions. We summarize the set-up of our application in Section 4.1 and discuss the results in Section 4.2.

[1]There are a variety of estimators possible for Σ_r, and the choice of such an estimator impacts the finite sample behavior of (9). We choose to apply a HAC-consistent covariance matrix estimator on the r (i.e. the total number factors driving the dynamics in x_t) estimated factors to proxy Σ_r - this makes sense as each factor is a linear combination of the individual predictor series whose dynamics is not explicitly modeled. Also, we use a HAC estimator for $\hat{\sigma}^2_{\hat{e}}$ in c_r of (9) when $h > 1$. In particular, we found that both in the Monte Carlo and the empirical applications using the Den Haan and Levin (1997) VAR-HAC estimator based on BIC lag selection resulted in the most accurate performance of the FPE criterion.

Table 1: Monte Carlo Results for standard BIC

		$k=1$					$k=2$					$k=4$					$k=6$				
R^2	N / T	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400
	30	3.77	4.11	5.04	6.08	6.81	4.21	4.62	5.39	6.12	6.91	5.16	5.40	5.96	6.48	6.91	6.06	6.39	6.60	7.02	7.25
	50	3.05	3.04	3.71	4.88	5.88	3.43	3.66	4.24	5.14	6.22	4.67	4.79	5.35	5.91	6.43	5.88	6.09	6.42	6.75	7.02
0.66	100	2.25	2.19	2.45	3.49	4.33	2.82	2.84	3.24	3.92	4.80	4.27	4.35	4.84	5.11	5.75	5.63	5.96	6.18	6.47	6.70
	200	1.82	1.71	1.73	2.16	2.74	2.44	2.37	2.52	2.76	3.51	4.07	4.01	4.23	4.55	5.01	5.49	5.76	6.00	6.18	6.45
	400	1.55	1.47	1.41	1.50	1.87	2.17	2.13	2.18	2.28	2.66	3.88	3.86	4.10	4.23	4.45	5.55	5.68	5.90	6.07	6.17
	30	3.23	3.49	4.29	5.46	6.39	3.70	4.00	4.58	5.68	6.57	4.71	4.99	5.57	6.21	6.81	5.77	6.01	6.46	6.84	7.13
	50	2.46	2.67	3.11	4.24	5.24	3.16	3.15	3.72	4.76	5.53	4.49	4.55	4.93	5.68	6.19	5.66	5.98	6.35	6.62	6.93
0.5	100	2.04	1.79	2.10	2.84	3.81	2.61	2.49	2.79	3.47	4.28	4.01	4.11	4.47	4.85	5.37	5.56	5.70	6.01	6.31	6.59
	200	1.66	1.57	1.56	1.85	2.34	2.20	2.22	2.27	2.63	3.06	3.88	3.91	4.05	4.37	4.73	5.36	5.72	5.88	6.08	6.29
	400	1.58	1.35	1.33	1.39	1.61	2.17	2.08	2.09	2.20	2.44	3.74	3.80	3.91	4.12	4.28	5.34	5.56	5.83	5.98	6.11
	30	1.91	1.91	2.19	3.11	4.45	2.36	2.50	2.81	3.68	4.77	3.50	3.78	4.26	4.86	5.82	4.74	5.08	5.74	6.22	6.48
	50	1.78	1.64	1.70	2.28	3.29	2.14	2.12	2.35	2.90	3.82	3.45	3.66	3.94	4.43	5.09	4.66	5.02	5.56	5.91	6.33
0.2	100	1.56	1.40	1.33	1.56	2.06	2.03	1.94	2.02	2.21	2.77	3.37	3.44	3.74	4.04	4.39	4.59	4.99	5.45	5.80	6.13
	200	1.46	1.27	1.23	1.26	1.46	1.99	1.83	1.91	2.00	2.17	3.25	3.42	3.60	3.81	4.11	4.56	5.09	5.43	5.77	5.94
	400	1.36	1.24	1.16	1.19	1.17	1.89	1.75	1.76	1.87	1.98	3.21	3.31	3.58	3.75	3.95	4.66	5.02	5.36	5.67	5.89
	30	1.53	1.45	1.45	1.80	2.68	1.94	1.85	2.01	2.36	3.23	2.84	3.00	3.39	4.04	4.63	3.74	4.09	4.95	5.53	5.98
	50	1.48	1.34	1.27	1.51	1.89	1.93	1.77	1.90	1.99	2.51	2.78	2.96	3.30	3.73	4.15	3.72	4.22	4.82	5.41	5.88
0.1	100	1.39	1.19	1.19	1.21	1.32	1.75	1.71	1.73	1.90	2.05	2.77	2.88	3.25	3.62	3.85	3.88	4.21	4.85	5.47	5.73
	200	1.30	1.19	1.14	1.11	1.16	1.75	1.65	1.70	1.83	1.93	2.77	2.85	3.24	3.51	3.76	3.79	4.24	4.86	5.39	5.70
	400	1.26	1.17	1.11	1.08	1.05	1.71	1.61	1.68	1.76	1.83	2.57	2.83	3.24	3.41	3.69	3.81	4.41	5.00	5.41	5.64

Notes: The entries are the average number of factors selected based on the BIC criterion across 1,000 Monte Carlo replications. The variables are generated through the DGPs in (8) for different values of the time series dimension of all series, T, and the cross-section dimension, N, of series in x_t. We also impose different levels for the asymptotic fit of the prediction regression, as indicated by the R^2's in the table.

Table 2: Monte Carlo Results for standard HQIC

		$k=1$					$k=2$					$k=4$					$k=6$				
R^2	N / T	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400
	30	5.00	5.36	6.18	6.84	7.29	5.30	5.68	6.27	6.82	7.32	5.91	6.14	6.62	7.02	7.30	6.53	6.71	6.95	7.28	7.44
	50	4.35	4.44	5.30	6.18	6.82	4.62	5.01	5.55	6.34	6.97	5.48	5.54	6.08	6.61	7.04	6.35	6.55	6.77	7.09	7.29
0.66	100	3.51	3.35	3.93	4.94	5.97	3.83	3.91	4.51	5.19	6.11	4.98	5.13	5.50	6.00	6.58	6.11	6.32	6.57	6.80	7.09
	200	2.85	2.76	2.91	3.65	4.65	3.44	3.29	3.50	4.18	5.00	4.76	4.61	5.00	5.33	5.86	5.96	6.11	6.31	6.55	6.82
	400	2.40	2.16	2.24	2.69	3.42	3.01	2.87	3.10	3.21	3.95	4.45	4.41	4.69	4.86	5.24	6.00	6.02	6.24	6.40	6.54
	30	4.61	4.91	5.67	6.48	7.11	4.91	5.21	5.67	6.61	7.13	5.51	5.78	6.36	6.79	7.27	6.31	6.52	6.86	7.16	7.42
	50	3.67	4.00	4.68	5.71	6.47	4.32	4.43	4.96	5.97	6.57	5.30	5.33	5.75	6.39	6.79	6.23	6.42	6.73	6.96	7.23
0.5	100	3.03	2.95	3.50	4.40	5.55	3.60	3.44	3.95	4.84	5.61	4.83	4.85	5.15	5.72	6.18	6.00	6.18	6.35	6.66	6.94
	200	2.61	2.39	2.48	3.21	4.10	3.09	3.05	3.22	3.81	4.55	4.58	4.56	4.68	5.14	5.65	5.84	6.07	6.22	6.46	6.69
	400	2.28	2.00	2.15	2.33	2.95	2.86	2.81	2.78	3.00	3.66	4.40	4.33	4.50	4.73	4.99	5.77	5.98	6.11	6.26	6.44
	30	2.91	2.94	3.59	4.83	5.88	3.36	3.63	3.99	5.17	6.02	4.36	4.71	5.17	5.84	6.62	5.51	5.84	6.28	6.67	6.93
	50	2.75	2.63	2.91	3.75	5.20	3.10	3.07	3.43	4.25	5.39	4.38	4.42	4.70	5.34	6.01	5.50	5.72	6.03	6.41	6.81
0.2	100	2.38	2.08	2.15	2.65	3.58	2.76	2.62	2.68	3.15	4.16	4.12	4.08	4.34	4.81	5.26	5.27	5.59	5.88	6.22	6.53
	200	2.13	1.86	1.76	2.05	2.48	2.72	2.33	2.50	2.70	3.17	4.05	3.93	4.12	4.34	4.87	5.22	5.59	5.85	6.11	6.29
	400	1.94	1.67	1.56	1.59	1.90	2.57	2.22	2.21	2.37	2.64	3.83	3.83	4.00	4.13	4.50	5.32	5.57	5.77	6.00	6.22
	30	2.27	2.19	2.34	3.24	4.38	2.72	2.56	2.87	3.58	4.88	3.62	3.83	4.18	5.01	5.68	4.60	4.92	5.65	6.07	6.52
	50	2.09	1.88	1.96	2.47	3.43	2.54	2.48	2.60	2.90	3.83	3.63	3.72	3.96	4.46	5.15	4.63	5.09	5.43	6.00	6.38
0.1	100	2.06	1.67	1.65	1.80	2.30	2.39	2.24	2.21	2.53	2.93	3.33	3.57	3.92	4.21	4.56	4.63	4.99	5.43	5.87	6.14
	200	1.84	1.62	1.44	1.51	1.76	2.37	2.12	2.11	2.22	2.52	3.48	3.39	3.73	3.99	4.31	4.63	5.01	5.38	5.75	6.02
	400	1.82	1.57	1.44	1.37	1.36	2.23	2.07	2.05	2.05	2.17	3.25	3.46	3.65	3.81	4.01	4.63	5.09	5.49	5.76	5.97

Notes: See the notes for Table 1, albeit that factor selection is now based on the HQIC criterion.

Table 3: Monte Carlo Results for Modified BIC

		$k=1$					$k=2$					$k=4$					$k=6$				
R^2	N / T	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400
	30	1.61	1.30	1.15	1.07	1.04	2.06	1.81	1.72	1.60	1.57	3.33	3.15	3.08	3.04	3.01	4.55	4.55	4.45	4.38	4.21
	50	1.62	1.36	1.17	1.09	1.03	2.17	1.93	1.79	1.71	1.67	3.51	3.44	3.47	3.47	3.27	5.08	5.04	5.13	5.07	5.05
0.66	100	1.60	1.40	1.21	1.12	1.08	2.29	2.00	1.97	1.83	1.78	3.70	3.70	3.70	3.59	3.64	5.16	5.41	5.47	5.54	5.55
	200	1.51	1.35	1.24	1.14	1.08	2.18	2.05	1.95	1.88	1.92	3.78	3.71	3.76	3.75	3.82	5.30	5.54	5.68	5.69	5.76
	400	1.45	1.33	1.20	1.12	1.09	2.04	2.02	2.00	1.93	1.93	3.75	3.69	3.88	3.87	3.86	5.47	5.58	5.73	5.80	5.84
	30	1.36	1.19	1.07	1.04	1.01	1.81	1.69	1.63	1.54	1.53	3.02	2.95	2.81	2.79	2.73	4.06	4.06	3.95	4.13	3.89
	50	1.41	1.24	1.12	1.07	1.02	1.95	1.77	1.71	1.68	1.66	3.34	3.24	3.22	3.21	3.24	4.61	4.80	4.86	4.81	4.89
0.5	100	1.45	1.27	1.15	1.10	1.03	2.02	1.89	1.83	1.77	1.73	3.54	3.45	3.56	3.56	3.55	5.11	5.15	5.31	5.41	5.41
	200	1.44	1.27	1.15	1.10	1.04	2.02	1.92	1.89	1.87	1.81	3.62	3.65	3.66	3.73	3.73	5.10	5.47	5.52	5.67	5.65
	400	1.49	1.23	1.15	1.11	1.07	2.09	1.91	1.90	1.88	1.91	3.61	3.64	3.71	3.81	3.81	5.24	5.42	5.67	5.76	5.78
	30	1.12	1.04	1.02	1.00	1.00	1.43	1.38	1.38	1.39	1.35	2.10	2.17	2.15	2.16	2.02	2.81	2.80	2.80	2.71	2.55
	50	1.19	1.07	1.02	1.00	1.00	1.51	1.49	1.47	1.49	1.49	2.58	2.58	2.65	2.60	2.59	3.40	3.51	3.83	3.78	3.68
0.2	100	1.25	1.10	1.03	1.01	1.01	1.67	1.56	1.59	1.58	1.61	2.84	2.82	3.02	3.13	3.21	3.93	4.25	4.62	4.88	4.89
	200	1.27	1.14	1.07	1.03	1.01	1.78	1.69	1.67	1.68	1.72	3.00	3.15	3.25	3.38	3.46	4.25	4.60	5.02	5.25	5.35
	400	1.24	1.17	1.09	1.04	1.02	1.77	1.67	1.69	1.73	1.76	3.07	3.18	3.41	3.55	3.67	4.49	4.85	5.17	5.42	5.57
	30	1.05	1.02	1.00	1.00	1.00	1.25	1.23	1.24	1.22	1.19	1.69	1.63	1.61	1.57	1.44	1.97	1.87	1.90	1.79	1.62
	50	1.10	1.04	1.00	1.00	1.00	1.38	1.34	1.36	1.33	1.34	1.91	2.00	2.06	2.12	2.02	2.48	2.59	2.69	2.71	2.53
0.1	100	1.16	1.05	1.01	1.00	1.00	1.47	1.45	1.43	1.49	1.47	2.32	2.38	2.54	2.67	2.69	3.12	3.43	3.71	3.94	4.08
	200	1.21	1.08	1.03	1.01	1.00	1.60	1.52	1.54	1.60	1.61	2.47	2.52	2.84	3.04	3.17	3.41	3.78	4.33	4.68	4.81
	400	1.22	1.10	1.05	1.03	1.00	1.61	1.54	1.59	1.63	1.65	2.45	2.70	3.06	3.22	3.43	3.61	4.12	4.67	5.04	5.26

Notes: See the notes for Table 1, albeit that factor selection is now based on the modified BIC criterion, BICM - see (5).

Table 4: Monte Carlo Results for Modified HQIC

		$k=1$					$k=2$					$k=4$					$k=6$				
R^2	N / T	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400
	30	2.46	1.91	1.55	1.32	1.20	2.98	2.51	2.17	1.96	1.80	4.18	3.89	3.68	3.62	3.47	5.44	5.37	5.27	5.24	5.17
	50	2.50	1.99	1.56	1.41	1.22	2.97	2.57	2.19	2.09	1.97	4.32	4.09	4.02	3.95	3.64	5.64	5.62	5.68	5.58	5.57
0.66	100	2.46	2.07	1.75	1.55	1.38	3.00	2.75	2.55	2.25	2.11	4.42	4.28	4.31	4.07	4.01	5.77	5.88	5.90	5.90	5.84
	200	2.35	2.05	1.74	1.66	1.45	2.94	2.72	2.52	2.33	2.32	4.48	4.23	4.24	4.23	4.17	5.77	5.89	6.00	6.02	6.06
	400	2.18	1.90	1.66	1.62	1.51	2.75	2.56	2.53	2.37	2.33	4.32	4.23	4.38	4.28	4.25	5.90	5.92	6.08	6.11	6.08
	30	2.02	1.56	1.36	1.19	1.09	2.47	2.21	1.96	1.75	1.71	3.71	3.60	3.36	3.32	3.23	4.99	4.92	5.00	4.97	4.90
	50	2.06	1.75	1.42	1.29	1.15	2.72	2.28	2.06	1.92	1.85	4.13	3.83	3.72	3.62	3.63	5.39	5.46	5.51	5.42	5.40
0.5	100	2.23	1.71	1.55	1.40	1.20	2.78	2.39	2.21	2.12	2.00	4.19	4.02	4.03	3.94	3.82	5.68	5.64	5.68	5.74	5.76
	200	2.15	1.82	1.56	1.42	1.26	2.68	2.50	2.27	2.27	2.08	4.26	4.15	4.05	4.05	4.01	5.64	5.85	5.88	5.93	5.91
	400	2.11	1.75	1.56	1.45	1.36	2.71	2.53	2.36	2.28	2.21	4.22	4.16	4.16	4.19	4.09	5.68	5.87	5.96	6.02	6.00
	30	1.36	1.14	1.07	1.02	1.01	1.70	1.55	1.50	1.53	1.49	2.60	2.62	2.68	2.62	2.65	3.76	3.74	3.84	3.79	3.74
	50	1.50	1.23	1.10	1.04	1.01	1.89	1.69	1.64	1.61	1.61	3.15	3.02	3.11	3.03	3.11	4.29	4.27	4.68	4.66	4.63
0.2	100	1.68	1.34	1.15	1.08	1.04	2.17	1.89	1.79	1.73	1.73	3.52	3.38	3.44	3.44	3.48	4.73	4.92	5.13	5.34	5.32
	200	1.75	1.44	1.23	1.14	1.05	2.39	1.97	1.91	1.83	1.84	3.66	3.64	3.61	3.63	3.72	4.97	5.27	5.43	5.58	5.61
	400	1.80	1.50	1.25	1.20	1.08	2.41	2.05	1.91	1.91	1.89	3.69	3.63	3.71	3.78	3.86	5.19	5.43	5.54	5.71	5.80
	30	1.17	1.08	1.01	1.01	1.00	1.44	1.38	1.35	1.33	1.33	2.16	2.10	2.05	2.12	1.96	2.74	2.70	2.68	2.69	2.61
	50	1.30	1.13	1.03	1.01	1.00	1.68	1.50	1.50	1.43	1.44	2.47	2.47	2.52	2.62	2.65	3.34	3.47	3.59	3.84	3.72
0.1	100	1.49	1.18	1.08	1.02	1.01	1.85	1.67	1.55	1.61	1.59	2.89	2.84	2.96	3.08	3.13	4.03	4.11	4.43	4.81	4.85
	200	1.54	1.28	1.14	1.05	1.02	2.02	1.80	1.71	1.74	1.71	3.16	3.03	3.25	3.33	3.45	4.29	4.54	4.88	5.20	5.33
	400	1.67	1.40	1.23	1.10	1.04	2.10	1.91	1.81	1.77	1.77	3.12	3.26	3.40	3.45	3.62	4.42	4.88	5.25	5.45	5.54

Notes: See the notes for Table 1, albeit that factor selection is now based on the modified HQIC criterion, HQICM - see (5).

Table 5: Monte Carlo Results for Bai-Ng FPE Criterion

		$k=1$					$k=2$					$k=4$					$k=6$				
R^2	N / T	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400	30	50	100	200	400
	30	2.61	2.81	3.80	5.03	6.12	2.86	3.02	3.76	4.77	5.65	3.52	3.77	4.29	4.92	5.59	4.24	4.63	5.08	5.63	5.97
	50	2.28	2.34	3.15	4.18	5.39	2.67	2.79	3.43	4.18	5.41	3.75	3.93	4.42	4.99	5.73	4.72	5.05	5.60	6.02	6.44
0.66	100	1.92	1.90	2.10	2.94	3.96	2.42	2.47	2.71	3.35	4.34	3.88	4.02	4.39	4.66	5.24	5.14	5.51	5.87	6.09	6.43
	200	1.79	1.64	1.64	2.10	2.69	2.28	2.19	2.42	2.68	3.35	3.81	3.90	4.09	4.36	4.74	5.24	5.58	5.85	6.06	6.29
	400	1.45	1.39	1.40	1.55	1.85	2.17	2.10	2.12	2.33	2.57	3.82	3.81	3.99	4.17	4.41	5.34	5.61	5.83	6.01	6.14
	30	2.18	2.40	3.27	4.60	5.68	2.64	2.79	3.47	4.35	5.45	3.46	3.60	4.29	4.85	5.53	4.21	4.61	5.19	5.58	6.10
	50	1.93	2.01	2.55	3.66	5.03	2.47	2.54	3.07	3.82	5.01	3.62	3.80	4.26	4.85	5.55	4.68	5.11	5.63	5.98	6.30
0.5	100	1.78	1.69	1.88	2.48	3.51	2.27	2.27	2.50	3.03	3.84	3.65	3.85	4.04	4.48	5.05	5.06	5.26	5.75	6.04	6.35
	200	1.62	1.51	1.56	1.68	2.21	2.13	2.14	2.20	2.43	2.90	3.65	3.80	3.97	4.24	4.54	5.13	5.47	5.75	5.97	6.22
	400	1.53	1.31	1.36	1.39	1.65	2.12	2.04	1.99	2.18	2.47	3.63	3.74	3.85	4.11	4.25	5.22	5.49	5.75	5.92	6.08
	30	1.55	1.53	1.75	2.48	3.79	1.93	1.94	2.24	2.91	4.03	2.79	3.13	3.60	4.21	4.97	3.71	4.28	5.07	5.62	6.08
	50	1.55	1.48	1.54	1.92	2.89	1.92	1.96	2.11	2.56	3.35	2.97	3.25	3.67	4.10	4.65	4.03	4.49	5.22	5.76	6.08
0.2	100	1.42	1.32	1.31	1.48	1.86	1.86	1.84	1.94	2.21	2.54	3.05	3.30	3.58	3.87	4.21	4.27	4.83	5.29	5.76	5.98
	200	1.39	1.29	1.19	1.27	1.37	1.87	1.77	1.85	1.97	2.11	3.05	3.30	3.54	3.78	4.06	4.44	4.92	5.36	5.68	5.96
	400	1.29	1.18	1.16	1.12	1.19	1.86	1.77	1.79	1.88	1.96	3.20	3.29	3.57	3.74	3.87	4.48	4.97	5.37	5.67	5.82
	30	1.32	1.26	1.35	1.49	2.08	1.65	1.69	1.81	2.11	2.75	2.33	2.68	3.08	3.50	4.15	2.97	3.60	4.41	5.07	5.67
	50	1.27	1.23	1.25	1.34	1.58	1.67	1.63	1.75	1.89	2.32	2.44	2.75	3.19	3.56	4.00	3.24	3.88	4.58	5.25	5.61
0.1	100	1.31	1.20	1.14	1.19	1.28	1.61	1.67	1.71	1.82	2.00	2.52	2.75	3.17	3.46	3.84	3.39	3.89	4.89	5.32	5.63
	200	1.28	1.16	1.12	1.09	1.12	1.65	1.60	1.72	1.74	1.87	2.54	2.91	3.13	3.46	3.72	3.51	4.16	4.91	5.36	5.63
	400	1.23	1.17	1.10	1.07	1.07	1.62	1.59	1.67	1.74	1.80	2.67	2.89	3.23	3.49	3.65	3.68	4.19	4.83	5.32	5.64

Notes: See the notes for Table 1, albeit that factor selection is now based on the Bai and Ng (2008) FPE criterion - see (9).

4.1 Set-Up

We focus in Section 4.2 on the performance of direct forecasts from factor-augmented regressions for a number of macroeconomic variables. It is standard practice in the macroeconomic forecasting literature to use as forecasting benchmarks for factor-augmented regressions an autoregressive (AR) model and the unconditional mean. The AR benchmark model in the context of direct forecasting can be writing as

$$\Delta y_{t+h,t} = \alpha^h + \sum_{i=1}^{p} \rho_i \Delta y_{t-i+1,t-i} + \epsilon_{t+h,t}, \quad t = 1, \ldots, T \tag{10}$$

with $\Delta y_{t+h,t} = y_{t+h} - y_t$ for $h > 0$ and $\Delta y_{t-i+1,t-i} = y_{t-i+1} - y_{t-i}$ for $i = 1, \ldots, p$. The number of lagged first differences p in (10) is determined by sequentially applying the standard Schwarz (1978)'s BIC starting with a maximum lag order of $p = p_{max}$ down to $p = 1$. The unconditional mean benchmark is simply

$$\Delta y_{t+h,t} = \alpha^h + \epsilon_{t+h,t}, \tag{11}$$

which implies a random walk (RW) forecast for the level of the forecast variable y_t. The assessment of the forecasting performance relative to pure AR-based and random walk-based forecasts is based on the square root of the mean of the squared forecast errors (RMSE). In Section 4.2 we will report ratios of the RMSE of factor-augmented regressions *vis-à-vis* the RMSE based on either (10) or (11). Obviously, superior out-of-sample performance of a factor-augmented regression relative to these benchmarks is indicated by a RMSE ratio smaller than one and *vice versa*.

Our factor-augmented regressions adhere to the following specification:

$$\Delta y_{t+h,t} = \alpha^h + \sum_{i=1}^{\hat{r}} \beta_i^h f_{i,t} + \sum_{j=1}^{\hat{p}} \rho_i j \Delta y_{t-i+j,t-j} + \epsilon_{t+h,t}. \tag{12}$$

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 $\hat{r}$ eigenvectors of the $N \times N$ matrix $\tilde{X}'\tilde{X}$ that correspond to the first $\hat{r}$ largest eigenvalues of that matrix and post-multiplying $\tilde{X}$ with these eigenvectors results in the estimated factors used in (12).

It is, of course, the aim of this exercise to evaluate the finite sample performance of different selection criteria than can be used to determine the optimal dimensions of a factor-augmented regression like (12). We will use the same set of selection criteria analyzed in the Monte Carlo study in Section 3.2. Given (12), we use the BICM and HQICM criteria outlined in (7) to search for that combination of $\hat{p}$ and $\hat{r}$ in (12) that minimizes either of these metrics, with the search done across the range $j = 0, \ldots, p^{max}$ and $i = 1, \ldots, r^{max}$. In addition, we do similar searches using the standard BIC and HQIC criteria as well as the Bai-Ng FPE criterion (9) (with $k \geq 1$ in (9)). In the end this results in five different versions of (12) for each forecast horizon that we will assess relative to our two benchmark models. The forecasting models will be updated based on an expanding window of historical data:

1. First forecast for all h is generated on t_0.
2. Extract r^{max} principal components from the N predictor variables over the sample $t = 1, \ldots, t_0 - h$.

3. Determine for each h over the sample $t = 1, \ldots, t_0 - h$ the optimal lag order and optimal number of factors for (12) for each of our five criteria: BICM, HQICM, BIC, HQIC and Bai-Ng FPE across the range $j = 0, \ldots, p^{max}$ and $i = 1, \ldots, r^{max}$. This results in $(\hat{p}_{BICM}, \hat{r}_{BICM})$, $(\hat{p}_{HQICM}, \hat{r}_{HQICM})$, $(\hat{p}_{BIC}, \hat{r}_{BIC})$, $(\hat{p}_{HQIC}, \hat{r}_{HQIC})$ and $(\hat{p}_{FPE}, \hat{r}_{FPE})$. In a similar vein, determine also the optimal lag order for the AR benchmark based on BIC.

4. Given the outcome of step 3, estimate (10), (11) and (12) over the sample $t = 1, \ldots, t_0 - h$ for each h.

5. Extract r^{max} principal components from the N predictor variables N over the sample $t = 1, \ldots, t_0$.

6. Generate for h the forecast $\Delta\hat{y}_{t+h,t}$ using the estimated dimensions from step 3 and the parameter estimates from step 4 as well as, in case of (12), the common factors from step 5.

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

4.2 Empirical Results

We base our empirical exercise on a large panel of monthly macroeconomic, financial and survey-based indicator variables for the United States, which is similar to that used Stock and Watson (2007) but updated by us up to mid-2008. This panel consists of 108 monthly series, which before transformation span a sample starting in January 1959 and ending in July 2008. It spans real variables (sectoral industrial production, employment, subcomponents of unemployment and hours worked), nominal variables (subcomponents of consumer price index, producer price indexes, deflators, wages, money and credit aggregates), asset prices (interest rates, stock prices and exchange rates) and surveys. Of these 108 series, we use 106 as predictor variables that are transformed such that they are $I(0)$. In general this 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 and unemployment series; see Appendix B for more details. We transform the nominal variables 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 predictor variable panel, as for example motivated in D'Agostino and Giannone (2006, Appendix B).[2] Hence, after transforming the predictor variables we end up with an effective span of the data that starts in February 1960 (i.e. 1960.2) and ends in July 2008 (i.e. 2008.07).

The aforementioned panel is used to forecast appropriate transformations of inflation based on the U.S. personal consumption expenditures (PCE) price index as well as the federal funds rate – see Table 6 for an overview of the appropriate transformation of each forecast variable – and we deliberately keep these two variables separate from the panel of predictor series. 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. The Board, like any other central bank, does that based on information extracted across a wide range of data series, so factor methods could potentially be very useful in predicting this variable. Inflation based on the PCE price index is of interest, as the expenditure weights of the individual consumption goods in this price index vary

[2] This 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.

Table 6: Transformation of the forecast variables

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

Notes: The table illustrates the transformation of a forecast variable Y_t, indicated in the first column, for use in the prediction regression (12).

from period to period and it is a chain-linked index. As such, one would expect that PCE inflation better reflects the effects of substitution across goods by consumers when relative prices change than other inflation measures, such as CPI inflation. Also, the Federal Reserve Board has made it clear that it views PCE inflation as its primary measure of inflation.[3]

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 2 horizons (in months): $h = 1$ and $h = 12$, which are horizons commonly analyzed in the literature. In each update we determine five versions of the factor-augmented regression (12) using our modified information criteria in (7), BIC, HQIC and the Bai-Ng FPE measure (9). For each criterion we simultaneously select the optimal lag order from $p = 0, \ldots, 12$ as well as the optimal number of factors across $i = 1, \ldots, 8$ such that a particular criterion is minimized. In case of the AR benchmark (10) we select that lag order from $p = 1, \ldots, 12$ that minimizes the BIC criterion for (10). The forecast evaluation spans three samples: January 1972 - July 2008, January 1972 - December 1984 and January 1985 - July 2008. The latter two subsamples split the first sample in two around the start of the 'Great Moderation'; see, e.g., McConnell and Perez-Quiros (2000) and Sensier and van Dijk (2004) who find evidence for a downward, exogenous, shift in the volatility of a large number of U.S. macroeconomic time series around 1985.

The Monte Carlo-based results in Section 3.2 indicate potentially large divergences across the different selection criteria in terms of the selected number of factors. Therefore, it might be of interest to see if this also occurs in real world data before we move on to discuss the out-of-sample forecasting performance of our factor-augmented regressions. Figures 1 and 2 plot the recursively selected number of factors for our factor-augmented regression model using our five selection criteria at horizons $h = 1$ and $h = 12$ across the January 1972 - July 2008 sample for both PCE inflation and the fed funds rate. When we focus first on PCE inflation, see Figure 1, it becomes clear that the standard BIC and HQIC criteria have a tendency to select, on average, a large number of factors, much larger than our modified versions of these selection criteria, BICM and HQICM. BICM and HQICM, except for the volatile 1970s in case of HQICM, consistently select only the first factor to model PCE inflation dynamics at $h = 1$, whereas the regular criteria jump around between one and eight factors and converging towards 7 factors at the end of the sample. Similar conclusions can be drawn at the one-year ahead horizon. The Bai and Ng (2008) FPE criterion performs in the same vein as our modified information criteria when it comes to number of selected factors at $h = 1$, but tends to select more factors at the one-year horizon. Also in the case of the federal funds rate (Figure 2) we observe the pattern of BIC and HQIC criteria that select, in a very volatile manner, a substantially larger number of factors than our BICM and HQICM criteria. This is fully consistent with the Monte Carlo

[3]See 'Monetary Policy Report to the Congress', February 2000, Federal Reserve Board.

evidence in Section 3.2 for cases when the time series dimension of an underlying panel of predictors grows faster than its cross-section dimension, which indicated that standard selection criteria overestimate the number of factors in a factor-augmented regression. It also corroborates our theoretical insights from Section 2: factor estimation error matters for the corresponding factor-augmented regression and thus model selection criteria need to take that into account in order to get a consistent estimate of the number of factors entering such a regression.

Let us now turn to the out-of-sample forecasting results for both PCE inflation and the federal funds rate, which are reported in Table 7. In case of PCE inflation the factor-augmented model selection strategy based on our HQICM criterion most frequently results in the best performing inflation forecast at the one-quarter horizon, except for the post-1984 sub-sample when the BICM-based strategy yields the most precise inflation forecast. The latter approach generally performs best in terms of one-year ahead inflation predictions. The exception, again, is the 1985-2008 sub-sample, in which all $h = 12$ forecasts seem to perform in a more or less similar fashion - this is a period for which it is known that changes in inflation are very hard to forecast. Our BICM criterion applied to a factor-augmented regression appears to be the most useful in forecasting future changes in the federal funds rate. Only for the 1972-1984 sub-sample, the strategy based on the Bai-Ng FPE criterion yields the most accurate fed fund rate forecasts at $h = 1$ whereas for one-year ahead the standard BIC approach seems to be a bit better than other model selection strategies.

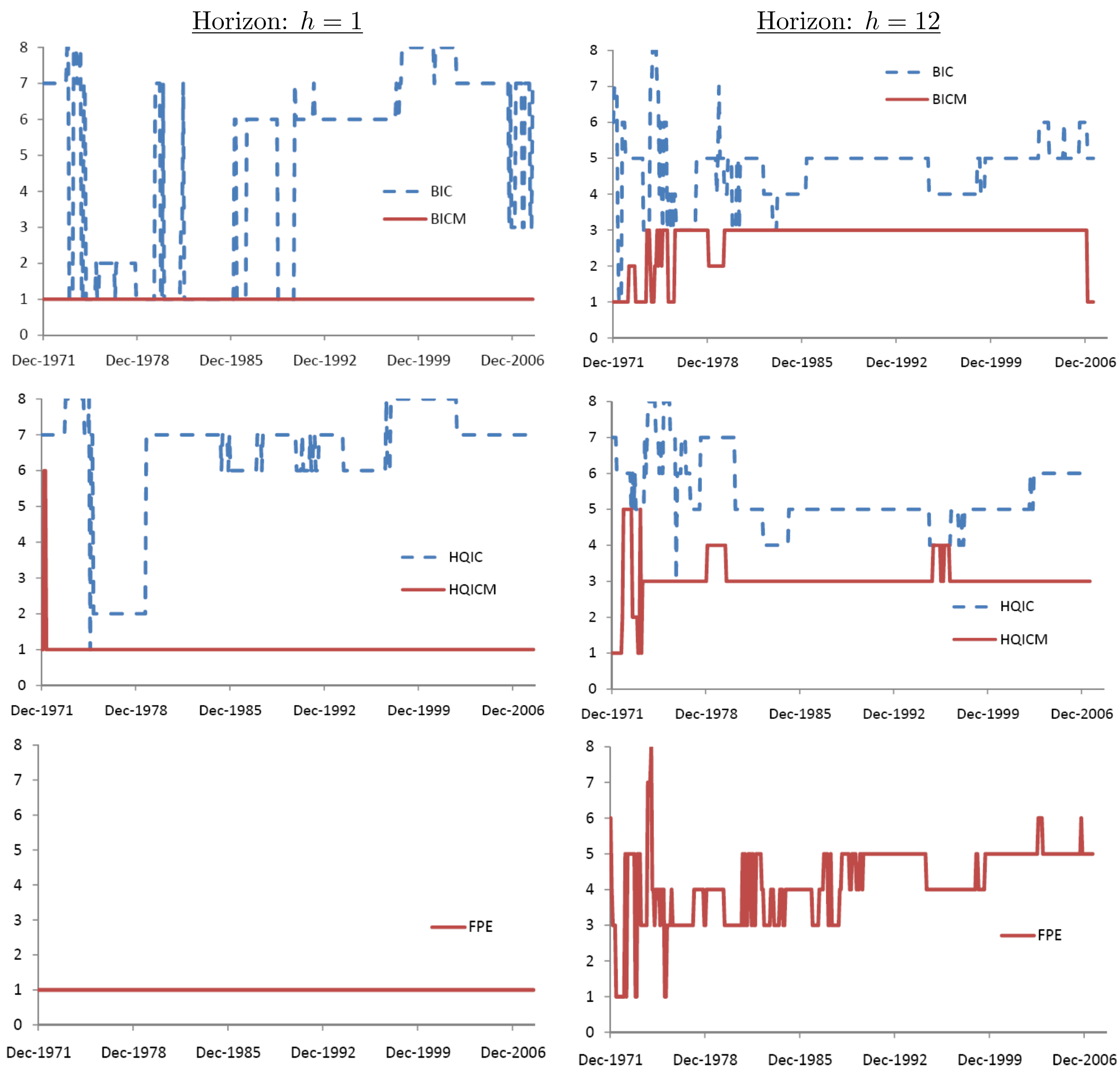
The empirical results in this section confirm our earlier insights from theory and Monte Carlo experiments. By taking into account factor estimation error when selecting the dimensions of a factor-augmented regression, one ends up with parsimonious and relatively stable regression models. These perform at least as well as, and in an overwhelming number of cases improves upon, factor-augmented regressions whose dimensions are determined through standard model selection criteria.

5 Conclusions

Factor-augmented regressions are often used for macroeconomic forecasting and analysis as a parsimonious way of basing the forecast or the analysis on information from a large number of variables. This paper focused on the issue of how to determine the appropriate number of factors that are relevant for such a factor-augmented regression, whereas existing work has been more focused on criteria that can consistently estimate the appropriate number of factors that drive the dynamics in a large-dimensional panel of explanatory variables. However, the resulting number of factors are not necessarily all relevant for modeling a specific dependent variable within a factor-augmented regression.

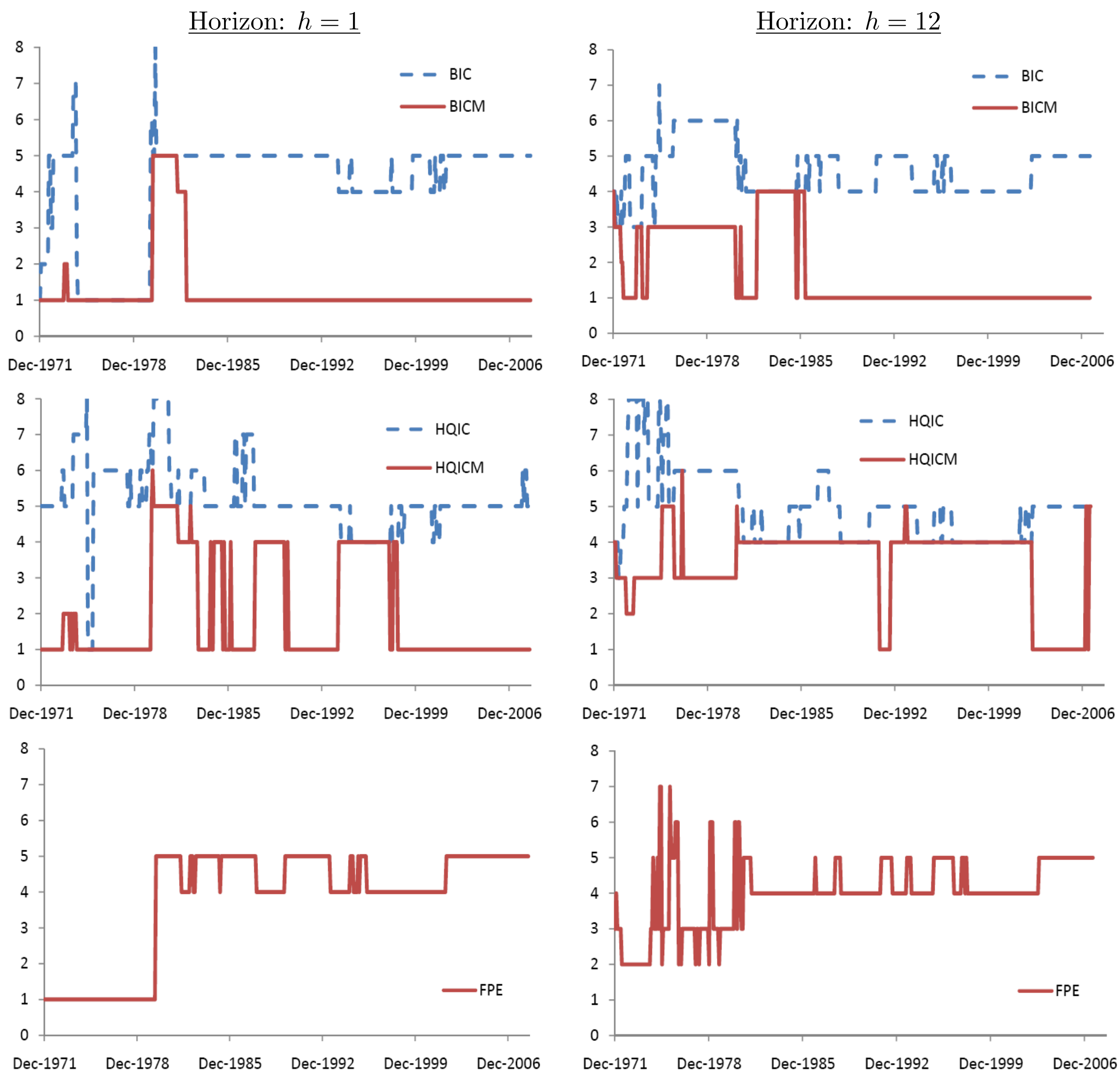
Factor estimation error is an important issue in determining the dimensions of a factor-augmented regression, particularly when the time series dimension of the underlying panel of predictor series is growing at a faster rate than the cross-section dimension. We develop a number of theoretical conditions selection criteria have to fulfill in order to estimate the factor dimension that is relevant for such a regression in a consistent manner. The framework does not hinge on a particular factor estimation methodology and can also provide a template for developing selection criteria for regressions that include standard generated regressors. Based on this framework it is clear that standard model selection criteria like AIC, BIC and HQIC are not guaranteed to provide consistent estimates of the dimensions of a factor-augmented regression. We propose alternative selection criteria that do fulfill the conditions set by our theoretical framework and thus are consistent for factor-augmented regressions. Our criteria essentially take standard information criteria that are

Figure 1: Number of factors selected in case of PCE inflation; 1972.01 - 2008.07



Notes: In the first column of this figure, we depict for PCE inflation the recursively selected number of factors for (12) with $h = 1$ using BIC and BICM (first row), HQIC and HQICM (second row) and the Bai-Ng FPE (last row) criteria. Similarly, in the second column we depict for PCE inflation these recursively selected number of factors for (12) with $h = 12$.

Figure 2: Number of factors selected in case of the federal funds rate; 1972.01 - 2008.07



Notes: See the notes for Figure 1.

commonly used in time series econometrics and modify these such that the corresponding penalty function for dimensionality also penalizes factor estimation error. We show through Monte Carlo applications and empirically, through forecast evaluations for PCE inflation and the fed funds rate, that our model selection criteria are useful in determining the dimensions of factor-augmented regressions and yield models that in most cases outperform factor-augmented regressions selected using existing criteria.

References

Akaike, H., 1974, A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control* **19**, 716–723.

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

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

Bai, J. and S. Ng, 2006, Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions, *Econometrica* **74**, 1133–1150.

Bai, J. and S. Ng, 2008, Boosting Diffusion Indices, *Journal of Applied Econometrics* . forthcoming.

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

Den Haan, W. J. and A. Levin, 1997, A Practitioner's Guide to Robust Covariance Matrix Estimation, *Handbook of Statistics*, Vol. 25, pp. 291–341.

Doz, C., D. Giannone and L. Reichlin, 2006, A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models, *Discussion Paper 5724*, Centre for Economic Policy Research.

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.

Groen, J. J. J. and G. Kapetanios, 2008, Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting, *Staff Reports 327*, Federal Reserve Bank of New York.

Hannan, E. J. and B. G. Quinn, 1979, The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society, Series B* **41**, 190–195.

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, 1998, Diffusion Indexes, *Working Paper 6702*, National Bureau for Economic Research.

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.

Table 7: Out-of-sample forecasting results

h	*BIC*	*HQIC*	*BICM*	*HQICM*	*FPE*	*BIC*	*HQIC*	*BICM*	*HQICM*	*FPE*
	PCE Inflation					Federal Funds Rate				
	January 1972 - July 2008					January 1972 - July 2008				
	Benchmark: RW					*Benchmark: RW*				
1	0.832	0.812	0.815	**0.806**	0.816	0.957	0.995	**0.933**	1.003	0.941
12	1.007	1.041	**0.964**	1.010	0.985	**0.925**	0.940	0.943	0.942	0.953
	Benchmark: AR					*Benchmark: AR*				
1	1.007	0.983	0.986	**0.975**	0.988	0.951	0.989	**0.928**	0.997	0.936
12	0.933	0.965	**0.894**	0.936	0.913	**0.858**	0.872	0.875	0.874	0.884
	January 1985 - July 2008					January 1985 - July 2008				
	Benchmark: RW					*Benchmark: RW*				
1	0.846	0.841	**0.838**	0.838	0.838	1.167	1.146	**0.894**	0.972	1.220
12	1.129	1.132	1.145	1.143	**1.129**	0.995	0.981	**0.898**	0.962	1.004
	Benchmark: AR					*Benchmark: AR*				
1	1.000	0.994	**0.991**	0.991	0.991	1.283	1.261	**0.984**	1.069	1.342
12	0.974	0.976	0.988	0.986	**0.974**	0.966	0.952	**0.872**	0.934	0.974
	January 1972 - December 1984					January 1972 - December 1984				
	Benchmark: RW					*Benchmark: RW*				
1	0.816	0.778	0.787	**0.767**	0.791	0.936	0.981	0.937	1.006	**0.913**
12	0.961	1.006	**0.900**	0.961	0.932	**0.871**	0.907	0.932	0.917	0.912
	Benchmark: AR					*Benchmark: AR*				
1	1.016	0.969	0.980	**0.955**	0.984	0.923	0.967	0.923	0.991	**0.900**
12	0.911	0.953	**0.852**	0.911	0.883	**0.797**	0.830	0.853	0.839	0.834

Notes: The table reports the ratio of the RMSE of a version of (12) *vis-à-vis* the random walk model (11) or the autoregressive model (10) for PCE inflation and fed funds rate (see Table 6)) at each horizon h (in months). Versions of (12) depend on which selection criterion has been used to select both the number of lagged dependent variables as well as the number of principal components; in case of the former the optimal lags are picked from a range between 0 and 12, whereas for the latter they are selected across a range from 1 to 8. Column *BIC* (*HQIC*) indicates the results when the optimal number of lags and factors are chosen to minimize the BIC criterion (Hannan-Quinn IC criterion), columns *BICM* and *HQICM* report results when the optimal number of lags and factors are chosen to minimize our modified BIC and Hannan-Quinn IC measures, see (7), and, finally, *FPE* show the results when the optimal number of lags and factors are chosen to minimize the Bai-Ng FPE criterion (9). The method that performs relatively best *vis-à-vis* the benchmark is highlighted in **bold**.

Appendices

A Proofs

Proof of Theorem 1

Let $\hat{\mathbf{F}} = (\hat{f}_1, \ldots, \hat{f}_T)'$ and $\mathbf{F} = (f_1, \ldots, f_T)'$ where $\hat{f}_t$ denotes a generic set of estimated factors and f_t denote its probability limit. We abstact from the fact that $\hat{\mathbf{F}} \rightarrow_p \mathbf{PF}$ rather than $\hat{\mathbf{F}} \rightarrow_p \mathbf{F}$ by assuming without loss of generality that $\mathbf{P} = \mathbf{I}$. From now on when the matrices $\mathbf{F}$ and $\mathbf{M} = I - \mathbf{F}'(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}$ have superscript 0, they are constructed using the true set of factors, $\mathbf{F}^0$. If they have no superscript then they are constructed using the generic set of factors. When the coefficient vector α has superscript 0 then it refers to a model using the true set of factors. Hats indicate estimated parameters. We denote the penalty term for the generic set of factors by $C_{T,N}$ and the penalty term for the true set of factors by $C^0_{T,N}$. The feasible information criterion takes the following form

$$\widehat{IC}(\alpha, C_{T,N}) = \frac{T}{2} \ln\{\frac{1}{T}(\mathbf{y} - \hat{\mathbf{F}}\alpha)'(\mathbf{y} - \hat{\mathbf{F}}\alpha)\} + C_{T,N}$$

We also introduce the infeasible criterion given by

$$IC(\alpha, C_{T,N}) = \frac{T}{2} \ln\{\frac{1}{T}(\mathbf{y} - \mathbf{F}\alpha)'(\mathbf{y} - \mathbf{F}\alpha)\} + C_{T,N}$$

We decompose $\widehat{IC}(\alpha, C_{T,N})$ as follows

$$\widehat{IC}(\alpha, C_{T,N}) = IC(\alpha, C_{T,N}) + \frac{T}{2} \ln\left[\frac{\frac{1}{T}(\mathbf{y} - \hat{\mathbf{F}}\alpha)'(\mathbf{y} - \hat{\mathbf{F}}\alpha)}{\frac{1}{T}(\mathbf{y} - \mathbf{F}\alpha)'(\mathbf{y} - \mathbf{F}\alpha)}\right]$$

At first, we consider the case $\mathbf{F}$ does not span $\mathbf{F}^0$ but $\mathbf{F}^0$ spans $\mathbf{F}$. Then, consistency requires that

$$\lim_{T\to\infty} P\{\widehat{IC}(\hat{\alpha}, C_{T,N}) - \widehat{IC}(\hat{\alpha}^0, C^0_{T,N}) < 0\} = 0 \tag{A.1}$$

By substitution and using standard regression results this becomes

$$\lim_{T\to\infty} P\left\{\begin{array}{c} \frac{T}{2}\ln\left[\frac{\frac{1}{T}\mathbf{y}'\mathbf{M}\mathbf{y}}{\frac{1}{T}\mathbf{y}'\mathbf{M}^0\mathbf{y}}\right] + \frac{T}{2}\ln\left[\frac{\frac{1}{T}(\mathbf{y}-\hat{\mathbf{F}}\alpha)'(\mathbf{y}-\hat{\mathbf{F}}\alpha)}{\frac{1}{T}(\mathbf{y}-\mathbf{F}\alpha)'(\mathbf{y}-\mathbf{F}\alpha)}\right] + \\ \frac{T}{2}\ln\left[\frac{\frac{1}{T}(\mathbf{y}-\hat{\mathbf{F}}^0\alpha)'(\mathbf{y}-\hat{\mathbf{F}}^0\alpha)}{\frac{1}{T}(\mathbf{y}-\mathbf{F}^0\alpha)'(\mathbf{y}-\mathbf{F}^0\alpha)}\right] < C^0_{T,N} - C_{T,N} \end{array}\right\} = 0 \tag{A.2}$$

We next examine the first term of the LHS of the inequality within the probability statement. By expanding $\mathbf{y}$ we get that

$$\frac{T}{2}\ln\left[\frac{\frac{1}{T}\mathbf{y}'\mathbf{M}\mathbf{y}}{\frac{1}{T}\mathbf{y}'\mathbf{M}^0\mathbf{y}}\right] = \frac{T}{2}\ln\left[\frac{\frac{1}{T_1}(\alpha^{0\prime}\mathbf{F}^{0\prime}\mathbf{M}\mathbf{F}^0\alpha^0 + \epsilon\mathbf{F}\epsilon + 2\epsilon'\mathbf{M}\mathbf{F}^0\alpha^0)}{\frac{1}{T_1}(\alpha^{0\prime}\mathbf{F}^{0\prime}\mathbf{M}^0\mathbf{F}^0\alpha^0 + \epsilon\mathbf{F}^0\epsilon + 2\epsilon'\mathbf{M}^0\mathbf{F}^0\alpha^0)}\right]$$

But idempotency implies positive-definiteness and as a result $\alpha^{0\prime}\mathbf{F}^{0\prime}\mathbf{M}\mathbf{F}^0\alpha^0 > 0$. Further ,

$$\alpha^{0\prime}\mathbf{F}^{0\prime}\mathbf{M}^0\mathbf{F}^0\alpha^0 = \alpha^{0\prime}\mathbf{F}^{0\prime}\mathbf{F}^0\alpha^0 - \alpha^{0\prime}\mathbf{F}^{0\prime}\mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0\alpha^{0\prime} = 0$$

Further, by the assumed stationarity of the model and using the theorem assumptions we get

$$0 < plim\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{F}^0 < \infty$$

$$0 < plim \frac{1}{T}\mathbf{F}'\mathbf{F} < \infty$$

and

$$\frac{1}{T}\mathbf{F}^0\epsilon \rightarrow_p 0$$

and

$$\frac{1}{T}\mathbf{F}\epsilon \rightarrow_p 0$$

Thus

$$(\frac{1}{T}\mathbf{F}^{0\prime}\epsilon)'(\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}(\frac{1}{T}\mathbf{F}^{0\prime}\epsilon) \rightarrow_p 0$$

and

$$(\frac{1}{T}\mathbf{F}'\epsilon)'(\frac{1}{T}\mathbf{F}'\mathbf{F})^{-1}(\frac{1}{T}\mathbf{F}'\epsilon) \rightarrow_p 0$$

Thus,

$$\frac{1}{T_1}\epsilon\mathbf{M}^0\epsilon \rightarrow_p \sigma^2 \tag{A.3}$$

and

$$\frac{1}{T_1}\epsilon\mathbf{M}\epsilon \rightarrow_p \sigma^2 \tag{A.4}$$

As a result of all the above $\ln\left[\frac{\frac{1}{T}\mathbf{y}'\mathbf{M}\mathbf{y}}{\frac{1}{T}\mathbf{y}'\mathbf{M}^0\mathbf{y}}\right]$ is positive and $O_p(1)$ and, therefore,

$$\frac{T}{2}\ln\left[\frac{\frac{1}{T}\mathbf{y}'\mathbf{M}\mathbf{y}}{\frac{1}{T}\mathbf{y}'\mathbf{M}^0\mathbf{y}}\right] = O_p(T)$$

By Theorem 3 of Bai and Ng (2006) we know that

$$\frac{T}{2}\ln\left[\frac{\frac{1}{T}(\mathbf{y}-\hat{\mathbf{F}}\alpha)'(\mathbf{y}-\hat{\mathbf{F}}\alpha)}{\frac{1}{T}(\mathbf{y}-\mathbf{F}\alpha)'(\mathbf{y}-\mathbf{F}\alpha)}\right] = O_p\left(T\min(N,T)^{-1}\right) = o_p(T) \tag{A.5}$$

as long as $N \rightarrow \infty$. Since $C^0_{T,N} - C_{T,N} = o(T)$, (A.2) holds proving the result when $\mathbf{F}$ does not span $\mathbf{F}^0$ but $\mathbf{F}^0$ spans $\mathbf{F}$.

Now, we want to prove that (A.1) holds, when $\mathbf{F}^0$ does not span $\mathbf{F}$ but $\mathbf{F}$ spans $\mathbf{F}^0$. By standard regression analysis we know that if $\mathbf{F}$ spans $\mathbf{F}^0$, then $\mathbf{M}F^0 = 0$. Then,

$$\alpha^{0\prime}\mathbf{F}^{0\prime}\mathbf{M}\mathbf{F}^0\alpha^0 = \epsilon'\mathbf{M}\mathbf{F}^0\alpha^0 = \alpha^{0\prime}\mathbf{F}^{0\prime}\mathbf{M}^0\mathbf{F}^0\alpha^0 = \epsilon'\mathbf{M}^0\mathbf{F}^0\alpha^0 = 0$$

Thus,

$$\frac{T}{2}\ln\left[\frac{\frac{1}{T}\mathbf{y}'\mathbf{M}\mathbf{y}}{\frac{1}{T}\mathbf{y}'\mathbf{M}^0\mathbf{y}}\right] = \frac{T}{2}\ln\left[\frac{\frac{1}{T}\epsilon'\mathbf{M}\epsilon}{\frac{1}{T}\epsilon'\mathbf{M}^0\epsilon}\right]$$

But

$$\ln\left[\frac{\frac{1}{T}\epsilon'\mathbf{M}\epsilon}{\frac{1}{T}\epsilon'\mathbf{M}^0\epsilon}\right] = O_p(T_1^{-1})$$

and therefore

$$\frac{T}{2}\ln\left[\frac{\frac{1}{T}\epsilon'\mathbf{M}\epsilon}{\frac{1}{T}\epsilon'\mathbf{M}^0\epsilon}\right] = O_p(1)$$

As a result

$$\lim_{T\rightarrow\infty} P\left\{\frac{T}{2}\ln\left[\frac{\frac{1}{T}\mathbf{y}'\mathbf{M}\mathbf{y}}{\frac{1}{T}\mathbf{y}'\mathbf{M}^0\mathbf{y}}\right] < C^0_{T,N} - C_{T,N}\right\} = 0 \tag{A.6}$$

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

as long as $C^0_{T,N} - C_{T,N} \to -\infty$. But, (A.5) implies this is not enough for (A.1) to hold. For (A.1) to hold, and given (A.6), we need that

$$\lim_{T\to\infty} P \left\{ \begin{array}{c} \frac{T}{2} \ln \left[\frac{\frac{1}{T}(\mathbf{y}-\hat{\mathbf{F}}\alpha)'(\mathbf{y}-\hat{\mathbf{F}}\alpha)}{\frac{1}{T}(\mathbf{y}-\mathbf{F}\alpha)'(\mathbf{y}-\mathbf{F}\alpha)} \right] + \frac{T}{2} \ln \left[\frac{\frac{1}{T}(\mathbf{y}-\hat{\mathbf{F}}^0\alpha)'(\mathbf{y}-\hat{\mathbf{F}}^0\alpha)}{\frac{1}{T}(\mathbf{y}-\mathbf{F}^0\alpha)'(\mathbf{y}-\mathbf{F}^0\alpha)} \right] \\ < C^0_{T,N} - C_{T,N} + C \end{array} \right\} = 0 \tag{A.7}$$

for all positive finite constants C. But (A.7) holds, if $\frac{C^0_{T,N} - C_{T,N}}{T \min(N,T)^{-1}} \to -\infty$. Given we assume this, the result is proven.

Proof of Theorem 2

The result follows immediately from the proof of Theorem 1 once we note that (A.5) can be replaced by

$$\frac{T}{2} \ln \left[\frac{\frac{1}{T}(\mathbf{y} - \hat{\mathbf{F}}\alpha)'(\mathbf{y} - \hat{\mathbf{F}}\alpha)}{\frac{1}{T}(\mathbf{y} - \mathbf{F}\alpha)'(\mathbf{y} - \mathbf{F}\alpha)} \right] = O_p\left(T q_{NT}\right) = o_p(T). \tag{A.8}$$

B Data Set

The data set used for forecasting are the monthly series from the panel of U.S. indicator series as employed in Stock and Watson (2007), extended by us up to July 2008. Our two dependent variables, PCE inflation and the (effective) federal funds rate, are excluded from this panel. In order to be sure that these predictor variables are $I(0)$, the underlying raw series need to be transformed such that this is the case; generally we employ the same transformation as Stock and Watson (2007), except for the bulk of the nominal series where we follow, e.g., D'Agostino and Giannone (2006) and use first differences of twelve-month transformations of the raw series. Table B.1 summarizes our potential transformations for the raw series.

Hence, we are using as predictor variables the following 106 series, which span the sample January 1959 - July 2008 before the appropriate transformations are applied, and we refer to Stock and Watson (2007) for more details regarding data construction and sources:

Series Y_t	**Transformation:** ***(See Table B.1)***
INDUSTRIAL PRODUCTION INDEX - PRODUCTS, TOTAL	5
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: AVERAGE(MEAN)DURATION IN WEEKS (SA)	2
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