Variable Selection and Inference for Multi-period Forecasting Problems

M. HASHEM PESARAN
ANDREAS PICK
ALLAN TIMMERMANN

An electronic version of the paper may be downloaded
- *from the SSRN website:*   *www.SSRN.com*
- *from the RePEc website:*   *www.RePEc.org*
- *from the CESifo website:*   *www.CESifo-group.org/wp*

Variable Selection and Inference for Multi-period Forecasting Problems

Abstract

This paper conducts a broad-based comparison of iterated and direct multi-step forecasting approaches applied to both univariate and multivariate models. Theoretical results and Monte Carlo simulations suggest that iterated forecasts dominate direct forecasts when estimation error is a first-order concern, i.e. in small samples and for long forecast horizons. Conversely, direct forecasts may dominate in the presence of dynamic model misspecification. Empirical analysis of the set of 170 variables studied by Marcellino, Stock and Watson (2006) shows that multivariate information, introduced through a parsimonious factor-augmented vector autoregression approach, improves forecasting performance for many variables, particularly at short horizons.

M. Hashem Pesaran
Faculty of Economics
Cambridge University
Sidgwick Avenue
Cambridge, CB3 9DD
United Kingdom
mhp1@econ.cam.ac.uk

Andreas Pick
De Nederlandsche Bank and
Faculty of Economics
Cambridge University
Sidgwick Avenue
Uk - Cambridge, CB3 9DD
ap236@econ.cam.ac.uk

Allan Timmermann
University of California San Diego
Rady School of Management
Otterson Hall, Room 4S146
9500 Gilman Dr. #0553
USA - La Jolla, CA 92093-0553
atimmerm@ucsd.edu

1 Introduction

Economists are commonly asked to forecast uncertain outcomes multiple periods ahead in time. For example, in a recession state a policy maker may want to know when the economy will recover and so is interested in forecasts of output growth for, say, horizons of 1, 3, 6, 12, and 24 months. Similarly, fixed-income investors are interested in comparing forecasts of spot rates multiple periods ahead against current long-term interest rates in order to arrive at an optimal investment strategy.

Given the importance of the horizon to many forecasting problems, it is not surprising that a substantial literature has considered the multi-step forecasting problem, see Cox (1961), Brown and Mariano (1989), Clements and Hendry (1998), Findley (1983), Schorfheide (2005), and Ullah (2004), with Bhansali (1999) providing a survey, and Ing (2003) giving asymptotic results for stationary ARs.

Two basic strategies exist for generating multi-period forecasts. The first approach is to estimate a dynamic model for data observed at the highest available frequency, e.g. monthly, and then use the chain rule to generate a forecast at the desired horizon, $h \geq 1$. Under this "iterated" or "indirect" approach, the forecasting specification is the same across all forecast horizons; only the number of iterations changes with h. In the context of univariate specifications, typically autoregressive models are used to this end, while if multivariate specifications are considered, vector autoregressions must be used. The second approach is to estimate a model for the variable measured h-periods ahead as a function of current information. This leads to so-called "direct" forecasts. Under this approach, the forecasting model and its estimates will typically vary across different forecast horizons.

Each approach has advantages and drawbacks. For a given model the iterated approach leads to more efficient estimates since it makes use of data recorded at the highest available frequency and so uses the largest available sample size. Conversely, if the model is misspecified, due, for example, to using an incorrect lag order, iterating the model multiple steps ahead can attenuate any existing biases. Direct forecasts are less efficient, but also more likely to be robust to model misspecification as they are typically linear projections of current realizations on past data. Direct forecasts introduce new problems, however, due to the overlap in data resulting from $h > 1$, which affect the covariance of the forecast errors.

Given such trade-offs, which approach delivers the best forecasts, the direct or the iterated, is therefore ultimately an empirical question. For univariate forecasting problems, as pointed out by Marcellino, Stock and Watson (2006) the performance of the two methods will depend on the sample size, forecast horizon and method used to select lag lengths for the forecasting models. In a multivariate setting it also becomes important how the potentially high-dimensional variable selection search is conducted and how multi-step forecasts of additional predictor variables are generated.

For multivariate specifications of even modest dimension, a global model specification search very rapidly becomes intractable unless the problem is further constrained: with d potential regressors, there are 2^d different linear models and with d easily in the hundreds it is infeasible to evaluate every possible model.

To deal with this dimensionality problem, we propose a factor-augmented VAR approach to iterated forecasting that builds on the work by Bernanke, Boivin and Eliasz (2005) and Stock and Watson (2005). This limits the model specification search to consider inclusion of only a few common factors extracted from different categories of economic variables such as income/output, employment, construction/inventories, interest rates/asset prices and nominal prices/wages. This means that in addition to past values of the predicted variable itself relatively few potential predictors need to be considered.

How close the best forecasting models approximate the "true" data generating process also plays a role since model misspecification can change the relative performance of the iterated and direct forecasts. Finally, we show that how the direct forecasts account for data overlap can affect the results. We propose modifications to existing model selection methods that account for such data overlap.

Empirically, we confirm the finding reported by Marcellino, Stock and Watson (2006) that the iterated forecasts are best overall among the univariate forecasting methods. However, we also find that forecasts generated by the factor-augmented VARs generally perform better than the univariate forecasts, the important exception being variables tracking prices and wages. This suggests that it is helpful to extend the forecasting models beyond purely univariate schemes and include multivariate information.

The main contributions of the paper are as follows: First, for simple dynamic models such as the AR(1) specification we present a new theoretical result establishing conditions under which both the iterated and the direct forecasts are unconditionally unbiased. A key condition turns out to be symmetry of the innovations. Second, we extend existing model selection criteria such as the AIC and the BIC to account for the overlap introduced by the direct forecasting method which affects the covariance of the forecast errors and so can lead to different models being chosen in small samples. Third, we undertake a comprehensive study of model selection methods and shed light, both through Monte Carlo simulations and through empirical results, on which approach works best in the recursive modeling of a variety of economic variables. We consider univariate models as well as multivariate factor augmented vector autoregressive specifications hitherto not used in the context of direct and indirect multi-period forecast comparisons.

The outline of the paper is as follows. Section 2 discusses the theoretical trade-offs associated with direct versus iterated forecasts. Section 3 sets up the forecasting problem for the univariate and multivariate case, while Sec-

3

tion 4 deals with the model selection issue. Section 5 presents some Monte Carlo results, while Section 6 describes our empirical findings using the data originally analyzed by Marcellino, Stock and Watson (2006). Section 7 concludes. Technical results are provided in appendices.

2 Comparison of Direct and Iterated Forecasts for the AR(1) Model

In this section we introduce the iterated and direct forecasting approaches in the context of the simplest possible dynamic model, namely an AR(1) specification. Next, we establish a new result showing that both approaches lead to forecasts that are unconditionally unbiased provided that the innovations to the underlying data generating process are symmetrically distributed. Finally, we characterize the relative efficiency of the two approaches in finite samples.

The analysis is conducted under the assumption that a stationary AR(1) model of the following form is the correct specification:

$$y_t = a + \phi y_{t-1} + \varepsilon_t, \quad |\phi| < 1, \ \varepsilon_t \sim iid(0, \sigma_\varepsilon^2). \tag{1}$$

Letting $\mu = a/(1 - \phi)$, we can use the equivalent representation

$$
\begin{aligned}
y_t &= \mu + u_t, \quad \text{where} \\
u_t &= \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}, \ \ b_i = \phi^i.
\end{aligned}
\tag{2}
$$

Alternatively, for purposes of the direct forecasting approach that projects y_t on y_{t-h}, we can write

$$
\begin{aligned}
y_t &= a \left(\frac{1 - \phi^h}{1 - \phi} \right) + \phi^h y_{t-h} + v_t, \\
&\equiv a_h + \phi_h y_{t-h} + v_t, \quad \text{where} \\
v_t &= \sum_{j=0}^{h-1} \phi^j \varepsilon_{t-j}.
\end{aligned}
\tag{3}
$$

Notice that v_t will follow an $MA(h-1)$ process even when ε_t is serially uncorrelated due to the data overlap resulting from $h > 1$.

2.1 Iterated and Direct Approach

Suppose that we are interested in forecasting y_{T+h} based on information up to period T and that the $T + h$ observations y_t, $t = -h+1, -h+2, \ldots, T$ are available for model estimation. Denote the h-step ahead forecast of y_{T+h} by

$\hat{y}_{T+h}$ if the iterated procedure is used, and by $\tilde{y}_{T+h}$ if the direct method is employed.

Under the iterated procedure

$$\hat{y}_{T+h} = \hat{a}_T \left(\frac{1 - \hat{\phi}_T^h}{1 - \hat{\phi}_T} \right) + \hat{\phi}_T^h y_T, \tag{4}$$

where $\hat{a}_T$ and $\hat{\phi}_T$ are the estimators of a and ϕ obtained from the OLS regression (1) of y_t on an intercept and y_{t-1} using the observations y_t, $t = -h+1, -h+2, \ldots, T$.

The direct forecast of y_{T+h} is given by

$$\tilde{y}_{T+h} = \tilde{a}_{h,T} + \tilde{\phi}_{h,T}\, y_T, \tag{5}$$

where $\tilde{a}_{h,T}$ and $\tilde{\phi}_{h,T}$ are the OLS estimators of a_h and ϕ_h obtained from the regression (3) of y_t on an intercept and y_{t-h} using the same sample observations, y_t, $t = -h+1, -h+2, \ldots, T$.

Specifically, under the iterated method,

$$(T + h - 1)^{-1} \sum_{t=-h+2}^{T} y_t = \hat{a}_T + \hat{\phi}_T (T + h - 1)^{-1} \sum_{t=-h+2}^{T} y_{t-1},$$

or, equivalently,

$$\bar{y}_{h:T} = \hat{a}_T + \hat{\phi}_T\, \bar{y}_{h:T,-1}, \tag{6}$$

where

$$\hat{\phi}_T = \frac{\sum_{t=-h+2}^{T} y_t (y_{t-1} - \bar{y}_{h:T,-1})}{\sum_{t=-h+2}^{T} (y_{t-1} - \bar{y}_{h:T,-1})^2}.$$

Similarly, for the direct method we have

$$T^{-1} \sum_{t=1}^{T} y_t = \tilde{a}_{h,T} + \tilde{\phi}_{h,T} T^{-1} \sum_{t=1}^{T} y_{t-h},$$

or, equivalently (noting that $\bar{y}_T = \bar{y}_{1:T}$),

$$\bar{y}_T = \tilde{a}_{h,T} + \tilde{\phi}_{h,T} \bar{y}_{T,-h}, \tag{7}$$

where

$$\tilde{\phi}_{h,T} = \frac{\sum_{t=1}^{T} y_t (y_{t-h} - \bar{y}_{T,-h})}{\sum_{t=1}^{T} (y_{t-h} - \bar{y}_{T,-h})^2}. \tag{8}$$

We next establish conditions under which both the direct and indirect forecasts are unconditionally unbiased:

Proposition 1 *Suppose data is generated by the stationary AR(1) process, (1) and define the h-step ahead forecast errors from the iterated and direct methods,*

$$\hat{e}_{T+h} = y_{T+h} - \hat{y}_{T+h},$$

and

$$\tilde{e}_{T+h} = y_{T+h} - \tilde{y}_{T+h},$$

where the iterated h-step forecast, $\hat{y}_{T+h}$, and the direct forecast, $\tilde{y}_{T+h}$, are given, for example, by (4) and (5), respectively. Assume that u_t and v_t, defined in (2) and (3), are symmetrically distributed around zero, have finite second-order moments and expectations of $\hat{\phi}_T$ and $\tilde{\phi}_{h,T}$ exist. Then for any finite T and h we have

$$\mathrm{E}(\hat{e}_{T+h}) = \mathrm{E}(\tilde{e}_{T+h}) = 0.$$

A proof of this proposition is provided in Appendix A.

The result is surprising since neither set of estimators $(\hat{a}_T, \hat{\phi}_T)$ and $(\tilde{a}_{h,T}, \tilde{\phi}_{h,T})$ is unbiased. However, the main intuition behind the results follows from the fact that, under the stated assumptions,

$$\mathrm{E}(\hat{\mu}_T - \mu) = 0,$$

and

$$\mathrm{E}\left(\frac{\hat{a}_T}{1 - \hat{\phi}_T}\right) = \frac{a}{1 - \phi}.$$

The proposition generalizes the known result in the literature for $h = 1$ established, for example, by Fuller (1996) to multi-step ahead forecasts. For $h = 1$, Pesaran and Timmermann (2005) also show that forecast errors are unconditionally unbiased for symmetrically distributed error processes even in the presence of breaks in the autocorrelation coefficient, ϕ, so long as μ is stable over the estimation sample.

Notice that the result holds only unconditionally and that the direct and iterated forecasts are not typically conditionally unbiased for a given value of $y_T \neq \mu$, as noted by Phillips (1987).

In comparing the iterated and direct forecasts it is worth noting that when ϕ is positive and not too close to unity, for even moderately large values of h, $\hat{\phi}_T^h \approx 0$, since $\hat{\phi}_T < \phi$. It follows that in such cases, $\hat{e}_{T+h} = (\mu - \hat{\mu}_T) + v_{T+h} + o(\phi^h)$. Similarly, $\tilde{e}_{T+h} = -\bar{v}_T + v_{T+h} + o(\phi^h)$. Hence, for h moderately large and ϕ not too close to the unit circle, a measure of the relative efficiency of the two forecasting approaches can be obtained:

$$\frac{\mathrm{E}(\hat{e}_{T+h}^2)}{\mathrm{E}(\tilde{e}_{T+h}^2)} = \frac{\mathrm{E}(\hat{\mu}_T - \mu)^2 + \mathrm{E}(v_{T+h}^2) + o(\phi^h)}{\mathrm{E}(\bar{v}_T^2) + \mathrm{E}(v_{T+h}^2) + o(\phi^h)}, \tag{9}$$

6

since $(\mu - \hat{\mu}_T)$ and $\bar{v}_T$ are uncorrelated with v_{T+h}. But $E(\mu - \hat{\mu}_T)^2 = O(T^{-1})$ and does not depend on h. To derive $E(\bar{v}_T^2)$, recall that $v_t = \sum_{j=0}^{h-1} \phi^j \varepsilon_{t-j}$, and hence after some algebra

$$T \; \bar{v}_T = \sum_{j=1}^{h-1} \left(\frac{1 - \phi^j}{1 - \phi} \right) \phi^{h-j} \varepsilon_{-h+j+1} + \left(\frac{1 - \phi^h}{1 - \phi} \right) \sum_{t=1}^{T-h+1} \varepsilon_t + \sum_{j=1}^{h-1} \left(\frac{1 - \phi^j}{1 - \phi} \right) \varepsilon_{T-j+1},$$

and

$$E(\bar{v}_T^2) = \text{Var}(\bar{v}_T) = \frac{\sigma_\varepsilon^2}{T^2} \left\{ \sum_{j=1}^{h-1} \left(\frac{1 - \phi^j}{1 - \phi} \right)^2 \left[1 + \phi^{2(h-j)} \right] + \left(\frac{1 - \phi^h}{1 - \phi} \right)^2 (T - h + 1) \right\}.$$

Clearly, $E(\bar{v}_T^2) = O(T^{-1})$ if h is fixed. But it is easily seen that we continue to have $E(\bar{v}_T^2) = O(T^{-1})$ even if $h \to \infty$ so long as $h/T \to \kappa$, where κ is a fixed finite fraction in the range $[0, 1)$. Therefore,

$$\frac{E(\hat{e}_{T+h}^2)}{E(\tilde{e}_{T+h}^2)} = 1 + O(T^{-1}) + o(\phi^h),$$

and for T sufficiently large there will be little to choose between the iterated and the direct procedures.

This analysis suggests two points. First, we should expect to find the greatest difference between the performance of the two forecasting methods in small samples (T) or in situations where h is large, i.e. when h/T is large. Second, in the context of correctly specified dynamic models there is little reason to choose one approach over the other in large samples or when h/T is small.

To illustrate these results, Figure 1 plots the ratio of RMSE values of the iterated over the direct forecasting approaches for $h = 1, 2, \ldots, 12$, assuming $\phi = 0.9$ and using sample sizes of $T = 50$, 100, 500, and 1000. The relative efficiency of the iterated approach is very pronounced when T is small or h is large. This is to be expected since the estimation error associated with the direct approach is larger in this situation. Notice also that the advantage of the iterated approach rapidly gets reduced as the sample size is expanded or the forecast horizon is shortened.

In line with the existing literature, the above results conditioned on a given model, namely the AR(1) specification in (1), while ignoring the effect of model misspecification and model uncertainty. It is not clear how to obtain finite-sample analytical results without conditioning on a particular model specification. To deal with this important issue, we next discuss the types of forecasting models under consideration and then turn to a variety of methods for model selection.

Figure 1: Ratios of RMSFE of iterated over direct forecast
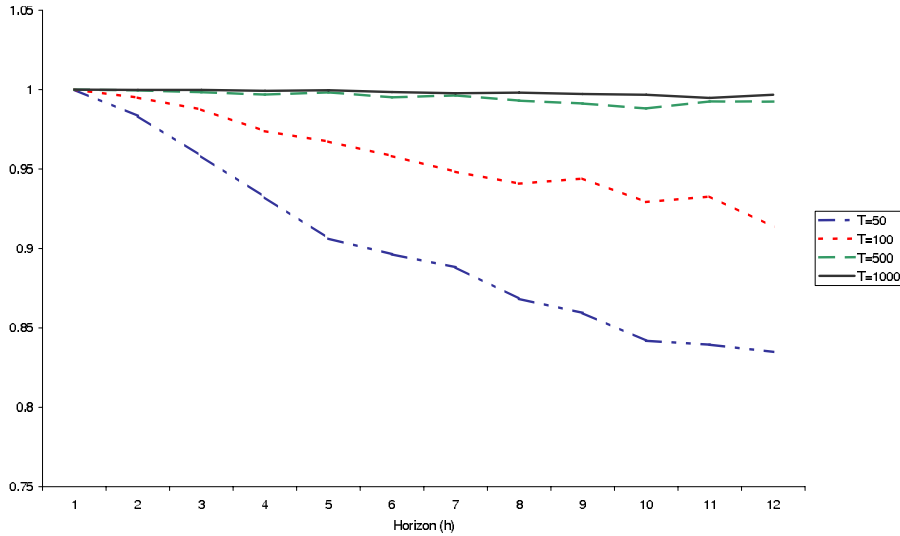
3 Methods for Multi-period Forecasting

Suppose a forecaster is interested in predicting a $k \times 1$ vector of target variables $\boldsymbol{y}_t = (y_{1t}, y_{2t}, \ldots, y_{kt})'$ by means of their own past values and the past values of an additional set of potentially relevant predictor variables, $\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Mt})'$. Typically k is small, often 1 or 2, but M could be large.

For purposes of calculating one-step-ahead forecasts under the iterated approach, the regressors are treated as conditional information and so how they are generated is not a concern. This also holds under the direct forecasting approach irrespective of the forecast horizon. In contrast, when applying the iterated forecasting approach to multi-period horizons, $h > 1$, the regressors themselves need to be predicted since such values in turn are required to predict future values of $\boldsymbol{y}$.

3.1 Factor-Augmented VARs

In cases where M is relatively small, say less than ten, one approach is to treat all variables simultaneously, i.e. model $\mathbf{z}_t = (\boldsymbol{y}_t', \boldsymbol{x}_t')'$. Multi-period forecasts of y_t can then be obtained iteratively using a conventional VAR of the form

$$\boldsymbol{z}_t = \boldsymbol{\mu}_z + \begin{pmatrix} \boldsymbol{A}_p(L) & \boldsymbol{B}_q(L) \\ \boldsymbol{C}_r(L) & \boldsymbol{D}_s(L) \end{pmatrix} \boldsymbol{z}_{t-1} + \boldsymbol{\xi}_t, \tag{10}$$

where p and q are the lag order of $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ in the equation for $\boldsymbol{y}_t$ and r and s is the lag order of $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ in the equation for $\boldsymbol{x}_t$.

However, in the common situation where the time-series dimension of the data is limited while M is large, this approach is unlikely to be successful due to the high dimension of the coefficient matrices of the VARs, i.e. $\boldsymbol{C}_r(L)$ and $\boldsymbol{B}_q(L)$ in particular.

To deal with this issue, a conditional factor-augmentation approach can be used. In this approach, the large-dimensional $\boldsymbol{x}_t$-vector is condensed into a subset of factors, $\hat{\boldsymbol{f}}_t$, of dimension $m < M$, used to summarize the salient features of the larger-dimensional data. A factor-augmented VAR based on the variables $\tilde{\boldsymbol{z}}_t = (\boldsymbol{y}_t', \hat{\boldsymbol{f}}_t')'$ can then be used:

$$\tilde{\boldsymbol{z}}_t = \boldsymbol{\mu}_z + \begin{pmatrix} \boldsymbol{A}_p(L) & \boldsymbol{C}_q(L) \\ \boldsymbol{0} & \boldsymbol{D}_s^*(L) \end{pmatrix} \tilde{\boldsymbol{z}}_{t-1} + \boldsymbol{\xi}_t. \tag{11}$$

Notice the asymmetric treatment of $\boldsymbol{y}_t$ and $\hat{\boldsymbol{f}}_t$ under this approach. Due to the zero in the lower left corner of the lag polynomial matrix pre-multiplying $\tilde{\boldsymbol{z}}_{t-1}$, future values of the factors are generated using only current and past values of the factors themselves. The y-variables are therefore not used to predict the factors, while the factors are used to predict the y-variables.

Only those factors that help predict $\boldsymbol{y}_t$ are relevant and should be included in this model. This may be a subset of the full set of factors under consideration, which we denote by $\hat{\boldsymbol{f}}_t^*$. One then has to choose whether to use the full set of factors $\hat{\boldsymbol{f}}_t$ to predict the subset of factors, $\hat{\boldsymbol{f}}_t^*$, selected when forecasting $\boldsymbol{y}_{t+h}$, or whether to use only lagged values of $\hat{\boldsymbol{f}}_t^*$ to predict their future values. The choice would depend on the number of available factors. In the empirical application below where we consider five factors we shall be using all the factors in forecasting future values of $\hat{\boldsymbol{f}}_t^*$.

3.2 Multi-step Forecasts with Dynamic Factor Models

Assume that the $k \times 1$ vector of target variables, $\boldsymbol{y}_t$, is generated according to the following dynamic factor model

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \boldsymbol{f}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{12}$$

$$\boldsymbol{\varepsilon}_t \sim \text{iid}(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{f}_t$ is a vector of unobserved common factors, while $\boldsymbol{A}$, $\boldsymbol{\Gamma}_1$, and $\boldsymbol{\Sigma}$ are unknown coefficient matrices.

In order to forecast $\boldsymbol{y}_{T+h}$ given information at time T we also need to forecast the factors. Despite the dynamic nature of the above model, we follow Stock and Watson (2002) and estimate $\boldsymbol{f}_t$ by the principal component (PC) procedure, although one could equally employ the dynamic factor approach of Forni et al. (2005). The key question when using factor models is the choice of the number of factors. This can be determined, for example, using the information criteria (IC) proposed by Bai and Ng (2002). In practice there is a considerable degree of uncertainty surrounding the number of

factors to be used, and as pointed out by Bai and Ng (2009), the IC assume that the factors are ordered as predictors of the regressors $\mathbf{x}_t$, an ordering that might not be appropriate for predicting the target variables, y_t. The resultant factors are also often difficult to associate with readily understood economic concepts.

In view of these concerns we adopt a hierarchical approach where we first divide the 170 variables into five economically distinct groups and then select the first PC from each of the categories. All the computations are carried out recursively, so that no future information is used in the construction of the factors. We denote the recursively estimated PC's by $\hat{\boldsymbol{f}}_t(T-w+1:T)$, where $T \geq t$ is the time at which factors are computed and w is the size of the estimation window used to do the computations. To simplify notations we denote $\hat{\boldsymbol{f}}_t(T-w+1:T)$ by $\hat{\boldsymbol{f}}_t$, although strictly speaking both subscripts are needed for appropriate timing when carrying out the computations.[1]

3.3 Iterated System Forecasts

For multivariate forecasts we first select the relevant subset of factors and the lag lengths p and q using the conditional model:

$$\boldsymbol{y}_t = \boldsymbol{\mu}_y + \boldsymbol{A}_p(L)\boldsymbol{y}_{t-1} + \boldsymbol{C}_q(L)\hat{\boldsymbol{f}}_{t-1} + \boldsymbol{u}_t, \tag{13}$$

where

$$\begin{aligned} \boldsymbol{A}_p(L) &= \mathbf{A}_0 + \mathbf{A}_1 L \ldots + \mathbf{A}_{p-1} L^{p-1}, \\ \boldsymbol{C}_q(L) &= \boldsymbol{C}_0 + \boldsymbol{C}_1 L + \ldots + \boldsymbol{C}_{q-1} L^{q-1}. \end{aligned}$$

We determine the lag orders p and q and the subset of m^* factors from the total set of m factors by applying information criteria to the likelihood of $\boldsymbol{y}_t$.

For simplicity, we obtain forecasts only from the full set of m factors. Hence we first select a VAR(s) model in $\hat{\boldsymbol{f}}_t$ so the value of s will be determined by AIC or BIC, again applied recursively. Then

$$\hat{\boldsymbol{f}}_t = \boldsymbol{\mu}_f + \boldsymbol{D}_s(L)\hat{\boldsymbol{f}}_{t-1} + \boldsymbol{\nu}_t, \tag{14}$$

where

$$\boldsymbol{D}_s(L) = \boldsymbol{D}_0 + \mathbf{D}_1 L + \ldots + \mathbf{D}_{s-1} L^{s-1}.$$

To compute $h-$step-ahead forecasts of $\boldsymbol{y}_t$, the conditional and marginal models need to be combined:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\mu}_y + \boldsymbol{A}_p(L)\boldsymbol{y}_{t-1} + \boldsymbol{C}_q^*(L)\hat{\boldsymbol{f}}_{t-1} + \boldsymbol{u}_t, \\ \hat{\boldsymbol{f}}_t &= \boldsymbol{\mu}_f + \boldsymbol{D}_s(L)\hat{\boldsymbol{f}}_{t-1} + \boldsymbol{\nu}_t, \end{aligned}$$

---

[1]In the case of an expanding estimation window (with $w = T$) one would set $T = t$ and work with $\hat{\boldsymbol{f}}_t(1:t)$.

or, defining $\boldsymbol{z}_t = (\boldsymbol{y}'_t, \hat{\boldsymbol{f}}'_t)'$ and $\boldsymbol{\xi}_t = (\boldsymbol{u}'_t, \boldsymbol{\nu}'_t)'$, we have

$$\boldsymbol{z}_t = \left(\begin{array}{c} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_f \end{array} \right) + \left(\begin{array}{cc} \boldsymbol{A}_p(L) & \boldsymbol{C}^*_q(L) \\ \boldsymbol{0} & \boldsymbol{D}_s(L) \end{array} \right) \boldsymbol{z}_{t-1} + \boldsymbol{\xi}_t. \qquad (15)$$

where the selection of factors in the conditional model (13) is reflected in zero-restrictions in $\boldsymbol{C}^*_1, \boldsymbol{C}^*_2, \dots \boldsymbol{C}^*_q$, where the columns corresponding to the factors that are not selected are set to zeros. The factor-augmented VAR in $\boldsymbol{z}_t$, (15), can then be iterated forward.[2]

3.4 Univariate Forecasts

Univariate forecasts are a special case of the multivariate forecasts described above. Nevertheless, it is worth briefly clarifying how the forecasts are computed for these models.

For example, with p lags the iterated forecast of y_{T+h} given information at time T is

$$\hat{y}_{T+h} = \hat{\alpha} + \hat{\beta}_1 \hat{y}_{T+h-1} + \hat{\beta}_2 y_{T+h-2} + \dots + \hat{\beta}_p y_{T+h-p},$$

where $\hat{y}_{T+h-1} = y_{T+h-j}$ if $h \leq j$ and the parameters are estimated from the regression model

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \varepsilon_t,$$

$$t = T - w + 1 + p, T - w + 2 + p, \dots, T.$$

Under the direct approach, for a given lag length, p, the parameters are estimated from the regression

$$y_t = \alpha + \beta_1 y_{t-h} + \beta_2 y_{t-h-1} + \dots + \beta_p y_{t-h-p} + \varepsilon_t,$$

and the forecast is

$$\hat{y}_{T+h} = \hat{\alpha} + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1} + \dots + \hat{\beta}_p y_{T-p}.$$

4 Model Selection

Given the large dimension of the set of potentially relevant predictor variables, how a particular model is chosen by the forecaster is of great importance. We consider two basic model selection criteria, namely AIC and BIC. Both criteria are commonly used in forecasting studies and have well known properties: AIC achieves a good approximate model as the sample

---

[2]Alternatively, one could model only the factors that have been selected in the conditional model (13). This may be less efficient than using the full VAR in (14), but a smaller number of parameters needs to be estimated from a finite number of observations. It is therefore not clear *a priori* which approach will perform better.

size expands even if the true model is not contained in the universe of models under consideration. However, AIC is not a consistent criterion in the sense that, asymptotically, it does not select the true model with probability one if it happens to be included in the search. In contrast, BIC is a consistent model selection criterion.

The iterated forecasting models can be selected either on the basis of single equations or using a system of equations. For the direct forecasting models a decision has to be made whether or not to correct for the overlap in the observations that affects the sample covariance matrix of the forecast errors, although such corrections will be of second order importance in large samples.

4.1 Iterated Forecasts

Models used to generate the iterated forecasts are selected based on the following criteria. First, we consider applying AIC or BIC to the single equation ($k = 1$) containing the variable of interest, i.e. for AIC:

$$\text{AIC} = \ln[\hat{\boldsymbol{u}}_i' \hat{\boldsymbol{u}}_i/(T-1)] + \frac{2d}{T-1}, \tag{16}$$

where $\hat{\boldsymbol{u}}_i$ is a $(T-1) \times 1$ vector of estimated residuals with typical element

$$\hat{u}_{it} = y_{1,t} - \hat{\boldsymbol{\beta}}_{iT}' \boldsymbol{x}_{t-1}, \qquad t = 2, 3, \ldots, T,$$

while the $d \times 1$ vector $\boldsymbol{x}_{t-1}$ contains subset i of y_{1t-1} and $\boldsymbol{f}_{t-1}$ and their lags and $\hat{\boldsymbol{\beta}}_T$ is estimated using observations up to T. Similarly, for BIC,

$$\text{BIC} = \ln[\hat{\boldsymbol{u}}_i' \hat{\boldsymbol{u}}_i/(T-1)] + \frac{\ln(T-1)d}{T-1}, \tag{17}$$

where $\hat{\boldsymbol{u}}_i$ is defined as above.

For $k > 1$, we have

$$\text{AIC} = \ln|\hat{\boldsymbol{\Sigma}}_i| + \frac{2kd}{T-1} \tag{18}$$

where $\hat{\boldsymbol{\Sigma}}$ is the $k \times k$ estimated error covariance matrix

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{T-1}(\boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{Y}),$$

$\boldsymbol{Y}$ is the $T - 1 \times k$ matrix of stacked $\boldsymbol{y}_t = (y_{1t}, y_{2t}, \ldots, y_{kt})'$, and $\boldsymbol{X}_i$ is the observation matrix formed from stacking $\boldsymbol{x}_{it-1}$ over the observations, $t = 2, 3, \ldots, T$. Similarly,

$$\text{BIC} = \ln|\hat{\boldsymbol{\Sigma}}_i| + \frac{\ln(T-1)kd}{T-1}. \tag{19}$$

4.2 Direct Forecasts

The direct forecasting models are selected based on either AIC

$$\text{AIC} = \ln[\hat{\boldsymbol{u}}_h' \hat{\boldsymbol{u}}_h / (T - h)] + \frac{2d}{T - h} \qquad (20)$$

where $\hat{\boldsymbol{u}}_h$ is a $(T - h) \times 1$ vector of estimated residuals with typical element

$$\hat{u}_{it} = y_{1t} - \hat{\boldsymbol{\beta}}_{iT}' \boldsymbol{x}_{t-1}, \quad t = h + 1, \dots, T,$$

or, using BIC,

$$\text{BIC} = \ln[\hat{\boldsymbol{u}}_h' \hat{\boldsymbol{u}}_h / (T - h)] + \frac{\ln(T - h)d}{T - h}, \qquad (21)$$

where $\hat{\boldsymbol{u}}_h$ is defined as above.

For $h > 1$ the overlap of the forecasts will produce considerable autocorrelation in the residuals. This should be accounted for in small samples when calculating the information criteria. Here we consider two ways to calculate the corrections with further details given in Appendix B. The first approach is to make use of the modified information criteria defined by

$$
\begin{aligned}
\text{AIC}_\Pi &= \ln[\hat{\boldsymbol{u}}_h' \hat{\boldsymbol{u}}_h / (T - h)] + \frac{2\text{tr}\left(\hat{\boldsymbol{\Pi}}_S\right)}{T - h} \\
\text{BIC}_\Pi &= \ln[\hat{\boldsymbol{u}}_h' \hat{\boldsymbol{u}}_h / (T - h)] + \frac{\ln(T - h)\text{tr}\left(\hat{\boldsymbol{\Pi}}_S\right)}{T - h},
\end{aligned}
\qquad (22)
$$

using

$$\hat{\boldsymbol{\Pi}}_S = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{S}\boldsymbol{X}/h$$

where $\boldsymbol{S}$ is a matrix with h on the diagonal, $h-1$ on the first diagonal above and below the main diagonal, $h-2$ on the second diagonal above and below the main diagonal, and so on, i.e.

$$
\boldsymbol{S} = \begin{pmatrix}
h & h-1 & h-2 & \dots & & 0 \\
h-1 & h & h-1 & & \dots & 0 \\
h-2 & h-1 & h & & & \vdots \\
\vdots & & & \ddots & h-1 & h-2 \\
0 & \dots & & h-1 & h & h-1 \\
0 & \dots & & h-2 & h-1 & h
\end{pmatrix}.
$$

This formulation aims at capturing the $MA(h-1)$ form of the error process of the overlapping regressions but assumes that the serial correlation of the underlying (non-overlapping) observations is negligible.

The second approach uses an estimated covariance matrix. In particular, we use the Newey-West (1987) covariance matrix to obtain the correction. This yields the modified AIC and BIC:

$$
\begin{aligned}
\text{AIC}_{\tilde{\Pi}} &= \ln[\hat{\boldsymbol{u}}_h'\hat{\boldsymbol{u}}_h/(T-h)] + \frac{2\text{tr}\left(\tilde{\boldsymbol{\Pi}}\right)}{T-h} \qquad\qquad (23) \\
\text{BIC}_{\tilde{\Pi}} &= \ln[\hat{\boldsymbol{u}}_h'\hat{\boldsymbol{u}}_h/(T-h)] + \frac{\ln(T-h)\text{tr}\left(\tilde{\boldsymbol{\Pi}}\right)}{T-h},
\end{aligned}
$$

where $\tilde{\boldsymbol{\Pi}} = \hat{\boldsymbol{\Sigma}}_{xx}^{-1}\hat{\boldsymbol{\Omega}}$, $\hat{\boldsymbol{\Sigma}}_{xx} = \hat{\sigma}^2\left(\frac{\boldsymbol{X}'\boldsymbol{X}}{T-h}\right)$, $\hat{\boldsymbol{\Omega}}$ is the long-run variance of the residuals as estimated by the Newey-West covariance matrix with bandwidth set to $\min(h, T^{1/3})$.

5 Monte Carlo Simulations

Having described the forecasting setup and the model selection methods, we next turn to Monte Carlo simulation as a means to evaluate the performance of the various model selection and estimation approaches under two data generating processes (DGPs). Both data generating processes assume the following VAR model for $\boldsymbol{y}_t = (y_{1t}, y_{2t}, \ldots, y_{kt})'$:

$$
\boldsymbol{y}_t = \boldsymbol{\mu} + \alpha\boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{H}_\varepsilon\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{I}_k). \qquad (24)
$$

The goodness of fit is controlled by the scaling parameter α, which is set such that the population $R^2 = 0.2$ or $R^2 = 0.8$, representing low and high predictability scenarios, respectively. An alternative would be to vary $\boldsymbol{H}_\varepsilon$. However, for the purpose of controlling the goodness of fit of the equations and given the choices to be considered for $\mathbf{A}$, it has proved more convenient to vary α and set $\mathbf{H}_\varepsilon = \boldsymbol{I}_k$. Data are generated for samples of $T = 60$,120, and 240 observations which in turn are used to generate forecasts for period $T+h$, so T corresponds to w, the length of a rolling window used for estimation. We consider forecast horizons of $h = 1, 3, 6, 12$, and 24 periods.

The first DGP assumes that all variables contain useful information for predicting the variable of interest (always the first variable) and so the one-step-ahead forecast should select all variables. Under the second DGP, iterated multi-step forecasts can be expected to be inefficient because they select models that produce good one-step-ahead forecasts and a variable in the DGP is only helpful for longer horizon forecasts.

Data Generating Process 1

Under the first DGP, we set $k = 3$ with $\boldsymbol{\mu} = \boldsymbol{0}$, and choose the $\boldsymbol{A}$-matrix as follows

$$\boldsymbol{A} = \begin{pmatrix} 0.8 & 0.5 & 0.5 \\ 0 & 0.95 & 0 \\ 0 & 0 & 0.7 \end{pmatrix},$$

so that both y_2 and y_3 help predict y_1, but y_{2t} and y_3 are in turn not themselves predictable by means of past values of y_1. Moreover, y_2 is highly persistent while y_3 is not, suggesting that for large values of h in particular, y_2 should play more of a role in forecasting y_1 as compared to y_3.

Data Generating Process 2

The parameters for the second DGP are set to $k = 4$, $\boldsymbol{\mu} = \boldsymbol{0}$, and

$$\boldsymbol{A} = \begin{pmatrix} 0.8 & 0.5 & 0 & 0 \\ 0 & 0.8 & 0.5 & 0 \\ 0 & 0 & 0.9 & 0.5 \\ 0 & 0 & 0 & 0.95 \end{pmatrix},$$

Notice that y_2 helps predict y_1, but y_2 is in turn not itself predictable by means of past values of y_1. Moreover, y_3 neither predicts nor is predicted by y_1 but y_3 predicts y_2 and therefore may help predict y_1 over medium horizons. Finally, the most persistent variable, y_4, indirectly helps predict y_1 through its ability to predict y_3.

5.1 Forecasts

We generate forecasts from both univariate and multivariate models. The univariate forecasts are based on AR models with lag length up to $p_{\max} = 12$. The multivariate models consider all regressors in the DGPs with the maximum lag length restricted to $p_{\max} = 2$.

In each case forecasts are based on the model selected by one of the criteria discussed in the previous section. Iterated forecasts are then calculated as follows:

$$\hat{\boldsymbol{Z}}_{T+h}^* = \hat{\boldsymbol{\mu}}(\boldsymbol{I} + \hat{\boldsymbol{A}} + \cdots + \hat{\boldsymbol{A}}^{h-1}) + \hat{\boldsymbol{A}}^h \boldsymbol{z}_T^*, \tag{25}$$

where $\boldsymbol{z}_t^*$ is the sub-vector of $\boldsymbol{z}_t$ chosen in the model selection step, $\boldsymbol{Z}_t^* = \boldsymbol{z}_t^*$ if $\boldsymbol{z}_t^*$ includes y_{1t} or $\boldsymbol{Z}_t^* = (y_{1t}, \boldsymbol{z}_t^*)$ if $\boldsymbol{z}_t^*$ does not include y_{1t}, and $\hat{\boldsymbol{\mu}} = \boldsymbol{0}$ if the intercept is not selected.

Direct forecasts are obtained from

$$\tilde{y}_{1,T+h} = \hat{\boldsymbol{\beta}}_h' \boldsymbol{x}_{s,T}, \tag{26}$$

where $\boldsymbol{x}_{s,T}$ is the subset of regressors that are selected, i.e., $\boldsymbol{x}_{s,T} = \boldsymbol{z}_T^*$ if the intercept is not selected or $\boldsymbol{x}_{s,T} = (1, \boldsymbol{z}_T^*)$ if the intercept is selected.

The forecast errors are calculated as

$$
\begin{aligned}
\hat{e}_{T+h} &= y_{1,T+h} - \hat{y}_{1,T+h} \quad \text{or} \\
\tilde{e}_{T+h} &= y_{1,T+h} - \tilde{y}_{1,T+h}
\end{aligned}
\tag{27}
$$

depending on whether the iterated or direct forecasts are used. Finally, the forecasting performance is measured by the mean squared forecast error (MSFE) computed as

$$
\text{MSFE} = \frac{1}{R} \sum_{r=1}^{R} e_{T+h,r}^2,
\tag{28}
$$

where $r = 1, 2, \ldots, R$ denotes the Monte Carlo replications and e_{T+h} is either $\hat{e}_{T+h}$ or $\tilde{e}_{T+h}$.

5.2 Summary of Results

Results from the Monte Carlo simulations are reported in Tables 1 and 2. In order to study how the degree of predictability affects the findings, each table contains a panel with $R^2 = 0.2$ and a panel with $R^2 = 0.8$. The former is closer to the empirical results that we obtain, while the second scenario is more relevant for some of the highly persistent variables.

First consider the results for the data generated under DGP 1. For both the univariate models and the VARs, the iterated approach dominates the direct approach. This is a robust finding that holds across sample sizes ($T = 60$, 120, and 240), information criteria (AIC versus BIC), and forecast horizons ($h = 3, 6, 12$, and 24). As expected the relative performance of the direct to the iterated forecasts improves with the sample size, T, since it becomes less costly to use an inefficient estimation method in the larger samples.

Turning to a comparison of the univariate and multivariate forecasts, the picture is less clear-cut. In many cases the best overall forecast is generated by the VAR models, although this depends on the degree of predictability assumed for the DGP. Assuming weak predictability ($R^2 = 0.2$), under the iterated approach VAR-models selected by the simple AIC or BIC generally produce better forecasts than their univariate counterparts. This holds across all forecast horizons when $T = 60$ or $T = 120$. The performance of the VAR forecasts gets even better relative to the univariate forecasts when the level of predictability is raised by setting $R^2 = 0.8$. In this case the VAR-forecasts generate lower MSFE-values than the corresponding univariate forecasts for the vast majority of horizons and sample sizes. These results suggest that the multivariate VAR forecasts perform relatively better when the forecast signal is strong and it becomes easier to correctly identify the relevant predictor variables.

The modifications to the AIC and BIC work well in many cases for $h \geq 6$ as the data overlap becomes more pronounced. Interestingly, the

band diagonal approach is generally more successful at reducing the MSFE-values than the Newey-West approach.

Results for the second DGP are presented in Table 2. While the results do not change fundamentally in the case with low predictability ($R^2 = 0.2$), under high predictability ($R^2 = 0.8$) the direct forecasting approach applied to the multivariate models now begins to dominate in the larger estimation samples ($T = 120$ or 240) at the medium-term horizons ($h = 3$ and 6 periods). The iterated approach continues to dominate, however, at the longest horizons ($h = 12$ and 24) and for the smallest estimation sample ($T = 60$). It also dominates at all horizons for the univariate models. These results are easy to explain since the deterioration in the (relative) performance of the direct forecasting approach can be expected to be greatest when fewer data points are available due either to a small sample size or due to a long horizon which leads to greater estimation errors under this approach.

6   Empirical Results

In their analysis, Marcellino, Stock and Watson (2006) (MSW, henceforth) found that iterated univariate forecasts typically outperform direct univariate forecasts. Furthermore, they found that the performance of the iterated univariate forecasts generally improves with the forecast horizon.

MSW studied univariate and bivariate VARs with lag orders either fixed or selected by AIC or BIC. Apart from the search over lag orders, they did not, however, conduct a broad model specification search involving multivariate models. This leaves open the issue of how robust their findings are to a broader model specification search. In this section we consider this question using the same data as in the MSW study.[3] This data set comprises 170 U.S. macroeconomic time series measured at the monthly frequency over the period 1959-2002 (528 months).

6.1   Data Transformations

Following MSW, all variables are transformed by differencing a suitable number of times to achieve stationarity for estimation and model selection. In a second step, forecasts are transformed back to levels and compared to level variables. We briefly explain how the forecasts are computed under the direct and iterated approaches using the autoregressive models as an illustration.

Denote the variables in levels by x_t and differenced variables by y_t. AR forecasts can be computed as follows. Under the iterated approach, y_{T+h} is modelled and $\hat{x}_{T+h}$ is constructed from x_T and $\hat{y}_{T+h}$. The forecasts of the AR model are based on the sample $y_{T-w+1}, y_{T-w+2}, \ldots, y_T$. Here $y_t = x_t$

---

[3]We are grateful to Mark Watson for making this data set publicly available.

if the variable is I(0), $y_t = \Delta x_t$ if the variable is I(1), and $y_t = \Delta^2 x_t$ if the variable is I(2), where for each variable we use the same order of integration as MSW.

Under the iterated approach the forecast of x_{T+h} is constructed from the forecast of y_{T+h}, x_T, and x_{T-1} as follows

$$\hat{x}_{T+h} = \begin{cases} \hat{y}_{T+h} & \text{if } x_t \text{ is I(0)} \\ x_T + \sum_{i=1}^{h} \hat{y}_{T+i} & \text{if } x_t \text{ is I(1)} \\ x_T + h\Delta x_T + \sum_{i=1}^{h} \sum_{j=1}^{i} \hat{y}_{T+j} & \text{if } x_t \text{ is I(2)} \end{cases}$$

Similarly, under the direct approach, the forecast of x_{T+h} is constructed from the forecast of y_{T+h}, x_T, and x_{t-1} as follows

$$\hat{x}_{T+h} = \begin{cases} \hat{y}_{T+h} & \text{if } x_t \text{ is I(0)} \\ x_T + \hat{y}_{T+h} & \text{if } x_t \text{ is I(1)} \\ x_T + h\Delta x_T + \hat{y}_{T+h} & \text{if } x_t \text{ is I(2)} \end{cases}$$

6.2   Setup

Forecasting is performed recursively and begins in 1979M1 (with a minimum of w observations before forecasting), running until the end of the sample 2002M12. This yields up to 288 forecasts for $h = 1$, 286 for $h = 3$, and so on. No allowance is made for the possibility of structural breaks. Forecasts are again reported for horizons of $h = 1, 3, 6, 12$, and 24 months. Two window lengths are used for estimation, namely $w = 120$ and $w = 240$, that is, 10 and 20 years of data. Fixing the window length allows us to better understand the relative performance of the various model selection criteria which can be quite sensitive to the sample size.

To address the effect of model selection on the multivariate forecasts, we extract factors from the 170 series in five groups, namely (A) one factor for "income, output, sales, capacity utilization" (38 variables); (B) one factor for "employment and unemployment" (27 variables); (C) one factor for "construction, inventories, and orders" (37 variables); (D) one factor for "interest rates and asset prices" (33 variables); and (E) one factor for "nominal prices, wages, and money" (35 variables).

To avoid any look-ahead biases, as noted earlier the factors are estimated recursively, i.e. $f_{it} = f(\boldsymbol{X}_i(1:t))$. We then obtain forecasts of the factors from VARs fitted to all five factors with lag length chosen by AIC or BIC and $p_{\max} = 2$.

The search over multivariate models is thus conducted over specifications that include own lags as well as factors. The space of models is limited as follows. For the univariate autoregressive models the possible lag lengths are $p = 0, 1, 2, \ldots, 12$, where $p = 0$ is an intercept only model. For the factor-augmented VAR models we search across five factors with one or two lags in addition to an intercept. For computational simplicity the lag length is set to be the same for y_{it} and f_{it}.

6.3 Forecasting Performance

Empirical results are presented in Tables 3-8. We first show MSFE results averaged across all 170 variables (Table 3), then present evidence on how many variables get included under the different approaches (Table 4) and report the proportion of cases for which the univariate iterated approach based on models selected by AIC produces lower MSFE-values than the alternative forecasting methods (Table 5). We also present test results from formal model comparisons (Table 6). To gain further insights into the forecasting performance for different types of economic variables, Tables 7 and 8 present MSFE values and test results for subcategories of the variables described above.

6.3.1 Univariate Models

Table 3 shows the relative forecasting performance (measured by MSFE) of the iterated and direct methods averaged over all 170 variables included in the MSW data set. For the shortest estimation window ($w = 120$), the iterated forecasts are better on average than the direct ones, particularly at long horizons ($h = 12$ and 24 months) and when using AIC. To understand this note that the iterated forecasts require a smaller pre-sample for the estimation and therefore tend to have a lower parameter estimation error. Such errors are most important in large models (AIC penalizes large models less than the BIC) and when the sample size is short. The relatively more efficient use of data by the iterated approach also explains why this performs better than the direct approach even with the large sample size ($w = 240$) when selecting models by the AIC.

In the large estimation sample ($w = 240$), the finding that the iterated forecasts produce the best performance is overturned across all forecast horizons when the BIC is used as the model selection criterion. This is partly explained by the fact that BIC chooses fewer parameters and so estimation error is less of an issue.

The modified information criteria, in particular the band-diagonal approach, are generally quite successful in improving the average forecasting performance relatively to using the simple version of the AIC. This finding holds less frequently when the BIC is used to select models.

In the short sample ($w = 120$), the best overall short-term ($h = 1$ or 3) forecasting performance is produced by the models selected by the BIC. At all other horizons, the univariate iterated models selected by the AIC deliver the best average performance. In the larger sample ($w = 240$), however, the best forecasting performance for horizons of 3 and 6 months is produced by the direct forecast approaches. The AIC modified to using a band-diagonal covariance matrix does particularly well. Once again the univariate iterated approach based on the AIC dominates overall when $h = 12$ and 24.

Table 4 shows the number of predictor variables included by the different approaches, averaged across time and across target variables. Univariate models selected by the AIC generally include 3-4 variables in the small sample and 4-5 variables in the larger sample. For the univariate models selected by the BIC, this number declines to only 1-2 variables in the small sample and two variables on average in the larger sample.

The average MSFE-values reported in Table 3 may be dominated by the most volatile variables and could provide an incomplete picture of relative forecasting performance. To deal with this, Table 5 shows the proportion of the 170 variables for which the iterated univariate AR forecasts based on models selected by the AIC generate a larger MSFE than the various alternatives. We use the iterated univariate forecasts selected by the AIC as our benchmark given the earlier evidence that this approach is the best univariate model, a finding supported by the results reported by MSW.

Among the univariate forecasts, in the small sample ($w = 120$), only the short-term iterated forecasts based on models selected by the BIC produce a majority of cases that outperform the iterated AIC. In the longer sample ($w = 240$), however, the iterated AIC approach is clearly best on average, particularly at the long horizons.

6.3.2 Multivariate Models

Turning to the multivariate models, the factor-augmented iterated VAR forecasts that use models selected by the AIC produce better forecasts than their univariate counterparts for a majority of the variables. This holds across all horizons and in both sample sizes. However, the factor-augmented iterated VAR forecasts that use models selected by the BIC produce the best forecasts for most variables at the four shortest horizons when $w = 120$ and for two horizons, $h = 1$ and $h = 12$, when $w = 240$.

Table 3 also shows that in many cases the proposed modifications to the information criteria continue to lead to improved performance for the direct forecasting models.

Table 4 shows that, as expected, the multivariate models include more predictor variables than their univariate counterparts. Under the AIC, on average 5-7 regressors get included in the small sample, rising to 6-8 variables in the larger sample. Once again, the BIC leads to somewhat smaller models with 3-4 predictor variables.

6.4 Model Comparisons

Table 6 provides test results based on a formal comparison of the benchmark iterated AIC model with the alternative approaches listed in each row. To this end we adopt the test methodology advocated by Giacomini and White (2006) which is ideally suited for our purpose since we are conducting pair-

wise model comparisons and use rolling window estimators. The table lists the proportion of model comparisons (out of 170) for which the null of equal predictive accuracy is rejected against the alternative that the two models' accuracy differ at a particular significance level.

Excluding the zeros which arise when forecasting methods are essentially identical, the proportion of rejections ranges between 0.15 and 0.40, with values near 0.15 and 0.20 most common. Asterisks indicate when the iterated AR model with lags selected by the AIC produces a smaller MSFE in fewer than 50% or 25% of the cases where the null of equal predictive accuracy is rejected.

The test results provide statistical evidence that on average the univariate iterated AIC approach performs very well relative to the other univariate methods and there is little evidence on which to prefer alternative univariate methods.

Among the factor-augmented VAR models, the iterated forecasts do better on average than the direct forecasts when a short estimation window ($w = 120$) is used. These results are overturned, however, when the long estimation window ($w = 240$) is used. In the latter case, the direct forecasting method is better for most forecast horizons irrespective of which information criterion is used.

For the models selected by the AIC or BIC, in many cases the Newey-West and band-diagonal modifications to the covariance matrix help improve on the average performance of the direct forecasts.

Overall, the factor models tend to generate smaller average forecast errors than the best univariate approach for the short horizons, $h = 1$ and $h = 3$, but do worse than the iterated AIC when $h = 6, 12$ and 24 months. Our results thus suggest that including more variables favors the direct method at short forecast horizons.

6.5   Results by Variable Categories

The empirical results turn out to be quite similar for four of the five categories of economic variables, namely (A) income, output, sales and capacity utilization, (B) employment and unemployment, (C) Construction, inventories and orders, and (D) interest rates and asset prices. In contrast, quite different results are obtained for the fifth category, namely (E) Nominal prices, wages and money. For this reason, Table 7 presents results averaged across variables in categories A-D, while Table 8 shows separate results for category-E variables.

Table 7 shows that the benefit from using the multivariate factor-based approach comes out very strongly for the first four categories. For these variables, across the vast majority of sample sizes and forecast horizons, the multivariate iterated AIC produces lower MSFE-values than the univariate iterated AIC approach. Moreover, the modifications to the AIC or BIC

methods applied to the multivariate models often generate the best overall forecasting performance.

Across the first four categories of variables, the multivariate iterated approach based on the AIC generates the best results on average. In fact, for these variables, the multivariate iterated AIC approach dominates the univariate iterated AIC across all forecast horizons and for both long and short estimation samples. The multivariate iterated approach based on BIC also performs well, particularly for the shortest estimation window ($w = 120$).

Table 8 shows that it is only for the final group of variables, (E) nominal prices, wages and money, that the univariate iterated AIC approach performs best on average.

Overall, these findings clearly demonstrate the value from utilizing multivariate information and also provide evidence that our proposed refinements to the information criteria work in many cases.

7 Conclusion

Our empirical results show an interesting interaction between the length of the estimation window, how strongly a particular model selection method penalizes the inclusion of additional variables, the forecast horizon, and the relative performance of the direct versus iterated approaches. Estimation error turns out to be key to understanding our findings. The iterated forecasting approach performs well relative to the direct approach when the sample size is small, when using an information criterion such as the AIC that does not penalize additional parameters too heavily and when the forecast horizon gets large. This is what we would expect to find since the iterated approach makes more efficient use of sample data than the direct forecasting approach. In this regard the theoretical results obtained for the simple AR(1) model also provide useful insights. Conversely, when the BIC is used, leading to smaller models and reduced parameter estimation error, the direct approach outperforms the iterated approach on average.

We confirm the finding reported by Marcellino, Stock and Watson (2006) that iterated forecasts generated by univariate models selected by the AIC produce good forecasts on average at long horizons. However, we also find that multivariate VAR-based forecasts outperform the univariate forecasts at short horizons ($h = 1$ and 3 months) and across all horizons for four of the five categories of economic variables studied here. This suggests that, for most economic variables and particularly at short horizons, information beyond what is contained in the past history of the variables themselves can be helpful in producing better forecasts.

Appendix: Mathematical details

Appendix A: Proof of Proposition 1

Direct forecast First consider the direct method, and note that from equation (3)

$$\bar{y}_T = a_h + \phi_h \bar{y}_{T,-h}, + \bar{v}_T$$

which together with equation (7) yields

$$(a_h - \tilde{a}_{h,T}) + (\phi_h - \tilde{\phi}_{h,T})\bar{y}_{T,-h} + \bar{v}_T = 0. \tag{A1}$$

Note that the forecast error for the direct method is given by

$$\tilde{e}_{T+h} = (a_h - \tilde{a}_{h,T}) + (\phi_h - \tilde{\phi}_{h,T})y_T + v_{T+h},$$

which, using (A1), can be written as

$$\tilde{e}_{T+h} = (\phi_h - \tilde{\phi}_{h,T})(y_T - \bar{y}_{T,-h}) - \bar{v}_T + v_{T+h}. \tag{A2}$$

But using equation (2), we have

$$(y_T - \bar{y}_{T,-h}) = u_T - \bar{u}_{T,-h},$$

where in a similar way $\bar{u}_{T,-h} = T^{-1}\sum_{t=1}^{T} u_{t-h}$. Also from equation (8)

$$
\begin{aligned}
\tilde{\phi}_{h,T} - \phi_h &= \frac{\sum_{t=1}^{T}(y_{t-h} - \bar{y}_{T,-h})v_t}{\sum_{t=1}^{T}(y_{t-h} - \bar{y}_{T,-h})^2} \\
&= \frac{\sum_{t=1}^{T}(u_{t-h} - \bar{u}_{T,-h})v_t}{\sum_{t=1}^{T}(u_{t-h} - \bar{u}_{T,-h})^2},
\end{aligned}
$$

and

$$\tilde{e}_{T+h} = (u_T - \bar{u}_{T,-h})\left(\frac{\sum_{t=1}^{T}(u_{t-h} - \bar{u}_{T,-h})v_t}{\sum_{t=1}^{T}(u_{t-h} - \bar{u}_{T,-h})^2}\right) - \bar{v}_T + v_{T+h}.$$

Recall that v_t and u_t are linear functions of the ε_t shocks. It follows that $\tilde{e}_{T+h}$ is an odd function of u_t and v_t, and so long as the errors are symmetrically distributed around zero, we have $E(\tilde{e}_{T+h}) = 0$, assuming that T is sufficiently large so that expectations exists.

Iterated forecasts Consider next the forecast errors of the iterated procedure and note that

$$\hat{e}_{T+h} = a\left(\frac{1 - \phi^h}{1 - \phi}\right) - \hat{a}_T\left(\frac{1 - \hat{\phi}_T^h}{1 - \hat{\phi}_T}\right) + \left(\phi^h - \hat{\phi}_T^h\right)y_T + v_{T+h},$$

which we re-write as

$$\hat{e}_{T+h} = -(\hat{\mu}_T - \mu)(1 - \hat{\phi}_T^h) - \left(\hat{\phi}_T^h - \phi^h\right) u_T + v_{T+h},$$

where $\mu = a/(1 - \phi)$ as before and $\hat{\mu}_T = \hat{a}_T / \left(1 - \hat{\phi}_T\right)$. But

$$
\begin{aligned}
\hat{a}_T - a &= \hat{\mu}_T - \mu - (\hat{\mu}_T \hat{\phi}_T - \mu\phi) \\
&= \hat{\mu}_T - \mu - (\hat{\mu}_T - \mu)\hat{\phi}_T - \mu(\hat{\phi}_T - \phi),
\end{aligned}
\tag{29}
$$

and from equation (6) we also have

$$a - \hat{a}_T + (\phi - \hat{\phi}_T)\,\bar{y}_{h:T,-1} + \bar{\varepsilon}_{h:T} = 0,$$

where $\bar{\varepsilon}_{h:T} = (T + h - 1)^{-1} \sum_{t=-h+2}^{T} \varepsilon_t$. Using this result in (29) we have

$$(\hat{\phi}_T - \phi)\,\bar{y}_{h:T,-1} - \bar{\varepsilon}_{h:T} = (\hat{\mu}_T - \mu)\,(1 - \hat{\phi}_T) - \mu(\hat{\phi}_T - \phi),$$

from which it follows that

$$
\begin{aligned}
\hat{\mu}_T - \mu &= \frac{(\hat{\phi}_T - \phi)\,(\bar{y}_{h:T,-1} - \mu) - \bar{\varepsilon}_{h:T}}{1 - \hat{\phi}_T} \\
&= \frac{(\hat{\phi}_T - \phi)\,\bar{u}_{h:T,-1} - \bar{\varepsilon}_{h:T}}{1 - \hat{\phi}_T} \\
&= \frac{(\hat{\phi}_T - \phi)\,\bar{u}_{h:T,-1} - \bar{\varepsilon}_{h:T}}{1 - \phi - (\hat{\phi}_T - \phi)},
\end{aligned}
$$

where $\bar{u}_{h:T,-1} = (T + h - 1)^{-1} \sum_{t=-h+2}^{T} u_{t-1}$. The small sample bias of $\hat{\phi}_T$ ensures that $1 - \hat{\phi}_T \neq 0$. Hence

$$\hat{e}_{T+h} = (\mu - \hat{\mu}_T)(1 - \phi^h) + \left(\hat{\phi}_T^h - \phi^h\right)(\hat{\mu}_T - \mu) - \left(\hat{\phi}_T^h - \phi^h\right) u_T + v_{T+h}.$$

Now let $\hat{\delta}_T = \hat{\phi}_T - \phi$ and note that

$$\hat{\phi}_T^h - \phi^h = \sum_{j=1}^{h} \binom{h}{j} \hat{\delta}_T^j \phi^{h-j},$$

and

$$\hat{\delta}_T = \frac{\sum_{t=-h+2}^{T}(u_{t-h} - \bar{u}_{T,-1})v_t}{\sum_{t=-h+2}^{T}(u_{t-h} - \bar{u}_{T,-1})^2}.$$

Therefore, it follows that since $\hat{\delta}_T$ is an even function of u_t and v_t, then $\hat{\phi}_T^h - \phi^h$ will also be an even function of the underlying shocks. Furthermore, since

$$\hat{\mu}_T - \mu = \frac{\hat{\delta}_T\,\bar{u}_{h:T,-1} - \bar{\varepsilon}_{h:T}}{1 - \phi - \hat{\delta}_T},$$

24

then $\hat{\mu}_T - \mu$ is also an odd function of the underlying shocks. Hence $\hat{e}_{T+h}$ will be an odd function of the shocks and assuming its expectation exists we have $\mathrm{E}(\hat{e}_{T+h}) = 0$ as required.

Specifically

$$
\begin{aligned}
\hat{e}_{T+h} &= -\left(\frac{\hat{\delta}_T \ \bar{u}_{h:T,-1} - \bar{\varepsilon}_{h:T}}{1 - \phi - \hat{\delta}_T}\right)(1 - \phi^h) \\
&+ \left(\sum_{j=1}^{h}\binom{h}{j}\hat{\delta}_T^j \phi^{h-j}\right)\left(\frac{\hat{\delta}_T \ \bar{u}_{h:T,-1} - \bar{\varepsilon}_{h:T}}{1 - \phi - \hat{\delta}_T} - u_T\right) + v_{T+h}.
\end{aligned}
$$

Appendix B: Derivation of modified information criteria

Consider the h-step ahead forecast in (3) and for simplicity of exposition assume that $\alpha_h = 0$ and drop the subscript from the parameter, so that we have

$$\boldsymbol{y}_{t+h} = \boldsymbol{\phi}\boldsymbol{y}_t + \boldsymbol{v}_t. \tag{B1}$$

Denote the loss evaluated at $\boldsymbol{\phi}$ by $\ell(\boldsymbol{\phi})$, while $\ell_0(\boldsymbol{\phi}) = \mathrm{E}[\ell(\boldsymbol{\phi})]$, and the jth derivative of the loss is given by $\ell^{(j)}(\boldsymbol{\phi})$. Moreover, let $\boldsymbol{\phi}_0$ be the true parameter and $\hat{\boldsymbol{\phi}}$ its estimate. Using these notations, we have that $\ell_0^{(1)}(\boldsymbol{\phi}_0) = 0$ and $\ell^{(1)}(\hat{\boldsymbol{\phi}}) = 0$. For $\ell(\boldsymbol{\phi}) = \frac{1}{2\sigma^2}(\boldsymbol{Y}_{T+h} - \boldsymbol{Y}_T\boldsymbol{\phi})'(\boldsymbol{Y}_{T+h} - \boldsymbol{Y}_T\boldsymbol{\phi})$ we also have that $\ell^{(2)}(\boldsymbol{\phi}) = \ell_0^{(2)}(\boldsymbol{\phi}_0) = \frac{1}{2\sigma^2}\boldsymbol{Y}_T'\boldsymbol{Y}_T$, where $\boldsymbol{Y}_{T+h} = (\boldsymbol{y}_{1+h}, \boldsymbol{y}_{2+h}, \ldots, \boldsymbol{y}_{T+h})'$ and $\boldsymbol{Y}_T = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)'$, which implies that a second order Taylor expansion is exact.

Define $\ell_0(\hat{\boldsymbol{\phi}})$ as the loss incurred by using the estimated parameter $\hat{\boldsymbol{\phi}}$ instead of the unknown true parameter. Then

$$\mathrm{E}[\ell_0(\hat{\boldsymbol{\phi}})] = \mathrm{E}[\ell(\hat{\boldsymbol{\phi}})] + \mathrm{E}[\ell_0(\hat{\boldsymbol{\phi}}) - \ell(\hat{\boldsymbol{\phi}})].$$

A second order Taylor expansion implies

$$\ell_0(\hat{\boldsymbol{\phi}}) = \ell_0(\boldsymbol{\phi}_0) + (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\ell_0^{(2)}(\boldsymbol{\phi}_0)(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$$

and

$$\ell(\hat{\boldsymbol{\phi}}) = \ell(\boldsymbol{\phi}_0) + (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\ell^{(1)}(\boldsymbol{\phi}_0) + (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\ell_0^{(2)}(\boldsymbol{\phi}_0)(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) \ .$$

Taking expectations of these expressions yields

$$
\begin{aligned}
\mathrm{E}[\ell(\hat{\boldsymbol{\phi}})] &= \mathrm{E}[\ell(\boldsymbol{\phi}_0)] + \mathrm{E}[(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\ell^{(1)}(\boldsymbol{\phi}_0)] + \mathrm{E}[(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\ell_0^{(2)}(\boldsymbol{\phi}_0)(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)] \\
&= \ell_0(\boldsymbol{\phi}_0) + \mathrm{E}[(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\ell^{(1)}(\boldsymbol{\phi}_0)] + \mathrm{E}[(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\ell_0^{(2)}(\boldsymbol{\phi}_0)(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)],
\end{aligned}
$$

and therefore

$$\mathrm{E}[\ell_0(\hat{\boldsymbol{\phi}}) - \ell(\hat{\boldsymbol{\phi}})] = -\mathrm{E}[(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\ell^{(1)}(\boldsymbol{\phi}_0)].$$

Furthermore, by Taylor expansion we have

$$(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = -[\ell^{(2)}(\boldsymbol{\phi}_0)]^{-1}\ell^{(1)}(\boldsymbol{\phi}_0),$$

and therefore

$$
\begin{aligned}
-\mathrm{E}[(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\ell^{(1)}(\boldsymbol{\phi}_0)] &= -\mathrm{E}[\mathrm{tr}(\sigma^{-2}\boldsymbol{U}_T'\boldsymbol{Y}_T(\boldsymbol{Y}_T'\boldsymbol{Y}_T)^{-1}\boldsymbol{Y}_T\boldsymbol{U}_T)] \\
&= -\mathrm{tr}(\sigma^{-2}(\boldsymbol{Y}_T'\boldsymbol{Y}_T)^{-1}\boldsymbol{Y}_T'\mathrm{E}[\boldsymbol{U}_T'\boldsymbol{U}_T]\boldsymbol{Y}_T),
\end{aligned}
$$

where $\boldsymbol{U}_T = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots \boldsymbol{v}_T)$.

If the errors $\boldsymbol{v}_t$ were iid, this would give the standard penalty term K. However, in overlapping forecasts the errors will be autocorrelated and the expression will not collapse to K.

References

[1] Bai, Jushan and Serena Ng, 2002, Determining the number of factors in approximate factor models. *Econometrica*, 70, 161–221.

[2] Bai, Jushan and Serena Ng, 2009, Boosting diffusion indices, *Journal of Applied Econometrics*, forthcoming.

[3] Bernanke, Ben S., Jean Boivin and Piotr Eliasz, 2005, Measuring the effect of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* 120, 387–422.

[4] Bhansali, Rajendra J., 1999, Parameter estimation and model selection for multistep prediction of a time series: a review. in *Asymptotics, Non-parametrics and Time Series* (ed. Subir Ghosh), 201–225, Marcel Dekker, New York.

[5] Brown, Bryan W. and Roberto S. Mariano, 1989, Measures of deterministic prediction bias in nonlinear models. *International Economic Review* 30(3), 667–684.

[6] Clements, Michael, and David Hendry, 1998, *Forecasting Economic Time Series*. Cambridge University Press, Cambridge.

[7] Cox, David R., 1961, Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society* B 23, 414–422.

[8] Findley, David F.,1983, On the use of multiple models for multi-period forecasting. *Proceedings of Business and Economic Statistics, American Statistical Association*, 528–531.

[9] Forni Mario, Hallin Marc, Lippi Marco, Reichlin Lucrezia, 2005, The generalized dynamic factor model, one sided estimation and forecasting. *Journal of the American Statistical Association*, 100, 830–840.

[10] Fuller, Wayne A., 1996, *Introduction to Statistical Time Series*. Wiley, New York.

[11] Giacomini, Raffaella and Halbert White, 2006, Tests of conditional predictive ability. *Econometrica* 74(6), 1545–1578.

[12] Ing, Ching-Kang, 2003, Multistep prediction in autoregressive processes. *Econometric Theory* 19, 254–279.

[13] Marcellino, Massimiliano, James H. Stock and Mark W. Watson, 2006, A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* 135, 499–526.

[14] Newey, Whitney K. and Mark W. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.

[15] Pesaran, M. Hashem, and Allan Timmermann, 2005, Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics* 129, 183–217.

[16] Phillips, Peter C. B., 1979, The sampling distribution of forecasts from a first order autoregression. *Journal of Econometrics*, 9, 241–261.

[17] Schorfheide, Frank, 2005, VAR forecasting under misspecification. *Journal of Econometrics*, 128, 99–136.

[18] Stock, James H., and Mark W. Watson, 2002, Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 97, 1167–1179.

[19] Stock, James H., and Mark W. Watson (2005) 'Implications of dynamic factor models for VAR Analysis.' *mimeo*.

[20] Ullah, Aman, 2004, *Finite Sample Econometrics*. Oxford University Press, Oxford.

Table 1: Forecasting performance of iterated and direct methods: Monte Carlo results under DGP 1

R² = 0.2

AR, T = 240 | VAR, T = 240

Method	AR h=1	AR h=3	AR h=6	AR h=12	AR h=24	VAR h=1	VAR h=3	VAR h=6	VAR h=12	VAR h=24
iterated AIC	1.0582	1.1077	1.1250	1.1394	1.1261	1.0155	1.1117	1.1287	1.1250	1.1159
direct AIC	1.0582	1.1098	1.1319	1.1451	1.1351	1.0156	1.1214	1.1384	1.1354	1.1282
mod. AIC(NW)	1.0583	1.1111	1.1317	1.1466	1.1358	1.0156	1.1218	1.1379	1.1358	1.1277
mod. AIC(diag)	1.0582	1.1095	1.1296	1.1450	1.1358	1.0156	1.1219	1.1368	1.1349	1.1271
iterated BIC	1.0550	1.1032	1.1241	1.1393	1.1261	1.0181	1.1111	1.1285	1.1251	1.1159
direct BIC	1.0550	1.1048	1.1252	1.1409	1.1281	1.0181	1.1189	1.1351	1.1315	1.1247
mod. BIC(NW)	1.0554	1.1047	1.1249	1.1413	1.1282	1.0177	1.1193	1.1338	1.1310	1.1239
mod. BIC(diag)	1.0550	1.1039	1.1243	1.1402	1.1290	1.0181	1.1193	1.1349	1.1302	1.1224

AR, T = 120 | VAR, T = 120

Method	AR h=1	AR h=3	AR h=6	AR h=12	AR h=24	VAR h=1	VAR h=3	VAR h=6	VAR h=12	VAR h=24
iterated AIC	1.0618	1.1331	1.1384	1.1237	1.1289	1.0302	1.1252	1.1224	1.1163	1.1253
direct AIC	1.0618	1.1389	1.1478	1.1423	1.1497	1.0302	1.1437	1.1444	1.1425	1.1542
mod. AIC(NW)	1.0631	1.1403	1.1509	1.1469	1.1530	1.0299	1.1442	1.1435	1.1416	1.1535
mod. AIC(diag)	1.0618	1.1374	1.1454	1.1430	1.1582	1.0302	1.1425	1.1414	1.1376	1.1532
iterated BIC	1.0564	1.1277	1.1358	1.1236	1.1285	1.0364	1.1237	1.1199	1.1144	1.1240
direct BIC	1.0564	1.1289	1.1371	1.1268	1.1342	1.0378	1.1355	1.1365	1.1362	1.1452
mod. BIC(NW)	1.0551	1.1284	1.1380	1.1279	1.1354	1.0361	1.1349	1.1365	1.1362	1.1456
mod. BIC(diag)	1.0564	1.1281	1.1368	1.1263	1.1365	1.0378	1.1349	1.1345	1.1333	1.1440

AR, T = 60 | VAR, T = 60

Method	AR h=1	AR h=3	AR h=6	AR h=12	AR h=24	VAR h=1	VAR h=3	VAR h=6	VAR h=12	VAR h=24
iterated AIC	1.1088	1.1595	1.1499	1.1501	1.1387	1.0613	1.1450	1.1247	1.1379	1.1232
direct AIC	1.1088	1.1729	1.1814	1.2042	1.2793	1.0621	1.1828	1.1794	1.2040	1.2142
mod. AIC(NW)	1.1137	1.1812	1.1881	1.2228	1.3325	1.0623	1.1855	1.1817	1.2086	1.2197
mod. AIC(diag)	1.1088	1.1678	1.1741	1.2075	1.4121	1.0621	1.1813	1.1749	1.2011	1.2292
iterated BIC	1.0935	1.1416	1.1376	1.1439	1.1342	1.0674	1.1390	1.1201	1.1346	1.1199
direct BIC	1.0935	1.1445	1.1447	1.1562	1.1714	1.0703	1.1699	1.1659	1.1881	1.1970
mod. BIC(NW)	1.0881	1.1458	1.1518	1.1646	1.2102	1.0645	1.1712	1.1671	1.1898	1.1982
mod. BIC(diag)	1.0935	1.1416	1.1419	1.1622	1.2925	1.0703	1.1701	1.1618	1.1844	1.2064

R² = 0.8

AR, T = 240 | VAR, T = 240

Method	AR h=1	AR h=3	AR h=6	AR h=12	AR h=24	VAR h=1	VAR h=3	VAR h=6	VAR h=12	VAR h=24
iterated AIC	1.2455	1.8920	2.2142	2.2864	2.2788	1.0136	1.6107	2.0772	2.2855	2.2482
direct AIC	1.2455	1.9004	2.2399	2.3243	2.3283	1.0136	1.6210	2.1031	2.3341	2.3002
mod. AIC(NW)	1.2454	1.9009	2.2425	2.3176	2.3215	1.0135	1.6217	2.1038	2.3329	2.2954
mod. AIC(diag)	1.2455	1.9000	2.2469	2.3067	2.3120	1.0136	1.6238	2.1059	2.3298	2.2833
iterated BIC	1.2415	1.8880	2.2092	2.2872	2.2787	1.0107	1.6081	2.0761	2.2854	2.2483
direct BIC	1.2415	1.8890	2.2208	2.2974	2.3029	1.0107	1.6194	2.1022	2.3259	2.2902
mod. BIC(NW)	1.2420	1.8891	2.2414	2.2897	2.2931	1.0108	1.6226	2.1064	2.3235	2.2824
mod. BIC(diag)	1.2415	1.8893	2.2565	2.2854	2.2842	1.0107	1.6338	2.1200	2.3193	2.2705

AR, T = 120 | VAR, T = 120

Method	AR h=1	AR h=3	AR h=6	AR h=12	AR h=24	VAR h=1	VAR h=3	VAR h=6	VAR h=12	VAR h=24
iterated AIC	1.2463	1.9314	2.2907	2.3418	2.3145	1.0240	1.6578	2.1116	2.2718	2.3016
direct AIC	1.2463	1.9577	2.3480	2.4539	2.4557	1.0240	1.6856	2.1782	2.3928	2.4481
mod. AIC(NW)	1.2470	1.9599	2.3613	2.4435	2.4529	1.0255	1.6917	2.1816	2.3867	2.4392
mod. AIC(diag)	1.2463	1.9589	2.3567	2.4140	2.4225	1.0240	1.6998	2.1799	2.3708	2.4176
iterated BIC	1.2340	1.9148	2.2719	2.3354	2.3132	1.0195	1.6552	2.1105	2.2712	2.3020
direct BIC	1.2340	1.9321	2.3165	2.3807	2.3847	1.0195	1.6978	2.1711	2.3696	2.4232
mod. BIC(NW)	1.2361	1.9369	2.3234	2.3724	2.3796	1.0193	1.7073	2.1773	2.3604	2.4144
mod. BIC(diag)	1.2340	1.9450	2.3256	2.3486	2.3444	1.0195	1.7466	2.1914	2.3473	2.3858

AR, T = 60 | VAR, T = 60

Method	AR h=1	AR h=3	AR h=6	AR h=12	AR h=24	VAR h=1	VAR h=3	VAR h=6	VAR h=12	VAR h=24
iterated AIC	1.3102	2.0168	2.3697	2.4584	2.4166	1.0514	1.7210	2.2013	2.3834	2.3552
direct AIC	1.3102	2.0812	2.5517	2.8518	3.3870	1.0513	1.7919	2.3706	2.6873	2.8453
mod. AIC(NW)	1.3151	2.1055	2.5644	2.8646	3.4856	1.0529	1.8018	2.3691	2.6866	2.8518
mod. AIC(diag)	1.3102	2.0942	2.5325	2.7801	3.5934	1.0513	1.8214	2.3692	2.6432	2.8363
iterated BIC	1.2825	1.9739	2.3213	2.4153	2.3844	1.0494	1.7173	2.1948	2.3783	2.3522
direct BIC	1.2825	2.0295	2.4285	2.6450	3.0082	1.0494	1.8104	2.3436	2.6447	2.7905
mod. BIC(NW)	1.2861	2.0422	2.4457	2.6760	3.1836	1.0498	1.8166	2.3509	2.6344	2.7960
mod. BIC(diag)	1.2825	2.0893	2.3886	2.5287	3.2233	1.0494	1.8727	2.3610	2.5743	2.7439

The table reports the MSFE of the different forecasting methods. The 'mod. AIC(NW)' and 'mod. BIC(NW)' forecasts are based on the modified AIC and BIC, where the covariance matrix is estimated using the Newey-West estimator; the 'mod. AIC(diag)' and 'mod. BIC(diag)' forecasts are based on the modified AIC and BIC, where the band diagonal covariance matrix is used. The Monte Carlo results are from 10000 iterations. Details of DGPs 1 and 2 are given in Section 5.

Table 2: Forecasting performance of iterated and direct methods: Monte Carlo results under DGP 2

Forecast horizon	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
$R^2 = 0.2$															
	AR, T = 60					*AR, T = 120*					*AR, T = 240*				
iterated AIC	1.0853	1.1568	1.1590	1.1511	1.1411	1.0491	1.1303	1.1375	1.1260	1.1302	1.0425	1.1139	1.1285	1.1402	1.1255
direct AIC	1.0853	1.1724	1.1880	1.2039	1.2831	1.0491	1.1351	1.1469	1.1445	1.1541	1.0425	1.1173	1.1331	1.1466	1.1348
mod. AIC(NW)	1.0925	1.1796	1.2007	1.2221	1.3374	1.0499	1.1377	1.1489	1.1490	1.1548	1.0427	1.1178	1.1330	1.1465	1.1354
mod. AIC(diag)	1.0853	1.1655	1.1819	1.2110	1.4120	1.0491	1.1354	1.1432	1.1449	1.1606	1.0425	1.1170	1.1321	1.1452	1.1352
iterated BIC	1.0721	1.1397	1.1433	1.1438	1.1338	1.0402	1.1259	1.1348	1.1253	1.1300	1.0397	1.1093	1.1255	1.1403	1.1254
direct BIC	1.0721	1.1432	1.1505	1.1589	1.1723	1.0402	1.1277	1.1354	1.1288	1.1372	1.0397	1.1116	1.1262	1.1423	1.1277
mod. BIC(NW)	1.0691	1.1459	1.1553	1.1654	1.2137	1.0396	1.1279	1.1362	1.1290	1.1378	1.0399	1.1116	1.1260	1.1421	1.1277
mod. BIC(diag)	1.0721	1.1396	1.1469	1.1614	1.2943	1.0402	1.1270	1.1355	1.1286	1.1375	1.0397	1.1110	1.1256	1.1414	1.1277
	VAR, T = 60					*VAR, T = 120*					*VAR, T = 240*				
iterated AIC	1.0613	1.1498	1.1231	1.1405	1.1243	1.0276	1.1254	1.1244	1.1153	1.1212	1.0139	1.1110	1.1266	1.1272	1.1173
direct AIC	1.0627	1.2057	1.2011	1.2331	1.2526	1.0277	1.1468	1.1543	1.1489	1.1639	1.0140	1.1232	1.1417	1.1413	1.1326
mod. AIC(NW)	1.0625	1.2056	1.2029	1.2369	1.2614	1.0284	1.1464	1.1523	1.1479	1.1615	1.0148	1.1236	1.1404	1.1405	1.1312
mod. AIC(diag)	1.0627	1.2000	1.1898	1.2272	1.2767	1.0277	1.1452	1.1493	1.1429	1.1577	1.0140	1.1220	1.1395	1.1380	1.1291
iterated BIC	1.0544	1.1415	1.1178	1.1366	1.1210	1.0239	1.1209	1.1222	1.1123	1.1182	1.0121	1.1099	1.1263	1.1265	1.1174
direct BIC	1.0581	1.1846	1.1767	1.2053	1.2217	1.0254	1.1381	1.1424	1.1408	1.1472	1.0123	1.1197	1.1375	1.1371	1.1280
mod. BIC(NW)	1.0548	1.1836	1.1798	1.2117	1.2294	1.0239	1.1371	1.1437	1.1400	1.1474	1.0121	1.1205	1.1350	1.1359	1.1271
mod. BIC(diag)	1.0581	1.1807	1.1682	1.1985	1.2387	1.0254	1.1387	1.1412	1.1357	1.1475	1.0123	1.1194	1.1351	1.1347	1.1258
$R^2 = 0.8$															
	AR, T = 60					*AR, T = 120*					*AR, T = 240*				
iterated AIC	1.2559	1.9305	2.3303	2.4816	2.5400	1.2039	1.8418	2.2374	2.3345	2.3501	1.2076	1.8194	2.1415	2.2883	2.2744
direct AIC	1.2559	2.0087	2.5424	2.9233	3.5071	1.2039	1.8708	2.2983	2.4464	2.5206	1.2076	1.8313	2.1704	2.3262	2.3303
mod. AIC(NW)	1.2627	2.0218	2.5662	2.9546	3.6406	1.2042	1.8724	2.3091	2.4418	2.5192	1.2068	1.8313	2.1772	2.3251	2.3229
mod. AIC(diag)	1.2559	2.0219	2.5360	2.8431	3.7352	1.2039	1.8720	2.3197	2.4048	2.4925	1.2076	1.8311	2.1814	2.3108	2.3074
iterated BIC	1.2293	1.8801	2.2880	2.4393	2.4454	1.1923	1.8316	2.2264	2.3246	2.3476	1.2022	1.8120	2.1325	2.2855	2.2743
direct BIC	1.2293	1.9527	2.4516	2.6917	3.0919	1.1923	1.8456	2.2686	2.3800	2.4398	1.2022	1.8168	2.1572	2.3036	2.3026
mod. BIC(NW)	1.2321	1.9729	2.4516	2.7202	3.2596	1.1932	1.8474	2.2893	2.3676	2.4291	1.2025	1.8170	2.1774	2.2966	2.2926
mod. BIC(diag)	1.2293	2.0034	2.4227	2.5575	3.3184	1.1923	1.8543	2.3178	2.3414	2.3888	1.2022	1.8169	2.2036	2.2864	2.2827
	VAR, T = 60					*VAR, T = 120*					*VAR, T = 240*				
iterated AIC	1.0557	1.5778	2.0149	2.3811	2.4218	1.0250	1.5205	1.9623	2.2897	2.2988	1.0150	1.4639	1.9313	2.2602	2.2632
direct AIC	1.0557	1.6076	2.0159	2.6871	3.1261	1.0250	1.4833	1.8338	2.3397	2.4919	1.0150	1.4189	1.7681	2.2410	2.3415
mod. AIC(NW)	1.0567	1.6188	2.0278	2.6898	3.1266	1.0267	1.4869	1.8458	2.3351	2.4808	1.0146	1.4178	1.7702	2.2402	2.3323
mod. AIC(diag)	1.0557	1.6265	2.0242	2.6514	3.1473	1.0250	1.4929	1.8571	2.3181	2.4583	1.0150	1.4199	1.7771	2.2318	2.3274
iterated BIC	1.0369	1.5751	2.0493	2.3943	2.4198	1.0169	1.5281	1.9989	2.3038	2.2989	1.0090	1.4755	1.9689	2.2719	2.2647
direct BIC	1.0367	1.6124	1.9996	2.6273	3.0221	1.0169	1.4883	1.8428	2.3254	2.4610	1.0090	1.4152	1.7719	2.2350	2.3255
mod. BIC(NW)	1.0383	1.6239	2.0110	2.6271	3.0353	1.0171	1.5010	1.8642	2.3112	2.4452	1.0091	1.4198	1.7884	2.2301	2.3130
mod. BIC(diag)	1.0367	1.6467	2.0222	2.5676	2.9751	1.0169	1.5247	1.8949	2.2885	2.4127	1.0090	1.4301	1.8145	2.2184	2.3071

See footnote of Table 1.

Table 3: Relative forecasting performance of iterated and direct methods: Average ratio of MSFE across all 170 variables in the Marcellino, Stock and Watson (2006) data set

Forecast horizon	AR, $w = 120$					AR, $w = 240$				
	1	3	6	12	24	1	3	6	12	24
iterated AIC	1.0000	1.0000	1.0000	1.0000	1.0000	0.9809	0.9629	0.9470	0.9546	1.0018
direct AIC	1.0000	1.0152	1.0178	1.0821	1.2234	0.9809	0.9605	0.9455	0.9887	1.1083
mod. AIC(NW)	0.9932	1.0025	1.0138	1.0844	1.2450	0.9738	0.9537	0.9411	0.9858	1.1115
mod. AIC(diag)	1.0000	1.0124	1.0170	1.1049	1.2113	0.9809	0.9602	0.9451	0.9833	1.1219
iterated BIC	0.9892	1.0100	1.0368	1.0529	1.0460	0.9876	1.0090	1.0459	1.0906	1.1413
direct BIC	0.9892	1.0080	1.0162	1.0766	1.1731	0.9876	0.9674	0.9539	0.9900	1.1012
mod. BIC(NW)	0.9908	1.0065	1.0217	1.0832	1.1821	0.9892	0.9755	0.9573	0.9995	1.1050
mod. BIC(diag)	0.9892	1.0083	1.0296	1.1739	1.2218	0.9876	0.9659	0.9544	1.0265	1.1942
	Factors, $w = 120$					Factors, $w = 240$				
	1	3	6	12	24	1	3	6	12	24
iterated AIC	0.9855	0.9822	1.0397	1.0821	1.0658	0.9684	0.9743	1.0302	1.0597	1.0951
direct AIC	0.9855	0.9866	1.0500	1.1310	1.2829	0.9684	0.9464	0.9981	1.0510	1.1428
mod. AIC(NW)	0.9838	0.9827	1.0554	1.1411	1.2871	0.9675	0.9465	0.9968	1.0510	1.1446
mod. AIC(diag)	0.9855	0.9867	1.0462	1.1146	1.2517	0.9684	0.9443	0.9938	1.0358	1.1131
iterated BIC	0.9775	0.9810	1.0272	1.0642	1.1149	0.9694	0.9883	1.0502	1.0986	1.1556
direct BIC	0.9775	0.9815	1.0221	1.1076	1.2606	0.9694	0.9564	1.0002	1.0452	1.1264
mod. BIC(NW)	0.9729	0.9825	1.0356	1.1187	1.2729	0.9698	0.9671	1.0085	1.0383	1.1289
mod. BIC(diag)	0.9775	0.9866	1.0294	1.1595	1.3241	0.9694	0.9562	0.9947	1.0523	1.1958

The table reports the MSFE of the different forecasts as a ratio of the MSFE of the iterated AR forecast based on the model selected by AIC with $w = 120$, where w is the length of the estimation window. AR results are based on univariate autoregressive models and results labeled 'Factors' are based on multivariate factor-augmented VAR models. The MSFE is calculated only for those forecast periods where forecasts from all methods are available. The 'mod. AIC(NW)' and 'mod. BIC(NW)' forecasts are based on the modified AIC and BIC, where the covariance matrix is estimated using the Newey-West estimator; the 'mod. AIC(diag)' and 'mod. BIC(diag)' forecasts are based on the modified AIC and BIC, where the band diagonal covariance matrix is used.

Table 4: Average number of parameters included by the various forecasting approaches

forecast horizon	1	3	6	12	24	1	3	6	12	24
	AR, w = 120					*AR, w = 240*				
iterated AIC	3.6818	3.6739	3.6643	3.6528	3.6268	5.3329	5.3324	5.3268	5.3097	5.2644
direct AIC	3.6818	3.8997	3.6890	3.5925	3.6599	5.3329	5.4674	4.9338	4.7061	4.5515
mod. AIC(NW)	3.5544	3.8351	3.8012	3.8965	4.3633	4.5663	4.7830	4.5292	4.4382	4.5739
mod. AIC(diag)	3.6818	3.9146	3.6884	3.1524	2.6761	5.3329	5.4836	4.9307	4.4363	3.3783
iterated BIC	1.5772	1.5761	1.5750	1.5749	1.5722	2.0762	2.0741	2.0708	2.0671	2.0564
direct BIC	1.5772	1.6242	1.6727	1.5076	1.0992	2.0762	2.5768	2.5133	2.1432	1.4824
mod. BIC(NW)	1.4430	1.5965	1.7039	1.6698	1.4564	1.7677	2.1656	2.3122	2.0125	1.5315
mod. BIC(diag)	1.5772	1.6596	1.6283	0.9403	0.6774	2.0762	2.6165	2.5194	1.7473	0.9544
	Factors, w = 120					*Factors, w = 240*				
iterated AIC	5.6457	5.6479	5.6524	5.6632	5.6738	6.7936	6.7982	6.8044	6.8130	6.8327
direct AIC	5.6457	6.2418	6.6317	7.1535	7.4837	6.7936	7.4745	7.7807	8.2464	8.6145
mod. AIC(NW)	5.5718	5.9080	6.3140	6.9700	7.3650	6.4328	6.6605	6.9104	7.3851	7.9594
mod. AIC(diag)	5.6457	5.5882	5.3345	5.3025	4.9973	6.7936	6.7860	6.3810	6.1154	5.6833
iterated BIC	2.9691	2.9697	2.9714	2.9765	2.9852	3.6786	3.6838	3.6902	3.6995	3.7182
direct BIC	2.9691	3.5073	3.8466	4.4481	4.7807	3.6786	4.3675	4.6880	5.2442	5.6612
mod. BIC(NW)	2.8732	3.2199	3.5118	4.2826	4.7391	3.2386	3.6748	3.7356	4.3888	4.8309
mod. BIC(diag)	2.9691	2.9437	2.7650	2.5562	2.0400	3.6786	3.8031	3.5475	3.2871	2.4881

The table reports the average number of parameters chosen by the different model selection procedures, averaged over time and across all 170 variables in the Marcellino et al. (2006) data set. For the model selection methods see the footnote to Table 3.

Table 5: Forecasting performance measured relative to the iterated AR forecasting model selected by AIC

Forecast horizon	1	3	6	12	24	1	3	6	12	24
	AR, w = 120					*AR, w = 240*				
direct AIC	—	0.4353	0.4294	0.2294	0.1412	—	0.5294	0.5118	0.3647	0.1765
mod. AIC(NW)	0.5059	0.5235	0.4529	0.2471	0.1235	0.5118	0.5471	0.4941	0.3706	0.2059
mod. AIC(diag)	—	0.4235	0.4000	0.2412	0.1471	—	0.5529	0.5118	0.3706	0.1824
iterated BIC	0.6294	0.5294	0.5294	0.4765	0.4941	0.5118	0.5059	0.4059	0.3765	0.4059
direct BIC	0.6294	0.4235	0.4353	0.2176	0.1824	0.5118	0.4412	0.4235	0.3235	0.1882
mod. BIC(NW)	0.5706	0.4118	0.4294	0.2471	0.1765	0.4824	0.4176	0.3824	0.2882	0.1941
mod. BIC(diag)	0.6294	0.4588	0.4235	0.2353	0.1765	0.5118	0.4471	0.4353	0.2647	0.1529
	Factors, w = 120					*Factors, w = 240*				
iterated AIC	0.6118	0.6176	0.5353	0.5294	0.5059	0.6588	0.6471	0.6000	0.5647	0.5941
direct AIC	0.6118	0.5647	0.4412	0.3412	0.2059	0.6588	0.6647	0.5294	0.4176	0.4118
mod. AIC(NW)	0.6235	0.5706	0.4118	0.2941	0.2000	0.6824	0.6529	0.5294	0.4294	0.4294
mod. AIC(diag)	0.6118	0.5706	0.4353	0.3588	0.2588	0.6588	0.6529	0.5588	0.4824	0.4353
iterated BIC	0.6529	0.7059	0.6118	0.5412	0.4588	0.6824	0.6118	0.5647	0.5706	0.5647
direct BIC	0.6529	0.6235	0.4941	0.3412	0.2000	0.6824	0.6529	0.5353	0.4235	0.4235
mod. BIC(NW)	0.6647	0.6118	0.4882	0.3235	0.2235	0.6824	0.6353	0.5176	0.4588	0.4353
mod. BIC(diag)	0.6529	0.5941	0.5176	0.3059	0.2706	0.6824	0.6353	0.5353	0.4824	0.4294

The table reports the proportion of variables for which the iterated AR forecasts based on the model selected by AIC has a larger MSFE than the forecasts based on the respective information criteria. Hence values above 0.50 suggest that a particular method dominates the univariate iterated AIC forecasts for a majority of the variables. For the model selection methods see the footnote to Table 3.

Table 6: Pairwise model comparison tests

Forecast horizon	1	3	6	12	24	1	3	6	12	24
	AR	forecasts	against	iterated		AR	forecasts	against	iterated	
	AR(AIC), $w = 120$					AR(AIC), $w = 240$				
direct AIC	—	0.1588	0.1765	0.2118	0.3588	—	0.1176*	0.1824*	0.1647	0.1588
mod. AIC(NW)	0.1647*	0.1588	0.1471	0.2118	0.3176	0.1588*	0.1353*	0.1353*	0.1235	0.1647
mod. AIC(diag)	—	0.1647	0.1824	0.2765	0.3176	—	0.1176*	0.1706*	0.1765	0.1824
iterated BIC	0.1529*	0.2235	0.2412	0.1529	0.1706	0.2294	0.2941	0.2824	0.2353	0.1941
direct BIC	0.1529*	0.1765	0.1765	0.2647	0.3353	0.2294	0.2059	0.1235	0.1235	0.1471
mod. BIC(NW)	0.1471	0.2412	0.2000	0.2353	0.3176	0.2059	0.2647	0.1471	0.1882	0.1588
mod. BIC(diag)	0.1529*	0.1824	0.2294	0.4118	0.3000	0.2294	0.2059	0.1412	0.2235	0.2294
	Factor	forecasts	against	iterated		Factor	forecasts	against	iterated	
	AR(AIC), $w = 120$					AR(AIC), $w = 240$				
iterated AIC	0.1706*	0.2294*	0.2294	0.1765	0.1706	0.2294*	0.2353	0.2176	0.1765	0.1882*
direct AIC	0.1706*	0.1647	0.1412	0.1294	0.1353	0.2294*	0.2000	0.1706	0.1471	0.1706
mod. AIC(NW)	0.1824*	0.1412*	0.1529	0.1471	0.1529	0.2353*	0.1941*	0.1471	0.1588	0.1706
mod. AIC(diag)	0.1706*	0.1471*	0.1471	0.1118	0.1294	0.2294*	0.1882	0.1588	0.1529	0.1353
iterated BIC	0.1824*	0.2412	0.2059	0.1941	0.1882	0.2353*	0.2529	0.2824	0.2235	0.1824
direct BIC	0.1824*	0.1647	0.1706	0.1118	0.1235	0.2353*	0.2353*	0.1765	0.1588	0.1471
mod. BIC(NW)	0.1647*	0.1706	0.1647	0.1235	0.1412	0.2765*	0.2529*	0.1647	0.1294	0.1529
mod. BIC(diag)	0.1824*	0.1588	0.1294	0.1647	0.2059	0.2353*	0.2647*	0.1882	0.2235	0.2176

The table reports the results of applying the Giacomini and White (2006) test to the null that the forecast precision of the iterated AR based on AIC and the method in the respective row are identical. We show the proportion of rejections of the null at the 5% level. * (**) indicates that the iterated AR based on AIC has a smaller MSFE in less than 50% (25%) of the forecasts. The test statistics are calculated only for those periods where forecasts from all methods are available. For the model selection methods see the footnote to Table 3.

Table 7: Relative forecasting performance among variables in Category (A)–(D) (averages over 135 variables)

Forecast horizon	1	3	6	12	24	1	3	6	12	24
	AR, w = 120					*AR, w = 240*				
iterated AIC	1.0000	1.0000	1.0000	1.0000	1.0000	0.9776	0.9652	0.9470	0.9457	0.9720
direct AIC	1.0000	1.0230	1.0255	1.0928	1.2061	0.9776	0.9647*	0.9499	0.9841	1.0726
mod. AIC(NW)	0.9915*	1.0070	1.0200	1.0944	1.2252	0.9680*	0.9555*	0.9435*	0.9805	1.0771
mod. AIC(diag)	1.0000	1.0197	1.0244	1.1187	1.1781	0.9776	0.9644*	0.9497	0.9786	1.0837
iterated BIC	0.9809**	0.9844*	0.9915*	0.9944*	0.9868*	0.9695*	0.9611	0.9560	0.9560*	0.9731*
direct BIC	0.9809**	0.9975	1.0088	1.0721	1.1385	0.9695*	0.9595	0.9490	0.9731	1.0396
mod. BIC(NW)	0.9822*	0.9976	1.0150	1.0819	1.1461	0.9719*	0.9613	0.9525	0.9819	1.0456
mod. BIC(diag)	0.9809**	0.9994	1.0238	1.1721	1.1082	0.9695*	0.9583	0.9511	1.0172	1.0692
	Factors, w = 120					*Factors, w = 240*				
iterated AIC	0.9686*	0.9337*	0.9633*	0.9870*	0.9754	0.9351**	0.8859**	0.8911*	0.8744*	0.8751*
direct AIC	0.9686*	0.9473**	0.9889**	1.0713	1.2533	0.9351**	0.8779**	0.9120*	0.9800*	1.1020
mod. AIC(NW)	0.9683*	0.9435**	0.9969**	1.0828	1.2589	0.9349**	0.8749**	0.9089**	0.9767*	1.1048
mod. AIC(diag)	0.9686*	0.9488**	0.9944*	1.0729	1.1980	0.9351**	0.8749**	0.9049**	0.9601*	1.0557
iterated BIC	0.9643*	0.9356*	0.9521*	0.9667**	1.0352	0.9372**	0.8969*	0.9002*	0.8958*	0.9253*
direct BIC	0.9643*	0.9429**	0.9658**	1.0519	1.2192	0.9372**	0.8830**	0.9071**	0.9640*	1.0730
mod. BIC(NW)	0.9600**	0.9489*	0.9864*	1.0725	1.2291	0.9379*	0.8941*	0.9163*	0.9548*	1.0700
mod. BIC(diag)	0.9643*	0.9491**	0.9875	1.1089	1.2337	0.9372**	0.8843**	0.9027*	0.9666	1.0909

The table reports the MSFE of the different forecasts as a ratio of the MSFE of the iterated AR forecast based on the model selected by AIC with $w = 120$. The MSFE is calculated only for those periods where forecasts from all methods are available. For the model selection methods see the footnote to Table 3. * (**) indicates that the iterated AR based on AIC has a smaller MSFE in less than 50% (25%) of the forecasts that are significantly different according to the Giacomini-White (2006) test.

Table 8: Relative forecasting performance among variables in category (E) "Nominal prices, wages, and money" (averages over 35 variables)

Forecast horizon	1	3	6	12	24	1	3	6	12	24
	AR, w = 120					*AR, w = 240*				
iterated AIC	1.0000	1.0000	1.0000	1.0000	1.0000	0.9935	0.9542	0.9471	0.9890	1.1168
direct AIC	1.0000	0.9851*	0.9881*	1.0408	1.2903	0.9935	0.9444	0.9285	1.0063	1.2457
mod. AIC(NW)	1.0001	0.9849*	0.9899*	1.0461	1.3211	0.9964	0.9471	0.9320	1.0062	1.2441
mod. AIC(diag)	1.0000	0.9844*	0.9885	1.0516	1.3396	0.9935	0.9437	0.9276	1.0014	1.2692
iterated BIC	1.0212	1.1084	1.2117	1.2787	1.2745	1.0572	1.1938	1.3929	1.6096	1.7898
direct BIC	1.0212	1.0485	1.0448	1.0936	1.3066	1.0572	0.9982	0.9727	1.0551	1.3388
mod. BIC(NW)	1.0239	1.0409	1.0473	1.0883	1.3207	1.0559	1.0301	0.9759	1.0674	1.3342
mod. BIC(diag)	1.0212	1.0430	1.0516	1.1805	1.6596	1.0572	0.9955	0.9669	1.0626	1.6760
	Factors, w = 120					*Factors, w = 240*				
iterated AIC	1.0507	1.1695	1.3346	1.4491	1.4145	1.0969	1.3154	1.5669	1.7742	1.9435
direct AIC	1.0507	1.1385*	1.2858**	1.3614	1.3969	1.0969	1.2109	1.3299	1.3248	1.3001
mod. AIC(NW)	1.0437*	1.1339	1.2813**	1.3662	1.3961	1.0935	1.2229	1.3360	1.3377	1.2978
mod. AIC(diag)	1.0507	1.1325*	1.2461**	1.2753	1.4590	1.0969	1.2119	1.3367	1.3278	1.3342
iterated BIC	1.0280	1.1559	1.3167	1.4406	1.4221	1.0935	1.3407	1.6288	1.8809	2.0442
direct BIC	1.0280	1.1303	1.2392	1.3224	1.4204	1.0935	1.2396	1.3591	1.3585	1.3324
mod. BIC(NW)	1.0228	1.1119	1.2251	1.2968	1.4422	1.0928	1.2489	1.3644	1.3604	1.3558
mod. BIC(diag)	1.0280	1.1315	1.1911	1.3550	1.6727	1.0935	1.2335	1.3494	1.3826	1.6004

See footnote of Table 7.

CESifo Working Paper Series

for full list see www.cesifo-group.org/wp
(address: Poschingerstr. 5, 81679 Munich, Germany, office@cesifo.de)

_____

2513 Dirk Schindler and Benjamin Weigert, Educational and Wage Risk: Social Insurance vs. Quality of Education, December 2008

2514 Bernd Hayo and Stefan Voigt, The Relevance of Judicial Procedure for Economic Growth, December 2008

2515 Bruno S. Frey and Susanne Neckermann, Awards in Economics – Towards a New Field of Inquiry, January 2009

2516 Gregory Gilpin and Michael Kaganovich, The Quantity and Quality of Teachers: A Dynamic Trade-off, January 2009

2517 Sascha O. Becker, Peter H. Egger and Valeria Merlo, How Low Business Tax Rates Attract Multinational Headquarters: Municipality-Level Evidence from Germany, January 2009

2518 Geir H. Bjønnes, Steinar Holden, Dagfinn Rime and Haakon O.Aa. Solheim, ‚Large' vs. ‚Small' Players: A Closer Look at the Dynamics of Speculative Attacks, January 2009

2519 Jesus Crespo Cuaresma, Gernot Doppelhofer and Martin Feldkircher, The Determinants of Economic Growth in European Regions, January 2009

2520 Salvador Valdés-Prieto, The 2008 Chilean Reform to First-Pillar Pensions, January 2009

2521 Geir B. Asheim and Tapan Mitra, Sustainability and Discounted Utilitarianism in Models of Economic Growth, January 2009

2522 Etienne Farvaque and Gaël Lagadec, Electoral Control when Policies are for Sale, January 2009

2523 Nicholas Barr and Peter Diamond, Reforming Pensions, January 2009

2524 Eric A. Hanushek and Ludger Woessmann, Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation, January 2009

2525 Richard Arnott and Eren Inci, The Stability of Downtown Parking and Traffic Congestion, January 2009

2526 John Whalley, Jun Yu and Shunming Zhang, Trade Retaliation in a Monetary-Trade Model, January 2009

2527 Mathias Hoffmann and Thomas Nitschka, Securitization of Mortgage Debt, Asset Prices and International Risk Sharing, January 2009

2528 Steven Brakman and Harry Garretsen, Trade and Geography: Paul Krugman and the 2008 Nobel Prize in Economics, January 2009

2529 Bas Jacobs, Dirk Schindler and Hongyan Yang, Optimal Taxation of Risky Human Capital, January 2009