Variable Selection, Estimation and Inference for Multi-period Forecasting Problems*

M. Hashem Pesaran
Cambridge University and USC

Andreas Pick
Erasmus University Rotterdam, De Nederlandsche Bank and CIMF

Allan Timmermann
UC San Diego and CREATES

24th April 2010

Abstract

This paper conducts a broad-based comparison of iterated and direct multi-period forecasting approaches applied to both univariate and multivariate models in the form of parsimonious factor-augmented vector autoregressions. To account for serial correlation in the residuals of the multi-period direct forecasting models we propose a new SURE-based estimation method and modified Akaike information criteria for model selection. Empirical analysis of the 170 variables studied by Marcellino, Stock and Watson (2006) shows that information in factors helps improve forecasting performance for most types of economic variables although it can also lead to larger biases. It also shows that finite-sample modifications to the Akaike information criterion can modestly improve the performance of the direct multi-period forecasts.

JEL Classifications: C22, C32, C52, C53
Key Words: Multi-period forecasts, direct and iterated methods, factor augmented VARs

1 Introduction

Economists are commonly asked to forecast uncertain outcomes multiple periods ahead in time. For example, when the economy is in a recession a policy maker may wish to know how long it might take before a recovery begins, and so is interested in forecasts of output growth for, say, horizons of 1, 3, 6, 12, and 24 months ahead. 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, and stock market investors may consider the effect of demographic variables on expected returns and risks at both short and long investment horizons (Favero and Tamoni, 2010).

Two basic strategies have been proposed 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 forecasts at longer horizons. Under this "iterated" or "indirect" approach, the model specification is the same across all forecast horizons; only the number of iterations changes with the horizon. Univariate ARMA models or their multivariate VARMA equivalent, are usually used in the iterations. The second approach is to estimate a separate model for each horizon, regressing future realizations on current information. Such "direct" forecasts dispense with the need for forward iteration. Under this approach, both the model specification and estimates can vary across different forecast horizons.

Both approaches have advantages and drawbacks. For a given model specification the iterated approach leads to more efficient parameter estimates since it includes data recorded at the highest available frequency and so uses the largest available sample size. If the model is misspecified, due, for example, to an omitted variable or because of an incorrect lag order, iterating the model multiple steps ahead can either attenuate or reduce 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 when the forecast horizon exceeds one which affects the covariance of the forecast errors.

Given the importance of the horizon to many forecasting problems, it is not surprising that a substantial theoretical literature has considered the multi-step forecasting problem, including Bao (2007), Cox (1961), Brown and Mariano (1989), Clements and Hendry (1998), Findley (1983), Hoque, Magnus and Pesaran (1988), Ing (2003), Schorfheide (2005), and Ullah (2004), with Bhansali (1999) providing a survey. This literature has examined the bias-efficiency trade-off in the context of specific models such as stationary first-order or higher-order autoregressive models. Whether the direct or iterated approach can be expected to produce the best forecasts

will generally depend on the sample size, forecast horizon, the (unknown) underlying data generating process and the methods used to select lag length for the forecasting models (Ing, 2003). In general, no approach can be shown to uniformly dominate the other so, as pointed out by Marcellino, Stock and Watson (2006) (MSW, henceforth), the relative merit of the iterated versus direct forecasting methods is an empirical question.

For multivariate forecasting models additional issues complicate the comparison of the direct and iterated forecasting approaches. First, it becomes important how the potentially high-dimensional variable selection search is conducted and how multi-step forecasts of additional predictor variables are generated under the iterated approach. 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. In addition to past values of the predicted variable itself, relatively few potential predictors therefore need to be considered.

A second issue that has not previously received much attention in this context is the serial correlation in the errors of the direct forecast models, which arises due to the use of overlapping data. This raises issues at the estimation and the model selection stage. We propose a SURE estimation approach that reorganizes the data in non-overlapping blocks of observations spaced apart by the length of the forecast horizon. We show how to compute the resulting covariance matrix under this approach which holds the potential of efficiency gains over conventional direct forecasts. We also propose modifications to the Akaike information criterion that account for serial correlation in residuals from the forecast models. Monte Carlo simulations confirm that the modifications to the AIC and the SURE approach both lead to improvements in the performance of the direct forecasting models.

In an empirical exercise we consider the 170 variables studied by MSW. We confirm their finding that the iterated forecasts are best overall among the univariate forecasting methods, particularly at long horizons where the inefficiency of the direct forecasting method is most prominent. We also find that forecasts generated by the factor-augmented VARs generally perform better than the univariate forecasts, an 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 the multivariate information that is embedded in common factors. Among the direct forecasts, in the majority of cases, both, the modified Akaike information

3

criteria and the SURE approach, help improve forecasting performance.

In summary, the main contributions of the paper are as follows: First, we propose a factor-augmented forecasting approach that extends to the multivariate setting iterated methods previously considered in a univariate context. Second, we propose a new SURE estimation method that accounts for the data overlap that arises at multi-period horizons under the direct forecast approach. Third, we extend the AIC 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. Fourth, we study the forecasting performance of both extant and new model selection and forecasting methods through Monte Carlo simulations. Finally, we present an empirical application that considers recursively generated forecasts of the economic variables included in the study by MSW and extends their study to a multivariate setting.

The outline of the paper is as follows. Section 2 sets up the multi-period forecasting problem for univariate and multivariate cases, while Section 3 deals with model selection and estimation issues. Section 4 presents Monte Carlo results, while Section 5 describes our empirical findings using the Marcellino et al. data set. Section 6 concludes.

2 Methods for Multi-period Forecasting

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

The forecaster's horizon, h, may be a single period, $h = 1$, or may involve several periods, $h > 1$. Iterated forecasts use a single model fitted to the shortest horizon and then iterate on this model to obtain multi-step forecasts. Direct forecasts regress realizations h periods into the future on current information and estimate and select a separate forecasting model for each horizon.

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 $\boldsymbol{y}_t$.

4

2.1 Multi-step Forecasts with Factor-Augmented VARs

In cases where M is relatively small one approach is to treat all variables simultaneously, i.e. model $(\boldsymbol{y}_t', \boldsymbol{x}_t')'$ jointly. Multi-period forecasts of $\mathbf{y}_t$ can then be obtained by iterating on a VAR of the form

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{x}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix} + \Lambda(L) \begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{x}_{t-1} \end{pmatrix} + \psi_t, \tag{1}$$

where $\Lambda(L)$ is a matrix lag polynomial of finite order. In the common situation where M is large while the time-series dimension of the data is limited, this approach is unlikely to be successful due to the high dimension of $\Lambda(L)$, particularly the parts tracking dynamics in the large-dimensional vector $\mathbf{x}_t$.

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

$$\boldsymbol{z}_t = \boldsymbol{\mu}_z + \begin{pmatrix} \boldsymbol{A}_p(L) & \boldsymbol{B}_q(L) \\ \mathbf{0} & \boldsymbol{D}_s(L) \end{pmatrix} \boldsymbol{z}_{t-1} + \boldsymbol{\xi}_t, \tag{2}$$

where the finite-order matrix lag polynomials are

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

Notice the asymmetric treatment of $\boldsymbol{y}_t$ and $\boldsymbol{f}_t$ under this approach: 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.

For illustration, suppose that the $K \times 1$ vector of target variables, $\boldsymbol{y}_t$, is generated according to the following factor model

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{B}\boldsymbol{f}_{t-1} + \boldsymbol{u}_t, \quad t = 1, ..., T, \tag{3}$$

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

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

Using this model to predict $\boldsymbol{y}_{t+h}$ given information at time t requires a forecast of the factors whenever $h \geq 2$. 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). A key question when using factor models is the choice of the number of factors. This can be determined, for example, by using the information criteria (IC) proposed by Bai and Ng (2002). In practice there is considerable uncertainty surrounding the number of factors to be used, and as pointed out by Bai and Ng (2009), the use of IC is based on the assumption 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.

In view of these concerns we adopt a hierarchical approach where we first divide all variables into economically distinct groups and then select the first PC from each of the categories. All the computations are carried out recursively with rolling estimation windows of length w, so no future information is used in the construction of the factors. We denote the recursively estimated PCs by $\hat{\boldsymbol{f}}_t$, $t = R, \ldots, T - h$, where t is the point in time where the factors are computed, $R \geq w$ is the length of the initial estimation sample, and T is the total sample length. Hence, at time t we use data over the sample $t - w + 1 : t$ to extract $\hat{\boldsymbol{f}}_t$.

Only those factors that help predict $\boldsymbol{y}_t$ are relevant and should be included in the model. This may be a subset, $\hat{\boldsymbol{f}}_{1t}$, of the full set of factors, $\hat{\boldsymbol{f}}_t$, under consideration. One then has to choose whether to use the full set of factors $\hat{\boldsymbol{f}}_t = (\hat{\boldsymbol{f}}_{1t}, \hat{\boldsymbol{f}}_{2t})$ to predict the subset of factors, $\hat{\boldsymbol{f}}_{1t}$, selected when forecasting $\boldsymbol{y}_{t+h}$, or whether to use only lagged values of $\hat{\boldsymbol{f}}_{1t}$ 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 use all factors to forecast future values of $\hat{\boldsymbol{f}}_{1t}$.

2.2 Iterated System Forecasts

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

$$\boldsymbol{y}_t = \boldsymbol{\mu}_y + \boldsymbol{A}_p(L)\boldsymbol{y}_{t-1} + \boldsymbol{B}_q(L)\hat{\boldsymbol{f}}_{t-1} + \boldsymbol{u}_t. \tag{4}$$

We determine the lag orders p and q and the subset of m_1 factors from the total set of m factors by applying IC to the likelihood of $\boldsymbol{y}_t$.

For simplicity, we obtain forecasts only from the full set of $m = m_1 + m_2$ factors. Hence we first select a VAR(s) model in $\hat{\boldsymbol{f}}_t$ where the value of s is determined by some IC, again applied recursively:

$$\hat{\boldsymbol{f}}_t = \boldsymbol{\mu}_f + \boldsymbol{D}_s(L)\hat{\boldsymbol{f}}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{5}$$

Next, to compute h−step-ahead forecasts of $\boldsymbol{y}_t$, we simply combine the conditional and marginal models:

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

which, recalling that $\boldsymbol{z}_t = (\boldsymbol{y}'_t, \hat{\boldsymbol{f}}'_t)'$ and $\boldsymbol{\xi}_t = (\boldsymbol{u}'_t, \boldsymbol{\nu}'_t)'$, is consistent with equation (2). Notice that the selection of factors in the conditional model (4) is reflected in zero-restrictions on $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, ... , $\boldsymbol{B}_q$, where the columns corresponding to the factors that are not selected are set to zeros. The factor-augmented VAR in $\boldsymbol{z}_t$ can then readily be iterated forward.[1]

2.3 Univariate Forecasts

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

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

$$\hat{y}_{t+h} = \hat{\alpha} + \hat{\beta}_1 \hat{y}_{t+h-1} + \hat{\beta}_2 \hat{y}_{t+h-2} + \ldots + \hat{\beta}_p \hat{y}_{t+h-p}, \quad t = R, \ldots, T - h$$

where $\hat{y}_{t+h-p} = y_{t+h-p}$ if $h \leq p$ and the parameters are estimated from the regression model by OLS

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + u_t, \quad t = R, \ldots, T - h.$$

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

$$y_t = \alpha_h + \beta_{1h} y_{t-h} + \beta_{2h} y_{t-h-1} + \ldots + \beta_{ph} y_{t-h-p} + u_{ht},$$

and the forecast is

$$\hat{y}_{t+h} = \hat{\alpha}_h + \hat{\beta}_{1h} y_t + \hat{\beta}_{2h} y_{t-1} + \ldots + \hat{\beta}_{ph} y_{t-p}.$$

3 Estimation and Model Selection

Two important econometric issues arise in the context of multi-period forecasting. First, the direct forecast models introduce overlaps in the observations and give rise to a particular dependence structure which, if imposed on the estimation, could lead to efficiency gains. Second, for both iterative and direct forecasts, how a particular model is chosen by the forecaster is of great importance given the large dimension of the set of potentially relevant predictor variables. In this section we discuss both issues, first presenting a new SURE estimation procedure that may lead to efficiency gains and next considering a variety of information criteria, including modified ones that deal with serial dependence in the errors.

---

[1] Alternatively, one could model only the factors that have been selected in the conditional model, (4). This may be less efficient than using the full VAR in (5), 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.

3.1 Estimation

Overlaps in the data associated with the direct forecasting models introduce serial dependence in the errors. Even if the underlying errors are serially uncorrelated, the errors associated with an h-period overlap typically follow an $MA(h-1)$ process. This suggests that estimating VARMA models could be beneficial. However, we do not follow this direction for two reasons. First, estimation of VARMA models is not a very common undertaking in the forecasting literature and thus goes against our focus on evaluating forecasting methods in common use. Second, VARMA models have stability and convergence problems for the types of multivariate models considered in our paper, which can be of large dimension, and require extensive specification searches; see Athanasopoulos and Vahid (2008) for further discussion of these points.

Instead, we propose using SURE estimation which leads to some efficiency gains as it exploits the information of the MA structure of the error even if it implies a more heavily parameterized model. Consider estimation of the following direct forecasting model

$$y_t = \boldsymbol{\beta}' \mathbf{z}_{t-h} + u_t, \ t = R, \ldots, T, \tag{6}$$

and suppose that u_t follows an $(h-1)$ order moving average process

$$u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_{h-1} \varepsilon_{t-h+1}, \quad \varepsilon_t \sim iid(0, \sigma^2). \tag{7}$$

The regressors, $\mathbf{z}_{t-h}$, include $y_{t-h}, y_{t-h-1}, \ldots, y_{t-h-p}$ for some order p, and estimated factors dated $t-h$ or earlier, i.e. $\hat{\boldsymbol{f}}_{t-h}, \ldots, \hat{\boldsymbol{f}}_{t-h-p}$. When the direct regression is derived from an underlying $VAR(p)$ in y_t and $\hat{\boldsymbol{f}}_t$ the regression coefficients, $\boldsymbol{\beta}$, and the MA coefficients $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{h-1})'$ are related, and fully efficient estimation of (6) must allow for such cross-parameter restrictions. This type of restrictions can be implemented by assuming that $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are both functions of a set of deeper parameters, $\boldsymbol{\phi}$, with estimation of $\boldsymbol{\beta}(\boldsymbol{\phi})$ and $\boldsymbol{\theta}(\boldsymbol{\phi})$ carried out directly in terms of $\boldsymbol{\phi}$. To simplify the notations, we suppress the explicit dependence of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ on $\boldsymbol{\phi}$.

Imposing such restrictions on the system of direct forecasting models simply helps recover the original parameter values from the iterated forecasting model. Since the latter efficiently uses data at the highest frequency, maximum likelihood estimates from joint estimation of the mutually consistent h-step direct forecasting models is identical to estimation of the parameter estimates from the one-step-ahead model, which of course is far more easily achieved. In scenarios where the direct forecast would dominate the iterated forecast, e.g., because the forecasting model is misspecified, using the parameters of the iterated forecast model for the covariance matrix may be undesirable. As for the model parameters, it may therefore be better

8

to estimate the covariance matrix directly without imposing the parameter restrictions on the iterated forecast.

Asymptotically efficient estimates of $\boldsymbol{\beta}$ can be computed by applying maximum likelihood directly to the overlapping regressions for $t = 1, 2, \ldots, T$, allowing for the MA$(h-1)$ process of the errors. Alternatively, one could consider estimating $\boldsymbol{\beta}$ from pooled regressions of h non-overlapping regressions. Suppose that T is an exact multiple of h and set $n = T/h$. Let $\tilde{z}_{ij} = y_{j+(i-1)h}$ and decompose the overlapping regressions in (6) into the following h non-overlapping regressions

$$\tilde{z}_{ij} = \boldsymbol{\beta}'\mathbf{w}_{i-1,j} + v_{ij}, \text{ for } i = 1, 2, \ldots, n, \text{ and } j = 1, 2, \ldots, h, \qquad (8)$$

where $\mathbf{w}_{i-1,j} = \mathbf{w}_{i,j-h} = \mathbf{x}_{j+(i-1)h-h}$, and $v_{ij} = u_{j+(i-1)h}$. When $h = 2$, we have two non-overlapping regressions: one for the odd observations, $\tilde{z}_{i1}$, and another for the even observations, $\tilde{z}_{i2}$, for $i = 1, 2, \ldots, n$.

For each j the errors v_{ij} are serially uncorrelated across i and the least squares regression of $\tilde{\mathbf{z}}_j = (\tilde{z}_{1,j}, \tilde{z}_{2,j}, \ldots, \tilde{z}_{n,j})'$ on $\mathbf{W}_j = (\mathbf{w}_{0,j}, \mathbf{w}_{1,j}, \ldots, \mathbf{w}_{n-1,j})'$ yields a consistent estimate of $\boldsymbol{\beta}$ which we denote by $\hat{\boldsymbol{\beta}}_j$. However, this estimate is not efficient and a pooled estimate that utilizes all the h non-overlapping regressions can be more efficient. The estimates across the h non-overlapping regressions can be pooled in a number of ways, e.g., with a simple or a weighted average of $\hat{\boldsymbol{\beta}}_j$, over $j = 1, 2, \ldots, h$, with weights based on the relative precision of the different estimates.

Alternatively, the h regressions in (8) can be viewed as a set of seemingly unrelated regression equations (SURE), allowing for the cross dependence of the errors v_{ij} across j for each i. Specifically, consider the regressions

$$\tilde{\mathbf{z}}_j = \mathbf{W}_j\boldsymbol{\beta} + \mathbf{v}_j, \text{ for } j = 1, 2, \ldots, h, \qquad (9)$$

which in stacked form can be written as

$$\tilde{\mathbf{z}} = \mathbf{W}\boldsymbol{\beta} + \mathbf{v},$$

where $\tilde{\mathbf{z}} = (\tilde{\mathbf{z}}_1', \tilde{\mathbf{z}}_2', \ldots, \tilde{\mathbf{z}}_h')'$, $\mathbf{W} = (\mathbf{W}_1', \mathbf{W}_2', \ldots, \mathbf{W}_h')'$, and $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2', \ldots, \mathbf{v}_h')'$. To derive the covariance matrix of $\mathbf{v}$ we first note that $\mathbf{v}_j = (u_j, u_{j+h}, u_{j+2h}, \ldots, u_{j+(n-1)h})'$, and denote the autocovariance of $\{u_t\}$, which follows an MA$(h-1)$ process, by $\gamma(s)$, where $\gamma(s) = \gamma(-s) = 0$ for $s \geq h$. It is now easily seen that

$$E(\mathbf{v}_j\mathbf{v}_j') = \gamma(0)\mathbf{I}_n,$$

where $\mathbf{I}_n$ is an identity matrix of order n. Similarly, if $s > r$,

$$E(\mathbf{v}_r\mathbf{v}_s') \equiv \boldsymbol{\Psi}_{rs} =$$

$$\left(\begin{array}{cccccc} \gamma(s-r) & 0 & 0 & \cdots & 0 & 0 \\ \gamma(h-s+r) & \gamma(s-r) & 0 & \cdots & 0 & 0 \\ 0 & \gamma(h-s+r) & \gamma(s-r) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \gamma(h-s+r) & \gamma(s-r) & 0 \\ 0 & 0 & \cdots & 0 & \gamma(h-s+r) & \gamma(s-r) \end{array}\right),$$

while if $r > s$,

$$E(\mathbf{v}_r \mathbf{v}_s') \equiv \boldsymbol{\Psi}_{rs} =$$

$$\left(\begin{array}{cccccc} \gamma(r-s) & \gamma(h-s+r) & 0 & \cdots & 0 & 0 \\ 0 & \gamma(r-s) & \gamma(h-r-s) & \cdots & 0 & 0 \\ 0 & 0 & \gamma(r-s) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \gamma(r-s) & \gamma(h-r+s) \\ 0 & 0 & \cdots & 0 & 0 & \gamma(r-s) \end{array}\right).$$

Therefore, the covariance matrix of $\mathbf{v}$ is given by

$$\boldsymbol{\Sigma}_v(\boldsymbol{\theta}) = \left(\begin{array}{cccccc} \gamma(0)\mathbf{I}_n & \boldsymbol{\Psi}_{12} & \boldsymbol{\Psi}_{13} & \cdots & & \boldsymbol{\Psi}_{1h} \\ \boldsymbol{\Psi}_{21} & \gamma(0)\mathbf{I}_n & \boldsymbol{\Psi}_{23} & \cdots & & \\ \vdots & \vdots & \vdots & & & \\ \boldsymbol{\Psi}_{h-1,1} & \boldsymbol{\Psi}_{h-1,2} & & \cdots & \gamma(0)\mathbf{I}_n & \boldsymbol{\Psi}_{h-1,h} \\ \boldsymbol{\Psi}_{h,1} & \boldsymbol{\Psi}_{h,2} & & \cdots & \boldsymbol{\Psi}_{h,h-1} & \gamma(0)\mathbf{I}_n \end{array}\right). \qquad (10)$$

The log-likelihood function of the SURE specification is now given by

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) \propto -\frac{1}{2}\ln|\boldsymbol{\Sigma}_v(\boldsymbol{\theta})| - \frac{1}{2}(\tilde{\mathbf{z}} - \mathbf{W}\boldsymbol{\beta})'\boldsymbol{\Sigma}_v^{-1}(\boldsymbol{\theta})(\tilde{\mathbf{z}} - \mathbf{W}\boldsymbol{\beta}). \qquad (11)$$

The unknown parameters can be obtained by maximization of the log-likelihood function. In particular

$$\hat{\boldsymbol{\beta}}_{SURE} = \left[\mathbf{W}'\boldsymbol{\Sigma}_v^{-1}(\boldsymbol{\theta})\mathbf{W}\right]^{-1}\mathbf{W}'\boldsymbol{\Sigma}_v^{-1}(\boldsymbol{\theta})\tilde{\mathbf{z}}. \qquad (12)$$

This maximum likelihood procedure is equivalent to the maximum likelihood estimation of the original overlapping regression in equation (6) with errors following the MA$(h-1)$ process, (7). To see this let $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$, and note that since the elements of $\mathbf{z}$ are selected from the elements of $\mathbf{y}$ without repetition then there exists a non-singular $T \times T$ selection matrix $\mathbf{P}$ such that $\tilde{\mathbf{z}} = \mathbf{P}\mathbf{y}$. Hence the log-likelihood functions based on $\mathbf{y}$ and $\tilde{\mathbf{z}}$ must be the same, and the covariance matrix in (10) is equivalent to using a GLS covariance matrix after reordering the observations.

Maximum likelihood estimation of $\boldsymbol{\beta}$ under the SURE approach suggests intermediate procedures that are computationally less extensive and relatively easy to implement. One possible approach would be to estimate

$\boldsymbol{\Sigma}_v(\boldsymbol{\theta})$ using consistent estimates of $\gamma(s)$ obtained from $\hat{u}_t = y_t - \hat{\boldsymbol{\beta}}'_{LS}\mathbf{z}_{t-h}$ where $\hat{\boldsymbol{\beta}}_{LS}$ is the first-stage least squares estimates of $\boldsymbol{\beta}$ computed from the overlapping regressions. Hence[2]

$$\hat{\gamma}(s) = \frac{\sum\limits_{t=1}^{T} \hat{u}_t \hat{u}_{t-s}}{T}, \text{ for } s = 0, 1, 2, \ldots, h.$$

Alternatively, we could apply the standard SURE estimation to (9) subject to the restrictions $\boldsymbol{\beta}_j = \boldsymbol{\beta}$ for all $j = 1, 2, \ldots, h$ and use the covariance matrix $\boldsymbol{\Sigma}_h \otimes \boldsymbol{I}_{T/h}$, where $\boldsymbol{\Sigma}_h$ is the covariance matrix of the SURE system (9) without accounting for the restrictions in (10). This procedure provides an efficient way of pooling the h different consistent estimates of $\boldsymbol{\beta}$ but does not take account of the specific form of the cross dependence of the errors from the different non-overlapping regressions. Monte Carlo experiments not reported here but available from the authors suggests that this leads to an inferior forecast performance and we will therefore not consider this approach.

In the Monte Carlo simulations and empirical analysis we use the first approach and the likelihood function (11) to calculate the information criteria. In order to obtain robust estimates of $\gamma(s)$ we estimate them nonparametrically using a Bartlett window. Additionally, we down-weigh the covariances $\gamma(s)$ by $(1 - \frac{s}{h})$. To robustify the calculation of the forecasts, one can use the SURE approach in the model selection stage, but then use OLS estimation for the selected model. Numbers reported in subsequent tables are based on this combination, but results are very similar without this modification.

3.2 Model selection

We consider two model selection criteria, namely the AIC and BIC. Both are commonly used in forecasting studies and have well known properties: AIC achieves a good approximate model as the sample size expands even if the true model is not contained in the universe of models under consideration. However, it is not a consistent criterion and so does not select the true model with probability one, asymptotically, if it happens to be included in the search. In contrast, BIC is a consistent model selection criterion if the true model is one of the models under consideration.

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.

---

[2]For relatively large values of h, the estimates of $\hat{\gamma}(s)$ can be down-weighted using Bartlett of Parzen windows.

3.2.1 Iterated Forecasts

Individual models in the specification search select different subsets of $\mathbf{z}_{t-1} = (\mathbf{y}_{t-1}'\ \hat{\mathbf{f}}_{t-1}')'$. We consider the following criteria. First, we apply AIC or BIC recursively to the single equation $(K = 1)$ containing the variable of interest, i.e., at time t, for a particular model we have:

$$
\begin{aligned}
\text{AIC}_t &= \ln[\hat{\boldsymbol{u}}_{t-w+2:t}'\hat{\boldsymbol{u}}_{t-w+2:t}/(w-1)] + \frac{2d}{w-1}, & (13) \\
\text{BIC}_t &= \ln[\hat{\boldsymbol{u}}_{t-w+2:t}'\hat{\boldsymbol{u}}_{t-w+2:t}/(w-1)] + \frac{d\ \ln(w-1)}{w-1},
\end{aligned}
$$

where d is the dimension of the vector containing the subset of $\mathbf{z}_{t-1}$ selected by the model under consideration, $\hat{\boldsymbol{u}}_{t-w+2:t}'$ is a $(w-1)\times 1$ vector of estimated residuals from the rolling-window estimation with typical element $\hat{u}_t = y_{1,t} - \hat{\boldsymbol{\beta}}_t'\mathbf{z}_{t-1}$, and $\hat{\boldsymbol{\beta}}_t$ is estimated using the most recent w observations up to period t.[3]

Similarly, for $K > 1$, we have

$$
\begin{aligned}
\text{AIC}_t &= \ln|\hat{\boldsymbol{\Sigma}}_t| + \frac{2Kd}{w-1} & (14) \\
\text{BIC}_t &= \ln|\hat{\boldsymbol{\Sigma}}_t| + \frac{Kd\ \ln(w-1)}{w-1}
\end{aligned}
$$

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

$$
\hat{\boldsymbol{\Sigma}}_t = \frac{1}{w-1}\left[\mathbf{Y}_t'\mathbf{Y}_t - \mathbf{Y}_t'\mathbf{Z}_{t-1}(\mathbf{Z}_{t-1}'\mathbf{Z}_{t-1})^{-1}\mathbf{Z}_{t-1}'\mathbf{Y}_t\right],
$$

$\mathbf{Y}_t$ is the $(w-1) \times K$ matrix of stacked $\boldsymbol{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Kt})'$, and $\mathbf{Z}_{t-1}$ is the observation matrix formed by stacking $\boldsymbol{z}_{t-1}$ over the $w-1$ observations indexed by $t - w + 2, t - w + 3, \ldots, t$.

3.2.2 Direct Forecasts

The direct forecasting models that are not based on the SURE system (9) are selected based on either AIC or BIC

$$
\begin{aligned}
\text{AIC}_t &= \ln[\hat{\boldsymbol{u}}_{t-w+1+h:t}'\hat{\boldsymbol{u}}_{t-w+1+h:t}/(w-h)] + \frac{2d}{w-h}, \\
\text{BIC}_t &= \ln[\hat{\boldsymbol{u}}_{t-w+1+h:t}'\hat{\boldsymbol{u}}_{t-w+1+h:t}/(w-h)] + \frac{d\ \ln(w-h)}{w-h}, & (15)
\end{aligned}
$$

where $\hat{\boldsymbol{u}}_{t-w+1+h:t}$ is a $(w-h) \times 1$ vector of estimated residuals with typical element $\hat{u}_t = y_{1t} - \hat{\boldsymbol{\beta}}_t'\boldsymbol{z}_{t-h}$. For $h > 1$, the overlap in the forecasts will

---

[3]For simplicity, we shall refer to $\mathbf{z}_{it-1}$ as comprising both the most recent lag as well as any additional lags. This notation therefore corresponds to using the companion form of the model.

produce 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 for the case where $K = 1$. To motivate these, assume that the $h-$step forecast model takes the form in equation (6) estimated using the observations $t = 1, 2, \ldots, T$, and denote the least squares criterion for estimating $\boldsymbol{\beta}$ by

$$Q(\boldsymbol{\beta}) = \frac{1}{2\sigma_u^2}(\boldsymbol{y}_T - \boldsymbol{Z}_{T,h}\boldsymbol{\beta})'(\boldsymbol{y}_T - \boldsymbol{Z}_{T,h}\boldsymbol{\beta}),$$

where $\sigma_u^2 = Var(u_t)$, $\boldsymbol{y}_T = (y_1, y_2, \ldots, y_T)'$, and $\boldsymbol{Z}_{T,h} = (\boldsymbol{z}_{1-h}, \boldsymbol{z}_{2-h}, \ldots, \boldsymbol{z}_{T-h})'$. Let $Q_0(\boldsymbol{\beta}) = \mathrm{E}[Q(\boldsymbol{\beta})]$, where expectations are taken with respect to the true conditional density of $\mathbf{y}_t$, and denote the jth derivative of $Q(\boldsymbol{\beta})$ by $Q^{(j)}(\boldsymbol{\beta})$. Moreover, let $\boldsymbol{\beta}_0$ be the true value of $\boldsymbol{\beta}$, and its estimate given by $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} [Q(\boldsymbol{\beta})]$. Using these notations, we have that $Q_0^{(1)}(\boldsymbol{\beta}_0) = 0$, $Q^{(1)}(\hat{\boldsymbol{\beta}}) = 0$, $Q^{(2)}(\boldsymbol{\beta}) = Q_0^{(2)}(\boldsymbol{\beta}_0) = \frac{1}{\sigma_u^2}\boldsymbol{Z}'_{T,h}\boldsymbol{Z}_{T,h}$.

Define $Q_0(\hat{\boldsymbol{\beta}})$ as the loss incurred by using the estimated parameter $\hat{\boldsymbol{\beta}}$ instead of the unknown true parameter, $\boldsymbol{\beta}_0$. Then

$$\mathrm{E}[Q_0(\hat{\boldsymbol{\beta}})] = \mathrm{E}[Q(\hat{\boldsymbol{\beta}})] + \mathrm{E}[Q_0(\hat{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}})].$$

Second order Taylor expansions yield

$$\begin{aligned}
Q_0(\hat{\boldsymbol{\beta}}) &= Q_0(\boldsymbol{\beta}_0) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'Q_0^{(2)}(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \\
Q(\hat{\boldsymbol{\beta}}) &= Q(\boldsymbol{\beta}_0) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'Q^{(1)}(\boldsymbol{\beta}_0) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'Q_0^{(2)}(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) .
\end{aligned}$$

Taking expectations of the last equation yields

$$\begin{aligned}
\mathrm{E}[Q(\hat{\boldsymbol{\beta}})] &= \mathrm{E}[Q(\boldsymbol{\beta}_0)] + \mathrm{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'Q^{(1)}(\boldsymbol{\beta}_0)] + \mathrm{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'Q_0^{(2)}(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)] \\
&= Q_0(\boldsymbol{\beta}_0) + \mathrm{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'Q^{(1)}(\boldsymbol{\beta}_0)] + \mathrm{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'Q_0^{(2)}(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)],
\end{aligned}$$

and therefore

$$\mathrm{E}[Q_0(\hat{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}})] = -\mathrm{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'Q^{(1)}(\boldsymbol{\beta}_0)].$$

Furthermore

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

and therefore (for a given set of regressors)

$$\begin{aligned}
-\mathrm{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'Q^{(1)}(\boldsymbol{\beta}_0)] &= -\mathrm{E}[\mathrm{tr}(\sigma_u^{-2}\boldsymbol{u}'_T\boldsymbol{Z}_{T,h}(\boldsymbol{Z}'_{T,h}\boldsymbol{Z}_{T,h})^{-1}\boldsymbol{Z}_{T,h}\boldsymbol{u}_T)] \\
&= -\mathrm{tr}\left[\sigma_u^{-2}(\boldsymbol{Z}'_{T,h}\boldsymbol{Z}_{T,h})^{-1}\boldsymbol{Z}'_{T,h}\mathrm{E}\left(\boldsymbol{u}_T\boldsymbol{u}'_T\right)\boldsymbol{Z}_{T,h}\right],
\end{aligned}$$

where $\boldsymbol{u}_T = (u_1, u_2, \ldots, u_T)'$. If the errors $\boldsymbol{u}_t$ 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 in small samples.

Building on this result, we first consider a band diagonal modified AIC that takes the form

$$\text{AIC}_{\hat{\boldsymbol{\Pi}}_S} = \ln[\hat{\boldsymbol{u}}'_{t-w+1+h:t}\hat{\boldsymbol{u}}_{t-w+1+h:t}/(w-h)] + \frac{2\text{tr}\left(\hat{\boldsymbol{\Pi}}_S\right)}{w-h}, \qquad (16)$$

where

$$\hat{\boldsymbol{\Pi}}_S = (\boldsymbol{Z}'_{T,h}\boldsymbol{Z}_{T,h})^{-1}\boldsymbol{Z}'_{T,h}\boldsymbol{S}_h\boldsymbol{Z}_{T,h}/h.$$

$\boldsymbol{S}_h$ 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 etc., i.e.

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

This formulation aims at capturing the $MA(h-1)$ form of the error process in the overlapping regressions but assumes that the serial correlation in 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:

$$\text{AIC}_{\tilde{\boldsymbol{\Pi}}} = \ln[\hat{\boldsymbol{u}}'_{t-w+1+h:t}\hat{\boldsymbol{u}}_{t-w+1+h:t}/(w-h)] + \frac{2\text{tr}\left(\tilde{\boldsymbol{\Pi}}\right)}{w-h}, \qquad (17)$$

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

When selecting the models based on the SURE system in (9), standard formulations of AIC and BIC with likelihood (11) can be used. The unknown parameters of the covariance matrix need not be included in the penalty term as models are compared only for the same forecast horizon, so the number of parameters in the covariance matrix will be the same across models.

4 Monte Carlo Simulations

We next turn to Monte Carlo simulations as a means to evaluate the performance of the various model selection and estimation approaches under

two data generating processes (DGPs). For both DGPs we consider situations with a single target variable, $K = 1$, and m factors which yields the following VAR model for $\boldsymbol{z}_t = (y_{1t}, f_{1t}, \ldots, f_{mt})'$:

$$\begin{pmatrix} y_{1t}^{(b)} \\ \boldsymbol{f}_t^{(b)} \end{pmatrix} = \begin{pmatrix} \alpha\rho & \alpha\boldsymbol{\gamma}' \\ \mathbf{0} & \mathbf{A} \end{pmatrix} \begin{pmatrix} y_{1t-1}^{(b)} \\ \boldsymbol{f}_{t-1}^{(b)} \end{pmatrix} + \boldsymbol{\varepsilon}_t^{(b)}, \quad \boldsymbol{\varepsilon}_t^{(b)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{m+1}), \quad (18)$$

where $b = 1, 2, \ldots, B$ tracks the replications in the Monte Carlo experiments. We set $\boldsymbol{\Sigma}_{m+1}$ to a block-diagonal matrix, where the first block corresponds to the target variable and the second block to the $m \times m$ covariance matrix of the factors. The goodness of fit of the prediction equation (the first row of (18)) is controlled by the parameters α, ρ, $\boldsymbol{\gamma}$, $\mathbf{A}$ and $\boldsymbol{\Sigma}_{m+1}$. We set α such that the population R^2 for y_{1t}, denoted R_y^2, is either 0.2 or 0.8, representing low and high predictability scenarios, respectively. Assuming that the eigenvalues of $\mathbf{A}$ lie inside the unit circle, it is readily seen that $\mathrm{Cov}(\boldsymbol{f}_t^{b)}) = \mathbf{I}_m$ provided that the part of the covariance matrix $\boldsymbol{\Sigma}_{m+1}$ that corresponds to the factors is set to $(\mathbf{I}_m - \mathbf{A}\mathbf{A}')$. From (18) we see that

$$R_y^2 = \frac{\alpha^2\rho^2 + \alpha^2 \left(\boldsymbol{\gamma}'\boldsymbol{\gamma}\right)/\sigma_{\varepsilon_1}^2}{1 + \alpha^2 \left(\boldsymbol{\gamma}'\boldsymbol{\gamma}\right)/\sigma_{\varepsilon_1}^2},$$

where $\sigma_{\varepsilon_1}^2$ is the variance of the innovation to y_{1t}. For a given choice of R_y^2, ρ and $\boldsymbol{\gamma}$ we then have (setting $\sigma_{\varepsilon 1}^2 = 1$)

$$\alpha^2 = \frac{R_y^2}{\rho^2 + \boldsymbol{\gamma}'\boldsymbol{\gamma}(1 - R_y^2)}.$$

In practice, the factors are unobserved and forecasters will extract estimates of these from a panel of observed variables. We generate M_j variables for each factor j, $j = 1, 2, \ldots m$, in a hierarchical fashion, which corresponds to the estimation of the factors in the empirical analysis,

$$x_{jit}^{(b)} = \lambda_{ji}f_{jt}^{(b)} + \psi_{jit}^{(b)}, \quad \psi_{jit}^{(b)} \sim \mathrm{N}(0, 1), \quad i = 1, 2, \ldots, M_j, \quad j = 1, 2, \ldots, m,$$

where $M_j = 30$ for all j. λ_{ji} is set such that $R_{jx}^2 = 0.5$ in the low predictability scenario and 0.8 in the high predictability scenario:

$$R_{jx}^2 = \frac{\boldsymbol{\lambda}_j'\boldsymbol{\lambda}_j}{1 + \boldsymbol{\lambda}_j'\boldsymbol{\lambda}_j}, \; j = 1, 2, \ldots, m,$$

where $\boldsymbol{\lambda}_j = (\lambda_{j1}, \lambda_{j2}, \ldots, \lambda_{jM_j})'$.

Data are generated for window sizes of $w = 60, 120$, and 240 observations which in turn are used to compute forecasts for period $w + h$. We consider forecast horizons of $h = 1, 3, 6, 12$, and 24 periods.

15

Data Generating Process 1

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. Moreover, we set $m = 2$ and choose the remaining parameters as follows

$$\mathbf{C} \equiv \left(\begin{array}{cc} \rho & \boldsymbol{\gamma}' \\ \mathbf{0} & \mathbf{A} \end{array} \right) = \left(\begin{array}{ccc} 0.8 & 0.5 & 0.5 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.8 \end{array} \right),$$

so that both f_1 and f_2 help predict y_1, but f_1 and f_2 are in turn not themselves predictable by means of past values of y_1. Moreover, f_2 is quite persistent while f_1 is not, suggesting that for large values of h, f_2 should play more of a role in forecasting y_1 than f_1.

Data Generating Process 2

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 factor in the DGP is only helpful for longer horizon forecasts. Specifically, the parameters for the second DGP are set to $m = 3$ and

$$\mathbf{C} \equiv \left(\begin{array}{cc} \rho & \boldsymbol{\gamma}' \\ \mathbf{0} & \mathbf{A} \end{array} \right) = \left(\begin{array}{cccc} 0.1 & 0.5 & 0 & 0 \\ 0 & 0.2 & 0.6 & 0 \\ 0 & 0 & 0.2 & 0.6 \\ 0 & 0 & 0 & 0.75 \end{array} \right).$$

Notice that f_1 helps predict y_1, but f_1 is in turn not itself predictable by means of past values of y_1. Moreover, f_2 neither predicts nor is predicted by y_1 but f_2 predicts f_1 and therefore may help predict y_1 over medium horizons. Finally, the most persistent factor, f_3, indirectly helps predict y_1 through its ability to predict f_2.

4.1 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:

$$\hat{\mathbf{Z}}_{t+h}^{(b)*} = \left[\boldsymbol{I}_{m+1} + \hat{\mathbf{C}}^{(b)} + \cdots + \left[\hat{\mathbf{C}}^{(b)} \right]^{h-1} \right] \hat{\boldsymbol{\mu}} + \left[\hat{\mathbf{C}}^{(b)} \right]^{h} \boldsymbol{z}_t^{(b)*}, \qquad (19)$$

where $\boldsymbol{Z}_t^{(b)*} = \boldsymbol{z}_t^{(b)*}$ if $\boldsymbol{z}_t^{(b)*}$ includes $y_{1t}^{(b)}$ or $\boldsymbol{Z}_t^{(b)*} = (y_{1t}^{(b)}, \boldsymbol{z}_t^{(b)*})$ if $\boldsymbol{z}_t^{(b)*}$ does not include $y_{1t}^{(b)}$. $\boldsymbol{z}_t^{(b)*}$ is the subset of $\boldsymbol{z}_t^{(b)}$ chosen in the model selection procedure, $\boldsymbol{z}_t^{(b)} = (\hat{f}_{1t}, \hat{f}_{2t}, \ldots, \hat{f}_{mt})$, and $\hat{f}_{jt}$ is the first principle component extracted from the set of M_j regressors, x_{jit}, $i = 1, 2, \ldots, M_j$. $\hat{\mathbf{C}}^{(b)}$ is the estimate of $\mathbf{C}$ defined above for the bth Monte Carlo replication. The iterated h-step ahead forecast of $y_{1t}^{(b)}$ is denoted by $\hat{y}_{1,t+h}^{(b)}$.

Direct forecasts are obtained from

$$\tilde{y}_{1,t+h}^{(b)} = \hat{\mu}_h + \hat{\boldsymbol{\beta}}_h' \boldsymbol{z}_{ht}^{(b)}, \tag{20}$$

where $\boldsymbol{z}_{ht}^{(b)}$ is the subset of regressors that are selected for the h-step ahead forecast.

Forecast errors are calculated as

$$
\begin{aligned}
\hat{e}_{t+h}^{(b)} &= y_{1,t+h}^{(b)} - \hat{y}_{1,t+h}^{(b)}, \\
\tilde{e}_{t+h}^{(b)} &= y_{1,t+h}^{(b)} - \tilde{y}_{1,t+h}^{(b)}.
\end{aligned}
\tag{21}
$$

Forecasting performance is measured by the mean squared forecast error (MSFE) computed as

$$\text{MSFE} = \frac{1}{B} \sum_{b=1}^{B} \left[e_{t+h}^{(b)} \right]^2, \tag{22}$$

where $e_{t+h}^{(b)}$ is either $\hat{e}_{t+h}^{(b)}$ or $\tilde{e}_{t+h}^{(b)}$.

4.2   Summary of Monte Carlo Results

Results from the Monte Carlo simulations are reported in Tables 1 and 2. To study how the degree of predictability affects the findings, each table contains a panel with $R_y^2 = 0.2$ and $R_x^2 = 0.5$, and a panel with $R_y^2 = R_x^2 = 0.8$. The former is closer to the empirical results that we obtain, while the second scenario is more relevant for highly persistent variables.

First consider the results in Table 1 for the data generated under DGP1. For both the univariate models and the FAVARs, the iterated approach dominates the direct approach. This is a robust finding that holds across estimation sample sizes ($w = 60$, 120, and 240), information criteria (AIC versus BIC), and forecast horizons ($h = 3, 6, 12$, and 24). Moreover, in this case the performance of the different methods can largely be explained by the effect of parameter estimation error: the relative performance of the direct to the iterated forecasts improves with the length of the estimation window, w, since it becomes less costly to use an inefficient estimation method in the larger samples. Conversely, for a fixed estimation window, w, the relative

17

performance of the direct approach worsens as h increases since fewer observations are effectively available to estimate the parameters of the direct forecast model.

Turning to a comparison of the univariate and multivariate forecasts, under the iterated approach FAVAR-models selected by the simple AIC or BIC produce better forecasts on average than their univariate counterparts. The dominance of the iterated FAVAR models selected by the AIC is more pronounced in the high predictability scenario ($R_y^2 = 0.8$) where the parameters of the model are more precisely estimated. Interestingly, the direct forecasts from the FAVAR models only dominate the direct univariate forecasts in the high predictability scenario.

For the longer horizons (i.e., $h \geq 6$), the modifications to the AIC work well as the data overlap becomes more pronounced and serial correlation in the errors is attenuated. In particular the band diagonal modified AIC, defined by (16), improves over the simple direct AIC in 3 of 6 of the FAVAR cases shown in Table 1 when $h = 6$, while this number rises to 5 out of 6 cases when $h = 24$. Interestingly, the band diagonal approach is generally more successful at reducing the MSFE-values than the Newey-West approach. The SURE approach provides an even better forecast performance and improves in 4 out of 6 cases over the simple direct AIC when $h = 6$ and in 6 out of 6 cases when $h = 24$.

Turning to the second DGP, Table 2 now shows a few cases for $h = 3$ where the direct FAVAR forecasts produce better performance than the best iterated forecasts. It is clear from this DGP, however, that the degree of misspecification has to be quite large for this to happen. The table also shows continued gains from using the SURE approach and the band diagonal covariance adjustment to the conventional AIC.

5   Empirical Results

In their empirical analysis, Marcellino, Stock and Watson (2006) (MSW) found that iterated univariate forecasts generally outperform direct univariate forecasts. Furthermore, they found that the relative performance of the iterated univariate forecasts 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. We consider this question using the same data as in the MSW study which comprises 170 U.S. macroeconomic time series measured at the monthly frequency over the period 1959-2002 (528 months).[4]

---

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

5.1 Data Transformations

Following MSW, all variables are transformed by differencing a suitable number of times to achieve stationarity for estimation and model selection.[5] 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 predicted and $\hat{x}_{t+h}$ is constructed from x_t and $\hat{y}_{t+h}$. The forecasts from the AR model are based on the sample $y_{t-w+1}, y_{t-w+2}, \ldots, y_t$. Here $y_t = x_t$ 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). 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 Δx_t as follows

$$
\hat{x}_{t+h} = \left\{
\begin{array}{ll}
\hat{y}_{t+h} & \text{if } x_t \text{ is I(0)} \\
x_t + \sum_{i=1}^{h} \hat{y}_{t+i} & \text{if } x_t \text{ is I(1)} \\
x_t + h\Delta x_t + \sum_{i=1}^{h}\sum_{j=1}^{i} \hat{y}_{t+j} & \text{if } x_t \text{ is I(2)}
\end{array}
\right. .
$$

Similarly, under the direct approach, the forecast of x_{t+h} is constructed from the forecast of y_{t+h}, x_t, and Δx_t as follows

$$
\hat{x}_{t+h} = \left\{
\begin{array}{ll}
\hat{y}_{t+h} & \text{if } x_t \text{ is I(0)} \\
x_t + \hat{y}_{t+h} & \text{if } x_t \text{ is I(1)} \\
x_t + h\Delta x_t + \hat{y}_{t+h} & \text{if } x_t \text{ is I(2)}
\end{array}
\right. .
$$

5.2 Setup

Forecasting is performed recursively, begins in 1979M1 (with a minimum of w observations before forecasting) and runs until the end of the sample 2002M12. This yields up to 286 forecasts for $h = 3$, and so on. Forecasts are reported for horizons of $h = 3, 6, 12$, 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 role of estimation error in the relative performance of the various approaches.

To address the effect of model selection on the multivariate forecasts, we extract factors from the 170 series arranged into 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

---

[5]We ignore structural breaks; see Pesaran and Timmermann (2005) for an analysis of this on forecasts from autoregressive models.

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, the factors are estimated recursively. We then obtain forecasts of the factors from VARs fitted to all five factors with lag orders chosen by AIC or BIC and $p_{\max} = 2$. The search over FAVAR models is thus conducted over specifications that include own lags as well as those of the 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 zero, one or two lags in addition to an intercept. For computational simplicity the lag length is restricted to be the same for y_{it} and $\hat{f}_{it}$.

5.3 Forecasting Performance

Empirical results are summarized in Tables 3-9. We present MSFE-values averaged across all 170 variables (Table 3) as well as subsets of these (Table 8). Since these could be dominated by extreme values for individual variables, we also report the proportion of cases (again out of the 170 variables) where a modeling approach either dominates the benchmark univariate iterated forecasting model selected by the AIC (Table 4), or an approach is best overall for a given variable (Table 5). To evaluate statistical significance, we conduct pairwise comparisons of the forecast precision of various approaches against the benchmark univariate iterated models selected by the AIC using the approach suggested by Giacomini and White (2006) (Table 7). Finally, we report the separate magnitude of the squared bias component in the MSFE in order to understand whether the performance of a given forecasting approach is driven by its bias or by imprecision in the forecasts (Tables 6 and 9).

5.3.1 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. The iterated univariate forecasts based on models selected by the AIC are better on average than the direct ones, particularly at long horizons ($h = 12$ and 24 months) and when the estimation window is short ($w = 120$). Conversely, in the large estimation sample ($w = 240$) the direct univariate forecasting models selected by the more parsimonious BIC perform better than the iterated forecasts selected by this criterion. To understand this, recall that the iterated forecasts are more efficient and therefore tend to have a lower estimation error. Such errors are most important for large models (AIC penalizes large models less than the BIC) and

when the sample size is short.[6]

In the short sample ($w = 120$), the iterated models selected by the AIC deliver the best average forecasting performance among the univariate models. In the larger sample ($w = 240$), however, the best forecasting performance for horizons of 3 and 6 months is produced by the direct forecast approach that uses the AIC modified by using a Newey-West covariance matrix. Once again the univariate iterated approach based on the AIC dominates on average when $h = 12$ and 24. For the direct univariate forecast models there is only very limited evidence that the SURE estimation approach helps reducing average MSFE-values.

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 4 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 generally selects good univariate models, a finding corroborated by the results reported by MSW.

Among the univariate forecasts, in the small sample ($w = 120$), only the iterated forecasts based on models selected by the BIC produce a majority of cases that outperform the iterated AIC, and only then for $h = 3$ or $h = 6$ months. In the longer sample ($w = 240$) the direct forecasts based on the AIC, whether modified or not, also produce lower average MSFE-values for the majority of variables at horizons of 3 and 6 months.

Table 5 shows the proportion of cases (averaged across the 170 variables) where each of the respective methods produces the lowest MSFE value. Among the univariate approaches only the iterated AIC and the iterated BIC produce a sizeable proportion of variables with the lowest MSFE-value, particularly for the short estimation window ($w = 120$) and at the longest horizons.

From a theoretical perspective it is unclear whether the iterated approach leads to greater (squared) biases than the direct approach. To shed empirical light on this issue, Table 6 reports the squared forecast bias as a ratio of the MSFE of the benchmark iterated univariate forecast models selected by the AIC. For all methods the squared bias grows as a proportion of the MSFE of the benchmark model when the forecast horizon is extended. Interestingly, at short horizons ($h \leq 6$) the squared bias component of the iterated forecasts based on models selected by the AIC is slightly larger than that of the direct approaches, while conversely the relative bias of the

---

[6]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.

iterated AIC models is smaller at the two longest horizons ($h = 12, 24$). The iterated forecast models selected by the BIC generate a comparatively large bias that exceeds that generated by the direct forecast models selected by the BIC, suggesting that the parsimony of these models comes at the expense of a larger bias.

5.3.2 Multivariate Models

Turning to the factor-augmented VAR models, Table 3 shows that the iterated forecasts continue to do better on average than the direct FAVAR forecasts when a short estimation window ($w = 120$) is used. These results are frequently overturned, however, when the long estimation window ($w = 240$) is used. In the latter case, the direct forecasting method is better for forecast horizons of $h = 3, 6$ and 12 months irrespective of which information criterion is used, and also for $h = 24$ months under the diagonal modified AIC method or the SURE estimation approach.[7]

For the factor-augmented models, Table 3 shows that the band-diagonal modification to the AIC helps improve on average the performance of the direct forecasts across all horizons and for both estimation windows ($w = 120$ or $w = 240$). The Newey-West modification is less consistent in improving on the conventional AIC method. Moreover, in contrast with the univariate models, the SURE approach generally improves the direct FAVAR forecasts, particularly in the larger sample ($w = 240$).

Comparing the average forecasting performance across both univariate and multivariate models, at the shortest horizon ($h = 3$) the SURE method based on the BIC produces the lowest average MSFE-values when $w = 120$. Similarly, when $w = 240$, the direct forecast models estimated by SURE produce the best performance for $h = 3$ and $h = 6$ months. In all other cases, the univariate iterated forecast models selected by the AIC produce the lowest average MSFE values. These results are somewhat dominated by extreme cases, however. Table 4 shows that, for a majority of the 170 variables the iterated FAVAR forecasts based on models selected by AIC or BIC produce lower MSFE values than the univariate iterated forecasting models selected by AIC.

Table 5 shows additional evidence that the iterated FAVAR models perform well. The multivariate iterated forecasting models selected by the AIC or BIC produce the lowest MSFE-values around 40% of the 170 variables. This is a greater share than is recorded by other methods, although the SURE approach also performs well at the shortest horizon ($h = 3$).

Table 6 shows that the squared bias associated with the FAVAR models

---

[7]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.

tends to be relatively greater than the bias found for the univariate models. Moreover, for the multivariate models the iterated approach tends to produce relatively larger biases than the direct forecast approach.

In conclusion, for the majority of variables the iterated FAVAR models generate smaller forecast errors than the best univariate approach. There is less evidence in favor of the direct forecasting models: only for $w = 240$ and $h = 3$ or 6 months do we find that the direct approach leads to comparable forecasting performance. Overall, these findings demonstrate the value from utilizing multivariate information and also provide evidence that our proposed refinements to the information criteria work in many cases.

5.4   Model Comparisons

Table 7 provides test results based on a formal comparison of the benchmark univariate iterated model selected by the AIC with the alternative approaches listed in each row. Tests are based on the 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 percentage of model comparisons (out of 170) for which the null of equal predictive accuracy is rejected in a two-sided test at a 5% significance level against the alternative that the univariate iterated models selected by the AIC are best or, conversely, that the alternative model is best (listed in brackets).

The percentage of cases where the iterated univariate AIC method dominates other univariate forecast methods generally grows with the forecast horizon and is around 5-15% when $h = 3$ or $h = 6$ months and 10-20% for $h = 12$ or $h = 24$ months. We find far fewer cases where the iterated univariate forecasts selected by the AIC are rejected in favor of alternative univariate methods.

These test results provide statistical evidence that the univariate iterated AIC approach frequently performs significantly better than the other univariate methods and there is little evidence on which to prefer alternative univariate methods.

Turning to the factor-augmented models, the evidence is generally less clear-cut, with the proportion of significant cases where the iterated univariate AIC forecasts are preferred over other approaches such as the direct AIC forecasts, generally being more balanced, at least at the short horizon. In most cases, however, the iterated univariate AIC forecasts continue to reject alternative approaches more often than it gets rejected itself. Exceptions to this are the iterated FAVAR models based on either the AIC or BIC which reject about as often as they themselves get rejected by the univariate iterated models based on the AIC.

5.5   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, Tables 8 and 9 present separate results averaged across variables in categories A-D versus category-E variables.

Table 8 shows that the benefit from using the multivariate factor-based approach comes out very strongly for the first four categories. For these variables, across almost all sample sizes and forecast horizons, the models selected by the multivariate iterated AIC or BIC produce lower MSFE-values than the univariate iterated AIC approach. Among the direct FAVAR forecasts the modified AIC and SURE methods perform quite well, particularly in the large sample ($w = 240$.)

Across the first four categories of variables, the multivariate iterated approach based on the BIC performs best for the shortest estimation window ($w = 120$) when $h = 3, 6$ or 12 months. When $w = 240$ the iterated FAVAR approach based on the AIC generates the best results on average, except for when $h = 3$ where the SURE approach is best. Table 9 shows that forecasts based on these methods are only modestly biased.

In contrast, for the final group of variables, (E) nominal prices, wages and money, the FAVAR approach strongly underperforms against the univariate iterated AIC models. Iterated FAVAR models are particularly poor and are outperformed by their direct counterparts. This suggests that the iterated FAVAR models are heavily biased, a conjecture that is confirmed in Table 9 which reveals massive biases for the iterated FAVAR models at long horizons in particular. The biases associated with the direct forecast models are much smaller.

6   Conclusion

We compared the performance of iterated and direct forecasts generated by univariate and multivariate (factor-augmented VAR) models. Our simulations and 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, the method used to estimate model parameters and the relative performance of the direct versus iterated approaches.

Like Marcellino, Stock and Watson (2006), our results suggest that there is no single dominant approach and that the best forecasting method varies considerably across economic variables. The iterated factor-augmented VAR

approach performs considerably better than the best univariate forecasting approach for variables tracking income, output, employment, construction, interest rates and asset prices. Conversely, the univariate iterative models dominate among variables tracking nominal prices, wages and money for which the factor-augmented iterated models produce heavily biased forecasts.

In general, our empirical and simulation results suggest that the degree of model misspecification has to be quite large for the direct forecasts to start dominating the iterated forecasts and that the forecasts generated by autoregressive models of low order−whether factor-augmented or not−are difficult to beat for most economic variables. This is a result of the (squared) bias component generally playing a relatively minor role relative to the important effect of parameter estimation error in the composition of MSFE-values. Consistent with this, the iterated forecasting approach performs particularly 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.

References

[1] Athanasopoulos, G. and F. Vahid, 2008, VARMA versus VAR for macroeconomic forecasting. *Journal of Business and Economic Statistics*, 26, 237–251.

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

[3] Bai, Jushan and Serena Ng, 2009, Boosting diffusion indices, *Journal of Applied Econometrics*, 24, 607–629.

[4] Bao, Y. 2007, Finite sample properties of forecasts from the stationary first-order autoregressive model under a general error distribution. *Econometric Theory* 23, 767–773.

[5] 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.

[6] 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.

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

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

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

[10] Favero, Carlo A. and Andrea Tamoni, 2010, Demographics and the term structure of stock market risk. Mimeo, Bocconi University.

[11] 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.

[12] 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.

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

[14] Hoque, A., J.R. Magnus and B. Pesaran, 1988, The exact multi-period mean squared forecast error for the first-order autoregressive model. *Journal of Econometrics* 39, 327-346.

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

[16] 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.

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

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

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

[20] 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.

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

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

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

In the column groups below, the three blocks correspond to w = 60, w = 120 and w = 240; within each block the columns are the forecast horizons h = 3, 6, 12, 24.

h	3	6	12	24	3	6	12	24	3	6	12	24
	w = 60				w = 120				w = 240			
$R_y^2 = 0.2,\ R_x^2 = 0.5$ — AR												
iterated AIC	1.408	1.388	1.407	1.393	1.365	1.363	1.333	1.348	1.294	1.343	1.381	1.344
direct AIC	1.447	1.474	1.540	1.828	1.389	1.386	1.379	1.405	1.304	1.357	1.399	1.368
mod. AIC(NW)	1.468	1.517	1.595	1.956	1.394	1.390	1.385	1.418	1.305	1.358	1.399	1.368
mod. AIC(diag)	1.440	1.457	1.549	2.212	1.383	1.379	1.372	1.417	1.305	1.350	1.391	1.360
AIC(SURE)	1.435	1.449	1.556	1.871	1.375	1.380	1.375	1.424	1.303	1.356	1.403	1.368
iterated BIC	1.366	1.361	1.392	1.382	1.352	1.353	1.331	1.347	1.288	1.336	1.380	1.344
direct BIC	1.386	1.387	1.430	1.526	1.363	1.355	1.346	1.366	1.297	1.340	1.383	1.346
BIC(SURE)	1.385	1.389	1.471	1.585	1.362	1.355	1.354	1.392	1.294	1.339	1.388	1.356
$R_y^2 = 0.2,\ R_x^2 = 0.5$ — Factors												
iterated AIC	1.363	1.350	1.374	1.366	1.284	1.356	1.321	1.315	1.249	1.327	1.338	1.323
direct AIC	1.456	1.496	1.553	1.671	1.321	1.414	1.391	1.394	1.277	1.361	1.369	1.353
mod. AIC(NW)	1.459	1.502	1.569	1.702	1.323	1.412	1.389	1.391	1.279	1.360	1.366	1.352
mod. AIC(diag)	1.443	1.449	1.482	1.688	1.328	1.397	1.358	1.364	1.284	1.356	1.353	1.337
AIC(SURE)	1.444	1.435	1.510	1.644	1.327	1.395	1.360	1.373	1.293	1.353	1.357	1.345
iterated BIC	1.365	1.350	1.373	1.364	1.297	1.361	1.320	1.315	1.254	1.328	1.338	1.323
direct BIC	1.419	1.432	1.473	1.560	1.316	1.382	1.357	1.355	1.282	1.350	1.353	1.337
BIC(SURE)	1.410	1.400	1.467	1.564	1.326	1.379	1.346	1.360	1.303	1.341	1.349	1.341
$R_y^2 = 0.8,\ R_x^2 = 0.8$ — AR												
iterated AIC	5.218	9.221	12.624	14.564	4.728	8.408	10.720	11.060	4.473	7.799	10.017	10.583
direct AIC	5.659	11.309	19.233	30.316	4.830	8.896	12.105	13.385	4.526	8.023	10.486	11.277
mod. AIC(NW)	5.727	11.548	19.540	31.614	4.838	8.960	12.058	13.304	4.531	8.043	10.489	11.203
mod. AIC(diag)	5.705	11.590	18.211	32.994	4.828	9.068	11.773	12.799	4.527	8.040	10.499	11.004
AIC(SURE)	5.566	11.180	18.369	28.429	4.796	8.855	12.032	13.034	4.503	7.973	10.482	11.245
iterated BIC	4.945	8.683	12.085	13.794	4.635	8.240	10.687	11.106	4.468	7.805	10.068	10.647
direct BIC	5.307	10.318	16.050	24.196	4.732	8.651	11.444	12.486	4.490	7.902	10.267	10.985
BIC(SURE)	5.222	10.159	15.220	22.714	4.696	8.605	11.381	12.135	4.447	7.878	10.280	10.921
$R_y^2 = 0.8,\ R_x^2 = 0.8$ — Factors												
iterated AIC	3.687	7.370	11.159	12.370	3.364	6.861	9.972	10.808	3.204	6.500	9.584	10.361
direct AIC	4.008	8.930	14.837	20.110	3.429	7.362	11.276	12.917	3.234	6.701	10.035	11.227
mod. AIC(NW)	4.057	9.040	14.753	20.076	3.445	7.502	11.200	12.725	3.234	6.747	10.064	11.088
mod. AIC(diag)	4.159	9.249	13.938	19.820	3.448	7.686	11.074	12.228	3.236	6.827	10.139	10.858
AIC(SURE)	4.019	8.905	14.433	18.882	3.459	7.366	11.206	12.675	3.249	6.705	10.046	11.185
iterated BIC	3.695	7.367	11.127	12.344	3.344	6.840	9.951	10.813	3.194	6.493	9.585	10.362
direct BIC	4.057	8.862	14.302	19.194	3.435	7.481	11.084	12.544	3.218	6.740	10.035	10.990
BIC(SURE)	4.052	8.859	13.773	17.897	3.433	7.476	11.050	12.248	3.224	6.694	10.001	10.945

The table reports the MSFE for the different forecasting methods. The forecasts labeled 'mod. AIC(diag)' and 'mod. AIC(NW)' are based on the modified AIC with band diagonal or Newey-West covariance matrices. Forecasts 'AIC(SURE)' and 'BIC(SURE)' are based on models that allow for autocorrelation in the likelihood function and use OLS estimates once the model has been selected. The Monte Carlo results are based on 10,000 iterations. Details of DGP 1 and 2 are given in Section 4.

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

$R_y^2 = 0.2, R_x^2 = 0.5$

AR

h	$w = 60$				$w = 120$				$w = 240$			
	3	6	12	24	3	6	12	24	3	6	12	24
iterated AIC	1.335	1.319	1.307	1.292	1.313	1.301	1.265	1.270	1.247	1.274	1.318	1.267
direct AIC	1.362	1.372	1.387	1.558	1.321	1.315	1.293	1.311	1.252	1.286	1.334	1.277
mod. AIC(NW)	1.376	1.404	1.424	1.676	1.327	1.321	1.305	1.323	1.255	1.288	1.335	1.280
mod. AIC(diag)	1.360	1.376	1.431	1.947	1.320	1.315	1.306	1.338	1.251	1.284	1.334	1.283
AIC(SURE)	1.368	1.402	1.463	1.586	1.329	1.323	1.316	1.341	1.253	1.294	1.346	1.296
iterated BIC	1.306	1.290	1.297	1.287	1.296	1.293	1.265	1.270	1.239	1.270	1.318	1.267
direct BIC	1.314	1.305	1.320	1.343	1.296	1.297	1.271	1.281	1.239	1.272	1.321	1.268
BIC(SURE)	1.325	1.333	1.402	1.397	1.305	1.307	1.293	1.314	1.243	1.278	1.334	1.287

Factors

h	$w = 60$				$w = 120$				$w = 240$			
	3	6	12	24	3	6	12	24	3	6	12	24
iterated AIC	1.321	1.282	1.290	1.265	1.256	1.292	1.247	1.268	1.263	1.274	1.267	1.257
direct AIC	1.412	1.385	1.430	1.486	1.286	1.343	1.305	1.331	1.270	1.302	1.296	1.281
mod. AIC(NW)	1.416	1.407	1.450	1.525	1.287	1.346	1.307	1.337	1.273	1.303	1.297	1.285
mod. AIC(diag)	1.400	1.368	1.425	1.581	1.284	1.333	1.289	1.328	1.267	1.293	1.290	1.276
AIC(SURE)	1.414	1.383	1.427	1.486	1.287	1.341	1.300	1.331	1.272	1.300	1.297	1.280
iterated BIC	1.317	1.280	1.289	1.265	1.251	1.291	1.247	1.268	1.262	1.274	1.267	1.257
direct BIC	1.361	1.324	1.353	1.372	1.266	1.308	1.265	1.291	1.265	1.282	1.274	1.262
BIC(SURE)	1.358	1.321	1.357	1.382	1.262	1.303	1.261	1.294	1.267	1.279	1.272	1.264

$R_y^2 = 0.8, R_x^2 = 0.8$

AR

h	$w = 60$				$w = 120$				$w = 240$			
	3	6	12	24	3	6	12	24	3	6	12	24
iterated AIC	6.146	5.978	5.867	5.898	5.893	5.941	5.728	5.656	5.684	5.729	5.820	5.581
direct AIC	6.374	6.463	6.481	7.657	5.944	6.042	5.917	5.898	5.692	5.788	5.885	5.644
mod. AIC(NW)	6.446	6.596	6.680	8.250	5.942	6.059	5.927	5.939	5.701	5.785	5.885	5.644
mod. AIC(diag)	6.288	6.419	6.628	9.232	5.894	6.002	5.883	5.997	5.678	5.768	5.876	5.642
AIC(SURE)	6.229	6.251	6.478	7.750	5.885	5.979	5.915	5.987	5.686	5.769	5.904	5.694
iterated BIC	5.959	5.861	5.813	5.857	5.822	5.882	5.712	5.654	5.652	5.693	5.814	5.580
direct BIC	5.981	5.956	6.018	6.371	5.766	5.913	5.751	5.741	5.631	5.707	5.836	5.590
BIC(SURE)	5.960	5.971	6.134	6.622	5.787	5.905	5.774	5.853	5.634	5.716	5.844	5.637

Factors

h	$w = 60$				$w = 120$				$w = 240$			
	3	6	12	24	3	6	12	24	3	6	12	24
iterated AIC	5.734	5.963	5.718	5.771	5.460	5.887	5.587	5.551	5.441	5.635	5.560	5.559
direct AIC	5.649	6.552	6.586	7.254	5.154	6.020	5.859	5.937	4.886	5.648	5.727	5.731
mod. AIC(NW)	5.651	6.597	6.625	7.406	5.143	6.032	5.864	5.938	4.895	5.660	5.707	5.721
mod. AIC(diag)	5.623	6.453	6.472	7.555	5.130	6.031	5.765	5.890	4.919	5.668	5.671	5.659
AIC(SURE)	5.816	6.363	6.328	7.182	5.300	5.992	5.740	5.883	5.112	5.664	5.688	5.678
iterated BIC	5.768	5.927	5.712	5.770	5.517	5.868	5.587	5.549	5.496	5.625	5.558	5.559
direct BIC	5.626	6.252	6.215	6.678	5.144	5.941	5.718	5.753	4.918	5.614	5.622	5.636
BIC(SURE)	5.853	6.147	6.100	6.728	5.434	5.931	5.649	5.761	5.393	5.636	5.578	5.590

See the footnote to Table 1.

Table 3: Average of 170 variables forecasting performance measured by the MSFE relative to the corresponding value generated by the univariate iterated forecast models selected by the AIC

Forecast horizon	3	6	12	24	3	6	12	24
		AR, w = 120				*AR, w = 240*		
iterated AIC	1.000	1.000	1.000	1.000	0.962	0.945	0.952	0.994
direct AIC	1.015	1.018	1.084	1.208	0.964	0.945	0.987	1.087
mod. AIC(diag)	1.013	1.017	1.106	1.202	0.964	0.944	0.982	1.097
mod. AIC(NW)	1.003	1.015	1.086	1.227	0.957	0.939	0.986	1.090
AIC(SURE)	1.017	1.028	1.081	1.190	0.985	0.997	1.058	1.106
iterated BIC	1.009	1.037	1.053	1.049	1.007	1.042	1.085	1.148
direct BIC	1.009	1.017	1.077	1.157	0.970	0.953	0.989	1.081
BIC(SURE)	1.007	1.026	1.087	1.150	0.993	1.024	1.069	1.093
		Factors, w = 120				*Factors, w = 240*		
iterated AIC	0.977	1.032	1.075	1.069	0.974	1.028	1.058	1.117
direct AIC	0.985	1.045	1.129	1.252	0.949	0.995	1.049	1.134
mod. AIC(diag)	0.984	1.038	1.108	1.204	0.947	0.993	1.035	1.103
mod. AIC(NW)	0.981	1.046	1.133	1.259	0.948	0.992	1.046	1.141
AIC(SURE)	0.978	1.048	1.133	1.227	0.940	0.983	1.038	1.109
iterated BIC	0.974	1.015	1.046	1.086	0.983	1.041	1.083	1.156
direct BIC	0.981	1.018	1.106	1.223	0.960	0.998	1.043	1.117
BIC(SURE)	0.965	1.020	1.105	1.216	0.941	0.980	1.029	1.098

The table reports the MSFE of the different forecasts relative to the MSFE of the iterated AR forecast based on the models selected by AIC with $w = 120$, where w is the length of the estimation window. 'AR' results are based on univariate autoregressive models and 'Factors' results are based on multivariate factor-augmented VAR models. The MSFEs are calculated only for those periods where forecasts from all methods are available. The forecasts labeled 'mod. AIC(diag)' and 'mod. AIC(NW)' are based on the modified AIC with band diagonal or Newey-West covariance matrices. Forecasts 'AIC(SURE)' and 'BIC(SURE)' are based on models that allow for autocorrelation in the likelihood function and use OLS estimates once the model has been selected. Averages are computed across all 170 series in the Marcellino, Stock and Watson (2006) data set.

Table 4: Proportion of 170 variables for which different forecast methods generate a lower MSFE than the corresponding values generated by the univariate iterated forecast models selected by the AIC

Forecast horizon	3	6	12	24	3	6	12	24
		AR, w = 120				*AR, w = 240*		
direct AIC	0.441	0.429	0.218	0.176	0.518	0.506	0.359	0.235
mod. AIC(diag)	0.435	0.394	0.235	0.153	0.547	0.506	0.365	0.212
mod. AIC(NW)	0.500	0.418	0.229	0.165	0.547	0.506	0.365	0.235
AIC(SURE)	0.447	0.353	0.194	0.159	0.318	0.412	0.265	0.212
iterated BIC	0.541	0.535	0.482	0.471	0.512	0.424	0.388	0.424
direct BIC	0.412	0.418	0.200	0.165	0.447	0.429	0.324	0.241
BIC(SURE)	0.476	0.365	0.194	0.171	0.406	0.412	0.271	0.235
		Factors, w = 120				*Factors, w = 240*		
iterated AIC	0.641	0.541	0.500	0.488	0.659	0.576	0.565	0.600
direct AIC	0.565	0.441	0.341	0.212	0.665	0.518	0.435	0.429
mod. AIC(diag)	0.565	0.465	0.371	0.271	0.647	0.535	0.441	0.459
mod. AIC(NW)	0.600	0.418	0.300	0.212	0.647	0.547	0.435	0.441
AIC(SURE)	0.582	0.412	0.306	0.194	0.671	0.547	0.418	0.447
iterated BIC	0.724	0.618	0.565	0.500	0.635	0.553	0.588	0.594
direct BIC	0.612	0.500	0.335	0.212	0.647	0.529	0.424	0.429
BIC(SURE)	0.641	0.476	0.341	0.218	0.676	0.571	0.441	0.447

The table reports the proportion of series for which the iterated AR forecasts based on the models selected by AIC have larger MSFEs than forecasts based on the respective information criteria. Hence, values above 0.5 suggest that the respective method dominates univariate AIC forecasts for a majority of variables. For the model selection methods see the footnote to Table 3. Proportions are computed as averages across all 170 series in the Marcellino, Stock and Watson (2006) data set.

Table 5: Proportion of 170 variables for which the forecast methods produce the lowest MSFE-values

Forecast horizon	3	6	12	24	3	6	12	24
	AR, w = 120				*AR, w = 240*			
iterated AIC	0.041	0.094	0.124	0.206	0.071	0.094	0.129	0.165
direct AIC	0.018	0.029	0.029	0.006	0.035	0.035	0.035	0.024
mod. AIC(diag)	0.041	0.029	0.006	0.018	0.041	0.041	0.035	0.000
mod. AIC(NW)	0.071	0.065	0.035	0.041	0.041	0.076	0.035	0.018
AIC(SURE)	0.041	0.018	0.006	0.006	0.029	0.029	0.000	0.024
iterated BIC	0.065	0.106	0.135	0.106	0.053	0.035	0.059	0.024
direct BIC	0.024	0.024	0.047	0.012	0.006	0.024	0.006	0.018
BIC(SURE)	0.029	0.029	0.012	0.029	0.035	0.065	0.006	0.000
	Factors, w = 120				*Factors, w = 240*			
iterated AIC	0.247	0.171	0.188	0.247	0.106	0.129	0.176	0.247
direct AIC	0.012	0.018	0.041	0.018	0.024	0.024	0.047	0.076
mod. AIC(diag)	0.018	0.035	0.053	0.059	0.024	0.035	0.053	0.094
mod. AIC(NW)	0.029	0.041	0.024	0.012	0.053	0.029	0.041	0.029
AIC(SURE)	0.059	0.047	0.006	0.012	0.118	0.065	0.053	0.047
iterated BIC	0.182	0.188	0.229	0.171	0.165	0.218	0.253	0.141
direct BIC	0.047	0.047	0.041	0.053	0.053	0.035	0.029	0.059
BIC(SURE)	0.076	0.059	0.024	0.006	0.147	0.065	0.041	0.035

For each horizon and window length, the table reports the proportion of variables for which the respective forecast methods generate the lowest MSFE-value. For the model selection methods see the footnote to Table 3. Proportions are computed across all 170 series in the Marcellino, Stock and Watson (2006) data set.

Table 6: Average ratio of squared bias measured relative to the MSFE of the iterated univariate forecast models selected by the AIC over 170 variables

Forecast horizon	3	6	12	24	3	6	12	24
	AR, w = 120				*AR, w = 240*			
iterated AIC	0.014	0.028	0.049	0.090	0.019	0.040	0.074	0.154
direct AIC	0.012	0.023	0.048	0.114	0.018	0.037	0.079	0.194
mod. AIC(diag)	0.012	0.023	0.052	0.120	0.018	0.037	0.079	0.195
mod. AIC(NW)	0.012	0.023	0.049	0.114	0.019	0.038	0.078	0.195
AIC(SURE)	0.013	0.024	0.048	0.116	0.025	0.040	0.081	0.202
iterated BIC	0.025	0.050	0.081	0.129	0.041	0.092	0.166	0.294
direct BIC	0.016	0.029	0.056	0.117	0.022	0.042	0.085	0.198
BIC(SURE)	0.018	0.031	0.059	0.122	0.032	0.050	0.094	0.213
	Factors, w = 120				*Factors, w = 240*			
iterated AIC	0.036	0.068	0.106	0.163	0.056	0.117	0.206	0.365
direct AIC	0.027	0.043	0.070	0.152	0.038	0.061	0.104	0.226
mod. AIC(diag)	0.026	0.041	0.073	0.164	0.038	0.065	0.106	0.215
mod. AIC(NW)	0.026	0.043	0.070	0.152	0.037	0.061	0.104	0.230
AIC(SURE)	0.025	0.041	0.069	0.151	0.035	0.055	0.097	0.206
iterated BIC	0.033	0.063	0.098	0.152	0.058	0.124	0.217	0.374
direct BIC	0.028	0.042	0.073	0.157	0.041	0.067	0.109	0.227
BIC(SURE)	0.025	0.041	0.073	0.157	0.037	0.061	0.101	0.205

The table reports the squared bias of the different forecasts as a ratio of the corresponding MSFE of the iterated AR forecast based on the models selected by AIC with $w = 120$, where w is the length of the estimation window. For the model selection methods see the footnote to Table 3. Averages are computed across all 170 series in the Marcellino, Stock and Watson (2006) data set.

Table 7: Pairwise model comparison tests

Forecast horizon	AR $w = 120$				AR, $w = 240$			
	3	6	12	24	3	6	12	24
direct AIC	5.3(3.5)	4.7(2.9)	12.4(0.6)	22.4(0.0)	1.8(4.7)	4.1(4.1)	9.4(1.8)	8.8(1.2)
mod. AIC(diag)	4.7(3.5)	6.5(2.9)	17.1(0.6)	18.8(1.2)	1.2(5.3)	4.1(5.3)	9.4(1.8)	9.4(1.2)
mod. AIC(NW)	5.9(3.5)	4.1(3.5)	11.8(0.6)	21.8(0.6)	3.5(5.3)	4.1(5.9)	8.2(2.4)	9.4(1.2)
AIC(SURE)	8.2(3.5)	7.1(1.8)	14.7(0.6)	21.8(0.6)	12.4(0.0)	7.6(2.4)	13.5(0.0)	10.6(1.2)
iterated BIC	11.2(2.9)	12.4(3.5)	6.5(2.4)	5.9(2.4)	18.8(2.4)	18.2(2.9)	12.4(1.2)	10.6(1.2)
direct BIC	10.6(1.8)	8.2(1.2)	15.3(0.0)	18.2(0.6)	10.6(1.8)	4.1(1.2)	5.3(0.6)	10.0(1.8)
BIC(SURE)	8.2(2.9)	12.4(1.2)	17.1(0.6)	17.1(0.6)	12.9(4.1)	18.8(1.2)	13.5(0.0)	11.2(2.4)
	Factors, $w = 120$				Factors, $w = 240$			
iterated AIC	10.6(11.2)	12.4(8.8)	8.8(8.2)	7.1(4.7)	12.4(11.2)	12.4(10.6)	11.2(7.6)	7.6(11.2)
direct AIC	8.2(8.2)	9.4(5.3)	11.2(2.4)	11.8(0.0)	10.6(10.0)	8.8(7.1)	8.8(6.5)	10.0(7.1)
mod. AIC(diag)	6.5(9.4)	8.2(5.3)	10.0(1.8)	8.8(0.0)	10.0(9.4)	10.0(7.6)	9.4(6.5)	6.5(5.3)
mod. AIC(NW)	7.6(9.4)	8.2(4.7)	11.2(2.4)	14.1(0.0)	9.4(10.6)	8.2(6.5)	9.4(6.5)	9.4(7.1)
AIC(SURE)	7.1(9.4)	10.6(5.9)	14.1(1.8)	12.9(0.0)	9.4(14.1)	9.4(7.6)	9.4(6.5)	9.4(7.1)
iterated BIC	12.4(12.4)	11.8(10.0)	8.2(7.6)	8.2(5.3)	12.4(13.5)	11.8(12.4)	10.6(10.0)	8.2(9.4)
direct BIC	8.2(8.2)	8.2(8.2)	10.0(2.4)	12.4(0.0)	11.2(12.9)	9.4(8.8)	9.4(6.5)	8.8(5.9)
BIC(SURE)	7.6(10.0)	10.6(7.1)	8.8(2.4)	10.0(0.0)	10.0(14.1)	9.4(8.2)	10.0(5.9)	10.0(5.3)

The table reports the results of the Giacomini and White (2006) test with H_0: the iterated AR forecast based on AIC and the method in the respective row have equal forecasting ability. Reported are the proportions of rejections at the 5% level for the two-sided test where the iterated AR(AIC) forecast has the lowest MSFE and in brackets the proportion of rejections when the respective forecasting method has the lower MSFE. For the model selection methods see the footnote to Table 3.

Table 8: Average forecasting performance measured by the MSFE relative to the corresponding value generated by the univariate iterated forecasting models selected by the AIC in sub-categories of the Marcellino, Stock and Watson (2006) data set

Forecast horizon	3	6	12	24	3	6	12	24
Categories (A)-(D) (averages over 135 series)								
	AR, w = 120				*AR, w = 240*			
iterated AIC	1.000	1.000	1.000	1.000	0.964	0.945	0.944	0.966
direct AIC	1.024	1.027	1.096	1.185	0.969	0.949	0.983	1.042
mod. AIC(diag)	1.021	1.026	1.122	1.163	0.969	0.949	0.978	1.051
mod. AIC(NW)	1.008	1.023	1.099	1.201	0.960	0.941	0.981	1.046
AIC(SURE)	1.023	1.036	1.088	1.161	0.974	0.969	0.992	1.048
iterated BIC	0.984	0.993	0.997	0.995	0.960	0.953	0.953	0.968
direct BIC	0.999	1.011	1.074	1.124	0.963	0.949	0.972	1.012
BIC(SURE)	0.996	1.015	1.076	1.113	0.962	0.966	0.980	1.009
	Factors, w = 120				*Factors, w = 240*			
iterated AIC	0.928	0.954	0.978	0.982	0.886	0.888	0.872	0.877
direct AIC	0.945	0.983	1.069	1.231	0.881	0.909	0.977	1.094
mod. AIC(diag)	0.946	0.985	1.065	1.165	0.878	0.903	0.959	1.043
mod. AIC(NW)	0.942	0.989	1.074	1.241	0.877	0.903	0.971	1.099
AIC(SURE)	0.939	0.989	1.081	1.201	0.872	0.898	0.965	1.066
iterated BIC	0.927	0.937	0.942	1.001	0.890	0.888	0.875	0.895
direct BIC	0.943	0.961	1.051	1.189	0.887	0.905	0.961	1.061
BIC(SURE)	0.931	0.967	1.061	1.183	0.872	0.890	0.949	1.045
Category (E) (averages over 35 series)								
	AR, w = 120				*AR, w = 240*			
iterated AIC	1.000	1.000	1.000	1.000	0.954	0.946	0.986	1.101
direct AIC	0.983	0.984	1.038	1.296	0.943	0.927	1.004	1.258
mod. AIC(diag)	0.982	0.984	1.048	1.352	0.942	0.926	0.999	1.275
mod. AIC(NW)	0.983	0.984	1.036	1.326	0.946	0.933	1.002	1.259
AIC(SURE)	0.997	0.997	1.057	1.299	1.029	1.108	1.312	1.330
iterated BIC	1.106	1.205	1.270	1.260	1.191	1.384	1.595	1.842
direct BIC	1.046	1.038	1.088	1.284	0.997	0.971	1.053	1.346
BIC(SURE)	1.047	1.067	1.129	1.295	1.111	1.246	1.410	1.420
	Factors, w = 120				*Factors, w = 240*			
iterated AIC	1.168	1.332	1.449	1.401	1.315	1.566	1.773	2.041
direct AIC	1.139	1.281	1.360	1.332	1.210	1.328	1.323	1.291
mod. AIC(diag)	1.132	1.245	1.273	1.356	1.212	1.337	1.328	1.335
mod. AIC(NW)	1.131	1.264	1.361	1.330	1.222	1.334	1.334	1.301
AIC(SURE)	1.129	1.274	1.333	1.327	1.203	1.309	1.320	1.275
iterated BIC	1.155	1.316	1.443	1.413	1.341	1.631	1.884	2.162
direct BIC	1.129	1.237	1.320	1.353	1.239	1.358	1.357	1.333
BIC(SURE)	1.098	1.228	1.276	1.346	1.205	1.328	1.334	1.299

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 'Factors' results are based on multivariate factor-augmented VAR models. The MSFEs are calculated only for those periods where forecasts from all methods are available. For details see the footnote to Table 3.

Table 9: Average ratio of squared bias measured relative to the MSFE of the iterated univariate forecasts selected by the AIC in sub-categories of the Marcellino, Stock and Watson (2006) data set

Forecast horizon	3	6	12	24	3	6	12	24
Categories (A)-(D) (averages over 135 series)								
	AR, w = 120				*AR, w = 240*			
iterated AIC	0.008	0.017	0.027	0.054	0.013	0.027	0.047	0.097
direct AIC	0.008	0.015	0.032	0.068	0.013	0.026	0.053	0.129
mod. AIC(diag)	0.008	0.016	0.037	0.067	0.013	0.026	0.053	0.128
mod. AIC(NW)	0.009	0.016	0.033	0.068	0.013	0.026	0.053	0.129
AIC(SURE)	0.008	0.016	0.031	0.066	0.015	0.029	0.054	0.125
iterated BIC	0.009	0.019	0.028	0.053	0.015	0.031	0.051	0.099
direct BIC	0.008	0.016	0.033	0.064	0.014	0.027	0.054	0.117
BIC(SURE)	0.009	0.017	0.034	0.065	0.016	0.030	0.053	0.119
	Factors, w = 120				*Factors, w = 240*			
iterated AIC	0.019	0.032	0.044	0.078	0.031	0.056	0.082	0.137
direct AIC	0.014	0.020	0.040	0.108	0.024	0.042	0.088	0.221
mod. AIC(diag)	0.015	0.022	0.048	0.119	0.024	0.044	0.087	0.194
mod. AIC(NW)	0.014	0.021	0.041	0.110	0.023	0.041	0.087	0.223
AIC(SURE)	0.013	0.019	0.040	0.109	0.022	0.037	0.080	0.196
iterated BIC	0.016	0.026	0.035	0.065	0.027	0.046	0.068	0.121
direct BIC	0.016	0.021	0.041	0.108	0.025	0.044	0.088	0.210
BIC(SURE)	0.014	0.020	0.043	0.109	0.024	0.040	0.080	0.186
Category (E) (averages over 35 series)								
	AR, w = 120				*AR, w = 240*			
iterated AIC	0.035	0.074	0.134	0.228	0.041	0.088	0.180	0.372
direct AIC	0.026	0.054	0.108	0.292	0.039	0.081	0.178	0.445
mod. AIC(diag)	0.026	0.053	0.111	0.322	0.039	0.081	0.177	0.456
mod. AIC(NW)	0.025	0.052	0.108	0.293	0.040	0.083	0.179	0.447
AIC(SURE)	0.028	0.054	0.114	0.306	0.062	0.081	0.185	0.499
iterated BIC	0.083	0.171	0.286	0.421	0.138	0.323	0.609	1.042
direct BIC	0.048	0.077	0.146	0.323	0.051	0.096	0.207	0.508
BIC(SURE)	0.052	0.085	0.159	0.338	0.093	0.128	0.250	0.576
	Factors, w = 120				*Factors, w = 240*			
iterated AIC	0.100	0.206	0.344	0.490	0.151	0.353	0.683	1.241
direct AIC	0.076	0.130	0.186	0.321	0.090	0.133	0.167	0.247
mod. AIC(diag)	0.068	0.118	0.171	0.338	0.091	0.145	0.182	0.296
mod. AIC(NW)	0.072	0.127	0.181	0.315	0.092	0.135	0.170	0.253
AIC(SURE)	0.072	0.126	0.181	0.312	0.083	0.127	0.164	0.244
iterated BIC	0.099	0.205	0.340	0.485	0.178	0.425	0.791	1.350
direct BIC	0.076	0.124	0.194	0.345	0.105	0.155	0.188	0.293
BIC(SURE)	0.069	0.123	0.186	0.341	0.091	0.142	0.182	0.279

The table reports the squared bias 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. For the model selection methods see the footnote to Table 3.