Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Journal of Econometrics 135 (2006) 499–526

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series

Massimiliano Marcellino[a], James H. Stock[b], Mark W. Watson[c,*]

[a]*Istituto di Economia Politica, Universita Bocconi and IGIER, Italy*

[b]*Department of Economics, Harvard University, and the NBER, USA*

[c]*Department of Economics and Woodrow Wilson School, Princeton University, and the NBER, Princeton, NJ 08544, USA*

Available online 24 August 2005

Abstract

“Iterated” multiperiod-ahead time series forecasts are made using a one-period ahead model, iterated forward for the desired number of periods, whereas “direct” forecasts are made using a horizon-specific estimated model, where the dependent variable is the multiperiod ahead value being forecasted. Which approach is better is an empirical matter: in theory, iterated forecasts are more efficient if the one-period ahead model is correctly specified, but direct forecasts are more robust to model misspecification. This paper compares empirical iterated and direct forecasts from linear univariate and bivariate models by applying simulated out-of-sample methods to 170 U.S. monthly macroeconomic time series spanning 1959–2002. The iterated forecasts typically outperform the direct forecasts, particularly, if the models can select long-lag specifications. The relative performance of the iterated forecasts improves with the forecast horizon.

JEL: C32; E37; E47

Keywords: Multistep forecasts; Var forecasts; Forecast comparisons

*Corresponding author. Tel.: +1 609 258 4811; fax: +1 609 258 5533.
E-mail address: mwatson@princeton.edu (M.W. Watson).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.07.020

1. Introduction

A forecaster making a multiperiod time series forecast—for example, forecasting the unemployment rate six months hence—confronts a choice between using a one-period model iterated forward, or instead using a multiperiod model estimated with a loss function tailored to the forecast horizon. In the case of univariate linear models and quadratic loss, the ‘‘iterated’’ forecast (sometimes called a ‘‘plug-in’’ forecast) entails first estimating an autoregression, then iterating upon that autoregression to obtain the multiperiod forecast. In contrast, the forecast based on the multiperiod model—which, following the literature, we shall call the ‘‘direct’’ forecast—entails regressing a multiperiod-ahead value of the dependent variable on current and past values of the variable. For example, the direct forecast of the unemployment rate six months from now might entail the regression of the unemployment rate, six months hence, against a constant and current and past values of the unemployment rate. But which forecast, the iterated or the direct, should the forecaster use in practice?

The theoretical literature on this problem tends to emphasize the advantages of the direct over indirect forecasts. The idea that direct multiperiod forecasts can be more efficient than iterated forecasts dates at least to Cox (1961), who made the suggestion in the context of exponential smoothing, and to Klein (1968), who suggested direct multiperiod estimation of dynamic forecasting models. Contributions to the theory of iterated vs. direct forecasts include Findley (1983, 1985), Weiss (1991), Tiao and Xu (1993), Lin and Granger (1994), Tiao and Tsay (1994), Clements and Hendry (1996), Bhansali (1996, 1997), Kang (2003), Chevillon and Hendry (2005), and Schorfheide (2005). Bhansali (1999) provides a nice survey of this theoretical literature, and Ing (2003) gives a complete treatment of first-order asymptotics for stationary autoregressions.

Choosing between iterated and direct forecasts involves a trade-off between bias and estimation variance: the iterated method produces more efficient parameter estimates than the direct method, but it is prone to bias if the one-step-ahead model is misspecified. Ignoring estimation uncertainty, if both the iterated model and the direct model have p lags of the dependent variable but the true autoregressive order exceeds p, then the asymptotic mean squared forecast error (MSFE) of the direct forecast typically is less than (and cannot exceed) the MSFE of the iterated forecast (e.g. Findley, 1983). On the other hand, if the true autoregressive order is p or less, then (still ignoring estimation uncertainty) the MSFEs of the direct and iterated methods are the same; because the iterated parameter estimator is more efficient, the MSFE including estimation uncertainty is less for the iterated method when the autoregressive order is correctly specified. Because it seems implausible that typically low-order autoregressive models are correctly specified, in the sense of estimating the best linear predictor, the theoretical literature tends to conclude that the robustness of the direct forecast to model misspecification makes it a more attractive procedure than the bias-prone iterated forecast (Bhansali, 1999; Ing, 2003).

Because the relative efficiency of iterated vs. direct forecasts is theoretically ambiguous and depends on the unknown population best linear projection, the question of which method to choose is an empirical one. Given the practical importance of the question, there are surprisingly few empirical studies of the relative performance of iterated vs. direct forecasts. Findley (1983, 1985) studies univariate models of two of Box and Jenkins's (1976) series (chemical process temperature and sunspots), and Liu (1996) studies univariate autoregressive forecasts of four economic time series. Ang et al. (2005) find that, at least during the 1990s, iterated forecasts of U.S. GDP growth outperform direct forecasts using a measure of short-term interest rates and the term spread. The largest empirical study we are aware of is Kang (2003), who studied univariate autoregressive models of nine U.S. economic time series with mixed results, concluding that the direct method "may or may not improve forecast accuracy" relative to the iterated method (Kang, 2003, p. 398).

This paper undertakes a large-scale empirical comparison of iterated vs. direct forecasts using data on 170 U.S. macroeconomic time series variables, available monthly from 1959 to 2002. Rather than narrowing in on individual series, this study considers the larger questions of whether the iterated or direct forecasts are more accurate on average for the population of U.S. macroeconomic time series, and whether the distribution of MSFEs for direct forecasts is statistically and substantively below the distribution of MSFEs for iterated forecasts. Using these data, we compare iterated and direct forecasts based on univariate autoregressions and bivariate vector autoregressions; in both cases, we consider models with fixed lag order and models with data-dependent lag order choices, using the Akaike Information Criterion (AIC) or, alternatively, the Bayes Information Criterion (BIC).[1] Multiperiod forecasts are computed for horizons of 3, 6, 12, and 24 months.[2] The experimental design uses a pseudo-out-of-sample (or "recursive") forecasting framework; for example, forecasts for the 12 months from January 1985 to December 1985 are computed from models estimated and selected using only data available through December 1984.

This study yields surprisingly sharp results. First, iterated forecasts tend to have lower sample MSFEs than direct forecasts, particularly if the lag length in the one-period ahead model is selected by AIC. Second, these improvements tend to be modest, as one would expect if the main source of the improvements is reduction in estimating uncertainty of the parameters. Third, direct forecasts become increasingly less desirable as the forecast horizon lengthens; this too is consistent with the efficiency of the iterated forecasts outweighing the robustness of the direct forecasts. Fourth, for series measuring wages, prices, and money, direct forecasts improve upon iterated forecasts based on low-order autoregressions, but not upon iterated forecasts from high-order autoregressions, a finding that is consistent with these series having, in effect, a large moving average root (or long lags in the optimal linear

[1]Because possible model misspecification is central to this comparison, data-dependent lag order choice can play an important role: selecting a high-order one-period model can reduce bias but increase estimation uncertainty, and thus increase total MSFE, relative to a lower order direct model (Bhansali, 1997).

[2]Following the literature we consider direct h-step versus one-step-ahead iterated forecasts. In principle, it would also be possible to construct iterated forecasts from k-step-ahead models, where $k<h$, and h/k is an integer.

predictor), as suggested by Nelson and Schwert (1977) and Schwert (1987). In contrast, iterated forecasts from low-order autoregressive models outperform direct forecasts for real activity measures and the other macroeconomic variables in our data set.

2. Forecasting models and methods of comparison

This section describes the iterated and direct forecasting models and estimators. We begin with two general observations.

First, many macroeconomic time series appear to be nonstationary in the sense of having one or more unit roots, while the literature surveyed above focuses on stationary variables.[3] The strategy adopted here is to transform the series of interest to approximate stationarity by taking its first or second difference as needed, to estimate the forecasting model, then to compute the h-step-ahead forecast of the original series produced by that model. For example, the logarithm of real GDP is first transformed by taking its first difference, $\Delta \log \mathrm{GDP}_t$, the forecasting models are estimated using $\Delta \log \mathrm{GDP}_t$, and these models are then used to compute the forecast of the level of the logarithm of GDP, h periods ahead. The transformations used for each series are discussed in the next section and in the data appendix.

Second, all forecasts are recursive (pseudo-out-of-sample), that is, forecasts are based only on values of the series up to the date on which the forecast is made. Parameters are then reestimated in each period, for each forecasting model, using data from the beginning of the sample through the current forecasting date. For forecasts entailing data-based model selection, the order of the model is also selected recursively, and thus can change over the sample as new information is added to the forecast data set.

2.1. Univariate models

Let X_t denote the level or logarithm of the series of interest. The objective is to compute forecasts of X_{t+h}, using information at time t. Let y_t denote the stationary transformation of the series after taking first or second differences. Specifically, suppose that X_t is integrated of order d (is $I(d)$); then $y_t = \Delta^d X_t$, where $d = 0$, 1, or 2 as appropriate.

2.1.1. Iterated AR forecasts

The one-step-ahead AR model for y_t is

$$y_{t+1} = \alpha + \sum_{i=1}^{p} \phi_i y_{t+1-i} + \varepsilon_t. \tag{1}$$

[3] A notable exception is Chevillon and Hendry (2005), which compares iterated and direct forecasts for $I(1)$ autoregressions. Long-horizon iterated forecasts in the local-to-unity autoregression are studied in Stock (1997), and these methods appear well-suited for studying direct forecasts as well.

For the iterated AR forecasts, the parameters α, $\phi_1, \ldots, \phi_p$ in (1) are estimated recursively by OLS, and the forecasts of y_{t+h} are constructed recursively as,

$$\hat{y}^I_{t+h/t} = \hat{\alpha} + \sum_{i=1}^{p} \hat{\phi}_i \hat{y}^I_{t+h-i/t}, \tag{2}$$

where $\hat{y}_{j/t} = y_j$ for $j \leqslant t$. Forecasts of X_{t+h} are then computed by accumulating the values of $\hat{y}^I_{t+k/t}$ as appropriate in the $I(0)$, $I(1)$ and $I(2)$ cases:

$$\hat{X}^I_{t+h|t} = \begin{cases} \hat{y}^I_{t+h|t} & \text{if } X_t \text{ is } I(0), \\ X_t + \sum_{i=1}^{h} \hat{y}^I_{t+i|t} & \text{if} X_t \text{ is } I(1), \\ X_t + h\Delta X_t + \sum_{i=1}^{h}\sum_{j=1}^{i} \hat{y}^I_{t+j|t} & \text{if} X_t \text{ is } I(2). \end{cases} \tag{3}$$

2.1.2. Direct forecasts

The direct estimates of the parameters are the recursive minimizers of the mean squared error of the h-step-ahead criterion function. Accordingly, the parameters are estimated by the OLS regression in which the regressors are a constant and $y_t, \ldots, y_{t-p+1}$ and the dependent variable is y^h_{t+h}, where

$$y^h_{t+h} = \begin{cases} X_{t+h} & \text{if } X_t \text{ is } I(0), \\ X_{t+h} - X_t & \text{if} X_t \text{ is } I(1), \\ \sum_{i=1}^{h}\sum_{j=1}^{i} \Delta^2 X_{t+j} = X_{t+h} - X_t - h\Delta X_t & \text{if} X_t \text{ is } I(2). \end{cases} \tag{4}$$

The direct forecasting regression model is,

$$y^h_{t+h} = \beta + \sum_{i=1}^{p} \rho_i y_{t+1-i} + \varepsilon_{t+h}. \tag{5}$$

The direct estimator of the coefficients is obtained by the recursive estimation of (5) by OLS, where data through period t are used (so that the last observation includes y^h_t on the left-hand side of the regression). The direct forecasts of y^h_{t+h} are

$$\hat{y}^{D,h}_{t+h} = \hat{\beta} + \sum_{i=1}^{p} \hat{\rho}_i y_{t+1-i}. \tag{6}$$

Forecasts of X_{t+h} are then computed from the $\hat{y}^{D,h}_{t+h}$ as appropriate in the $I(0)$, $I(1)$ and $I(2)$ cases: $\hat{X}^D_{t+h/t} = \hat{y}^{D,h}_{t+h/t}$ for $I(0)$, $\hat{X}^D_{t+h/t} = \hat{y}^{D,h}_{t+h} + X_t$ for $I(1)$ and $\hat{X}^D_{t+h/t} = \hat{y}^{D,h}_{t+h} + X_t + h\Delta X_t$ for $I(2)$.[4]

[4]As an alternative, direct forecasts could be computed by first estimating regressions of y_{t+i} onto (1, y_t, y_{t-1}, …, y_{t+1-p}) for $i = 1$, …, h, and then accumulating the forecasts of y_{t+i} to form forecasts of y^h_{t+h}. Because each regression uses the same set of regressors, these forecasts will be identical to those in (6) when data over the sample period are used.

2.1.3. Lag-length determination

Four different methods were used to determine the lag order p: (1) $p = 4$ (fixed); (2) $p = 12$ (fixed); (3) p chosen by the AIC, with $0 \leqslant p \leqslant 12$, and (4) p chosen by the BIC, with $0 \leqslant p \leqslant 12$. For the iterated forecasts, the AIC and BIC were computed using the standard formulas based on the sum of squared residuals (SSR) from the one-step-ahead regression. For the direct forecasts, the AIC and BIC were computed using the SSR from the estimated h-step-ahead regression (5). The AIC and BIC were recomputed at each date, so the order of the selected forecasting model can change from one period to the next, where the model selection and parameter estimates are based only on data through the date of the forecast (period t).

These four choices cover leading cases of theoretical interest. If the true lag order p_0 is finite and if the maximum lag considered exceeds p_0, then the BIC provides a consistent estimator of p_0 and the iterated estimator with BIC is asymptotically efficient. If p_0 is infinite, then the direct estimator with AIC model selection achieves an efficiency bound for direct estimators and this bound is below that for all iterated estimators (see Bhansali (1996) for a precise statement of this result; he shows that the direct estimator bound also is achieved using Shibata's (1980) lag-length selector). In finite samples, however, BIC and AIC lag-length selection introduces additional sampling uncertainty and the short (4 lag) and long (12 lag) fixed-lag autoregressions provide benchmarks against which to compare the BIC and AIC forecasts.[5]

2.2. Multivariate models

We also consider iterated and direct forecasts computed using bivariate vector autoregressions (VARs). For two series i and j, the iterated VARs are specified in terms of the stationary transforms y_{it} and y_{jt}. The iterated forecast is then obtained by iterating forward the VAR and then applying the transformation (3). The h-step direct forecast for series i is obtained from the OLS regression of $y^h_{i,t+h}$ against a constant and p lags each of y_{it} and y_{jt}. In both the iterated and direct models, the same number of lags p is used for both regressors. The same four methods of lag determination are used as in the analysis of the univariate models.

2.3. Estimation and forecast sample periods

Let T_0 denote the first observation used in estimation of the regressions, T_1 denote the date at which the first pseudo-out-of-sample forecast is made, and T_2 denote the date at which the final pseudo-out-of-sample forecast is made. The date T_0 is the date at which the first observation is available (for most series, 1959:1), plus 12 (because 12 lags are used for the long-lagged models), plus the order of integration of the

[5]Other possible lag-length selection methods are possible but are not pursued here. For example, Bhansali (1999) and Schorfheide (2005) suggest selecting the order of the iterated model based on the h-step-ahead SSR of the iterated forecasts, rather than (as is conventional and as we do) based on the one-step-ahead SSR.

series (to allow for first and second differences). For most series, the initial forecast date T_1 is 1979:1; for series that start after 1959:1, T_1 is the later of 1979:1 or the first observation for which all regressions can be estimated using a minimum of 120 observations. The final forecast date depends on the forecast horizon, and is the date of the last available observation (2002:12) minus the forecast horizon h. Thus, for most series, pseudo-out-of-sample forecasts $\hat{X}_{t+h}$ were computed for $t = 1979{:}1$ to 2002:12–h.

The pseudo-out-of-sample forecast error is $e_{t+h} = \hat{X}_{t+h} - X_{t+h}$, and the sample MSFE is,

$$\text{MSFE} = \frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} e_{t+h}^2. \tag{7}$$

The sample MSFE is computed for each series (170 series), for each forecasting method (iterated, with 4 lag choices, and direct, with 4 lags choices), and for each horizon (3, 6, 12, and 24 months). For a given series and horizon, the empirical efficiency of comparable direct and indirect forecasts is assessed by comparing the respective MSFEs.

2.4. Parametric bootstrap method for comparing iterated vs. direct forecasts

The sample MSFE might be less for a direct than an iterated forecast either because the direct forecast is more efficient in population or because of sampling variability. For a single series, the null hypothesis that a direct forecast fails to improve upon an indirect forecast can be tested using suitable versions of tests proposed by West (1996) and Clark and McCracken (2001) for comparing simulated out-of-sample forecasts. Our focus, however, is on whether the direct method improves upon the iterated method on average over the population of macroeconomic variables of interest. Thus the objects of interest in this study are summary measures of the distribution of the relative MSFEs, for example, the mean relative MSFE of the direct estimator, relative to the iterated estimator, across the population of U.S. macroeconomic series, from which we have a sample of 170 series. This comparison of empirical distributions of direct and iterated estimators goes beyond the theoretical results available in the forecast evaluation literature.

To assess the statistical significance of our summary statistics, we therefore implemented a parametric bootstrap that examined the spread of the distribution of relative MSFEs under the null hypothesis that the iterated forecasting model is correctly specified, so that the iterated forecast is efficient. The parametric bootstrap has the following steps:

(1) For each series i, $i = 1, \ldots, 170$, an autoregressive model of order p_i is estimated using the full sample, producing the (one-step ahead) residuals e_{it}.
(2) Previous research suggests that these series are well-modeled by a factor model with a small number of factors (e.g. Stock and Watson, 2002a, b). Accordingly, a static factor model with four factors is fit to these residuals, where the factor

loadings and error variances are estimated using principal components. Separate factor models (different factor loadings and idiosyncratic variances) were estimated in the pre-1982:12 and post-1983:1 periods, where the break point was chosen approximately to coincide with the decline in volatility of many U.S. macroeconomic time series (McConnell and Perez-Quiros, 2000; Kim and Nelson, 1999).

(3) Using the estimated parameters from the dynamic factor models, a pseudo-random data set consisting of 170 series was computed, where the sample periods for the pseudo-data are the same as the actual data. From these pseudo-random data, recursive iterated and direct forecasts are computed as described above, along with their MSFEs. This process is repeated 200 times. This produces an empirical distribution of direct MSFEs, relative to iterated MSFEs, under the hypothesis that the true AR lag length is p_i, $i = 1, \ldots, 170$.

(4) This procedure is repeated for each of the four-lag selection methods. For $p = 4$ (fixed), p_i is fixed at 4, and similarly for $p = 12$ (fixed). For the AIC method, p_i is determined by AIC prior to estimation in step #1, and similarly for the BIC method.

This algorithm provides an estimate of the distribution of relative MSFEs under the null hypothesis that the iterated model is correctly specified (so that the iterated model is asymptotically efficient), where this distribution allows for both sampling uncertainty in the MSFEs and heterogeneity among and time variation of the autoregressive processes. This distribution allows for a comparison of the observed distribution of relative MSFEs to their null distribution. This distribution also can be used to compute bootstrap p-values. For example, consider the comparison of the direct estimator with $p = 4$ to the iterated estimator with $p = 4$. The bootstrap p-value of the hypothesis that the median relative MSFE (where the median is computed across all 170 series for the given horizon) equals its population value that would obtain were the iterated model correctly specified so that the iterated estimator is efficient, against the alternative that the direct estimator is more efficient, is the fraction of the 200 bootstrap draws of the median that are less than the median ratio actually observed in the data.

3. The data

The data set consists of 170 major monthly U.S. macroeconomic time series. The full data set spans 1959:1–2002:12, and most series are available over this full sample. The data set consists of five categories of series:

(A) Income, output, sales, and capacity utilization (38 series);
(B) Employment and unemployment (27 series);
(C) Construction, inventories and orders (37 series);
(D) Interest rates and asset prices (33 series); and
(E) Nominal prices, wages, and money (35 series).

The series and their spans are listed in the data appendix.

The series were subject to three transformations and manipulations. First, series that represent quantities, indexes, and price levels were transformed to logarithms; interest rates, unemployment rates, etc. were left in the original levels; this yields the X_{it} series in the notation of Section 2.

Second, these series were then differenced so that the resulting series were integrated of order zero, yielding the y_{it} series in the notation of Section 2. Generally speaking, real quantities and real prices were treated as $I(1)$. For our primary set of results, we treated nominal prices, wages, and money as $I(1)$. There is a disagreement among practitioners about whether it is best to treat these series as $I(1)$ or $I(2)$; however, so we repeated the analysis treating the series in category (E), prices, wages, and money, as $I(2)$. The results of this sensitivity analysis are discussed briefly in Section 4.

Third, a few of the resulting y_{it} series contained large outliers. So that these outliers would not dominate the results, observations were dropped when $|y_{it}|$ exceeded its median by more than six times its interquartile range.

4. Results for univariate autoregressions

Table 1 summarizes the distributions of the ratios of the MSFE of the direct forecast to the MSFE of the iterated forecast, where the forecasts are based on the same method of lag selection, for different forecast horizons. For example, across the 170 series, when $p = 4$ lags are used for both the iterated and direct forecast and the forecast horizon is $h = 3$, the mean relative MSFE is 0.99, indicating that the direct estimator on average makes a very slight improvement over the indirect estimator, at least by this measure. In 10% of the 170 series, relative MSFE is less than 0.97 at this horizon, while in 10% of the series the relative MSFE exceeds 1.02. The numbers in parentheses in Table 1 are the bootstrap p-values for the test of the hypothesis that the iterated estimator is efficient, computed as described in Section 2.4. For example, the bootstrap p-value of the mean relative MSFE for the $p = 4$ lag model at horizon $h = 3$ is <0.005; according to the bootstrap null distribution, were the iterated model correctly specified, the probability of observing a mean relative MSFE of 0.99 or less is less than 0.5%.[6]

Inspection of Table 1 suggests that whether the iterated or direct estimator is preferred depends on the method of lag selection. For the short-lag selection methods ($p = 4$ and BIC), the direct estimator is preferred; this is particularly true for the BIC, where the improvements in the lower tail of the distribution are substantial, at least through the 12-month horizon. According to the bootstrap p-values, these improvements generally are statistically significant. In contrast, within the long-lag models, the iterated estimator is preferable, and the direct estimator typically does not improve substantially upon the iterated estimator. At the longer 24-month horizon, the iterated forecast is generally preferable to the direct forecast for all four lag selection methods. Indeed, at this horizon the direct forecasts can be

[6]If the iterated model is correctly specified, then the direct estimator is inefficient and the relative MSFE ratio would tend to exceed one.

Table 1
Distributions of relative MSFEs of direct vs. iterated univariate forecasts based on the same lag selection method: all series

Lag selection	Mean/percentile	Forecast horizon			
		3	6	12	24
AR(4)	Mean	0.99 (<0.005)	0.99 (<0.005)	1.00 (<0.005)	1.05 (0.83)
	0.10	0.97 (<0.005)	0.92 (<0.005)	0.90 (<0.005)	0.85 (<0.005)
	0.25	0.99 (<0.005)	0.98 (<0.005)	0.98 (<0.005)	0.97 (0.04)
	0.50	1.00 (0.01)	1.00 (0.03)	1.01 (0.25)	1.05 (>0.995)
	0.75	1.01 (0.85)	1.02 (0.83)	1.04 (0.55)	1.12 (>0.995)
	0.90	1.02 (0.83)	1.04 (0.86)	1.08 (0.82)	1.23 (0.99)
AR(12)	Mean	1.01 (>0.995)	1.01 (>0.995)	1.03 (>0.995)	1.10 (>0.995)
	0.10	0.98 (>0.995)	0.97 (>0.995)	0.95 (>0.995)	0.93 (>0.995)
	0.25	1.00 (>0.995)	0.99 (>0.995)	1.00 (>0.995)	1.02 (>0.995)
	0.50	1.00 (>0.995)	1.01 (>0.995)	1.03 (>0.995)	1.09 (>0.995)
	0.75	1.01 (>0.995)	1.02 (>0.995)	1.06 (>0.995)	1.17 (>0.995)
	0.90	1.02 (0.99)	1.05 (>0.995)	1.11 (>0.995)	1.29 (>0.995)
AR(BIC)	Mean	0.98 (<0.005)	0.97 (<0.005)	0.99 (0.21)	1.05 (0.99)
	0.10	0.92 (<0.005)	0.86 (<0.005)	0.86 (0.01)	0.88 (0.06)
	0.25	0.97 (<0.005)	0.96 (<0.005)	0.97 (0.02)	0.98 (0.50)
	0.50	1.00 (<0.005)	1.00 (0.01)	1.01 (0.56)	1.04 (>0.995)
	0.75	1.01 (0.99)	1.02 (0.91)	1.03 (0.76)	1.12 (>0.995)
	0.90	1.03 (>0.995)	1.05 (>0.995)	1.10 (>0.995)	1.20 (0.98)
AR(AIC)	Mean	1.00 (>0.995)	1.01 (>0.995)	1.02 (>0.995)	1.09 (>0.995)
	0.10	0.97 (0.51)	0.95 (0.99)	0.94 (>0.995)	0.91 (0.97)
	0.25	0.98 (0.08)	0.98 (0.90)	0.98 (0.97)	1.00 (>0.995)
	0.50	1.00 (0.22)	1.00 (>0.995)	1.02 (>0.995)	1.07 (>0.995)
	0.75	1.01 (>0.995)	1.03 (>0.995)	1.06 (>0.995)	1.18 (>0.995)
	0.90	1.04 (>0.995)	1.06 (>0.995)	1.11 (>0.995)	1.29 (>0.995)

Notes: The first entry in each cell is the indicated summary measure of the distribution of the ratio of the MSFE for the direct forecast to the MSFE of the iterated forecast for the lag selection method listed in the first column and the horizon indicated in the column heading. For each cell, the distribution and summary measure are computed over the 170 series being forecasted. The entry in parentheses is the p-value of the test of the hypothesis that the iterated model is efficient, against the alterative that the direct model is more efficient, computed using the parametric bootstrap algorithm described in Section 2.

markedly worse than the iterated forecasts: the 90th percentile of the distribution of relative MSFEs at $h = 24$ exceeds 1.2 for all four lag methods. These results suggest that the robustness of the direct estimator is outweighed by its larger variance.[7]

[7]As a check of this interpretation of the results, a referee suggested that we compute the results separately for the first and second half of the out-of-sample period. The variance component of the MSFE should be smaller in the second half because of the increased sample used for estimation, so that the relative performance of the direct forecast should improve. Indeed, the forecast errors did show a slight improvement in the relative forecast performance of the direct forecast in the second half of the out-of-sample period.

Table 2 breaks down the overall results in Table 1 into two categories of series, the 35 series on nominal prices, wages, and money, and the remaining 135 series. The conclusions are substantially different for these two sets of series. Once the price, wage, and money series are excluded, the iterated forecast is universally preferred to the direct forecast at all horizons. Even in the few cases that the direct estimator has a small p-value, the actual MSFE ratio is one or very nearly so, indicating that the improvement from the direct forecast is too small to be of practical forecasting value. In contrast, for the price, wage, and money series, the direct estimator provides statistically significant improvements over the indirect estimator at all horizons, and at all points in the distribution, for both short-lag models; in some cases, these improvements are large from a practical perspective (for example, the mean relative MSFE at $h = 6$ and 12 for the BIC model is 0.86). But using longer lags in the iterated model eliminates most if not all of the advantages of the direct forecast; for example, at $h = 12$, the mean relative MSFE for the 4 lag forecasts is 0.87, but this rises to 1.00 for the 12-lag forecasts, a value that is statistically significant but provides no practical improvement from using the direct method.

Table 3 summarizes the mean and median relative MSFEs of the various forecasts, all relative to the iterated 4-lag forecasts (so the entry for the iterated AR(4) column is 1.00 by construction), for all series together (part A) and for the two groups of nonprice and price series separately (parts B and C). Also reported are the fraction of series among the 170 series for which a given forecast has the smallest MSFE at that horizon among the eight competitors. Several results stand out. If prices, wages, and money are excluded, then the iterated forecasts produce the lowest MSFEs in the clear majority of cases; the forecasts that are most frequently best are the short-lag iterated forecasts. On average, direct forecasts produce higher MSFEs than the iterated AR(4), sometimes by a substantial margin. The relative performance of the iterated forecasts improves as the horizon lengthens. For the price, wage, and money series, the short-lag iterated forecasts are not successful, and for nearly half these series the direct forecasts are better at short horizons. As the horizon lengthens, however, the iterated forecasts become more desirable.

The fact that short-lag iterated forecasts are most successful for the nonprice series and long-lag iterated forecasts are most successful for the price series suggests that iterated forecasts with a data-dependent lag choice that can select long-lagged models should be best in some average sense. This is in fact the case. For all series combined (Table 3, part A), the mean and median MSFE of the iterated AIC forecast, relative to the iterated AR(4), is as small or smaller than the relative MSFEs of all the other forecasts, at all horizons.

As a sensitivity check, the results for the price, wage, and money variables (the variables in category E in the data appendix) were recomputed, treating these variables as $I(2)$ instead of $I(1)$. The results are summarized in part D of Table 3; full results are available on the Web.[8] In the $I(2)$ specification, the iterated AR(4) forecasts have larger MSFEs, relative to the other forecasts, than they do in the $I(1)$ specification, so that the mean relative MSFEs are smaller in part D than in part C.

[8]www.wws.princeton.edu/~mwatson/

Table 2
Distributions of relative MSFEs of direct vs. iterated univariate forecasts based on the same lag selection method, by category of series

Model	Mean/percentile	Forecast horizon			
		3	6	12	24
(A) *Excluding prices, wages, and money*					
AR(4)	Mean	1.00 (0.01)	1.01 (0.51)	1.03 (0.97)	1.09 (>0.995)
	0.10	0.98 (<0.005)	0.97 (<0.005)	0.96 (0.07)	0.94 (0.25)
	0.25	1.00 (0.01)	0.99 (0.06)	0.99 (0.09)	1.01 (>0.995)
	0.50	1.00 (0.47)	1.01 (0.84)	1.02 (0.82)	1.06 (>0.995)
	0.75	1.01 (0.93)	1.02 (0.89)	1.05 (0.91)	1.14 (>0.995)
	0.90	1.02 (0.96)	1.05 (0.94)	1.10 (0.98)	1.33 (>.995)
AR(12)	Mean	1.01 (>0.995)	1.01 (>0.995)	1.03 (>0.995)	1.11 (>0.995)
	0.10	0.99 (>0.995)	0.97 (0.97)	0.96 (>0.995)	0.93 (0.79)
	0.25	1.00 (>0.995)	0.99 (>0.995)	1.00 (>0.995)	1.03 (>0.995)
	0.50	1.00 (>0.995)	1.01 (0.99)	1.03 (>0.995)	1.11 (>0.995)
	0.75	1.01 (0.96)	1.02 (0.95)	1.06 (>0.995)	1.18 (>0.995)
	0.90	1.02 (0.97)	1.04 (0.94)	1.12 (>0.995)	1.31 (>0.995)
BIC	Mean	1.00 (<0.005)	1.00 (0.01)	1.03 (0.94)	1.07 (0.99)
	0.10	0.96 (<0.005)	0.95 (<0.005)	0.97 (0.30)	0.94 (0.28)
	0.25	0.98 (<0.005)	0.99 (<0.005)	0.99 (0.14)	1.00 (0.98)
	0.50	1.00 (0.04)	1.01 (0.22)	1.02 (0.86)	1.05 (>0.995)
	0.75	1.01 (0.97)	1.02 (0.90)	1.05 (0.94)	1.13 (>0.995)
	0.90	1.03 (>0.995)	1.05 (0.98)	1.11 (>0.995)	1.26 (0.99)
AIC	Mean	1.01 (>0.995)	1.01 (>0.995)	1.04 (>0.995)	1.11 (>0.995)
	0.10	0.97 (0.08)	0.95 (0.17)	0.96 (0.88)	0.95 (0.83)
	0.25	0.99 (<0.005)	0.99 (0.77)	0.99 (0.78)	1.02 (>0.995)
	0.50	1.00 (0.20)	1.01 (0.98)	1.02 (>0.995)	1.10 (>0.995)
	0.75	1.02 (>0.995)	1.03 (>0.995)	1.07 (>0.995)	1.18 (>0.995)
	0.90	1.04 (>0.995)	1.06 (>0.995)	1.12 (>0.995)	1.32 (>0.995)
(B) *Prices, wages, and money only*					
AR(4)	Mean	0.96 (<0.005)	0.90 (<0.005)	0.87 (<0.005)	0.90 (<0.005)
	0.10	0.90 (<0.005)	0.68 (<0.005)	0.57 (<0.005)	0.64 (<0.005)
	0.25	0.95 (<0.005)	0.87 (<0.005)	0.78 (<0.005)	0.77 (<0.005)
	0.50	0.98 (<0.005)	0.95 (<0.005)	0.92 (<0.005)	0.95 (<0.005)
	0.75	0.99 (<0.005)	0.98 (<0.005)	1.00 (<0.005)	1.04 (0.04)
	0.90	1.01 (0.15)	1.01 (<0.005)	1.04 (0.04)	1.10 (0.17)
AR(12)	Mean	1.00 (>0.995)	1.01 (>0.995)	1.00 (>0.995)	1.04 (>0.995)
	0.10	0.98 (>0.995)	0.96 (>0.995)	0.92 (>0.995)	0.89 (>0.995)
	0.25	0.99 (>0.995)	0.98 (>0.995)	0.95 (>0.995)	0.96 (>0.995)
	0.50	1.00 (>0.995)	1.01 (>0.995)	1.01 (>0.995)	1.04 (>0.995)
	0.75	1.01 (>0.995)	1.03 (>0.995)	1.04 (>0.995)	1.13 (>0.995)
	0.90	1.02 (0.98)	1.06 (>0.995)	1.07 (0.96)	1.20 (0.99)
BIC	Mean	0.93 (<0.005)	0.86 (<0.005)	0.86 (<0.005)	0.96 (0.63)
	0.10	0.74 (<0.005)	0.56 (<0.005)	0.56 (<0.005)	0.68 (0.12)
	0.25	0.91 (<0.005)	0.81 (<0.005)	0.72 (<0.005)	0.79 (0.01)
	0.50	0.95 (<0.005)	0.88 (<0.005)	0.91 (0.02)	0.97 (0.20)

Table 2 (*continued*)

Model	Mean/percentile	Forecast horizon			
		3	6	12	24
	0.75	1.00 (0.01)	0.98 (<0.005)	1.00 (0.06)	1.09 (>0.995)
	0.90	1.04 (>0.995)	1.02 (0.82)	1.06 (0.87)	1.14 (0.92)
AIC	Mean	0.98 (0.86)	0.98 (>0.995)	0.96 (>0.995)	1.00 (>0.995)
	0.10	0.92 (0.39)	0.87 (00.99)	0.85 (>.995)	0.81 (0.98)
	0.25	0.95 (0.19)	0.96 (>0.995)	0.89 (0.99)	0.90 (0.95)
	0.50	0.99 (0.54)	0.99 (>0.995)	0.99 (>.995)	1.00 (0.99)
	0.75	1.01 (>0.995)	1.01 (>0.995)	1.03 (0.99)	1.07 (0.99)
	0.90	1.02 (>0.995)	1.06 (>0.995)	1.06 (0.94)	1.18 (0.97)

Notes: See the notes to Table 1.

Adjusting for this difference in the denominators, however, one can see that the general pattern in part D is the same as in the $I(1)$ specifications in part C. In particular, the long-lag specifications outperform the short-lag specifications, and the iterated long-lag forecasts tend to have the best average performance, especially as the horizon increases.

The different results for the wage and price series suggest that the population best linear projections for the nonprice series tend to be short, whereas they tend to be long for the price, wage, and money series. In particular, there could be large moving average root in ARIMA models of prices, wages, and money, where the number of autoregressive lags is short. This possibility has been previously suggested by Nelson and Schwert (1977) and Schwert (1987) and is consistent with the long-lag lengths for backward-looking Phillips curve specifications that Brayton et al. (1999) argue is appropriate for postwar U.S. data. To examine this possibility, Table 4 reports estimated ARIMA(2,1,1) and ARIMA(1,2,1) models for the eight wage and price inflation series for which a direct forecast exhibited the greatest improvement, relative to the iterated AR(4) forecast. In all cases, the MA coefficient is large, in a few cases exceeding 0.9. This large moving average root occurs in both the $I(1)$ specifications and the $I(2)$ specifications for these series, so it is not a simple consequence of overdifferencing. These large moving average coefficients are consistent with a slow decay in the coefficients of the optimal linear predictor for the price and wage series and are consistent with the relatively poor performance of the short-lag iterated estimators, and the relatively good performance of the long-lag direct and iterated estimators, for these series.

5. Results for bivariate forecasts

This data set contains a total of $170 \times 169 = 28{,}730$ different possible pairs of series. To keep the computations tractable, we used a stratified random subsample of

Table 3
Relative MSFEs of each univariate forecast method, relative to iterated AR(4), and the fraction of times each forecast method is best

Forecast horizon	Summary statistic	Iterated					Direct				
		AR(4)	AR(12)	BIC	AIC	Sum	AR(4)	AR(12)	BIC	AIC	Sum
(A) *All series*											
3	Mean	1.00	0.99	1.01	0.99		0.99	0.99	0.99	0.99	
	Median	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	
	Fraction best	0.15	0.22	0.21	0.12	0.70	0.06	0.14	0.06	0.08	0.33
6	Mean	1.00	0.97	1.00	0.97		0.99	0.98	0.98	0.98	
	Median	1.00	1.00	1.00	1.00		1.00	1.01	1.01	1.00	
	Fraction best	0.15	0.25	0.15	0.19	0.75	0.05	0.14	0.05	0.06	0.31
12	Mean	1.00	0.98	1.00	0.97		1.00	1.01	1.00	1.00	
	Median	1.00	1.01	1.01	1.00		1.01	1.03	1.02	1.02	
	Fraction best	0.25	0.23	0.14	0.17	0.79	0.07	0.09	0.05	0.05	0.25
24	Mean	1.00	1.01	1.00	1.00		1.05	1.10	1.05	1.08	
	Median	1.00	1.01	1.00	1.00		1.05	1.09	1.04	1.08	
	Fraction best	0.22	0.22	0.16	0.21	0.81	0.09	0.05	0.05	0.04	0.22
(B) *Excluding prices, wages, and money*											
3	Mean	1.00	1.02	1.01	1.02		1.00	1.03	1.01	1.02	
	Median	1.00	1.02	1.00	1.01		1.00	1.01	1.00	1.01	
	Fraction best	0.19	0.19	0.25	0.11	0.75	0.06	0.10	0.07	0.05	0.28
6	Mean	1.00	1.01	1.02	1.01		1.01	1.03	1.02	1.02	
	Median	1.00	1.02	1.00	1.00		1.01	1.02	1.02	1.01	
	Fraction best	0.19	0.21	0.18	0.21	0.79	0.06	0.10	0.05	0.05	0.27
12	Mean	1.00	1.03	1.01	1.01		1.03	1.06	1.04	1.05	
	Median	1.00	1.02	1.01	1.00		1.02	1.05	1.03	1.03	
	Fraction best	0.30	0.19	0.17	0.16	0.82	0.08	0.05	0.06	0.03	0.22
24	Mean	1.00	1.05	1.01	1.02		1.09	1.15	1.09	1.13	
	Median	1.00	1.01	1.00	1.00		1.06	1.12	1.06	1.10	
	Fraction best	0.26	0.19	0.16	0.21	0.81	0.10	0.03	0.05	0.04	0.22

(C) *Prices, wages, and money*											
3	Mean	1.00	0.85	0.98	0.88		0.96	0.85	0.91	0.86	
	Median	1.00	0.87	0.99	0.89		0.98	0.87	0.91	0.87	
	Fraction best	0.00	0.34	0.03	0.14	0.51	0.06	0.29	0.00	0.17	0.51
6	Mean	1.00	0.79	0.96	0.82		0.90	0.79	0.83	0.80	
	Median	1.00	0.81	0.98	0.83		0.95	0.82	0.85	0.83	
	Fraction best	0.00	0.40	0.06	0.14	0.60	0.03	0.29	0.06	0.09	0.46
12	Mean	1.00	0.80	0.95	0.83		0.87	0.79	0.83	0.80	
	Median	1.00	0.84	0.99	0.85		0.92	0.86	0.87	0.86	
	Fraction best	0.06	0.40	0.03	0.20	0.69	0.03	0.23	0.00	0.11	0.37
24	Mean	1.00	0.88	0.95	0.91		0.90	0.89	0.92	0.89	
	Median	1.00	0.88	0.99	0.90		0.95	0.92	0.95	0.92	
	Fraction best	0.09	0.34	0.17	0.20	0.80	0.03	0.11	0.03	0.06	0.23
(D) *Prices, wages, and money* (*I*(2) *specification*)											
3	Mean	1.00	0.85	0.91	0.85		0.97	0.85	0.87	0.85	
6	Mean	1.00	0.78	0.88	0.79		0.91	0.79	0.80	0.78	
12	Mean	1.00	0.77	0.88	0.78		0.89	0.77	0.79	0.77	
24	Mean	1.00	0.79	0.89	0.80		0.88	0.79	0.81	0.79	

Notes: The entries in the "mean" rows are the mean relative MSFE for the indicated group of series at the indicated horizon, for the column forecasting method, relative to the MSFE for the iterated AR(4) benchmark forecast, where the mean is computed across the 170 series. The entries in the "median" rows are the median of this relative MSFE across the 170 series. The "fraction best" row reports the fraction of the 170 series in which the column forecasting method has the smallest MSFE among the eight possibilities; the sum of these fractions is reported in the "sum" columns respectively for all iterated and for all direct forecasts. The sum of fraction best exceeds 1.0 in some cases because of ties.

Table 4
ARIMA(2,1,1) and ARIMA(1,2,1) models for selected price and wage series

Series	$(1-\phi_1 L-\phi_2 L^2)\Delta X_t = (1-\theta L)\varepsilon_t$			$(1-\phi L)\Delta^2 X_t = (1-\theta L)\varepsilon_t$	
	ϕ_1	ϕ_2	θ	ϕ	θ
Wages, construction (lehcc)	0.57 (0.04)	0.41 (0.04)	0.93 (0.02)	−0.42 (0.04)	0.93 (0.02)
Wages, trade and utilities (lehtu)	0.78 (0.05)	0.21 (0.05)	0.91 (0.02)	−0.21 (0.05)	0.92 (0.02)
PPI, int. materials (pwimsa)	0.76 (0.09)	0.13 (0.07)	0.50 (0.09)	−0.05 (0.06)	0.66 (0.05)
CPI, food (pu81)	1.27 (0.7)	−0.30 (0.06)	0.87 (0.05)	0.32 (0.05)	0.93 (0.02)
CPI, housing (puh)	1.12 (0.08)	−0.15 (0.07)	0.77 (0.06)	0.17 (0.07)	0.81 (0.04)
CPI, apparel (pu83)	1.04 (0.05)	−0.04 (0.05)	0.94 (0.02)	0.03 (0.05)	0.93 (0.02)
CPI, services (pus)	0.91 (0.07)	0.06 (0.06)	0.69 (0.05)	−0.02 (0.06)	0.76 (0.04)
PCE, durables (gmdcd)	1.04 (0.06)	−0.06 (0.06)	0.82 (0.04)	0.08 (0.05)	0.85 (0.03)

Notes: Entries are estimated ARIMA coefficients and standard errors (in parentheses); series mnemonics appear in parentheses in the first column.

these VARs. There are five categories of series, listed as (A) through (E), in Section 3. This produces 10 possible pairs of nonrepeated series categories (AB, AC, …, BC, BD,…,DE). From each pair of categories, 200 pairs of series are randomly drawn (one from each category, with replacement), for a total of 2000 pairs of series. This set of 2000 pairs of series constitutes the data set for the bivariate forecasts. At each horizon and for each forecasting method (iterated or direct, four-lag selection methods), a total of 4000 forecasts are computed from the 2000 pairs, one for each series in the pair.

The iterated and direct forecasts are compared, for the same lag-length selection method, in Table 5, for all the series combined (this is the bivariate counterpart of Table 1). The conclusions are similar to those for the univariate models. Generally speaking, the long-lag ($p = 12$ or AIC) direct forecasts offer little or no average improvements over the long-lag iterated forecasts. For a subset of the pairs, the direct forecasts have lower MSFEs than the iterated forecasts for the short-lag selection methods.

Table 6 summarizes the performance of the various forecasting methods, relative to the iterated VAR(4) benchmark (this is the bivariate counterpart of Table 3). The results are qualitatively similar to those found using the univariate models. For the pairs that do not contain a nominal price, wage or money series (part B of Table 3), the short-lag iterated methods are most frequently the best, and the iterated methods outperform the direct methods in approximately three-fourths of the series. For the price, wage, and money series (part D), the short-lag iterated methods are infrequently best, and are beaten by the long-lag iterated methods and, at short horizons, the long-lag direct methods. At long horizons, the direct methods still outperform the iterated AR(4) benchmark for these series, but do not outperform the long-lag iterated method. Looking across all variables, the iterated method with AIC lag selection tends to produce the lowest, or nearly lowest, MSFE on average across all horizons.

Table 5
Distributions of relative MSFEs of direct vs. iterated bivariate forecasts based on the same lag selection method: all series

Model	Mean/percentile	Forecast horizon			
		3	6	12	24
AR(4)	Mean	1.00	1.00	1.02	1.09
	0.10	0.96	0.90	0.85	0.82
	0.25	0.99	0.97	0.96	0.96
	0.50	1.00	1.01	1.02	1.06
	0.75	1.02	1.04	1.08	1.19
	0.90	1.03	1.07	1.16	1.37
AR(12)	Mean	1.02	1.04	1.07	1.16
	0.10	0.99	0.97	0.95	0.91
	0.25	1.00	1.00	1.01	1.03
	0.50	1.01	1.03	1.06	1.13
	0.75	1.02	1.06	1.12	1.28
	0.90	1.04	1.10	1.20	1.45
BIC	Mean	0.98	0.97	0.99	1.06
	0.10	0.88	0.79	0.78	0.79
	0.25	0.96	0.93	0.92	0.94
	0.50	1.00	1.00	1.00	1.04
	0.75	1.02	1.03	1.06	1.15
	0.90	1.05	1.08	1.15	1.31
AIC	Mean	1.01	1.02	1.05	1.15
	0.10	0.94	0.91	0.89	0.87
	0.25	0.98	0.98	0.98	1.00
	0.50	1.01	1.02	1.05	1.11
	0.75	1.04	1.07	1.13	1.26
	0.90	1.08	1.13	1.23	1.47

Notes: The entries are based on the 2000 randomly selected pairs of series (4000 forecasts for method and horizon), drawn as described in the text. The ‘‘mean’’ and ‘‘median’’ entries are those summary statistics for the relative MSFEs of the column forecasting method, relative to the iterated VAR(4). See the notes in Table 1.

6. Discussion

The main finding from this study is that, for our large data set of monthly U.S. macroeconomic time series, iterated forecasts tend to have smaller MSFEs than direct forecasts, particularly if the iterated forecasts are computed using AIC lag-length selection. The relative performance of the direct forecasts deteriorates as the forecast horizon increases. These findings are consistent with the view that the single-period models, upon which the iterated forecasts are based, are not badly misspecified in the sense that they provide good approximations to the best linear predictor; accordingly, the reduction in estimation variance arising from estimating the one-period ahead model outweighs the reduction in bias obtained from the direct multiperiod model.

Table 6
Relative MSFEs of each bivariate forecast method, relative to iterated VAR(4), and the fraction of times each forecast method is best

Forecast horizon	Percentile	Iterated forecasts					Direct forecasts				
		AR(4)	AR(12)	BIC	AIC	Sum	AR(4)	AR(12)	BIC	AIC	Sum
(A) *All variables*											
3	Mean	1.00	1.03	1.04	1.00		1.00	1.04	1.01	0.01	
	Median	1.00	1.04	1.01	1.00		1.00	1.05	1.01	0.02	
	Fraction best	0.15	0.14	0.27	0.13	0.69	0.08	0.07	0.10	0.08	0.33
6	Mean	1.00	1.00	1.06	0.99		0.99	1.03	1.01	0.01	
	Median	1.00	1.03	1.02	1.00		1.01	1.05	1.01	0.02	
	Fraction best	0.18	0.20	0.24	0.14	0.75	0.07	0.07	0.07	0.06	0.26
12	Mean	1.00	1.00	1.06	0.99		1.01	1.07	1.03	0.04	
	Median	1.00	1.03	1.03	1.00		1.02	1.09	1.03	0.05	
	Fraction best	0.21	0.21	0.19	0.16	0.77	0.06	0.08	0.06	0.04	0.25
24	Mean	1.00	1.03	1.04	0.99		1.09	1.19	1.09	0.15	
	Median	1.00	1.03	1.02	1.00		1.06	1.15	1.07	0.11	
	Fraction best	0.22	0.22	0.19	0.19	0.81	0.05	0.07	0.06	0.04	0.21
(B) *Pairs not including wages, prices, or money*											
3	Mean	1.00	1.06	1.03	1.02		1.01	1.08	1.02	0.04	
	Median	1.00	1.05	1.00	1.01		1.01	1.06	1.01	0.02	
	Fraction best	0.18	0.10	0.29	0.13	0.71	0.09	0.04	0.11	0.07	0.31
6	Mean	1.00	1.04	1.04	1.01		1.02	1.08	1.03	0.05	
	Median	1.00	1.05	1.01	1.01		1.01	1.07	1.02	0.03	
	Fraction best	0.22	0.16	0.25	0.14	0.77	0.08	0.05	0.08	0.04	0.25
12	Mean	1.00	1.05	1.04	1.01		1.05	1.12	1.05	0.08	
	Median	1.00	1.04	1.02	1.00		1.03	1.11	1.03	0.06	
	Fraction best	0.24	0.17	0.20	0.17	0.78	0.06	0.06	0.08	0.04	0.24
24	Mean	1.00	1.07	1.02	1.01		1.12	1.23	1.10	0.18	
	Median	1.00	1.04	1.01	1.00		1.08	1.18	1.07	0.12	
	Fraction best	0.23	0.17	0.22	0.19	0.81	0.05	0.06	0.07	0.03	0.22
(C) *Nonprice, wage, money variables in pairs that include a price, wage, money variable*											
3	Mean	1.00	1.08	1.01	1.03		1.01	1.09	1.01	1.06	
	Median	1.00	1.07	1.00	1.02		1.01	1.08	1.01	1.05	
	Fraction best	0.18	0.04	0.41	0.09	0.73	0.08	0.02	0.14	0.04	0.28
6	Mean	1.00	1.07	1.01	1.02		1.03	1.10	1.03	1.08	
	Median	1.00	1.06	1.00	1.02		1.02	1.09	1.02	1.06	
	Fraction best	0.22	0.07	0.41	0.13	0.82	0.06	0.02	0.06	0.04	0.18
12	Mean	1.00	1.08	1.02	1.03		1.07	1.16	1.07	1.13	
	Median	1.00	1.07	1.02	1.02		1.05	1.14	1.05	1.11	
	Fraction best	0.30	0.10	0.29	0.13	0.83	0.07	0.03	0.05	0.02	0.18
24	Mean	1.00	1.09	1.03	1.04		1.16	1.32	1.16	1.28	
	Median	1.00	1.07	1.02	1.02		1.13	1.26	1.12	1.23	
	Fraction best	0.31	0.14	0.23	0.18	0.86	0.04	0.04	0.04	0.03	0.16
(D) *Price, wage, money variables*											
3	Mean	1.00	0.88	1.11	0.92		0.97	0.88	1.01	0.89	
	Median	1.00	0.89	1.05	0.94		0.98	0.89	1.01	0.91	
	Fraction best	0.01	0.38	0.03	0.16	0.58	0.07	0.20	0.02	0.14	0.43

Table 6 (*continued*)

Forecast horizon	Percentile	Iterated forecasts					Direct forecasts				
		AR(4)	AR(12)	BIC	AIC	Sum	AR(4)	AR(12)	BIC	AIC	Sum
6	Mean	1.00	0.80	1.15	0.88		0.90	0.82	0.93	0.83	
	Median	1.00	0.82	1.11	0.89		0.92	0.84	0.95	0.84	
	Fraction best	0.01	0.47	0.03	0.14	0.64	0.04	0.17	0.04	0.12	0.37
12	Mean	1.00	0.79	1.15	0.87		0.87	0.81	0.92	0.82	
	Median	1.00	0.81	1.12	0.89		0.89	0.84	0.95	0.84	
	Fraction best	0.06	0.44	0.04	0.15	0.69	0.04	0.21	0.01	0.07	0.33
24	Mean	1.00	0.85	1.10	0.90		0.91	0.93	0.97	0.92	
	Median	1.00	0.83	1.08	0.91		0.92	0.93	0.98	0.92	
	Fraction best	0.13	0.44	0.04	0.17	0.78	0.03	0.12	0.03	0.07	0.25

Notes: The entries are based on the 2000 randomly selected pairs of series (4000 forecasts for method and horizon), drawn as described in the text. The "mean" and "median" entries are those summary statistics for the relative MSFEs of the column forecasting method, relative to the iterated VAR(4). See the notes to Table 3.

There is considerable heterogeneity in these data with respect to the best lag order of the one-period model: for nominal price, wage, and money series, a long-lag order is indicated, whereas for the other series a short-lag order is more appropriate. Overall, this heterogeneity appears to be handled adequately by using AIC lag-length selection when specifying the model for the iterated forecast.

It is interesting to note that these findings in favor of the iterated forecasts are at odds with some of the theoretical literature, which emphasizes the robustness and bias reduction of the direct forecasts in contrast to the special parametric, finite-lag assumptions that underlie optimality properties for the iterated forecasts (cf. Bhansali, 1999; Ing, 2003). It appears that, in practice, the robustness and bias reduction obtained using direct forecasts do not justify the price paid in terms of increased sampling variance.

Acknowledgments

The authors thank Jin-Lung Lin, Frank Schorfheide, Ken West, and two referees for comments. This research was funded in part by NSF grant SBR-0214131.

Appendix A. Data Appendix

This appendix lists the time series used in the empirical analysis. The series were either taken directly from the DRI-McGraw Hill Basic Economics database, in which case the original mnemonics are used, or they were produced by authors' calculations based on data from that database, in which case the authors' calculations and original DRI/McGraw series mnemonics are summarized in the data description field. Following the series name is a transformation code, the sample period for the data series, and a short data description. The transformations are (Lev) level of the series; (Δ) first difference; (Ln) logarithm of the series; (ΔLn) first difference of the logarithm. The following abbreviations appear in the data descriptions: SA = seasonally adjusted; NSA = not seasonally adjusted; SAAR = seasonally adjusted at an annual rate; AC = authors' calculations.

Series	Trans.	Sample period	Description
(A) Income, output, sales, capacity utilization			
Msmq	ΔLn	1967: 1–2001: 7	Sales, business—manufacturing (chained)
ips11	ΔLn	1959: 1–2002: 12	Industrial production index—products, total
ips299	ΔLn	1959: 1–2002: 12	Industrial production index—final products
ips12	ΔLn	1959: 1–2002: 12	Industrial production index—consumer goods
ips13	ΔLn	1959: 1–2002: 12	Industrial production index—durable consumer goods
ips18	ΔLn	1959: 1–2002: 12	Industrial production index—nondurable consumer goods
ips25	ΔLn	1959: 1–2002: 12	Industrial production index—business equipment
ipi	ΔLn	1959: 1–2002: 10	industrial production: intermediate products (1992 = 100, sa)
ips32	ΔLn	1959: 1–2002: 12	Industrial production index—materials
ips34	ΔLn	1959: 1–2002: 12	Industrial production index—durable goods materials
ips38	ΔLn	1959: 1–2002: 12	Industrial production index—nondurable goods materials
ips43	ΔLn	1959: 1–2002: 12	Industrial production index—manufacturing (sic)
ipd	ΔLn	1959: 1–2002: 10	Industrial production: durable manufacturing (1992 = 100, sa)
ipn	ΔLn	1959: 1–2002: 10	Industrial production: nondurable manufacturing (1992 = 100, sa)
ipmin	ΔLn	1959: 1–2002: 10	Industrial production: mining (1992 = 100, sa)
iput	ΔLn	1959: 1–2002: 10	Industrial production: utilities (1992 = 100, sa)
utl10	Lev	1967: 1–2002: 12	Capacity utilization—total index
utl11	Lev	1959: 1–2002: 12	Capacity utilization—manufacturing (sic)
utl13	Lev	1967: 1–2002: 12	Capacity utilization—durable manufacturing (naics)

utl25	Lev	1967: 1–2002: 12	Capacity utilization—nondurable manufacturing (naics)
utl35	Lev	1967: 1–2002: 12	Capacity utilization—mining naics = 21
utl36	Lev	1967: 1–2002: 12	Capacity utilization—electric and gas utilities
gmpyq	ΔLn	1959: 1–2002: 12	Personal income (chained, series #52, bil 92$, saar)
gmyxpq	ΔLn	1959: 1–2002: 12	Personal income less transfer payments (chained, series #51, bil 92$, saar)
gmcq	ΔLn	1967: 1–2002: 12	Personal consumption expenditure (chained)—total (bil 92$, saar)
gmcdq	ΔLn	1967: 1–2002: 12	Personal consumption expenditure (chained)—total durables (bil 1996$, saar)
gmcnq	ΔLn	1967: 1–2002: 12	Personal consumption expenditure (chained)—nondurables (bil 96$,saar)
gmcsq	ΔLn	1967: 1–2002: 12	Personal consumption expenditure (chained)—services (bil 92$, saar)
gmcanq	ΔLn	1967: 1–2002: 12	Personal consumption expenditure (chained)—new cars (bil 1996$, saar)
wtq	ΔLn	1959: 1–2001: 7	Merch wholesalers: total (mil of chained 1996 dollars, sa)
wtdq	ΔLn	1959: 1–2001: 7	Merch wholesalers: durable goods total (mil of chained 1996 dollars, sa)
msdq	ΔLn	1959: 1–2001: 7	Mfg. and trade: mfg.; durable goods (mil of chained 1996 dollars, sa)
msmtq	ΔLn	1959: 1—2001: 7	Mfg. and trade: total (mil of chained 1996 dollars, sa)
msnq	ΔLn	1959: 1–2001: 7	Mfg. and trade: mfg.;nondurable goods (mil of chained 1996 dollars, sa)
wtnq	ΔLn	1959: 1–2001: 7	Merch. wholesalers: nondurable goods (mil of chained 1996 dollars, sa)
rtdrq	ΔLn	1967: 1–2001: 4	Retail sales durables, real (rtdr/pucd, AC)
rtnrq	ΔLn	1967: 1–2001: 4	Retail sales nondurables, real (rtnr/pu882, AC)
ips10	ΔLn	1959: 1–2002: 12	Industrial production index—total index
(B) Employment and unemployment			
lpnag	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: total (thous., sa)
lhu26	Lev	1959: 1–2002: 12	Unemployed by duration: persons unemployed 15–26 wks (thous., sa)
lpgd	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: goods-producing (thous., sa)

lhu15	Lev	1959: 1–2002: 12	Unemployed by duration: persons unemployed 15 wks + (thous., sa)
lp	ΔLn	1959: 1–2002: 12	Employees on nonag payrolls: total, private (thous., sa)
lpcc	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: contract construction (thous., sa)
lhelx	Ln	1959: 1–2002: 12	Employment: ratio; help-wanted ads: no. unemployed clf
lhu5	Lev	1959: 1–2002: 12	Unemployed by duration: persons unemployed less than 5 wks (thous., sa)
lhu14	Lev	1959: 1–2002: 12	Unemployed by duration: persons unemployed 5–14 wks (thous., sa)
lpsp	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: service-producing (thous., sa)
lptu	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: trans. and public utilities (thous., sa)
lpt	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: wholesale and retail trade (thous., sa)
lpfr	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: finance, insur. and real estate (thous., sa)
lps	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: services (thous., sa)
lpgov	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: government (thous., sa)
lw	Dif	1964: 1–2002: 12	Avg. weekly hrs. of prod. wkrs.: total private (sa)
lphrm	Lev	1959: 1–2002: 12	Avg. weekly hrs. of production wkrs.: manufacturing (sa)
lpmosa	Lev	1959: 1–2002: 12	Avg. weekly hrs. of prod. wkrs.: mfg., overtime hrs. (sa)
lhu680	Lev	1959: 1–2002: 12	Unemployed by duration: average (mean) duration in weeks (sa)
lhur	Lev	1959: 1–2002: 12	Unemployment rate: all workers, 16 years and over (%, sa)
lpen	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: nondurable goods (thous., sa)
lpem	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: manufacturing (thous., sa)
lhel	ΔLn	1959: 1–2002: 12	Index of help-wanted advertising in newspapers (1967 = 100; sa)
lped	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: durable goods (thous., sa)
lhem	ΔLn	1959: 1–2002: 12	Civilian labor force: employed, total (thous., sa)

lhnag	ΔLn	1959: 1–2002: 12	Civilian labor force: employed, nonagric. industries (thous., sa)
lpmi	ΔLn	1959: 1–2002: 12	Employees on nonag. payrolls: mining (thous., sa)
(C) Construction, inventories and orders			
hssou	Ln	1959: 1–2002: 12	Housing starts: south (thous.u., s.a.)
contc	ΔLn	1964: 1–2002: 12	Construct.put in place: total priv and public 1987 $ (mil$, saar)
conpc	ΔLn	1964: 1–2002: 12	Construct. put in place: total private 1987 $ (mil$, saar)
conqc	ΔLn	1964: 1–2002: 12	New construction put in place—public (c30)
condo9	Ln	1963: 1–2002: 12	Construct.contracts: comm'l and indus.bldgs (mil.sq.ft.floor sp.;sa)
hniv	Ln	1963: 1–2002: 12	New 1-family houses for sale at end of month (thous, sa)
hnr	Ln	1963: 1–2002: 12	New 1-family houses, month's supply @ current sales rate (ratio)
hns	Ln	1963: 1–2002: 12	New 1-family houses sold during month (thous, saar)
hsbr	Ln	1959: 1–2002: 12	Housing authorized: total new priv housing units (thous., saar)
hswst	Ln	1959: 1–2002: 12	Housing starts: west (thous.u.) s.a.
hmob	Ln	1959: 1–2002: 12	Mobile homes: manufacturers' shipments (thous.of units, saar)
hsmw	Ln	1959: 1–2002: 12	Housing starts: midwest (thous.u.) s.a.
hsne	Ln	1959: 1–2002: 12	Housing starts: northeast (thous.u.) s.a.
hsfr	Ln	1959: 1–2002: 12	Housing starts: nonfarm (1947–58);total farm and nonfarm (1959-, thous., sa)
ivmtq	ΔLn	1959: 1–2001: 7	Mfg. and trade inventories: total (mil of chained 1996, sa)
ivmfgq	ΔLn	1959: 1–2001: 7	Inventories, business, mfg. (mil of chained 1996 dollars, sa)
ivmfdq	ΔLn	1959: 1–2001: 7	Inventories, business durables (mil of chained 1996 dollars, sa)
ivmfnq	ΔLn	1959: 1–2001: 7	Inventories, business, nondurables (mil of chained 1996 dollars, sa)
ivwrq	ΔLn	1959: 1–2001: 7	Mfg. and trade inv: merchant wholesalers (mil of chained 1996 dollars, sa)
ivrrq	ΔLn	1959: 1–2001: 7	Mfg. and trade inv: retail trade (mil of chained 1996 dollars, sa)
ivsrq	ΔLn	1959: 1–2001: 7	Ratio for mfg. and trade: inventory/sales (chained 1996 dollars, sa)

ivsrmq	ΔLn	1959: 1–2001: 7	Ratio for mfg. and trade: mfg.;inventory/ sales (1996 $ s.a.)
ivsrwq	ΔLn	1959: 1–2001: 7	Ratio for mfg. and trade: wholesaler; inventory/sales (1996 $ s.a.)
ivsrrq	ΔLn	1959: 1–2001: 7	Ratio for mfg. and trade: retail trade;inventory/sales(1996 $ s.a.)
pmi	Lev	1959: 1–2002: 12	Purchasing managers' index (sa)
pmp	Lev	1959: 1–2002: 12	Napm production index (percent)
pmno	Lev	1959: 1–2002: 12	Napm new orders index (percent)
pmdel	Lev	1959: 1–2002: 12	Napm vendor deliveries index (percent)
pmnv	Lev	1959: 1–2002: 12	Napm inventories index (percent)
pmemp	Lev	1959: 1–2002: 12	Napm employment index (percent)
pmcp	Lev	1959: 1–2002: 12	Napm commodity prices index (percent)
mocmq	ΔLn	1959: 1–2002: 12	New orders (net)—consumer goods and materials, 1996 dollars (bci)
msondq	ΔLn	1959: 1–2002: 12	New orders, nondefense capital goods, in 1996 dollars (bci)
moq	ΔLn	1959: 1–2001: 5	Mfg. new orders: all manufacturing industries, total, real (mo/pwfsa, AC)
mdoq	ΔLn	1959: 1–2001: 5	Mfg. new orders: durable goods industries, total, real (mdo/pwfsa, AC)
muq	ΔLn	1959: 1–2001: 5	Mfg. unfilled orders: all manufacturing industries, total (mu/pwfsa, AC)
mduq	ΔLn	1959: 1–2001: 5	Mfg. unfilled orders: durable goods industries, total (mdu/pwfsa, AC)
(D) Interest rates and asset prices			
fygt10	Δ	1959: 1–2002: 12	Interest rate: U.S. treasury const maturities, 10-yr.(% per ann, nsa)
fclnq	ΔLn	1959: 1–2002: 12	Commercial and industrial loans oustanding in 1996 dollars (bci)
fsncom	ΔLn	1959: 1–2002: 12	Nyse common stock price index: composite (12/31/65 = 50)
fsnin	ΔLn	1966: 1–2002: 12	Nyse common stock price index: industrial (12/31/65 = 50)
fsntr	ΔLn	1966: 1–2002: 12	Nyse common stock price index: transportation (12/31/65 = 50)
fsnut	ΔLn	1966: 1–2002: 12	Nyse common stock price index: utility (12/31/65 = 50)
fsnfi	ΔLn	1966: 1–2002: 12	Nyse common stock price index: finance (12/31/65 = 50)
fspcom	ΔLn	1959: 1–2002: 12	S&p's common stock price index: composite (1941-43 = 10)
fspin	ΔLn	1959: 1–2002: 12	S&p's common stock price index: industrials (1941-43 = 10)

fsdxp	Lev	1959: 1–2002: 12	S&p's composite common stock: dividend yield (% per annum)
fspxe	Lev	1959: 1–2002: 12	S&p's composite common stock: price-earnings ratio (%, nsa)
fyff	Δ	1959: 1–2002: 12	Interest rate: federal funds (effective) (% per annum, nsa)
fygm3	Δ	1959: 1–2002: 12	Interest rate: U.S. treasury bills, sec mkt, 3-mo. (% per ann, nsa)
fygm6	Δ	1959: 1–2002: 12	Interest rate: U.S. treasury bills, sec mkt, 6-mo. (% per ann, nsa)
fygt1	Δ	1959: 1–2002: 12	Interest rate: U.S. treasury const maturities, 1-yr. (% per ann, nsa)
fygt5	Δ	1959: 1–2002: 12	Interest rate: U.S. treasury const maturities, 5-yr. (% per ann, nsa)
fm2dq	ΔLn	1959: 1–2002: 2	Money supply—m2 in 1996 dollars (bci)
fyaaac	Δ	1959: 1–2002: 12	Bond yield: moody's aaa corporate (% per annum)
fybaac	Δ	1959: 1–2002: 12	Bond yield: moody's baa corporate (% per annum)
fymcle	Δ	1963: 1–2002: 12	Effective interest rate: conventional home mtge loans closed (%)
sfygm3	Lev	1959: 1–2002: 12	Fygm3-fyff (AC)
sfygm6	Lev	1959: 1–2002: 12	Fygm6-fyff (AC)
sfygt1	Lev	1959: 1–2002: 12	Fygt1-fyff (AC)
sfygt5	Lev	1959: 1–2002: 12	Fygt5-fyff (AC)
sfygt10	Lev	1959: 1–2002: 12	Fygt10-fyff (AC)
sfyaaac	Lev	1959: 1–2002: 12	Fyaaac-fyff (AC)
sfybaac	Lev	1959: 1–2002: 12	Fybaac-fyff (AC)
sfymcle	Lev	1963: 1–2002: 12	Fymcle-fyff (AC)
exrus	ΔLn	1959: 1–2002: 12	United states; effective exchange rate (merm, index no.)
exrsw	ΔLn	1959: 1–2002: 12	Foreign exchange rate: switzerland (swiss franc per U.S. $)
exrjan	ΔLn	1959: 1–2002: 12	Foreign exchange rate: japan (yen per U.S. $)
exruk	ΔLn	1959: 1–2002: 12	Foreign exchange rate: united kingdom (cents per pound)
exrcan	ΔLn	1959: 1–2002: 12	Foreign exchange rate: canada (canadian $ per U.S. $)
(E) Nominal prices, wages, and money			
fm1	ΔLn	1959: 1–2002: 12	Money stock: m1(curr, trav. cks, dem dep, other ck'able dep, bil$, sa)
fm2	ΔLn	1959: 1–2002: 12	Money stock: m2 (m1 + o'nite rps, euro$, g/p&b/d mmmfs & sav & sm time dep, bil$)

fm3	ΔLn	1959: 1–2002: 12	Money stock: m3 (m2 + lg time dep, term rp's & inst only mmmfs, bil$, sa)
fmfba	ΔLn	1959: 1–2002: 12	Monetary base, adj for reserve requirement changes (mil$, sa)
fmrra	ΔLn	1959: 1–2002: 12	Depository inst reserves: total, adj for reserve req chgs (mil$, sa)
leh	ΔLn	1964: 1–2002: 12	Avg hr earnings of prod. wkrs.: total private nonagric ($, sa)
lehcc	ΔLn	1959: 1–2002: 12	Avg hr earnings of constr. wkrs.: construction ($, sa)
lehm	ΔLn	1959: 1–2002: 12	Avg hr earnings of prod. wkrs.: manufacturing ($, sa)
lehtu	ΔLn	1964: 1–2002: 12	Avg hr earnings of nonsupv. wkrs.: trans & public util($, sa)
lehtt	ΔLn	1964: 1–2002: 12	Avg hr earnings of prod. wkrs.: wholesale & retail trade (sa)
lehfr	ΔLn	1964: 1–2002: 12	Avg hr earnings of nonsupv. wkrs.: finance, insur, real est ($, sa)
lehs	ΔLn	1964: 1–2002: 12	Avg hr earnings of nonsupv. wkrs.: services ($, sa)
pwfsa	ΔLn	1959: 1–2002: 12	Producer price index: finished goods (82 = 100, sa)
pwfcsa	ΔLn	1959: 1–2002: 12	Producer price index: finished consumer goods (82 = 100, sa)
pwimsa	ΔLn	1959: 1–2002: 12	Producer price index: intermed mat.supplies and components (82 = 100, sa)
pwcmsa	ΔLn	1959: 1–2002: 12	Producer price index: crude materials (82 = 100, sa)
pwfxsa	ΔLn	1967: 1–2002: 12	Producer price index: finished goods,excl. foods (82 = 100, sa)
psm99q	ΔLn	1959: 1–2002: 12	Index of sensitive materials prices (1990 = 100, bci-99a)
punew	ΔLn	1959: 1–2002: 12	Cpi-u: all items (82–84 = 100, sa)
pu81	ΔLn	1967: 1–2002: 12	Cpi-u: food and beverages (82–84 = 100, sa)
puh	ΔLn	1967: 1–2002: 12	Cpi-u: housing (82–84 = 100, sa)
pu83	ΔLn	1959: 1–2002: 12	Cpi-u: apparel and upkeep (82–84 = 100, sa)
pu84	ΔLn	1959: 1–2002: 12	Cpi-u: transportation (82–84 = 100, sa)
pu85	ΔLn	1959: 1–2002: 12	Cpi-u: medical care (82–84 = 100, sa)
pu882	ΔLn	1959: 1–2002: 12	Cpi-u: nondurables (1982–84 = 100, sa)
puc	ΔLn	1959: 1–2002: 12	Cpi-u: commodities (82–84 = 100, sa)
pucd	ΔLn	1959: 1–2002: 12	Cpi-u:durables (82–84 = 100, sa)
pus	ΔLn	1959: 1–2002: 12	Cpi-u: services (82–84 = 100, sa)

puxf	ΔLn	1959: 1–2002: 12	Cpi-u: all items less food (82–84 = 100, sa)
puxhs	ΔLn	1959: 1–2002: 12	Cpi-u: all items less shelter (82–84 = 100, sa)
puxm	ΔLn	1959: 1–2002: 12	Cpi-u: all items less midical care (82–84 = 100, sa)
gmdc	ΔLn	1959: 1–2002: 12	Pce, impl pr defl: pce (1987 = 100)
gmdcd	ΔLn	1959: 1–2002: 12	Pce, impl pr defl: pce; durables (1987 = 100)
gmdcn	ΔLn	1959: 1–2002: 12	Pce, impl pr defl: pce; nondurables (1996 = 100)
gmdcs	ΔLn	1959: 1–2002: 12	Pce, impl pr defl: pce; services (1987 = 100)

References

Ang, A., Piazzesi, M., Wei, M., 2005. What does the yield curve tell us about GDP growth? Journal of Econometrics (forthcoming).

Bhansali, R.J., 1996. Asymptotically efficient autogressive model selection for multistep prediction. Annals of the Institute of Statistical Mathematics 48 (3), 577–602.

Bhansali, R.J., 1997. Direct autoregressive predictions for multistep prediction: order selection and performance relative to the plug in predictors. Statistica Sinca 7, 425–449.

Bhansali, R.J., 1999. Parameter estimation and model selection for multistep prediction of a time series: a review. In: Ghosh, S. (Ed.), Asymptotics, Nonparametrics, and Time Series. Marcel Dekker, New York, pp. 201–225.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control, 2nd ed. Holden Day, New York.

Brayton, F., Roberts, J., Williams, J., 1999. What's Happened to the Phillips Curve? manuscript, FRB Working Paper 1999-49, U.S. Federal Reserve Board.

Chevillon, G., Hendry, D.F., 2005. Non-parametric direct multi-step estimation for forecasting economic processes. International Journal of Forecasting 21, 201–218.

Clark, T.E., McCracken, M.W., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85–100.

Clements, M.P., Hendry, D.F., 1996. Multi-step estimation for forecasting. Oxford Bulletin of Economics and Statistics 58, 657–684.

Cox, D.R., 1961. Prediction by exponentially weighted moving averages and related methods. JRSS, B 23, 414–422.

Findley, D.F., 1983. On the use of multiple models for multi-period forecasting. Proceedings of the Business and Statistics Section, American Statistical Association, 528–531.

Findley, D.F., 1985. Model selection for multi-step-ahead forecasting. In: Baker, H.A., Young, P.C. (Eds.), Proceedings of the Seventh Symposium on Identification and System Parameter Estimation. Pergamon, Oxford, pp. 1039–1044.

Ing., C.-K., 2003. Multistep prediction in autoregressive processes. Econometric Theory 19, 254–279.

Kang, I.-B., 2003. Multi-period forecasting using different models for different horizons: an application to U.S. economic time series data. International Journal of Forecasting 19, 387–400.

Kim, Chang-Jin., Charles, R. Nelson., 1999. Has the U.S. economy become more stable? a Bayesian approach based on a Markov-switching model of the business cycle. The Review of Economics and Statistics 81, 608–616.

Klein, L.R., 1968. An Essay on the Theory of Economics Prediction. Sanomaprint, Helsinki, Finland (Yrjö Jansen lectures).

Lin, J.-L., Granger, C.W.G., 1994. Forecasting from non-linear models in practice. Journal of Forecasting 13, 1–9.

Liu, S.-I., 1996. Model selection for multiperiod forecasts. Biometrika 83, 861–873.

McConnell, M.M., Perez-Quiros, G., 2000. Output fluctuations in the United States: what has changed since the early 1980s. American Economic Review 90 (5), 1464–1476.

Nelson, C.R., Schwert, G.W., 1977. On testing the hypothesis that the real rate of interest is constant. American Economic Review 67, 478–486.

Schorfheide, F., 2005. VAR forecasting under misspecification. Journal of Econometrics (forthcoming).

Schwert, G.W., 1987. Effects of model misspecification on tests for unit roots in macroeconomic data. Journal of Monetary Economics 20, 73–103.

Shibata, R., 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. Annals of Statistics 8 (1), 1464–1470.

Stock, J.H., 1997. Cointegration, long-run comovements, and long-horizon forecasting. In: Kreps, D., Wallis, K.F. (Eds.), Advances in Econometrics: Proceedings of the Seventh World Congress of the Econometric Society, vol. III. Cambridge, Cambridge University Press, pp. 34–60.

Stock, J.H., Watson, M.W., 2002a. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Stock, J.H., Watson, M.W., 2002b. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20 (2), 147–162.

Tiao, G.C., Xu, D., 1993. Robustness of MLE for multi-step predictions: the exponential smoothing case. Biometrika 80, 623–641.

Tiao, G.C., Tsay, R.S., 1994. Some advances in non-linear and adaptive modelling in time-series. Journal of Forecasting 13, 109–131.

Weiss, A.A., 1991. Multi-step estimation and forecasting in dynamic models. Journal of Econometrics 48, 135–149.

West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.