1
A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series

JAMES H. STOCK AND MARK W. WATSON

1 Introduction

This paper is inspired by four themes that run through Clive Granger's extraordinary body of research on time series analysis and economic forecasting. First, it is plausible that the complicated forces that drive economic events introduce nonlinear dynamics into aggregate time series variables, so an important research program is modeling and exploiting these nonlinearities for forecasting (Granger, 1993; Granger and Teräsvirta, 1993; Granger, Teräsvirta, and Anderson, 1993; Granger and Lin, 1994; Teräsvirta *et al.* 1994). Second, a dominant feature of economic time series data is the considerable persistence, or long-range dependence, of those series (Granger, 1966), and the correct modeling of this persistence is a critical step in constructing reliable forecasts over medium- and long-term forecasting horizons (Granger and Newbold, 1977). Third, because time series models are simplifications of complicated processes that are imperfectly understood, combinations of forecasts from different models using different information might well outperform the forecasts produced by any particular model (Bates and Granger, 1969; Granger and Ramanathan, 1984). Fourth, time series models and forecasting methods, however appealing from a theoretical point of view, ultimately must be judged by their performance in real economic forecasting applications.

Inspired by these themes, we tackle five specific questions in the context of forecasting US macroeconomic time series. First, do nonlinear time series models produce forecasts that improve upon linear models in real time? Second, if there are benefits to using nonlinear models, are the benefits greatest for relatively tightly parameterized models or for more nonparametric approaches?

This research was supported in part by National Science Foundation grant SBR–9409629 and SBR–9730489. We benefited from comments on this work by Frank Diebold, Robin Lumsdaine, Serena Ng, Pierre Perron, Norm Swanson, Andreas Weigend, Richard Zeckhauser and members of the NECSI Time Series Working Group.

Third, can forecasts at the six month or one year horizon be improved by using preliminary evidence on the persistence of the time series to select the forecasting model? Fourth, do combination forecasts outperform forecasts based on a single method across a range of time series, and, if so, how heavily should these combination forecasts weight the currently best performing forecasting methods? Finally, are the gains from using these advanced methods over simple autoregressive forecasts large enough to justify their use?

We conduct an experiment designed to answer these questions. In this experiment, various forecasts are compared at the one, six and twelve month horizons for 215 monthly US economic time series. The experiment simulates real-time implementation of these methods, that is, all forecasts (including all parameter estimates, all model selection rules, all pretests, all forecast combining weights, etc.) are based exclusively on data through the date of each forecast. The parameter estimates, model selection statistics, pretests, and forecast combining weights for all models are updated each month, and these updated statistics are used to make that month's simulated out of sample forecasts.

The forecasts studied here are produced by 49 forecasting methods. We refer to these as *methods* because many of these forecasts are based not on a single estimated model but on results from multiple models that are subject to model selection criteria or pretests. We shall refer to the underlying individual models used by these forecasting methods as *primitive models*, of which there are a total of 105. For example, one of our forecasting methods is an autoregression in levels with a constant term and lag order selection based on the Akaike Information Criterion (AIC), with lag length ranging from zero to twelve; in our terminology this forecasting method selects among thirteen primitive models. The primitive models fall into four classes: autoregressions (AR), exponential smoothing (EX), artificial neural networks (ANN), and logistic smooth transition autoregressions (LSTAR). As an additional benchmark, a "no change" forecast was also considered.

We also consider various procedures to combine information from these 49 forecasting methods. We refer to these as *forecast pooling procedures.* Bates and Granger (1969), Granger and Newbold (1977), and Granger and Ramanathan (1984) demonstrated that averaging forecasts from different models can improve forecast performance when all the models are approximations. The pooling procedures considered here differ by the amount of weight placed on the model with the currently best performance, including weighting all the forecasts equally, weighting the forecasts in inverse proportion to their current mean squared error (MSE), using median forecasts, and placing all weight on the forecasting method that currently has the lowest simulated real-time MSE; this final pooling procedure is simulated real-time model selection by predictive least squares (PLS).

The forecasting methods used in this study have been chosen in part to facilitate comparison with other large-scale "horse races" among time series models. Makridakis *et al.* (1982) studied performance of univariate methods in many series, some of which were economic time series, and concluded that exponential smoothing was often successful. Meese and Geweke (1984) compared various linear models using 150 macroeconomic time series and found that AR models with lag lengths selected by the AIC generally worked well. Interestingly, they also found that linear combination forecasts did not appreciably improve forecast quality. More recently, in a model comparison exercise conducted under the auspices of the Santa Fe Institute, Weigend and Gershenfeld (1994) compared linear models with a large number of nonlinear models; although they detected nonlinear dynamics in several non-economic time series, the nonlinear forecasting models fared relatively poorly for the economic time series they considered (exchange rates). Swanson and White (1995, 1997) compared multivariate ANN models to linear vector autoregressions, and found that the vector autoregressions generally had lower MSEs than the ANN models in simulated real time (their models are all multivariate however so their study does not compare directly to the exercise here).[1] Relative to this literature, the contributions of our study include the use of a large number of macroeconomic time series, the use of a large number of nonlinear models, the investigation of unit root pretest methods, and an extensive investigation of forecast pooling procedures.

The remainder of this paper is organized as follows. The experimental design and forecasting models are given in Section 2. The data are described briefly in Section 3 and in more detail in the Appendix. The results are presented and discussed in Section 4, and conclusions are summarized in Section 5.

2 Forecasting Methods and Experimental Design

2.1 General Considerations

Forecasting models. All the models investigated in this experiment are of the form:

$$y_{t+h} = f_i(Z_t;\theta_{ih}) + u_{it+h}, \tag{2.1}$$

where y_t is the series being forecast, h is the forecast horizon, i indexes the forecasting model ($i = 1, ..., 105$), θ_{ih} is a vector of unknown parameters, u_{it} is an error term, and Z_t is a vector of predictor variables. In general, $Z_t = (y_t, ..., y_{t-p}, \Delta y_t, ..., \Delta y_{t-p}, 1, t)$, where p is the maximal lag lengths. Typically, individual forecasting models use only a subset of the elements of Z_t.

All forecasts are made fully recursively, that is, forecasts of y_{t+h} are made using information in time periods 1, 2, ..., t. For the forecast of y_{t+h}, the parameter vector θ_{ih} is estimated using the data $(y_1, y_2, ..., y_t)$. In all models, the parameter vector is estimated by minimizing the sum of squared residuals of the h-step ahead forecast, that is, the estimate of θ_{ih} at time period t, $\hat{\theta}_{iht}$, solves, $\min_{\theta_{ih}} \sum_{s=t_0}^{t} [y_{t+h} - f_i(Z_t;\ \theta_{ih})]^2$, where t_0 denotes the first observation used for estimation for that model.

Note that in general each forecasting method, applied to a particular series, has different parameter values at different horizons (that is, the h-period ahead forecast is *not* computed by iterating forward for h periods the one-period ahead forecasting model). This has costs and benefits. If the errors are Gaussian and the one-period ahead forecasting model is correctly specified, then estimating it at the one-period horizon and iterating forward is more efficient than estimating the h-period ahead model directly. On the other hand, to the extent that the models are mis-specified, estimating the h-period ahead model directly permits the method to mitigate the effects of the mis-specification at the horizon at hand. From a practical perspective, forecasting the h-period ahead model directly requires more computer time for parameter estimation, but it simplifies the computation of multistep forecasts from the nonlinear models.

The h-step ahead forecast and the forecast error are:

$$y_{t+h|t,ih} = f_i(Z_t; \hat{\theta}_{iht}) \tag{2.2}$$

$$e_{t+h,ih} = y_{t+h} - y_{t+h|t,ih}. \tag{2.3}$$

Forecast trimming. For our main results, all forecasts were automatically trimmed so that a forecasted change that exceeded in absolute value any change previously observed for that series was replaced by a no-change forecast. This adjustment was adopted to simulate the involvement of a human forecaster, who would be present in actual applications but is absent from our computerized experiment. Because the forecasts in this experiment are made automatically, some models could (and do) make extreme forecasts. Possible sources of these extreme forecasts include parameter breaks, errors arising from incorrect inclusion of deterministic trends, and difficulties arising from multiple local optima for the nonlinear models. In true real time, such "crazy" forecasts would be noticed and adjusted by human intervention. Accordingly, our forecast trimming algorithm can be thought of as a rule of thumb that a human forecaster might use in real time to detect and address such problems. Although we focus primarily on the trimmed forecasts, some results for the untrimmed forecasts are also presented for the purpose of comparison.

Startup and forecast periods. For each series, there are three separate periods: a startup period over which initial estimates of the model are produced; an intermediate period over which forecasts are produced by the 105 primitive models and 49 forecasting methods, but not by the pooling procedures; and the simulated real-time forecast period over which recursive forecasts are produced by all models, methods, and pooling procedures. Let T_0 be the date of the first observation used in this study. Then the startup estimation period is $T_0 + 14$ to T_1, where $T_1 = T_0 + 134$ and the first 13 observations are used for initial conditions as needed. Thus 120 observations are used for the startup estimation period. The intermediate period is T_1 to $T_2 - 1$, where $T_2 = T_1 + 24$. The forecast period is T_2 to T_3, where T_3 is the date of the final observation (1996:12) minus the forecast horizon h.

All forecast performance results reported in the tables are from the simulated real-time forecast period, T_2 to T_3 (inclusive). For most series, the initial observation date is 1959:1, in which case $T_0 = 1959{:}1$, $T_1 = 1971{:}3$, $T_2 = 1973{:}3$, and $T_3 = 1996{:}12 - h$.

2.2 Forecasting Models and Methods

The forecasting methods are listed in Table 1.

Autoregressive (AR) models. Results are reported for 18 different autoregressive forecasting methods. These differ in their treatment of lag lengths (3 variants); in whether a constant, or a constant and a time trend, were included (2 variants); and in their treatment of persistence in the form of large autoregressive roots (3 variants).

Three alternative treatments of lag lengths were considered: a fixed lag length of 4; lag length determination by the BIC ($0 \leq p \leq 12$); and lag length determination by the AIC ($0 \leq p \leq 12$).

The possibility of persistence in the time series was handled by considering three alternatives. In the first, the autoregression was specified in levels, that is, y_{t+h} was forecast using y_t, ..., y_{t-p+1} with no restrictions on the coefficients. In the second, a unit root was imposed, so that the dependent variable was $y_{t+h} - y_t$ and the predictors were Δy_t, ..., Δy_{t-p+1}.

The third approach was to use a recursive unit root pretest to select between the levels or first differences specification. Theoretical and Monte Carlo evidence suggests that forecasting performance can be improved by using a unit root pretest rather than always using levels or always using differences, see for example Campbell and Perron (1991), Stock (1996), and Diebold and Kilian (1997). The unit root pretesting approach is widely used in practice, and many unit root tests statistics are available for this purpose. In a Monte Carlo study of unit root pretest autoregressive forecasts at moderate to long horizons, Stock (1996) compared several different pretest methods at various significance levels, and found that the best forecast performance across different values of the largest autoregressive root was obtained using the Elliott–Rothenberg–Stock (1996) DF-GLS test with a small significance level. We therefore computed the unit root pretest using the DF-GLS$^{\mu}$ statistic for the

Table 1. Summary of forecasting methods

Code	Description
A. Linear Methods	
AR(p, u, d)	Autoregressive methods p = number of lags = 4, A (AIC, $0 \le p \le 12$), or B (BIC, $0 \le p \le 12$) u = method of handling possible unit root = L (levels), D (differences), or P (unit root pretest: DF-GLS$^\mu$ if $d = C$, DF-GLS$^\tau$ if $d = T$) d = deterministic components included = C (constant only) or T (constant and linear time trend)
EX1	Single exponential smoothing
EX2	Double exponential smoothing
EXP	DF-GLS$^\mu$ pretest between EX1 and EX2
B. Nonlinear Methods	
NN(p, u, n_1, n_2)	Artificial Neural Net methods p = number of lags = 3, A (AIC, $p = 1, 3$), or B (BIC, $p = 1, 3$) (same number of lags in each hidden unit) u = L (levels), D (differences), or P (DF-GLS$^\mu$ unit root pretest) n_1 = number of hidden units in first hidden layer = 2, A (AIC, $1 \le n_1 \le 3$), or B (BIC, $1 \le n_1 \le 3$) n_2 = number of hidden units in second hidden layer = 0 (only one hidden layer), 1 or 2
LS(p, u, ξ)	LSTAR methods p = number of lags = 3, A (AIC, $p = 1, 3, 6$), or B (BIC, $p = 1, 3, 6$) u = L (levels), D (differences), or P (DF-GLS$^\mu$ unit root pretest) ξ = switching variable = L ($\xi_t = y_t$), D ($\xi_t = \Delta y_t$), M (either L or D depending on unit root pretest), D6 ($\xi_t = y_t - y_{t-6}$), A (AIC over $\xi_t = \{y_t, y_{t-2}, y_{t-5}, y_t - y_{t-6}$, and $y_t - y_{t-12}\}$ if levels specification, or $\xi_t = \{\Delta y_t, \Delta y_{t-2}, \Delta y_{t-5}, y_t - y_{t-6}$, and $y_t - y_{t-12}\}$ if differences specification), or B (BIC, same set as AIC)
C. No Change	
NOCHANGE	$y_{t+h\|t} = y_t$
D. Pooling Procedures	
C(ω, TW, Group)	Linear combination forecasts ω = exponent in (2.9) = $\{0, 1, 5\}$ (0 is equal weighting) TW = number of observations in rolling window to compute MSEs = REC (recursive—all past forecasts used), 120, 60 Group = A, B, A–C
Med(Group)	Median combination forecasts Group = A, B, A–C
PLS(TW, Group)	Predictive least squares combination forecasts TW = REC, 120, 60 Group = A, B, A–C, or PM (all Primitive Methods)
E. Pooled Over All Groups	
PLS(TW, A–D)	Predictive least squares combination forecasts over groups A–D TW = REC, 120, 60

selection between models that included a constant term only. For selection between models that included a linear time trend under the levels alternative, the DF-GLS$^{\tau}$ statistic was used.[2]

In all, a total of 52 primitive autoregressive models were estimated (2 specifications of deterministic terms, 13 lag choices, in either levels or differences). The 18 forecasting methods based on these 52 primitive models include recursive model selection using information criteria and/or recursive unit root pretests, as detailed in Table 1.

For some of the results, it is useful to normalize the performance of the models by comparison to a benchmark method. Throughout, we use a simple autoregression as the benchmark method, specifically, an AR(4) (fixed lag length) in levels with a constant term.

Exponential Smoothing (EX). Two primitive exponential smoothing models are considered. Single or simple exponential smoothing forecasts are given by:

$$y_{t+h|t} = \alpha y_{t+h-1|t-1} + (1-\alpha)y_t. \tag{2.4}$$

Double exponential smoothing forecasts are given by:

$$f_t = \alpha_1(f_{t-1} + g_{t-1}) + (1-\alpha_1)y_t, \tag{2.5a}$$
$$g_t = \alpha_2 g_{t-1} + (1-\alpha_2)(f_t - f_{t-1}), \tag{2.5b}$$

where the forecast is $y_{t+h|t} = f_t + hg_t$. The parameters α in (2.4) and (α_1, α_2) in (2.5) are estimated by recursive nonlinear least squares for each horizon (cf. Tiao and Xu, 1993).

Single exponential smoothing is conventionally intended for use with non-trending series and double exponential smoothing is conventionally intended for trending series. We therefore considered a unit root pretest version of these two, in which the single exponential smoothing forecast was used if the recursive DF-GLS$^{\mu}$ pretest (described above) rejected the null of a unit root, otherwise the double exponential smoothing forecast was used. The three forecasting methods based on these two primitive models therefore include the I(0) specification (2.4), the I(1) specification (2.5), and the specification selected by a recursive unit root pretest.

Artificial neural networks (ANN).[3] Neural network models with one and two hidden layers were considered. The single layer feedforward neural network models have the form:

$$\nu_{t+h} = \beta_0'\zeta_t + \sum_{i=1}^{n_1}\gamma_{1i}g(\beta_{1i}'\zeta_t) + u_{t+h}, \tag{2.6}$$

where $g(z)$ is the logistic function, $g(z) = 1/(1 + e^z)$. When y_t is modeled in levels, $\nu_{t+h} = y_{t+h}$ and $\zeta_t = (1, y_t, y_{t-1}, ..., y_{t-p+1})$. When y_t is modeled in first differences, $\nu_{t+h} = y_{t+h} - y_t$ and $\zeta_t = (1, \Delta y_t, \Delta y_{t-1}, ..., \Delta y_{t-p+1})$. The neural network models with two hidden layers have the form:

$$\nu_{t+h} = \beta_0'\zeta_t + \sum_{j=1}^{n_2} \gamma_{2j} g\left[\sum_{i=1}^{n_1} \beta_{2ji} g(\beta_{1i}'\zeta_t)\right] + u_{t+h}. \tag{2.7}$$

Note that all the neural nets are forced to include a linear component. We will refer to (2.6) as having n_1 hidden units, and to (2.7) as having n_1 and n_2 hidden units, plus a linear component. Alternatively, (2.6) could be thought of as having $n_1 + 1$ hidden units, with one of the hidden units forced to be linear.

The variants of (2.6) and (2.7) that are considered include different lag lengths p; the number of hidden units; and specification in levels and differences. The choices for single hidden layer ANNs are $n_1 = \{1, 2, 3\}$, $p = \{1, 3\}$, and levels/differences specification, for a total of 12 primitive models. (The restricted lag length choice of $p = \{1, 3\}$ was used to reduce computational requirements.) The choices for ANNs with two hidden layers are $n_1 = 2$, $n_2 = \{1, 2\}$, $p = \{1, 3\}$, and levels/differences specification, comprising 8 primitive models. The 15 forecasting methods based on these 20 primitive models include recursive model selection using information criteria and/or recursive unit root pretests, as detailed in Table 1.

In all ANN models, coefficients were estimated by recursive nonlinear least squares. For these models, multiple local minima are an important concern, so the objective function was minimized by an algorithm developed for this application. The algorithm uses a combination of random search methods and local Gauss–Newton optimization. The algorithm and its performance are discussed in the Appendix.

Logistic smooth transition autoregressions (LSTAR).[4] The LSTAR models have the form:

$$\nu_{t+h} = \alpha'\zeta_t + d_t\beta'\zeta_t + u_{t+h}, \tag{2.8}$$

where ν_{t+h} and ζ_t are defined following (2.7) and $d_t = 1/(1 + \exp[\gamma_0 + \gamma_1\xi_t])$, where ξ_t is a function of current and past y_t and is the variable used to define the smooth threshold.

The variants of the LSTAR models differ by the variable used to define the threshold; the specification in levels or differences or unit root pretest; and the lag length p. For models specified in levels, the following five alternatives were used for the threshold variable: $\xi_t = y_t$; $\xi_t = y_{t-2}$; $\xi_t = y_{t-5}$; $\xi_t = y_t - y_{t-6}$; and $\xi_t = y_t - y_{t-12}$. For models specified in first differences, the following five alternatives were used for the threshold variable: $\xi_t = \Delta y_t$; $\xi_t = \Delta y_{t-2}$; $\xi_t = \Delta y_{t-5}$; $\xi_t = y_t - y_{t-6}$; and $\xi_t = y_t - y_{t-12}$. In each case, lag lengths of $p = \{1, 3, 6\}$ were considered, for a total of 30 primitive models (15 in levels, 15 in differences). The 12 forecasting methods based on these 30 primitive models include recursive model selection using information criteria and/or recursive unit root pretests, as detailed in Table 1.

The parameters α, β and γ were estimated using the optimizer described in the Appendix.

No change forecast. The no change forecast is $y_{t+h|t} = y_t$.

2.3 Forecast Pooling Procedures

Linear combination forecasts. Pooled forecasts were computed as weighted averages of the forecasts produced by the 49 forecasting methods. These combination forecasts have the form:

$$\sum_{i=1}^{M} \kappa_{iht} y_{t+h|t,ih}, \text{ where } \kappa_{iht} = (1/\mathrm{MSE}_{iht})^{\omega} \Big/ \sum_{j=1}^{M} (1/\mathrm{MSE}_{jht})^{\omega}, \tag{2.9}$$

where i runs over the M methods and $\{\kappa_{iht}\}$ are the weights. The weighting schemes differ in the choice of ω, how the MSE is computed, and the sets of methods that are combined. The simplest scheme places equal weight on all the forecasts, which corresponds to setting $\omega = 0$ (in which case the MSE does not enter). As ω is increased, an increasing amount of emphasis is placed on those models that have been performing relatively well.

As shown by Bates and Granger (1969), if forecast error variances are finite then the optimal linear weighting scheme under quadratic loss involves the entire covariance matrix of forecast errors (see Granger and Newbold, 1977). With the large number of forecasts at hand, this scheme is impractical and would be unreliable because of the large number of covariances that would need to be estimated. Instead, we follow Bates and Granger's (1969) suggestion and drop the covariance term from our weighting expressions. Accordingly, the weights on the constituent forecasts are inversely proportional to their out-of-sample MSE, raised to the power ω. The weights with $\omega = 1$ correspond to Bates and Granger's (1969) suggestion. We also explore the possibility that more weight should be placed on the best performing models than would be indicated by inverse MSE weights, and this is achieved by considering $\omega > 1$. If $\omega \neq 0$ the weights $\{\kappa_{iht}\}$ differ from series to series.

Bates and Granger (1969) also stress that the relative performance of different models can change over time. This suggests computing MSEs over rolling windows. The MSEs were therefore computed in three ways: over 60 and 120 period rolling windows (more precisely, over the past $\min(t - T_1 + 1, 60)$ or $\min(t - T_1 + 1, 120)$ periods, respectively), and recursively (over the past $t - T_1 + 1$ periods).

The averages were computed over three different sets of forecasts: the linear methods (AR and EX); the nonlinear methods (ANN and LSTAR); and all the methods discussed above (linear, nonlinear, and no change). Note that the equal-weighted combinations do not depend on the rolling window; these are denoted as C(0, REC, Group) in Table 1 for the different groups.

Median combination forecasts. If forecast errors are non-Gaussian then linear combinations are no longer optimal. We therefore consider combination forecasts constructed as the median from a group of methods. In practice this guards against placing weight on forecasts that are badly wrong for method-specific reasons such as parameter breaks or parameter estimates achieving local but not global optima. The medians were computed over three different sets of forecasts: linear (AR and EX); nonlinear (ANN and LSTAR); and all the methods discussed above (linear, nonlinear, and no change). This median forecasts can be thought of as a consensus forecasts obtained by a vote of a panel of experts, where each expert (forecasting method) gets one vote: the consensus forecast is achieved when half the experts are on each side of the forecast.

Predictive least squares (PLS) forecasts. An alternative approach to pooling forecast information is to select the model that has produced the best forecasts (as measured by the lowest out-of-sample MSE) up to the forecast date. This constitutes selection across these models by predictive least squares. The PLS forecasts differ by the period over which the PLS criterion is computed and the sets of models for which it is computed.

The periods for which the PLS forecast were computed are the same as for the combination forecasts, specifically, over the past $\min(t - T_1 + 1, 60)$ periods; over the past $\min(t - T_1 + 1, 120)$ periods; and over the past $t - T_1 + 1$ periods.

The PLS forecasts were computed for five sets of models: all 49 models listed in Table 1 under the categories AR, EX, ANN, LSTAR, NOCHANGE; all linear models listed in Table 1 (AR and EX); all nonlinear models listed in Table 1 (ANN and LSTAR); all 105 primitive models; and all 49 methods plus the 36 linear combination, median, and PLS pooling forecasts. The purpose of examining this final group is to see whether the potential optimality of pooled forecasts could have been ascertained empirically in (simulated) real time.

3 Data

The data are monthly US macroeconomic time series. The 215 series fall into the following general categories: production (including personal income), employment and unemployment, wages (hours and earnings), construction (including housing starts), trade (wholesale and retail), inventories, orders, money and credit, stock returns, stock market dividends and volume, interest rates, exchange rates, producer price inflation, consumer price inflation, consumption, and miscellaneous (e.g. consumer confidence). In general, seasonally adjusted versions of these data were used for those series that, when unadjusted, have seasonal patterns. Non-seasonally adjusted data were generally used for inflation, interest rates, stock market variables, and exchange rates.

Some of these series were subjected to preliminary transformations. The series in dollars, real quantities and price deflators were transformed to their logarithms. Most other series (interest rates, the unemployment rate, exchange rates, etc.) were left in their native units.

In general, the first date used is either the first date for which the series is available or 1959:1, whichever is later. The exception to this rule is exchange rates; because exchange rates are essentially flat in the fixed exchange rate period, following Meese and Rogoff (1983) the first observation used for exchange rates is 1973:1.

A complete list of the series, their sources, the initial observation date used, whether the series were seasonally adjusted at the source, and the transformation used are given in the Appendix.

4 Results

4.1 Description of Tables

Table 2 contains statistics summarizing the performance of each forecasting method, relative to the benchmark method (an AR(4) specified with a constant term in levels). For each series, forecast method and horizon, the mean square of the $T_3 - T_2 + 1$ simulated out-of-sample forecast errors was computed; for forecasting method i, denote this $\mathrm{MSE}_{ij,h}$, $j = 1, \ldots, 215$ and $h = 1$, 6, 12. The relative mean square forecast error of the ith forecasting method is $\mathrm{MSE}_{ij,h}/\mathrm{MSE}_{1j,h}$, where $i = 1$ corresponds to the benchmark AR(4) forecast. Table 2 contains the averages and empirical quantiles of the distribution (across series) of this relative MSE, for each of 49 AR, EX, no change, ANN, and LSTAR methods listed in Table 1, and for various pooled forecasts. If, for example, the median of this distribution exceeds one for a candidate forecasting model and horizon, then for at least half the series the benchmark method had a lower simulated out-of-sample MSE at that horizon than the candidate forecasting model.

Table 3 compares forecasting methods by presenting the fraction of series for which each forecasting method is either best or among the top five. The forecasts compared in this table consist of the 49 methods in groups A–C in Table 1 and the pooling procedures that use the full recursive sample (“REC”). For example, at horizon $h = 12$, for 5 percent of the series, the AR(4, L, C) method (which is the benchmark method used in Table 2) had the lowest simulated out-of-sample MSE of all the forecasting methods; for 20 percent of the series, its MSE was among the lowest five.

Table 2. Mean and percentiles of relative MSEs of various forecasting methods
relative MSE = MSE of method i/MSE of benchmark model
benchmark model = AR(4, L, C) (AR(4) in levels with a constant term)
For each forecast, the first row corresponds to one-step ahead forecasts; the second row, to 6-step ahead forecasts; the third row, to 12-step ahead forecasts.

Method	Mean	2%	10%	25%	50%	75%	90%	98%
AR(4, L, C)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR(4, L, T)	1.02	0.96	0.99	1.00	1.01	1.03	1.04	1.10
	1.10	0.78	0.88	0.99	1.08	1.17	1.27	1.56
	1.26	0.44	0.77	1.02	1.19	1.38	1.76	2.55
AR(4, D, C)	1.00	0.90	0.95	0.98	1.00	1.02	1.04	1.08
	0.98	0.59	0.77	0.90	0.97	1.05	1.17	1.36
	0.99	0.35	0.64	0.81	0.94	1.15	1.36	1.76
AR(4, D, T)	1.01	0.95	0.97	0.99	1.01	1.02	1.04	1.09
	1.06	0.74	0.89	0.99	1.03	1.11	1.25	1.46
	1.15	0.52	0.83	0.97	1.06	1.18	1.42	1.89
AR(4, P, C)	1.00	0.90	0.95	0.98	1.00	1.01	1.03	1.07
	0.98	0.59	0.77	0.91	0.97	1.05	1.15	1.34
	0.98	0.35	0.64	0.81	0.94	1.11	1.33	1.76
AR(4, P, T)	1.00	0.90	0.95	0.98	1.00	1.02	1.04	1.07
	0.98	0.59	0.77	0.91	0.97	1.06	1.16	1.34
	0.99	0.35	0.64	0.81	0.95	1.14	1.36	1.76
AR(A, L, C)	1.02	0.83	0.95	1.00	1.02	1.04	1.07	1.14
	1.00	0.61	0.86	0.99	1.01	1.06	1.13	1.24
	0.98	0.63	0.87	0.98	1.00	1.02	1.08	1.18
AR(A, L, T)	1.03	0.85	0.96	1.00	1.04	1.06	1.10	1.16
	1.12	0.66	0.81	0.96	1.10	1.25	1.37	1.82
	1.29	0.45	0.75	0.95	1.20	1.41	1.82	3.13
AR(A, D, C)	1.01	0.77	0.94	0.98	1.02	1.05	1.09	1.15
	0.97	0.43	0.72	0.88	0.99	1.09	1.18	1.42
	0.98	0.33	0.58	0.83	0.95	1.15	1.36	1.74
AR(A, D, T)	1.03	0.84	0.96	1.00	1.03	1.06	1.10	1.17
	1.07	0.60	0.80	0.96	1.06	1.16	1.31	1.53
	1.16	0.46	0.72	0.96	1.07	1.23	1.49	1.97
AR(A, P, C)	1.01	0.77	0.94	0.98	1.02	1.05	1.08	1.15
	0.97	0.43	0.72	0.89	0.99	1.09	1.15	1.40
	0.97	0.33	0.58	0.83	0.95	1.13	1.35	1.74
AR(A, P, T)	1.02	0.77	0.94	0.98	1.02	1.05	1.09	1.15
	0.98	0.43	0.72	0.89	0.99	1.09	1.18	1.41
	0.98	0.33	0.58	0.83	0.95	1.15	1.35	1.74
AR(B, L, C)	1.01	0.91	0.97	0.99	1.01	1.02	1.04	1.12
	0.99	0.68	0.87	0.98	1.01	1.03	1.07	1.20
	0.99	0.71	0.93	0.99	1.00	1.02	1.05	1.11
AR(B, L, T)	1.02	0.93	0.97	1.00	1.02	1.05	1.08	1.14
	1.11	0.67	0.83	0.96	1.09	1.23	1.36	1.67
	1.27	0.48	0.75	0.98	1.20	1.42	1.74	2.99
AR(B, D, C)	1.00	0.83	0.94	0.98	1.01	1.03	1.07	1.13
	0.97	0.43	0.73	0.90	0.98	1.07	1.17	1.42
	0.99	0.33	0.58	0.83	0.94	1.15	1.37	1.66

Table 2. (*cont.*)

Method	Mean	2%	10%	25%	50%	75%	90%	98%
AR(B, D, T)	1.02	0.89	0.96	1.00	1.02	1.04	1.08	1.14
	1.06	0.64	0.79	0.99	1.05	1.12	1.26	1.57
	1.16	0.62	0.76	0.99	1.07	1.19	1.43	1.93
AR(B, P, C)	1.00	0.83	0.94	0.98	1.00	1.03	1.06	1.12
	0.96	0.43	0.73	0.90	0.98	1.07	1.16	1.36
	0.98	0.33	0.58	0.83	0.94	1.14	1.31	1.66
AR(B, P, T)	1.00	0.83	0.94	0.98	1.01	1.03	1.07	1.13
	0.97	0.43	0.73	0.90	0.99	1.08	1.16	1.36
	0.99	0.33	0.58	0.83	0.95	1.15	1.36	1.66
EX1	1.73	0.90	0.98	1.01	1.09	1.47	2.82	8.42
	2.12	0.81	0.90	0.97	1.20	1.78	4.61	11.47
	1.83	0.69	0.81	0.95	1.17	1.87	3.02	9.18
EX2	1.06	0.82	0.94	1.00	1.04	1.10	1.18	1.37
	1.16	0.37	0.76	0.94	1.11	1.32	1.64	2.30
	1.26	0.30	0.63	0.91	1.15	1.47	2.14	3.02
EX3	1.06	0.82	0.94	1.00	1.04	1.09	1.18	1.37
	1.15	0.37	0.76	0.94	1.11	1.30	1.55	2.30
	1.23	0.30	0.63	0.91	1.14	1.40	1.92	2.79
NN(3, L, 2, 0)	1.03	0.80	0.91	0.96	1.01	1.08	1.16	1.53
	1.29	0.57	0.92	1.00	1.12	1.32	1.84	3.19
	1.42	0.51	0.81	1.02	1.20	1.51	2.16	4.19
NN(3, D, 2, 0)	0.99	0.83	0.89	0.95	1.00	1.03	1.07	1.24
	0.99	0.63	0.78	0.89	0.98	1.08	1.20	1.63
	1.02	0.35	0.64	0.80	0.95	1.17	1.43	2.35
NN(3, P, 2, 0)	0.99	0.84	0.90	0.95	0.99	1.02	1.07	1.24
	0.99	0.63	0.78	0.89	0.98	1.06	1.18	1.63
	1.01	0.35	0.64	0.81	0.95	1.15	1.35	2.35
NN(3, L, 2, 1)	0.99	0.80	0.89	0.95	1.00	1.03	1.09	1.18
	1.07	0.45	0.79	0.95	1.05	1.16	1.34	1.70
	1.12	0.30	0.60	0.91	1.09	1.24	1.57	2.54
NN(3, D, 2, 1)	0.99	0.83	0.91	0.95	0.99	1.03	1.07	1.21
	1.00	0.63	0.77	0.89	0.98	1.08	1.22	1.64
	1.02	0.35	0.64	0.82	0.95	1.18	1.45	2.31
NN(3, P, 2, 1)	0.99	0.83	0.91	0.95	0.99	1.03	1.07	1.16
	0.99	0.63	0.77	0.89	0.98	1.07	1.19	1.64
	1.01	0.35	0.64	0.82	0.95	1.16	1.38	2.31
NN(3, L, 2, 2)	0.99	0.77	0.88	0.95	1.00	1.04	1.10	1.22
	1.09	0.46	0.80	0.96	1.07	1.19	1.38	1.84
	1.22	0.33	0.73	0.97	1.15	1.39	1.87	2.56
NN(3, D, 2, 2)	1.01	0.85	0.91	0.95	1.00	1.04	1.08	1.24
	1.00	0.62	0.78	0.88	0.99	1.07	1.23	1.61
	1.02	0.35	0.63	0.81	0.97	1.18	1.44	2.31
NN(3, P, 2, 2)	1.01	0.85	0.90	0.95	1.00	1.04	1.08	1.23
	0.99	0.62	0.78	0.88	0.99	1.07	1.19	1.61
	1.01	0.35	0.63	0.81	0.97	1.16	1.36	2.31
NN(A, L, A, 0)	1.03	0.81	0.92	0.96	1.02	1.08	1.16	1.26
	1.32	0.47	0.94	1.02	1.16	1.45	1.91	3.19
	1.50	0.59	0.88	1.05	1.30	1.64	2.16	3.59

Table 2. (*cont.*)

Method	Mean	2%	10%	25%	50%	75%	90%	98%
NN(A, D, A, 0)	1.02	0.84	0.91	0.95	1.00	1.05	1.10	1.31
	1.00	0.62	0.77	0.89	0.98	1.10	1.24	1.64
	1.03	0.37	0.63	0.81	0.96	1.17	1.44	2.32
NN(A, P, A, 0)	1.01	0.82	0.91	0.95	1.00	1.04	1.10	1.24
	1.00	0.62	0.77	0.89	0.98	1.08	1.20	1.64
	1.02	0.37	0.63	0.83	0.96	1.16	1.37	2.32
NN(B, L, B, 0)	1.03	0.83	0.94	0.98	1.02	1.07	1.16	1.24
	1.31	0.56	0.95	1.04	1.15	1.45	1.83	2.92
	1.49	0.59	0.95	1.06	1.31	1.65	2.15	3.29
NN(B, D, B, 0)	1.02	0.87	0.93	0.96	1.01	1.05	1.10	1.19
	1.01	0.64	0.80	0.92	0.98	1.10	1.24	1.56
	1.03	0.37	0.65	0.83	0.95	1.18	1.41	2.22
NN(B, P, B, 0)	1.02	0.87	0.93	0.96	1.01	1.05	1.10	1.18
	1.00	0.64	0.80	0.92	0.98	1.08	1.18	1.56
	1.02	0.37	0.65	0.83	0.96	1.16	1.38	2.22
LS(3, L, L)	1.07	0.91	0.98	1.01	1.05	1.10	1.17	1.31
	1.24	0.80	1.00	1.06	1.15	1.34	1.72	2.00
	1.34	0.56	0.92	1.07	1.19	1.45	1.95	2.89
LS(3, D, D)	1.06	0.90	0.95	1.00	1.04	1.09	1.16	1.38
	1.04	0.69	0.82	0.93	1.02	1.11	1.26	1.60
	1.05	0.40	0.67	0.84	0.98	1.20	1.44	2.21
LS(3, P, P)	1.05	0.90	0.96	1.00	1.04	1.08	1.15	1.33
	1.03	0.69	0.82	0.93	1.02	1.11	1.22	1.60
	1.04	0.40	0.67	0.85	0.98	1.18	1.41	2.21
LS(3, L, D6)	1.04	0.93	0.97	1.00	1.03	1.07	1.12	1.25
	1.09	0.75	0.92	1.00	1.06	1.14	1.28	1.52
	1.10	0.74	0.92	1.01	1.06	1.16	1.26	1.61
LS(3, D, D6)	1.03	0.85	0.95	0.99	1.02	1.06	1.11	1.27
	1.01	0.52	0.72	0.91	1.00	1.12	1.24	1.46
	1.04	0.34	0.60	0.83	0.96	1.20	1.45	1.99
LS(3, P, D6)	1.03	0.85	0.95	0.99	1.02	1.06	1.11	1.24
	1.00	0.52	0.72	0.91	1.00	1.11	1.22	1.42
	1.03	0.34	0.60	0.84	0.96	1.17	1.45	1.99
LS(A, L, A)	1.13	0.92	0.98	1.04	1.08	1.18	1.33	1.68
	1.42	0.77	0.96	1.11	1.29	1.57	2.10	2.99
	1.47	0.73	0.92	1.12	1.34	1.70	2.13	3.27
LS(A, D, A)	1.11	0.83	0.97	1.01	1.08	1.16	1.29	1.72
	1.07	0.47	0.80	0.95	1.06	1.18	1.35	1.61
	1.06	0.31	0.60	0.82	1.00	1.25	1.56	2.35
LS(A, P, A)	1.11	0.83	0.97	1.01	1.07	1.16	1.29	1.61
	1.07	0.47	0.80	0.96	1.06	1.17	1.33	1.58
	1.05	0.31	0.60	0.83	1.00	1.24	1.57	2.36
LS(B, L, B)	1.11	0.89	0.97	1.02	1.07	1.15	1.27	1.70
	1.41	0.74	0.96	1.11	1.26	1.59	1.97	3.02
	1.46	0.72	0.90	1.11	1.32	1.71	2.09	3.19
LS(B, D, B)	1.07	0.81	0.96	1.00	1.05	1.11	1.19	1.46
	1.04	0.47	0.77	0.92	1.03	1.15	1.31	1.61
	1.06	0.31	0.60	0.84	0.99	1.20	1.52	2.31

Table 2. (*cont.*)

Method	Mean	2%	10%	25%	50%	75%	90%	98%
LS(B, P, B)	1.06	0.81	0.96	1.00	1.04	1.11	1.19	1.44
	1.03	0.47	0.77	0.92	1.03	1.15	1.29	1.61
	1.05	0.31	0.60	0.84	0.99	1.19	1.52	2.31
NOCHANGE	1.76	0.89	0.99	1.04	1.12	1.49	2.82	8.42
	2.14	0.81	0.90	1.00	1.22	1.78	4.61	11.47
	1.83	0.69	0.82	0.97	1.21	1.77	2.93	9.18
C(0, REC, A–C)	0.95	0.79	0.87	0.93	0.96	0.98	1.00	1.01
	0.89	0.42	0.69	0.85	0.92	0.97	1.03	1.12
	0.87	0.27	0.56	0.78	0.89	1.00	1.08	1.32
C(0, REC, A)	0.98	0.81	0.91	0.96	0.99	1.01	1.02	1.06
	0.92	0.46	0.68	0.88	0.96	1.01	1.08	1.15
	0.90	0.37	0.57	0.81	0.93	1.05	1.16	1.33
C(0, REC, B)	0.94	0.79	0.86	0.92	0.95	0.97	1.00	1.05
	0.90	0.45	0.72	0.85	0.92	0.99	1.04	1.13
	0.88	0.28	0.59	0.79	0.89	1.00	1.08	1.35
C(1, REC, A–C)	0.95	0.76	0.88	0.93	0.96	0.98	1.00	1.01
	0.89	0.43	0.69	0.85	0.92	0.98	1.03	1.11
	0.87	0.34	0.59	0.78	0.90	1.01	1.08	1.34
C(1, REC, A)	0.98	0.82	0.91	0.97	0.99	1.01	1.02	1.05
	0.93	0.48	0.71	0.90	0.96	1.01	1.08	1.15
	0.91	0.43	0.57	0.81	0.94	1.07	1.15	1.40
C(1, REC, B)	0.94	0.79	0.86	0.92	0.95	0.97	1.00	1.05
	0.89	0.44	0.70	0.85	0.91	0.99	1.04	1.14
	0.87	0.32	0.60	0.79	0.89	1.00	1.11	1.38
C(1, 60, A–C)	0.95	0.75	0.88	0.93	0.96	0.98	1.00	1.01
	0.89	0.42	0.68	0.85	0.93	0.98	1.03	1.12
	0.88	0.31	0.58	0.78	0.91	1.03	1.10	1.34
C(1, 120, A–C)	0.95	0.76	0.88	0.93	0.96	0.98	1.00	1.01
	0.89	0.43	0.69	0.85	0.92	0.98	1.03	1.11
	0.87	0.34	0.58	0.78	0.90	1.01	1.09	1.34
C(5, 60, A–C)	0.94	0.74	0.88	0.93	0.96	0.98	1.00	1.02
	0.91	0.41	0.64	0.85	0.94	1.01	1.07	1.15
	0.97	0.28	0.60	0.82	0.98	1.12	1.30	1.49
C(5, 120, A–C)	0.95	0.74	0.88	0.93	0.96	0.98	1.00	1.01
	0.89	0.43	0.68	0.85	0.93	0.99	1.04	1.11
	0.91	0.35	0.59	0.80	0.92	1.04	1.16	1.41
C(5, REC, A–C)	0.95	0.74	0.88	0.93	0.96	0.98	1.00	1.01
	0.90	0.42	0.69	0.85	0.93	0.99	1.04	1.11
	0.91	0.35	0.61	0.80	0.92	1.04	1.14	1.40
C(1, 60, A)	0.98	0.82	0.91	0.97	0.99	1.01	1.02	1.05
	0.93	0.47	0.71	0.90	0.96	1.01	1.08	1.15
	0.92	0.37	0.57	0.81	0.95	1.07	1.17	1.40
C(1, 120, A)	0.98	0.82	0.91	0.97	0.99	1.01	1.02	1.05
	0.93	0.48	0.71	0.90	0.96	1.01	1.08	1.15
	0.91	0.43	0.57	0.80	0.94	1.05	1.15	1.40
C(5, 60, A)	0.98	0.81	0.92	0.97	0.99	1.01	1.02	1.04
	0.94	0.48	0.73	0.90	0.98	1.03	1.08	1.18
	0.98	0.40	0.64	0.87	1.01	1.14	1.27	1.50

Table 2. (*cont.*)

Method	Mean	2%	10%	25%	50%	75%	90%	98%
C(5, 120, A)	0.98	0.82	0.91	0.97	1.00	1.01	1.02	1.04
	0.93	0.51	0.72	0.89	0.97	1.01	1.07	1.16
	0.93	0.43	0.60	0.82	0.96	1.06	1.14	1.52
C(5, REC, A)	0.98	0.82	0.91	0.97	1.00	1.01	1.02	1.04
	0.93	0.52	0.72	0.89	0.97	1.02	1.07	1.16
	0.93	0.42	0.61	0.82	0.96	1.06	1.16	1.52
C(1, 60, B)	0.94	0.79	0.86	0.92	0.95	0.97	1.00	1.05
	0.89	0.45	0.69	0.85	0.92	0.99	1.04	1.13
	0.88	0.31	0.59	0.79	0.91	1.03	1.10	1.38
C(1, 120, B)	0.94	0.79	0.86	0.92	0.95	0.97	1.00	1.05
	0.89	0.44	0.69	0.84	0.91	0.99	1.04	1.14
	0.87	0.32	0.58	0.79	0.88	1.00	1.10	1.38
C(5, 60, B)	0.94	0.78	0.86	0.92	0.95	0.97	1.00	1.06
	0.91	0.45	0.69	0.85	0.94	1.01	1.07	1.26
	0.97	0.28	0.60	0.80	0.96	1.13	1.31	1.56
C(5, 120, B)	0.94	0.77	0.86	0.92	0.95	0.97	1.00	1.04
	0.90	0.41	0.69	0.86	0.92	0.99	1.05	1.18
	0.91	0.32	0.59	0.79	0.91	1.06	1.16	1.50
C(5, REC, B)	0.94	0.77	0.86	0.92	0.95	0.97	1.00	1.04
	0.90	0.41	0.72	0.86	0.92	0.99	1.05	1.17
	0.92	0.32	0.60	0.78	0.91	1.06	1.17	1.52
PLS(REC, PM)	1.02	0.79	0.92	0.98	1.02	1.07	1.12	1.26
	1.05	0.48	0.74	0.95	1.05	1.16	1.35	1.55
	1.14	0.38	0.65	0.91	1.09	1.33	1.59	2.32
PLS(REC, A–C)	1.01	0.77	0.90	0.97	1.02	1.05	1.11	1.21
	1.03	0.47	0.75	0.93	1.03	1.15	1.27	1.50
	1.10	0.35	0.65	0.88	1.07	1.24	1.51	2.15
PLS(REC, A)	1.00	0.81	0.94	0.99	1.01	1.03	1.06	1.12
	1.01	0.49	0.75	0.94	1.03	1.10	1.21	1.36
	1.05	0.48	0.65	0.92	1.04	1.19	1.38	1.92
PLS(REC, B)	1.01	0.77	0.92	0.97	1.01	1.05	1.11	1.27
	1.04	0.50	0.81	0.94	1.04	1.14	1.31	1.61
	1.08	0.34	0.64	0.88	1.02	1.22	1.51	2.44
PLS(REC, A–D)	1.00	0.78	0.91	0.95	1.00	1.04	1.10	1.17
	1.05	0.42	0.77	0.93	1.06	1.16	1.32	1.62
	1.13	0.38	0.63	0.87	1.12	1.33	1.59	2.05
PLS(60, PM)	1.01	0.76	0.90	0.97	1.01	1.07	1.13	1.21
	1.11	0.40	0.76	0.96	1.11	1.28	1.43	1.74
	1.23	0.37	0.63	0.94	1.23	1.46	1.73	2.29
PLS(120, PM)	1.02	0.77	0.91	0.97	1.02	1.07	1.11	1.26
	1.06	0.45	0.75	0.94	1.07	1.19	1.32	1.55
	1.17	0.39	0.68	0.93	1.12	1.33	1.59	2.40
PLS(60, A–C)	1.01	0.73	0.91	0.96	1.01	1.06	1.11	1.30
	1.08	0.39	0.71	0.95	1.10	1.23	1.40	1.56
	1.18	0.33	0.65	0.88	1.16	1.45	1.69	2.19
PLS(120, A–C)	1.01	0.77	0.90	0.96	1.01	1.05	1.10	1.21
	1.04	0.42	0.75	0.93	1.05	1.18	1.32	1.47
	1.12	0.34	0.65	0.90	1.10	1.27	1.54	2.39

Table 2. (*cont.*)

Method	Mean	2%	10%	25%	50%	75%	90%	98%
PLS(60, A)	1.01	0.77	0.95	0.99	1.02	1.04	1.07	1.11
	1.06	0.47	0.79	0.96	1.07	1.19	1.28	1.48
	1.12	0.43	0.65	0.92	1.11	1.34	1.51	1.73
PLS(120, A)	1.00	0.79	0.94	0.99	1.01	1.03	1.06	1.13
	1.02	0.50	0.75	0.94	1.03	1.12	1.23	1.36
	1.05	0.44	0.64	0.89	1.05	1.19	1.41	1.91
PLS(60, B)	1.01	0.78	0.92	0.96	1.01	1.06	1.12	1.28
	1.09	0.41	0.78	0.98	1.08	1.22	1.38	1.68
	1.16	0.33	0.70	0.90	1.13	1.38	1.64	2.22
PLS(120, B)	1.01	0.78	0.92	0.96	1.01	1.05	1.10	1.23
	1.05	0.47	0.80	0.94	1.05	1.16	1.30	1.58
	1.09	0.34	0.64	0.87	1.06	1.22	1.60	2.16
PLS(60, A–D)	1.00	0.75	0.90	0.96	1.00	1.05	1.10	1.31
	1.09	0.40	0.78	0.95	1.08	1.23	1.42	1.74
	1.21	0.37	0.67	0.90	1.21	1.48	1.71	2.21
PLS(120, A–D)	1.00	0.78	0.91	0.95	1.01	1.05	1.10	1.17
	1.05	0.42	0.74	0.94	1.06	1.18	1.32	1.62
	1.12	0.38	0.63	0.88	1.12	1.35	1.60	2.02
MED(A–C)	0.96	0.81	0.90	0.94	0.97	0.99	1.00	1.02
	0.91	0.44	0.71	0.87	0.94	0.99	1.05	1.15
	0.90	0.32	0.58	0.79	0.91	1.05	1.16	1.49
MED(A)	0.99	0.82	0.93	0.97	1.00	1.01	1.03	1.07
	0.94	0.44	0.73	0.89	0.97	1.03	1.11	1.19
	0.94	0.37	0.59	0.83	0.93	1.11	1.22	1.54
MED(B)	0.95	0.80	0.89	0.93	0.96	0.99	1.01	1.05
	0.92	0.46	0.74	0.85	0.93	0.99	1.07	1.26
	0.92	0.31	0.58	0.79	0.90	1.05	1.23	1.57

A natural question to ask in this comparison is which forecasting method is best overall. The answer to this question depends, among other things, on the attitude towards risk of the forecaster, that is, on the forecaster's loss function. Table 4 therefore reports rankings of the different methods for different loss functions. The loss functions are all of the form:

$$\text{Loss}_{i,h} = (1/215)\sum\nolimits_{\text{series }\{y\}} (T_3 - T_2 + 1)^{-1} \sum_{i=T_2}^{T_3} |\,(y_{t+h} - \hat{y}_{t+h|t,ih})/\sigma_h\,|^{\rho}, \qquad (4.1)$$

where σ_h is the estimated standard deviation of $y_{t+h} - y_t$.[5]

4.2 Highlights of the Results

Unit root pretests. Among the 215 series, a 5 percent DF-GLS$^{\mu}$ unit root test rejects the null in 13.5 percent of the series, and a 5 percent DF-GLS$^{\tau}$ test rejects the null in 10.2 percent of the series, using the full sample and six lags.

Table 3. Summary of rankings of various methods

Entries are fraction of series for which the indicated method performs in the top N

Method	1 step ahead		6 steps ahead		12 steps ahead	
	$N = 1$	5	$N = 1$	5	$N = 1$	5
AR(4, L, C)	0.00	0.03	0.04	0.10	0.05	0.20
AR(4, L, T)	0.00	0.02	0.00	0.06	0.00	0.07
AR(4, D, C)	0.00	0.03	0.00	0.03	0.00	0.03
AR(4, D, T)	0.00	0.00	0.00	0.01	0.01	0.04
AR(4, P, C)	0.00	0.02	0.00	0.03	0.00	0.03
AR(4, P, T)	0.00	0.03	0.01	0.03	0.00	0.03
AR(A, L, C)	0.01	0.03	0.04	0.11	0.02	0.14
AR(A, L, T)	0.01	0.01	0.04	0.06	0.04	0.07
AR(A, D, C)	0.00	0.04	0.00	0.06	0.01	0.07
AR(A, D, T)	0.01	0.02	0.00	0.02	0.01	0.03
AR(A, P, C)	0.00	0.04	0.01	0.07	0.02	0.07
AR(A, P, T)	0.01	0.05	0.01	0.06	0.01	0.07
AR(B, L, C)	0.00	0.03	0.00	0.11	0.03	0.16
AR(B, L, T)	0.00	0.01	0.01	0.07	0.02	0.07
AR(B, D, C)	0.00	0.03	0.00	0.04	0.00	0.04
AR(B, D, T)	0.00	0.01	0.01	0.02	0.00	0.02
AR(B, P, C)	0.00	0.01	0.00	0.03	0.00	0.03
AR(B, P, T)	0.01	0.03	0.00	0.04	0.00	0.03
EX1	0.03	0.06	0.04	0.12	0.05	0.13
EX2	0.01	0.03	0.00	0.05	0.01	0.06
EX3	0.00	0.03	0.04	0.06	0.03	0.07
NN(3, L, 2, 0)	0.05	0.17	0.01	0.07	0.01	0.08
NN(3, D, 2, 0)	0.02	0.13	0.00	0.06	0.00	0.05
NN(3, P, 2, 0)	0.05	0.13	0.00	0.07	0.00	0.04
NN(3, L, 2, 1)	0.02	0.18	0.07	0.15	0.05	0.12
NN(3, D, 2, 1)	0.02	0.09	0.00	0.05	0.00	0.05
NN(3, P, 2, 1)	0.03	0.13	0.01	0.06	0.01	0.05
NN(3, L, 2, 2)	0.06	0.19	0.04	0.13	0.02	0.08
NN(3, D, 2, 2)	0.00	0.12	0.01	0.08	0.01	0.04
NN(3, P, 2, 2)	0.01	0.12	0.00	0.07	0.00	0.04
NN(A, L, A, 0)	0.06	0.15	0.02	0.06	0.01	0.06
NN(A, D, A, 0)	0.05	0.12	0.04	0.08	0.05	0.09
NN(A, P, A, 0)	0.01	0.10	0.00	0.07	0.00	0.08
NN(B, L, B, 0)	0.00	0.07	0.01	0.05	0.01	0.03
NN(B, D, B, 0)	0.01	0.06	0.00	0.04	0.00	0.02
NN(B, P, B, 0)	0.01	0.06	0.00	0.04	0.00	0.02
LS(3, L, L)	0.00	0.03	0.00	0.03	0.01	0.04
LS(3, D, D)	0.00	0.00	0.00	0.02	0.00	0.02
LS(3, P, P)	0.00	0.01	0.01	0.02	0.00	0.02
LS(3, L, D6)	0.00	0.02	0.02	0.06	0.02	0.08
LS(3, D, D6)	0.01	0.04	0.01	0.05	0.00	0.06
LS(3, P, D6)	0.00	0.04	0.02	0.07	0.01	0.07
LS(A, L, A)	0.00	0.02	0.00	0.02	0.01	0.04
LS(A, D, A)	0.01	0.01	0.00	0.02	0.02	0.06

Table 3. (*cont.*)

Method	1 step ahead		6 steps ahead		12 steps ahead	
	N = 1	5	N = 1	5	N = 1	5
LS(A, P, A)	0.00	0.02	0.00	0.01	0.00	0.06
LS(B, L, B)	0.00	0.02	0.01	0.04	0.01	0.04
LS(B, D, B)	0.00	0.01	0.00	0.04	0.01	0.05
LS(B, P, B)	0.00	0.02	0.00	0.03	0.00	0.06
NOCHANGE	0.00	0.02	0.03	0.09	0.04	0.11
C(0, REC, A–C)	0.04	0.34	0.06	0.33	0.04	0.34
C(0, REC, A)	0.01	0.05	0.03	0.16	0.03	0.11
C(0, REC, B)	0.10	0.51	0.09	0.32	0.13	0.30
C(1, REC, A–C)	0.02	0.27	0.01	0.33	0.00	0.23
C(1, REC, A)	0.00	0.03	0.00	0.07	0.01	0.07
C(1, REC, B)	0.13	0.53	0.05	0.34	0.03	0.33
MED(A–C)	0.01	0.10	0.04	0.17	0.04	0.11
MED(A)	0.01	0.02	0.00	0.05	0.00	0.03
MED(B)	0.08	0.27	0.07	0.23	0.07	0.22
PLS(REC, PM)	0.00	0.04	0.01	0.05	0.01	0.06
PLS(REC, A–C)	0.00	0.06	0.00	0.04	0.00	0.05
PLS(REC, A)	0.00	0.03	0.00	0.06	0.00	0.06
PLS(REC, B)	0.00	0.04	0.01	0.06	0.01	0.04
PLS(REC, A–D)	0.01	0.06	0.00	0.07	0.00	0.03

When the DF-GLS unit root pretest is employed recursively with a critical value that depends on the sample size (see note 2), it generally improves forecast performance at all horizons as measured by mean or median relative MSEs in Table 2. The improvement is largest for EX methods. Among AR methods, this improvement is most pronounced when the levels specification includes a time trend. The improvement for ANN and LSTAR methods is small at $h = 1$ but increases with the forecast horizon. Evidently the AR methods in levels with time trends and the ANN and LSTAR methods in levels can produce forecasts which are quite poor, especially at the longer horizons, and pretesting to identify situations in which a unit root can be imposed reduces the frequency of extreme errors.

AIC- and BIC-based model selection. The performance of automatic lag length selection methods depends on the family of models being used, and it does not seem possible to reach general conclusions. Among autoregressions, on average automatic order selection yields only marginal improvements over the benchmark imposition of 4 lags. Comparison of the relative MSEs in Table 2 for autoregressive methods using AIC and BIC lag length choice indicates that

Table 4. Rankings of various methods, combined over all series, for different loss functions: Trimmed forecasts
Cost function = $E|e_t|^\rho$, e_t = forecast error

Rank	1 step ahead			6 steps ahead			12 steps ahead		
	$\rho = 1.0$	2.0	3.0	$\rho = 1.0$	2.0	3.0	$\rho = 1.0$	2.0	3.0
1	C(1, REC, B)	C(1, REC, B)	C(1, REC, B)	C(0, REC, A–C)	C(0, REC, A–C)	C(0, REC, A–C)	C(0, REC, A–C)	C(0, REC, A–C)	C(0, REC, A–C)
2	C(0, REC, B)	C(0, REC, B)	C(0, REC, B)	C(1, REC, A–C)	C(1, REC, B)	C(1, REC, B)	C(1, REC, B)	C(1, REC, B)	C(0, REC, B)
3	C(0, REC, A–C)	C(1, REC, A–C)	C(1, REC, A–C)	C(1, REC, B)	C(1, REC, A–C)	C(0, REC, B)	C(1, REC, A–C)	C(0, REC, B)	C(1, REC, B)
4	C(1, REC, A–C)	C(0, REC, A–C)	C(0, REC, A–C)	C(0, REC, B)	C(0, REC, B)	C(1, REC, A–C)	C(0, REC, B)	C(1, REC, A–C)	C(1, REC, A–C)
5	MED(B)	MED(B)	MED(B)	MED(A–C)	MED(A–C)	MED(A–C)	MED(A–C)	MED(A–C)	MED(A–C)
6	MED(A–C)	MED(A–C)	MED(A–C)	C(0, REC, A)	MED(B)	MED(B)	C(0, REC, A)	C(0, REC, A)	C(0, REC, A)
7	C(0, REC, A)	C(0, REC, A)	NN(3, L, 2, 1)	MED(B)	C(0, REC, A)	C(0, REC, A)	MED(B)	MED(B)	MED(B)
8	C(1, REC, A)	C(1, REC, A)	NN(3, L, 2, 2)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)
9	MED(A)	NN(3, L, 2, 1)	C(0, REC, A)	MED(A)	MED(A)	MED(A)	MED(A)	MED(A)	MED(A)
10	PLS(REC, A–D)	NN(3, P, 2, 0)	C(1, REC, A)	AR(B, P, C)	AR(B, P, C)	AR(B, P, C)	AR(A, P, C)	AR(A, P, C)	AR(A, D, C)
11	NN(3, P, 2, 0)	MED(A)	NN(3, P, 2, 0)	AR(A, P, C)	AR(B, P, T)	AR(B, P, T)	AR(A, D, C)	AR(A, D, C)	AR(A, P, C)
12	NN(3, P, 2, 1)	NN(3, L, 2, 2)	NN(3, D, 2, 0)	AR(B, P, T)	AR(B, D, C)	AR(B, D, C)	AR(A, P, T)	AR(A, P, T)	AR(A, P, T)
13	NN(3, D, 2, 0)	NN(3, D, 2, 0)	NN(3, P, 2, 1)	AR(A, D, C)	AR(A, P, C)	AR(4, P, C)	AR(B, P, C)	AR(B, P, C)	AR(B, P, C)
14	NN(3, D, 2, 1)	NN(3, P, 2, 1)	MED(A)	AR(B, D, C)	AR(A, D, C)	AR(A, P, C)	AR(B, P, T)	AR(B, D, C)	AR(B, D, C)
15	AR(4, P, C)	NN(3, D, 2, 1)	NN(3, D, 2, 1)	AR(A, P, T)	AR(A, P, T)	AR(A, D, C)	AR(B, D, C)	AR(B, P, T)	AR(B, P, T)
16	AR(4, D, C)	AR(4, P, C)	AR(4, P, C)	AR(4, P, C)	AR(4, P, C)	AR(A, P, T)	AR(4, P, C)	AR(4, P, C)	AR(4, P, C)
17	AR(4, P, T)	AR(4, P, T)	AR(4, P, T)	AR(4, D, C)	AR(4, D, C)	AR(4, D, C)	NN(3, P, 2, 0)	AR(4, D, C)	AR(4, D, C)
18	NN(3, P, 2, 2)	AR(4, D, C)	AR(4, L, C)	AR(4, P, T)	AR(4, P, T)	AR(4, P, T)	AR(4, D, C)	AR(4, P, T)	AR(4, P, T)
19	NN(3, L, 2, 1)	PLS(REC, A–D)	AR(4, D, C)	NN(3, P, 2, 0)	NN(3, P, 2, 0)	NN(3, P, 2, 0)	AR(4, P, T)	LS(3, P, D6)	LS(3, P, D6)
20	PLS(REC, A)	AR(4, L, C)	PLS(REC, A)	LS(3, P, D6)	NN(3, P, 2, 1)	NN(3, P, 2, 1)	NN(3, P, 2, 2)	NN(3, P, 2, 0)	LS(3, D, D6)
21	AR(B, P, C)	PLS(REC, A)	AR(B, P, C)	NN(3, P, 2, 2)	NN(3, P, 2, 2)	NN(3, D, 2, 0)	NN(3, P, 2, 1)	LS(3, D, D6)	NN(3, P, 2, 0)
22	NN(3, L, 2, 2)	AR(B, P, C)	AR(B, L, C)	NN(3, P, 2, 1)	NN(3, D, 2, 0)	NN(3, D, 2, 1)	NN(3, D, 2, 0)	NN(3, D, 2, 0)	NN(B, D, B, 0)
23	AR(B, D, C)	AR(B, P, T)	AR(B, P, T)	LS(3, D, D6)	NN(3, D, 2, 1)	NN(3, P, 2, 2)	NN(A, P, A, 0)	NN(3, P, 2, 2)	NN(B, P, B, 0)
24	AR(B, P, T)	AR(B, D, C)	NN(A, L, A, 0)	AR(B, L, C)	NN(3, D, 2, 2)	LS(3, P, D6)	LS(3, P, D6)	NN(3, P, 2, 1)	NN(3, D, 2, 0)
25	AR(4, L, C)	AR(B, L, C)	AR(B, D, C)	PLS(REC, A)	LS(3, P, D6)	AR(B, L, C)	NN(3, D, 2, 2)	NN(A, P, A, 0)	NN(3, P, 2, 1)
26	NN(A, P, A, 0)	PLS(REC, A–C)	AR(4, L, T)	NN(3, D, 2, 0)	AR(B, L, C)	NN(3, D, 2, 2)	NN(3, D, 2, 1)	NN(3, D, 2, 1)	NN(3, D, 2, 1)
27	NN(3, D, 2, 2)	AR(4, D, T)	AR(4, D, T)	NN(3, D, 2, 2)	NN(A, P, A, 0)	PLS(REC, A)	NN(A, D, A, 0)	NN(3, D, 2, 2)	LS(B, D, B)
28	PLS(REC, A–C)	PLS(REC, B)	PLS(REC, A–D)	AR(A, L, C)	LS(3, D, D6)	NN(A, P, A, 0)	LS(B, P, B)	NN(B, P, B, 0)	NN(3, P, 2, 2)
29	NN(A, D, A, 0)	NN(3, P, 2, 2)	AR(A, L, C)	NN(A, P, A, 0)	NN(A, D, A, 0)	AR(4, L, C)	NN(B, P, B, 0)	NN(A, D, A, 0)	NN(3, D, 2, 2)
30	AR(B, L, C)	AR(A, P, C)	AR(B, L, T)	NN(3, D, 2, 1)	PLS(REC, A)	LS(3, D, D6)	LS(3, D, D6)	NN(B, D, B, 0)	NN(A, P, A, 0)
31	PLS(REC, B)	AR(4, L, T)	AR(B, D, T)	LS(B, P, B)	NN(B, P, B, 0)	NN(A, D, A, 0)	LS(B, D, B)	LS(B, P, B)	LS(B, P, B)

Table 4. (*cont.*)

Rank	1 step ahead			6 steps ahead			12 steps ahead		
	$\rho = 1.0$	2.0	3.0	$\rho = 1.0$	2.0	3.0	$\rho = 1.0$	2.0	3.0
32	NN(B, P, B, 0)	AR(A, D, C)	AR(A, P, C)	NN(B, P, B, 0)	AR(A, L, C)	NN(B, P, B, 0)	NN(B, D, B, 0)	LS(B, D, B)	NN(A, D, A, 0)
33	NN(B, D, B, 0)	AR(A, L, C)	NN(B, L, B, 0)	NN(A, D, A, 0)	NN(B, D, B, 0)	NN(B, D, B, 0)	LS(A, P, A)	LS(A, P, A)	LS(A, P, A)
34	AR(A, P, C)	AR(A, P, T)	PLS(REC, A–C)	LS(B, D, B)	AR(4, L, C)	AR(A, L, C)	LS(A, D, A)	LS(A, D, A)	LS(A, D, A)
35	AR(A, D, C)	NN(A, L, A, 0)	NN(3, L, 2, 0)	NN(B, D, B, 0)	LS(B, P, B)	PLS(REC, A–C)	AR(A, L, C)	AR(A, L, C)	AR(A, L, C)
36	PLS(REC, PM)	AR(B, D, T)	AR(A, D, C)	PLS(REC, A–C)	LS(B, D, B)	LS(3, P, P)	LS(3, P, P)	LS(3, P, P)	LS(3, P, P)
37	AR(A, P, T)	NN(3, D, 2, 2)	AR(A, P, T)	AR(4, L, C)	PLS(REC, A–C)	AR(4, D, T)	AR(B, L, C)	AR(B, L, C)	LS(3, D, D)
38	AR(4, D, T)	NN(A, P, A, 0)	PLS(REC, B)	LS(A, P, A)	LS(3, P, P)	LS(B, P, B)	LS(3, D, D)	LS(3, D, D)	PLS(REC, A)
39	AR(A, L, C)	PLS(REC, PM)	PLS(REC, PM)	PLS(REC, B)	LS(3, D, D)	LS(B, D, B)	PLS(REC, B)	PLS(REC, A)	AR(B, L, C)
40	LS(3, P, D6)	NN(B, P, B, 0)	AR(A, L, T)	LS(3, P, P)	PLS(REC, B)	AR(B, D, T)	PLS(REC, A)	PLS(REC, B)	PLS(REC, A–D)
41	LS(3, D, D6)	AR(B, L, T)	LS(3, P, D6)	LS(A, D, A)	PLS(REC, PM)	LS(3, D, D)	PLS(REC, A–C)	PLS(REC, A–C)	PLS(REC, B)
42	AR(B, D, T)	NN(A, D, A, 0)	AR(A, D, T)	LS(3, D, D)	AR(B, D, T)	PLS(REC, B)	AR(4, L, C)	PLS(REC, A–D)	PLS(REC, A–C)
43	AR(4, L, T)	NN(B, D, B, 0)	LS(3, D, D6)	PLS(REC, A–D)	PLS(REC, A–D)	PLS(REC, PM)	PLS(REC, A–D)	AR(4, L, C)	NN(3, L, 2, 1)
44	NN(A, L, A, 0)	NN(3, L, 2, 0)	LS(3, L, D6)	PLS(REC, PM)	AR(4, D, T)	PLS(REC, A–D)	PLS(REC, PM)	NN(3, L, 2, 1)	PLS(REC, PM)
45	AR(A, D, T)	LS(3, P, D6)	EX3	AR(B, D, T)	LS(A, D, A)	AR(A, D, T)	NN(3, L, 2, 1)	PLS(REC, PM)	AR(4, L, C)
46	NN(3, L, 2, 0)	NN(B, L, B, 0)	EX2	AR(A, D, T)	LS(A, P, A)	NN(3, L, 2, 1)	LS(3, L, D6)	AR(A, D, T)	AR(A, D, T)
47	AR(B, L, T)	LS(3, D, D6)	LS(3, L, L)	AR(4, D, T)	AR(A, D, T)	AR(4, L, T)	AR(A, D, T)	EX3	EX3
48	LS(3, P, P)	AR(A, D, T)	NN(B, P, B, 0)	NN(3, L, 2, 1)	NN(3, L, 2, 1)	NN(3, L, 2, 2)	AR(B, D, T)	AR(B, D, T)	EX2
49	LS(3, L, D6)	AR(A, L, T)	NN(3, P, 2, 2)	LS(3, L, D6)	NN(3, L, 2, 2)	LS(A, D, A)	EX3	AR(4, D, T)	AR(B, D, T)
50	NN(B, L, B, 0)	LS(3, L, D6)	LS(3, P, P)	NN(3, L, 2, 2)	LS(3, L, D6)	AR(B, L, T)	AR(4, D, T)	LS(3, L, D6)	AR(4, D, T)
51	LS(3, D, D)	EX3	NN(B, D, B, 0)	EX3	AR(4, L, T)	LS(A, P, A)	EX2	EX2	LS(3, L, D6)
52	LS(B, P, B)	LS(3, P, P)	NN(A, P, A, 0)	EX2	AR(B, L, T)	LS(3, L, D6)	NN(3, L, 2, 2)	NN(3, L, 2, 2)	NN(3, L, 2, 2)
53	LS(B, D, B)	EX2	LS(3, D, D)	AR(B, L, T)	EX3	AR(A, L, T)	AR(B, L, T)	AR(B, L, T)	AR(A, L, T)
54	AR(A, L, T)	LS(3, D, D)	NN(3, D, 2, 2)	AR(4, L, T)	AR(A, L, T)	EX3	AR(4, L, T)	AR(A, L, T)	AR(B, L, T)
55	EX3	LS(B, P, B)	LS(B, P, B)	AR(A, L, T)	EX2	EX2	AR(A, L, T)	AR(4, L, T)	AR(4, L, T)
56	EX2	LS(3, L, L)	NN(A, D, A, 0)	NN(3, L, 2, 0)	LS(3, L, L)	LS(3, L, L)	NN(3, L, 2, 0)	NN(3, L, 2, 0)	LS(3, L, L)
57	LS(A, P, A)	LS(B, D, B)	LS(B, D, B)	LS(3, L, L)	NN(3, L, 2, 0)	NN(B, L, B, 0)	LS(3, L, L)	LS(3, L, L)	NN(3, L, 2, 0)
58	LS(3, L, L)	LS(B, L, B)	LS(B, L, B)	NN(A, L, A, 0)	NN(A, L, A, 0)	NN(A, L, A, 0)	NN(B, L, B, 0)	NN(B, L, B, 0)	LS(B, L, B)
59	LS(A, D, A)	LS(A, P, A)	LS(A, P, A)	NN(B, L, B, 0)	NN(B, L, B, 0)	NN(3, L, 2, 0)	NN(A, L, A, 0)	LS(B, L, B)	NN(B, L, B, 0)
60	LS(B, L, B)	LS(A, D, A)	LS(A, D, A)	LS(B, L, B)	LS(B, L, B)	LS(B, L, B)	LS(B, L, B)	NN(A, L, A, 0)	LS(A, L, A)
61	LS(A, L, A)	LS(A, L, A)	LS(A, L, A)	LS(A, L, A)	LS(A, L, A)	LS(A, L, A)	LS(A, L, A)	LS(A, L, A)	NN(A, L, A, 0)
62	EX1	EX1	EX1	EX1	EX1	EX1	NOCHANGE	NOCHANGE	NOCHANGE
63	NOCHANGE	NOCHANGE	NOCHANGE	NOCHANGE	NOCHANGE	NOCHANGE	EX1	EX1	EX1

BIC lag choice yields slightly lower average MSE than AIC-based methods. Among ANNs, average forecast performance was slightly better using BIC than AIC, and the worst AIC-based forecasts were worse than the worst BIC-based forecasts. Among LSTARs, neither the AIC nor the BIC methods have mean, median, or 10 percent and 90 percent percentile relative MSEs as good as some of the fixed methods (in particular the LS(3, D, D6) and LS(3, P, D6) methods).

On average, the MSE improvement over the benchmark method from using data-based model selection methods are modest. For example, adopting BIC lag selection and unit root pretesting in an autoregression with a constant produces a median relative MSE of 1.00 for $h = 1$, 0.98 for $h = 6$, and 0.94 for $h = 12$. However, for some series, large MSE gains are possible, relative to the benchmark forecast. For example, in 2 percent of series, MSE reductions of two-thirds were achieved at the 12 month horizon by introducing BIC lag selection and unit root pretests to the benchmark method. Comparison of the AR(4, L, C), AR(B, L, C), AR(4, P, C), and AR(B, P, C) results in Table 2 shows that most of these gains are achieved by the unit root pretest rather than BIC lag selection.

Performance of simple methods. The simplest methods performed poorly relative to the benchmark AR(4, L, C) method. For example, for approximately three-fourths of series, the no change forecast was worse than the benchmark forecast at all three horizons (Table 2). The exponential smoothing method EX1 went badly wrong for some series, and on average all exponential smoothing methods have relative MSEs exceeding one at all horizons.

ANN methods. Generally speaking, some ANN methods performed well at the one-month horizon but no ANN methods performed as well as autoregressions at the six and twelve month horizons. First consider the results for the one month horizon. Based on the $\rho = 2$ results in Table 4, for $h = 1$ the best ANN model is NN(3, L, 2, 1) and the best AR model is AR(4, P, C). At $h = 1$, ANN methods are best for 40 percent of the individual series (Table 3). Comparison of the $h = 1$ entries in Table 2 for these models reveals that, for the ANN methods, the relative MSE performance measure has heavier tails than for the AR methods: the successes and failures across series are more pronounced. However, these methods have the same median and approximately the same mean. On average the forecasting gains from using the ANN models over the AR models at $h = 1$ are small or negligible from the perspective of mean square error loss. Thus while it is intriguing that ANN methods rank highly at short horizons, their edge in performance over autoregressive models is slim.

The relative performance of the ANN methods deteriorates as the forecast horizon increases. For the twelve month horizon, the worst ANN forecasts are considerably worse than the worst AR forecasts, with all ANN methods having relative MSEs exceeding 2 for at least 2 percent of the series. At these longer horizons, the ANN methods specified in levels perform particularly poorly. This generally poor performance of feedforward ANN methods for economic data, relative to linear models, is consistent with the findings in Swanson and White (1995, 1997) and Weigend and Gershenfeld (1994).

LSTAR methods. Although the LSTAR methods were rarely best for any series, in some cases they provided average MSE improvements, relative to the benchmark method. The best-performing LSTAR methods were the LS(3, D, D6) and its pretest variant LS(3, P, D6). Although both have mean relative MSEs of at least one, their median relative MSEs are less than one at the twelve month horizon (Table 2). The LSTAR methods generally performed worse than the ANN methods.

Forecast pooling. One of the striking features in Tables 2–4 is the strong performance of various forecast pooling procedures. Simple average forecasts, forecasts weighted by inverse MSEs, and the median forecasts outperform the benchmark method. Indeed, based on the loss function comparisons in Table 4, the most attractive forecast at the six and twelve month horizons for $\rho =$ 1, 2 or 3 is the simple average of the forecasts from all methods, and this is nearly the best at the one month horizon as well. Among the various weighting schemes, simple averaging and weighting by inverse MSEs produce similar performance. Performance, as measured by mean relative MSE, deteriorates as ω increases, especially at long horizons. In fact, performance of the PLS forecasts, which are the limit as $\omega \to \infty$ of the weighted average forecasts, is worse than all weighted average forecasts and the median forecast for all horizons and all ρ. As measured by average relative MSEs, the PLS forecasts are never better than the benchmark forecast. Use of a shortened window (60 or 120 months) seems to have little effect on the combination forecasts based on inverse MSE weights.

For $h = 6$ and $h = 12$, the pooling procedures that combine forecasts from all 49 methods have a slight edge over these procedures applied to only the linear, or only the nonlinear, methods. Indeed, for one-third of the series at all three horizons, the equal-weighted linear combination forecast that averages the forecasts from all 49 methods produces forecasts that are among the top five in Table 3 at all horizons. For one-fourth of the series, at all horizons a linear combination procedure produces the best forecast.

Sensitivity to forecaster attitudes towards risk. Rankings are provided in Table 4 for three loss functions: mean absolute error loss ($\rho = 1$), quadratic loss ($\rho = 2$), and cubic absolute error loss ($\rho = 3$). Mean absolute error loss characterizes a forecaster who is equally concerned about small and large errors; cubic loss most heavily penalizes large errors.

The rankings among the various methods are surprisingly insensitive to the choice of risk parameter ρ. Linear combination procedures minimize average loss for all three loss functions and, given h, the best combination method does not depend on ρ. For a given horizon, the identity of the best individual method usually does not depend on ρ (the exception is $h = 1$, $\rho = 2$).

Table 4 establishes a clear ranking of classes of models and procedures. At the six and twelve month horizons, combination forecasts are first, followed by AR forecasts, followed by ANN forecasts, followed by LSTAR forecasts, followed by EX and no change. At the one month horizon, combination forecasts are first, followed (in order) by ANN, AR, LSTAR, EX, and no change. If pooling procedures are excluded, the best method at the six and twelve month horizon is an autoregression based on a unit root pretest, a data-dependent lag length, and a constant. At the one month horizon, an AR(4) with a unit root pretest is the best linear method, but it is slightly outperformed by several ANN methods, in particular NN(3, L, 2, 1).

Effect of forecast trimming. All results discussed so far are based on trimmed forecasts. The results for some methods are very different when the forecasts are not trimmed. The effects of trimming are most important for the nonlinear methods, which for some series at some dates produce forecasts that err by an order of magnitude. The trimming also considerably improves AR forecasts in levels with a time trend.

For comparison purposes, the rankings for the various forecasting methods based on the untrimmed forecasts are given in Table 5. The differences between the rankings based on the trimmed (Table 4) and untrimmed (Table 5) forecasts are attributable to the relatively few extremely large forecast errors made by the nonlinear methods and, to a lesser degree, by the AR methods in levels with time trends. Because of these outliers, the median pooled forecasts are optimal for the untrimmed forecasts, and because the large errors are most frequent in the nonlinear methods, the linear combination forecasts perform well only when computed over just the linear methods.

The rankings of the individual methods change somewhat for the untrimmed forecasts. Autoregressive methods are now best at all horizons for all ρ. Autoregressive methods work well if the series is specified in levels with a constant, in first differences with a constant and/or time trend, or if a pretest is used, but they work poorly for the levels/time trend specification. Exponential smoothing and no change methods rank relatively higher because they produce fewer extreme errors. Among nonlinear methods for $\rho = 2$, the best ranking at any horizon is for NN(3, P, 2, 1), which is fifteenth for $h = 6$.

Nonlinearities across categories of series. The relative performance of linear and nonlinear methods for different groups of series is explored in Table 6 for the trimmed forecasts. The first three columns compare the relative performance of a linear method, AR(B, P, C), to two nonlinear methods,

Table 5. Rankings of various methods, combined over all series, for different loss functions: Untrimmed forecasts

Cost function = $\mathrm{E}|e_t|^\rho$, e_t = forecast error

Rank	1 step ahead			6 steps ahead			12 steps ahead		
	$\rho = 1.0$	2.0	3.0	$\rho = 1.0$	2.0	3.0	$\rho = 1.0$	2.0	3.0
1	MED(B)	MED(B)	MED(B)	MED(A–C)	MED(A–C)	MED(A–C)	MED(A–C)	MED(A–C)	MED(A–C)
2	MED(A–C)	MED(A–C)	MED(A–C)	C(0, REC, A)	MED(B)	MED(B)	C(0, REC, A)	C(0, REC, A)	C(0, REC, A)
3	C(0, REC, A)	C(0, REC, A)	C(0, REC, A)	MED(B)	C(0, REC, A)	C(0, REC, A)	MED(B)	MED(B)	MED(B)
4	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)	C(1, REC, A)
5	MED(A)	MED(A)	MED(A)	MED(A)	MED(A)	MED(A)	MED(A)	MED(A)	MED(A)
6	AR(4, P, C)	AR(4, P, C)	AR(4, P, C)	AR(B, P, C)	AR(B, P, C)	AR(B, P, C)	AR(A, P, C)	AR(A, P, C)	AR(A, D, C)
7	AR(4, D, C)	AR(4, P, T)	AR(4, P, T)	AR(A, P, C)	AR(B, P, T)	AR(B, P, T)	AR(A, D, C)	AR(A, D, C)	AR(A, P, C)
8	AR(4, P, T)	AR(4, D, C)	AR(4, L, C)	AR(B, P, T)	AR(B, D, C)	AR(B, D, C)	AR(A, P, T)	AR(A, P, T)	AR(A, P, T)
9	PLS(REC, A)	AR(4, L, C)	AR(4, D, C)	AR(A, D, C)	AR(A, P, C)	AR(A, P, C)	AR(B, P, C)	AR(B, P, C)	AR(B, P, C)
10	NN(3, D, 2, 1)	PLS(REC, A)	PLS(REC, A)	AR(B, D, C)	AR(A, D, C)	AR(A, D, C)	AR(B, P, T)	AR(B, D, C)	AR(B, D, C)
11	NN(3, P, 2, 1)	AR(B, P, C)	AR(B, P, C)	AR(A, P, T)	AR(A, P, T)	AR(A, P, T)	AR(B, D, C)	AR(B, P, T)	AR(B, P, T)
12	AR(B, P, C)	AR(B, P, T)	AR(B, L, C)	AR(4, P, C)	AR(4, P, C)	AR(4, P, C)	AR(4, P, C)	AR(4, P, C)	AR(4, P, C)
13	AR(4, L, C)	AR(B, D, C)	AR(B, P, T)	AR(4, D, C)	AR(4, D, C)	AR(4, P, T)	AR(4, D, C)	AR(4, D, C)	AR(4, D, C)
14	AR(B, D, C)	AR(B, L, C)	AR(4, L, T)	AR(4, P, T)	AR(4, P, T)	AR(4, D, C)	AR(4, P, T)	AR(4, P, T)	AR(4, P, T)
15	AR(B, P, T)	AR(4, D, T)	AR(B, D, C)	PLS(REC, A)	NN(3, P, 2, 1)	AR(4, L, C)	NN(3, P, 2, 1)	AR(A, L, C)	AR(A, L, C)
16	AR(B, L, C)	AR(A, P, C)	AR(4, D, T)	AR(B, L, C)	AR(B, L, C)	PLS(REC, A)	AR(A, L, C)	NN(3, P, 2, 1)	NN(3, P, 2, 1)
17	NN(3, P, 2, 2)	AR(4, L, T)	AR(A, L, C)	AR(A, L, C)	PLS(REC, A)	NN(3, P, 2, 1)	NN(3, D, 2, 1)	NN(3, D, 2, 1)	NN(3, D, 2, 1)
18	AR(A, P, C)	AR(A, D, C)	AR(B, L, T)	NN(3, P, 2, 1)	NN(3, D, 2, 1)	AR(B, L, C)	AR(B, L, C)	AR(B, L, C)	AR(B, L, C)
19	AR(A, D, C)	AR(A, L, C)	AR(B, D, T)	NN(3, D, 2, 1)	AR(A, L, C)	NN(3, D, 2, 1)	PLS(REC, A)	PLS(REC, A)	PLS(REC, A)
20	AR(A, P, T)	AR(A, P, T)	AR(A, P, C)	AR(4, L, C)	AR(4, L, C)	AR(A, L, C)	NN(3, P, 2, 2)	AR(4, L, C)	AR(4, L, C)
21	AR(4, D, T)	AR(B, D, T)	AR(A, D, C)	AR(B, D, T)	AR(B, D, T)	AR(4, D, T)	NN(3, D, 2, 2)	PLS(REC, A–C)	AR(A, D, T)
22	PLS(REC, A–C)	AR(B, L, T)	AR(A, P, T)	AR(A, D, T)	AR(4, D, T)	AR(B, D, T)	AR(4, L, C)	AR(A, D, T)	PLS(REC, A–C)
23	NN(3, D, 2, 2)	AR(A, D, T)	AR(A, L, T)	AR(4, D, T)	AR(A, D, T)	AR(A, D, T)	PLS(REC, A–C)	AR(B, D, T)	AR(B, D, T)
24	AR(A, L, C)	AR(A, L, T)	AR(A, D, T)	PLS(REC, A–C)	AR(4, L, T)	AR(4, L, T)	AR(A, D, T)	AR(4, D, T)	AR(4, D, T)
25	AR(B, D, T)	EX3	EX3	NN(3, L, 2, 1)	AR(B, L, T)	AR(B, L, T)	AR(B, D, T)	EX3	EX3
26	AR(4, L, T)	EX2	EX2	EX3	EX3	AR(A, L, T)	AR(4, D, T)	EX2	EX2
27	AR(A, D, T)	NN(3, D, 2, 1)	EX1	PLS(REC, B)	AR(A, L, T)	EX3	EX3	AR(B, L, T)	AR(B, L, T)
28	AR(B, L, T)	NN(3, P, 2, 1)	NOCHANGE	EX2	EX2	EX2	EX2	AR(A, L, T)	AR(A, L, T)
29	AR(A, L, T)	NN(3, P, 2, 2)	NN(3, D, 2, 1)	PLS(REC, PM)	NN(3, L, 2, 1)	NN(3, L, 2, 1)	LS(3, L, D6)	AR(4, L, T)	AR(4, L, T)
30	EX3	NN(3, D, 2, 2)	NN(3, P, 2, 1)	LS(3, L, D6)	LS(3, L, D6)	EX1	LS(3, P, D6)	NOCHANGE	NOCHANGE
31	EX2	PLS(REC, A–C)	NN(3, D, 2, 2)	AR(4, L, T)	EX1	NOCHANGE	AR(B, L, T)	EX1	EX1

Table 5. (*cont.*)

Rank	1 step ahead			6 steps ahead			12 steps ahead		
	$\rho = 1.0$	2.0	3.0	$\rho = 1.0$	2.0	3.0	$\rho = 1.0$	2.0	3.0
32	LS(3, P, D6)	EX1	NN(3, P, 2, 2)	AR(B, L, T)	NOCHANGE	LS(3, L, D6)	AR(A, L, T)	LS(3, L, L)	LS(3, L, L)
33	NN(3, L, 2, 1)	NOCHANGE	PLS(REC, A–C)	NN(3, P, 2, 0)	LS(3, L, L)	LS(B, L, B)	AR(4, L, T)	NN(3, P, 2, 2)	NN(3, P, 2, 2)
34	EX1	LS(3, P, D6)	LS(3, P, D6)	AR(A, L, T)	LS(B, L, B)	LS(3, L, L)	LS(A, D, A)	NN(3, D, 2, 2)	NN(3, D, 2, 2)
35	NOCHANGE	NN(3, L, 2, 1)	NN(3, L, 2, 1)	NN(3, D, 2, 0)	LS(A, L, A)	LS(A, L, A)	LS(B, P, B)	NN(3, L, 2, 2)	NN(3, L, 2, 2)
36	NN(3, L, 2, 2)	NN(3, L, 2, 2)	NN(3, L, 2, 2)	LS(3, L, L)	PLS(REC, A–C)	PLS(REC, A–C)	LS(B, D, B)	LS(B, L, B)	LS(B, L, B)
37	LS(3, D, D6)	LS(3, D, D6)	LS(3, D, D6)	LS(3, P, P)	PLS(REC, B)	PLS(REC, B)	LS(3, D, D6)	LS(3, L, D6)	LS(3, L, D6)
38	LS(3, P, P)	LS(3, P, P)	LS(3, P, P)	LS(3, D, D)	PLS(REC, PM)	PLS(REC, PM)	NN(3, L, 2, 2)	LS(3, P, D6)	LS(3, P, P)
39	LS(3, D, D)	LS(3, D, D)	LS(3, D, D)	LS(B, L, B)	NN(3, P, 2, 0)	NN(3, P, 2, 0)	LS(A, P, A)	LS(A, D, A)	LS(3, D, D)
40	LS(B, P, B)	LS(B, P, B)	LS(B, P, B)	LS(A, L, A)	NN(3, D, 2, 0)	NN(3, D, 2, 0)	LS(3, L, L)	LS(B, D, B)	LS(3, P, D6)
41	LS(B, D, B)	LS(B, D, B)	LS(B, D, B)	EX1	LS(3, P, P)	LS(3, P, P)	LS(3, P, P)	LS(B, P, B)	LS(A, D, A)
42	LS(3, L, D6)	LS(3, L, D6)	LS(3, L, D6)	NOCHANGE	LS(3, D, D)	LS(3, D, D)	LS(3, D, D)	LS(A, L, A)	LS(B, D, B)
43	PLS(REC, B)	PLS(REC, B)	PLS(REC, B)	LS(B, P, B)	LS(B, P, B)	LS(B, P, B)	NOCHANGE	LS(3, P, P)	LS(B, P, B)
44	LS(3, L, L)	LS(3, L, L)	LS(3, L, L)	LS(B, D, B)	LS(A, P, A)	LS(A, P, A)	EX1	LS(3, D, D)	LS(A, L, A)
45	LS(B, L, B)	LS(B, L, B)	LS(B, L, B)	LS(A, P, A)	LS(B, D, B)	LS(B, D, B)	LS(B, L, B)	LS(3, D, D6)	LS(3, D, D6)
46	LS(A, L, A)	LS(A, L, A)	LS(A, L, A)	LS(A, D, A)	LS(A, D, A)	LS(A, D, A)	LS(A, L, A)	LS(A, P, A)	LS(A, P, A)
47	LS(A, P, A)	LS(A, P, A)	LS(A, P, A)	NN(3, P, 2, 2)	NN(3, L, 2, 2)	NN(3, L, 2, 2)	NN(3, L, 2, 1)	NN(3, L, 2, 1)	NN(3, L, 2, 1)
48	LS(A, D, A)	LS(A, D, A)	LS(A, D, A)	NN(3, D, 2, 2)	NN(3, P, 2, 2)	NN(3, P, 2, 2)	PLS(REC, B)	NN(B, P, B, 0)	NN(B, P, B, 0)
49	NN(B, P, B, 0)	NN(B, D, B, 0)	NN(B, D, B, 0)	NN(3, L, 2, 2)	NN(3, D, 2, 2)	NN(3, D, 2, 2)	NN(B, P, B, 0)	PLS(REC, B)	PLS(REC, B)
50	NN(B, D, B, 0)	NN(B, P, B, 0)	NN(B, P, B, 0)	LS(3, P, D6)	LS(3, P, D6)	LS(3, P, D6)	NN(B, D, B, 0)	NN(B, D, B, 0)	NN(B, D, B, 0)
51	PLS(REC, A–D)	PLS(REC, A–D)	PLS(REC, A–D)	LS(3, D, D6)	LS(3, D, D6)	LS(3, D, D6)	NN(A, P, A, 0)	NN(A, P, A, 0)	NN(A, P, A, 0)
52	PLS(REC, PM)	PLS(REC, PM)	PLS(REC, PM)	NN(A, P, A, 0)	NN(A, P, A, 0)	NN(A, P, A, 0)	NN(A, D, A, 0)	NN(A, D, A, 0)	NN(A, D, A, 0)
53	NN(3, L, 2, 0)	NN(3, L, 2, 0)	NN(B, L, B, 0)	NN(A, D, A, 0)	NN(A, D, A, 0)	NN(A, D, A, 0)	NN(3, P, 2, 0)	NN(3, D, 2, 0)	NN(3, D, 2, 0)
54	NN(B, L, B, 0)	C(1, REC, B)	C(1, REC, A–C)	PLS(REC, A–D)	PLS(REC, A–D)	PLS(REC, A–D)	NN(3, D, 2, 0)	NN(3, P, 2, 0)	NN(3, P, 2, 0)
55	NN(A, L, A, 0)	NN(B, L, B, 0)	C(0, REC, B)	NN(B, D, B, 0)	NN(B, D, B, 0)	NN(B, D, B, 0)	PLS(REC, A–D)	PLS(REC, A–D)	PLS(REC, A–D)
56	C(0, REC, A–C)	C(0, REC, A–C)	C(1, REC, B)	NN(B, P, B, 0)	NN(B, P, B, 0)	NN(B, P, B, 0)	PLS(REC, PM)	PLS(REC, PM)	PLS(REC, PM)
57	C(1, REC, A–C)	C(1, REC, A–C)	C(0, REC, A–C)	NN(B, L, B, 0)	NN(B, L, B, 0)	C(0, REC, A–C)	NN(B, L, B, 0)	NN(B, L, B, 0)	C(1, REC, A–C)
58	C(0, REC, B)	NN(A, L, A, 0)	NN(A, L, A, 0)	NN(3, L, 2, 0)	NN(3, L, 2, 0)	C(1, REC, A–C)	NN(3, L, 2, 0)	NN(3, L, 2, 0)	NN(A, L, A, 0)
59	C(1, REC, B)	NN(3, P, 2, 0)	NN(3, P, 2, 0)	C(1, REC, A–C)	C(1, REC, A–C)	C(1, REC, B)	C(1, REC, A–C)	C(1, REC, A–C)	C(1, REC, B)
60	NN(3, D, 2, 0)	C(0, REC, B)	NN(A, P, A, 0)	C(1, REC, B)	C(1, REC, B)	NN(B, L, B, 0)	C(1, REC, B)	C(1, REC, B)	C(0, REC, B)
61	NN(3, P, 2, 0)	NN(3, D, 2, 0)	NN(3, D, 2, 0)	C(0, REC, A–C)	C(0, REC, A–C)	NN(A, L, A, 0)	C(0, REC, A–C)	C(0, REC, B)	C(0, REC, A–C)
62	NN(A, P, A, 0)	NN(A, P, A, 0)	NN(3, L, 2, 0)	C(0, REC, B)	C(0, REC, B)	NN(3, L, 2, 0)	C(0, REC, B)	C(0, REC, A–C)	NN(B, L, B, 0)
63	NN(A, D, A, 0)	NN(A, D, A, 0)	NN(A, D, A, 0)	NN(A, L, A, 0)	NN(A, L, A, 0)	C(0, REC, B)	NN(A, L, A, 0)	NN(A, L, A, 0)	NN(3, L, 2, 0)

Table 6. Forecasting performance broken down by category of series

Numbers in parentheses are the number of time series in each category
For each forecast, the first row corresponds to one-step ahead forecasts; the second row, to 6-step ahead forecasts; the third row, to 12-step ahead forecasts.

Category	Individual Methods			Pooling Procedures	
	Linear	Nonlinear		Linear	Nonlinear
	AR(B, P, C)	NN(B, P, B, 0)	LS(B, P, B)	C(1, REC, A)	C(1, REC, B)
Production	0.54	0.25	0.21	0.08	0.92
(24)	0.38	0.29	0.33	0.21	0.79
	0.33	0.21	0.46	0.38	0.63
Employment	0.52	0.21	0.28	0.17	0.83
(29)	0.48	0.48	0.03	0.21	0.79
	0.28	0.41	0.31	0.17	0.83
Wages	0.57	0.00	0.43	0.43	0.57
(7)	0.71	0.00	0.29	0.29	0.71
	0.43	0.00	0.57	0.29	0.71
Construction	0.57	0.29	0.14	0.19	0.81
(21)	0.29	0.38	0.33	0.05	0.95
	0.48	0.29	0.24	0.14	0.86
Trade	0.60	0.30	0.10	0.00	1.00
(10)	0.60	0.30	0.10	0.50	0.50
	0.90	0.10	0.00	0.40	0.60
Inventories	0.50	0.50	0.00	0.30	0.70
(10)	0.60	0.30	0.10	0.30	0.70
	0.50	0.20	0.30	0.40	0.60
Orders	0.64	0.29	0.07	0.07	0.93
(14)	0.57	0.29	0.14	0.29	0.71
	0.57	0.36	0.07	0.29	0.71
Money & Credit	0.43	0.48	0.09	0.13	0.87
(21)	0.39	0.30	0.30	0.57	0.43
	0.52	0.39	0.09	0.39	0.61
Stock Prices	0.36	0.55	0.09	0.00	1.00
(11)	0.64	0.18	0.18	0.55	0.45
	0.55	0.27	0.18	0.82	0.18
Interest Rates	0.18	0.73	0.09	0.00	1.00
(11)	0.18	0.64	0.18	0.00	1.00
	0.45	0.45	0.09	0.45	0.55
Exchange Rates	0.17	0.50	0.33	0.33	0.67
(6)	1.00	0.00	0.00	0.33	0.67
	0.33	0.50	0.17	0.17	0.83
Producer Prices	0.31	0.44	0.25	0.19	0.81
(16)	0.69	0.25	0.06	0.50	0.50
	0.63	0.25	0.13	0.44	0.56
Consumer Prices	0.38	0.50	0.13	0.19	0.81
(16)	0.69	0.00	0.31	0.63	0.38
	0.31	0.25	0.44	0.44	0.56
Consumption	0.40	0.40	0.20	0.00	1.00
(5)	0.40	0.40	0.20	0.40	0.60
	0.40	0.00	0.60	0.80	0.20
Miscellaneous	0.50	0.21	0.29	0.21	0.79
(14)	0.57	0.29	0.14	0.50	0.50
	0.50	0.36	0.14	0.57	0.43

Notes: Comparisons are for trimmed forecasts. For each row, the entries in the first three columns sum to 1.00, as do the entries in the final two columns, up to rounding error.

ANN(B, P, B, 0) and LS(B, P, B), by reporting the fraction of times that the column forecasting method is best among these three methods for the category of series specified in that row, by horizon. These three methods are automatic methods and were chosen for comparability: they all entail a recursive unit root pretest, recursive BIC lag length selection, and, for the nonlinear methods, recursive BIC-selected nonlinear parameterization. The final two columns provide a similar comparison, computed for the two linear combination forecasts respectively based on the linear and nonlinear methods (in both cases, weights are recursive inverse MSE).

The results suggest that the importance of nonlinearities differs across horizons and series. At $h = 1$, the nonlinear methods have the greatest relative success for interest rates and exchange rates, and have the least success for trade and orders. Combinations of the nonlinear forecasts are better than combinations of linear forecasts at $h = 1$ for most categories, notably so for stock prices, trade, and consumption. At the twelve month horizon, nonlinear methods work best for production, employment, exchange rates, and consumer prices. Exchange rates are interesting because the combination of nonlinear forecasts outperforms the combination of linear forecasts for five of the six exchange rates at $h = 12$. This is in some contrast to previous studies which have found limited ability of nonlinear methods to forecast exchange rates (Brooks, 1997). Consistent with the previous findings, the LSTAR methods generally are not the best (although they are for wages); of the nonlinear methods, the ANN forecasts are first more often.

5 Discussion and Conclusions

The LSTAR and ANN models must be viewed as these models *cum* the optimizers with which they were fit. The optimizers are designed to achieve local optima and, by random searching, to compare several local optima and to select the best. However, the evidence presented in the Appendix indicates that the resulting sequence of recursively estimated local optima are not, in general, global optima, and moreover different repetitions of the optimizer using different randomly drawn starting parameter values produce different sequences of local optima and thus different sequences of forecast errors. At first blush, this sounds like an important deficiency in this study, but in fact this is not obvious because improvements in the in-sample objective function seem not to correspond to better out-of-sample forecasts. In fact, the evidence in the Appendix suggests that improving the in-sample objective function over the value obtained using our algorithms on average neither improves nor worsens out-of-sample forecast performance. Thus, it is not clear that using an optimizer that more reliably achieved higher in-sample fits would necessarily improve the out-of-sample performance. These issues appear to be most pronounced for ANN models in levels and least pronounced for LSTAR models in differences. One interpretation of this is that the highly nonlinear ANN models can overfit these data at the global optimum, and more reliable out-of-sample forecasts are produced when "sensible" local optima are used. This makes us reluctant to endorse ANN models, even for the application in which they perform best in this study, one month ahead forecasting.

Another issue is whether our use of seasonally adjusted data might favor nonlinear methods. It is known that seasonal adjustment procedures are nonlinear filters, and Ghysels, Granger and Siklos (1996) showed that for Census X-11 these nonlinearities are sufficiently important that they can be detected with nontrivial power using various tests for nonlinearities. This suggests that some of the forecast MSE reduction of nonlinear methods is attributable to seasonal adjustment. It should be borne in mind that, were this the case, its implications are not self-evident. On the one hand, to the extent that we are interested in empirical evidence of nonlinear dynamics to guide theoretical macroeconomic modeling, then it is important to know if these nonlinearities are spuriously introduced by seasonal adjustment. On the other hand, if our interest is in forecasting seasonally adjusted series, the source of the nonlinearity is of only academic interest and the relevant question is which forecasting method best handles this nonlinearity. In any event, because the nonlinear models performed relatively poorly at the six and twelve month horizon, and made only slight improvements over linear models at the one month horizon, spurious nonlinearity from seasonal adjustment seems not to be an important practical consideration for forecasting, at least on average over these series.

Some additional caveats are also in order. Although a large number of methods have been considered, we have only considered two classes of nonlinear methods, and within artificial neural networks we have only considered feedforward neural nets. It is possible that other nonlinear methods, for example feedforward neural nets with more lags or recurrent neural nets, could perform better than those considered here. Also, these results are subject to sampling error. Although the design has carefully adhered to a recursive (simulated real-time) structure, because there are many forecasting methods considered, the estimated performance of the best-performing single method for these data arguably overstates the population counterpart of this performance measure. This criticism is less likely to be a concern, however, for the combination forecasts. Finally, it is unlikely that the best performing forecasts could have been identified as such in real time. When PLS was applied to all forecasts (including all the combination forecasts), the resulting PLS forecasts performed considerably worse than the best combination forecast, and indeed on average it performed worse than the benchmark method as measured by its mean relative MSE.

Bearing these comments in mind, we turn to the implications of this forecasting experiment for the five questions raised in the introduction.

First, although some of the nonlinear forecasts improve upon the linear forecasts for some series, most of the nonlinear forecasting methods produce worse forecasts than the linear methods. Overall, AR methods have lower average loss than the LSTAR or ANN methods at the six and twelve month horizons. The ANN methods have lower average loss at the one month horizon than the AR methods, but the improvement is small and is only present after trimming the outlier forecasts.

Second, perhaps surprisingly the nonlinear models that perform the best are not necessarily the most tightly parameterized. Generally speaking the ANN models have more nonlinear parameters than the LSTAR models, yet the ANN models outperform the LSTAR models. Evidently the nonlinearities exploited by the ANN models go beyond the switching or threshold effects captured by the LSTAR specifications.

Third, forecasts at all horizons are improved by unit root pretests. Severe forecast errors are made in nonlinear models specified in levels and in linear models in levels with time trends, and these errors are reduced substantially by choosing a differences or levels specification based on a preliminary test for a unit root.

Fourth, pooled forecasts were found to outperform the forecasts from any single method. The pooled forecasts that performed best combined the forecasts from all methods. Interestingly, although individual nonlinear methods performed poorly, the median nonlinear forecast outperformed all individual methods at all horizons, as did the averages of the nonlinear forecasts after trimming. The pooling procedures that place weight on all forecasting methods (whether equal weighting, inverse MSE weighting, or median) proved most reliable, while those that emphasized the recently best performing methods (especially PLS) proved least reliable. At the twelve month horizon, the mean relative MSE of the pooled forecast computed by simple averaging of all 49 methods is .87, and the 2 percent percentile relative MSE is .27. There was little effect (positive or negative) of using a reduced or rolling sample for computing the combination weights. We find these gains from combining forecasts to be surprising. Bates and Granger's (1969) motivation for combining forecasts is that each forecast draws on a different information set, so that the information embodied in the combined forecast is greater than the information in any individual forecast. Here, however, all forecasts are univariate, and in this sense the information sets of the forecasts are the same. These issues are further explored in Chan, Stock, and Watson (1998).

Fifth, although the combination forecasts require considerable programming and computation time to produce, the gains might well be worth this cost. If, however, a macroeconomic forecaster is restricted to using a single method, then, for the family of loss functions considered here, he or she would be well advised to use an autoregression with a unit root pretest and data-dependent lag length selection.

Appendix

A.1 Nonlinear Optimization Methods

The ANN, exponential smoothing, and LSTAR models are nonlinear in the parameters. This Appendix describes the optimization methods used to minimize the least squares objective functions for the models.

Exponential Smoothing. The parameters of the exponential smoothing models were estimated using a Gauss–Newton optimizer. The parameters at date T_1 were estimated using 200 iterations of the optimizer from a starting value of 0.5. These estimates were updated in subsequent time periods using two iterations of the optimizer.

LSTAR models. The parameters of the LSTAR models were estimated using a modified random search algorithm. The initial parameter estimates at date T_1 were obtained as follows. The LSTAR models can be organized into families of models that have natural nestings, from least complicated (fewest parameters) to most complicated. For example, one such family is, in increasing order of complexity, {LS(1, L, L), LS(3, L, L), LS(6, L, L)}. For each family, the most restrictive version of the model was estimated first. For this most restrictive version the objective function was evaluated using 5,000 random draws of the parameter vector. The parameter vectors corresponding to the four smallest values of the objective function were then used as initial values for 250 Gauss–Newton iterations, and the minimizer was chosen from the resulting set of parameters. This parameter vector together with 1,000 additional random draws was used to evaluate the objective function associated with the next most complicated model in the family; the parameter vectors associated with the two smallest values of the function were used to initialize 100 Gauss–Newton iterations. This procedure was repeated for each larger model in the nesting sequence.

At subsequent dates ($T_1 < t \leq T_3$), with probability .99 the parameter values for each model were updated by taking three Gauss–Newton steps, using the parameter estimates from the previous date as starting values. With probability .01 the parameters were updated by using the minimum of these results and results obtained by completely reoptimizing from a set of 500 randomly selected initial parameter values (using the same method as at time T_1).

ANN models. ANN objective functions typically have multiple local minima. When the previous algorithm was applied to ANN models, local minima were obtained, and many of these local minima produced poor out of sample forecasts. A different algorithm was therefore used to fit the ANN models.

The algorithm used for the ANN models has two stages, a preprocessing phase and a recursive estimation phase. In the preprocessing phase, the objective function for each model was intensively minimized at three dates, T_1, T_3, and $T_{\mathrm{mid}} = (T_1 + T_3)/2$. For a model with one hidden unit the algorithm is: (i) fit an AR(p) with a constant or time trend, in levels or differences as appropriate, for the desired horizon, and save the residuals $\hat{u}_t^{(0)} = y_t - \hat{\beta}_0'\zeta_{t-h}$; (ii) using 10 randomly selected initial conditions for the nonlinear parameter vector β_{11} in the first hidden unit, compute 50 Gauss–Newton iterations on the objective function $\sum_t[\hat{u}_t^{(0)} - \gamma_{11}g(\beta_{11}'\zeta_t)]^2$, and retain the best of the resulting 10 sets of values of β_{11}; (iii) perform 15 Gauss–Seidel iterations over the full model. Each Gauss–Seidel iteration involves (iiia) fixing β_{11} and estimating β_0 and γ_{11} by OLS; (iiib) given the resulting values of β_0, reestimating β_{11} and γ_{11} by 10 Gauss–Newton iterations. At each Gauss–Newton step only updated values of β_{11} that improved the objective function were retained so that each step of this algorithm is guaranteed not to increase the objective function.

For models with $n_1 > 1$ hidden units and a single layer, the same procedure was used, except that (i) was omitted, (ii) used the residuals from the model with $n_1 - 1$ hidden units, and in (iii) the Gauss–Seidel steps moved sequentially over each hidden unit, estimating β_{1i} holding β_{1j} fixed, $j \neq i$, etc.

For models with two hidden units, this algorithm was used, with the modification that the nonlinear parameters $\{\beta_{2ji}, \beta_{1i}\}$ were estimated jointly by Gauss–Newton, given $\{\beta_0, \gamma_{2j}\}$, then given $\{\beta_{2ji}, \beta_{1i}\}$, $\{\beta_0, \gamma_{2j}\}$ were estimated by OLS, etc.

In the recursive estimation phase, for $t = T_1, \ldots, T_3$, in each time period a single Gauss–Seidel iteration (with one nested Gauss–Newton step) was used to update the parameters, with the initial estimates at $t = T_1$ obtained from the initial estimation phase. The objective function was also evaluated for the parameter values obtained in the initial phase for T_1, T_{mid} and T_3. If at date t either of these three produced a lower objective function value than the recursively updated parameters, the Gauss–Seidel step for date t was recomputed using the T_1, T_{mid} or T_3 parameter vector (as appropriate) as the initialization, and the resulting new parameter vector was retained as the recursive estimate and as the initial parameter vector for the Gauss–Seidel step at $t + 1$.

Performance of Algorithms. Several checks were performed on these algorithms to assess their performance. These checks entailed examining the performance of the optimizers (and variants on these optimizers) for different series and different models.

One such check, which examined four series (mdu, fygm3, hsbr, ivmtq) and four models, is discussed here. In this experiment, the optimizer described above was run 25 times for each model/series combination. This produced 25 series of recursive forecasts. The time series of optimized parameter values, and thus the time series of recursive forecasts, differ from one trial to the next solely because of different random draws of the starting values for the parameters. The four models investigated in this experiment are two LSTAR and two ANN models using both levels and differences of the data (LS(3, L, Δy_{t-2}), LS(3, D, Δy_{t-2}), NN(3, L, 0), and NN(3, D, 0)). For the purpose of this paper, the appropriate measure for assessing whether these algorithms converge to a common optimum is neither the value of the in-sample objective function nor the values of the estimated parameters, but rather the path of recursive forecast errors produced. If the same sequence of forecasts is produced by each trial, then for the purposes of this project the algorithms effectively converge to the same value, which can reasonably be taken to be the optimum.

The results for six month ahead forecasts are summarized in Table A1. The entries are summary statistics of the distribution of relative MSEs of the sequence of simulated out of sample forecast errors across the 25 trials; as in Table 2, the relative MSEs are standardized by the MSE of the AR(4, L, C). Shown in the final row of each block are results from forecasts constructed from more computationally intensive maximization algorithms. In the case of the LSTAR model increasing computation by a factor of 50 led to results that consistently achieved what appears to be the global optimum.[6] In the case of the ANN models, experiments suggested that there was little hope of consistently achieving a global optimum for each time period even with a significant increase in computational resources. In this case we simply computed a sequence of forecasts from the 25 trials by recursively choosing the forecast with the best in-sample fit: that is, at each date, the current objective functions of the 25 trials are compared, and the parameter values associated with the currently best of these in-sample objective functions is used to produce the forecast at that date. This sequence of forecasts will by construction have the lowest sequence of in-sample objective functions and in this sense is the closest to the global optimum.

The results in Table A1 suggest several conclusions. For some series and models, the trials resulted in essentially identical results (for example, ivmtq and LS(3, D, Δy_{t-2})), while in other cases there was considerable variation across the trials (mdu and NN(3, L, 0)). In general, the distribution of relative MSEs is tighter for LSTAR models than for ANN models, and is tighter in first differences than in levels. Strikingly, the relative MSE of the forecast based on the global optimum (LSTAR models) does about as well as a randomly selected forecast from the original group of 25: among the eight LSTAR cases in Table A1, the global optimum produces forecast MSEs that are greater than the median of the 25 trials in four cases and are less than or equal to the median in four cases. The same is true for the forecasts based on the recursive best fit for the ANN models. Evidently, among these model/series combinations, the parameters with the best in-sample fits do not in general provide the best out of sample forecasts. Rather, the forecasts with the lowest out of sample MSEs typically are obtained from in-sample fits that are at local but not global optima.

These findings are consistent with those from other checks we performed.

Table A1. Distribution of relative MSES of 6-month ahead forecasts across 25 optimization trials for selected series and models

Model	Series			
	mdu	fygm3	hsbr	ivmtq
LS(3, L, Δy_{t-2})				
Minimum	1.24	1.06	1.05	1.18
1st Quartile	1.30	1.10	1.06	1.18
Median	1.42	1.18	1.06	1.19
3rd Quartile	1.43	1.20	1.07	1.20
Maximum	2.41	1.43	1.16	1.27
In-Sample Optimum	1.76	1.35	1.20	1.18
LS(3, D, Δy_{t-2})				
Minimum	0.99	0.99	1.10	1.01
1st Quartile	1.00	1.03	1.12	1.01
Median	1.00	1.21	1.16	1.01
3rd Quartile	1.01	1.22	1.19	1.01
Maximum	1.02	1.27	1.19	1.01
In-Sample Optimum	0.99	1.12	1.17	1.01
NN(3, L, 2, 0)				
Minimum	1.71	0.91	0.97	1.62
1st Quartile	2.38	0.99	1.00	1.80
Median	2.46	1.07	1.04	1.93
3rd Quartile	2.77	1.16	1.05	1.97
Maximum	3.24	1.32	1.11	2.37
Rec. Best In-Sample	2.56	1.18	1.02	1.85
NN(3, D, 2, 0)				
Minimum	0.92	0.98	1.05	1.01
1st Quartile	0.97	1.07	1.13	1.03
Median	1.00	1.12	1.15	1.04
3rd Quartile	1.04	1.21	1.19	1.06
Maximum	1.10	1.30	1.24	1.09
Rec. Best In-Sample	1.01	1.09	1.18	1.07

Notes: Entries are summaries of the distribution of relative MSEs of recursive forecasts, where the benchmark MSE is the AR(4, L, C) model. "In-Sample Optimum" refers to forecasts constructed from parameters that achieve the global recursively calculated in-sample MSE. "Rec. Best In-Sample" refers to the simulated real time forecast error produced using the sequence of parameter values that, date by date, have the best in-sample fit selected from the 25 optimized parameter values in the different trials.

For many series, this indicates that there was a significant likelihood that these algorithms would not locate a global optimum over some fraction of the sample period. The probability of achieving a local and not global maximum appeared to be higher for the ANN models than for the LSTAR models and higher for series modeled in levels than in first differences. Finally, achieving a better in-sample value of the objective function does not necessarily imply producing forecasts with better out-of-sample performance.

A.2 Data Description

This Appendix lists the time series used. The data were obtained from the DRI BASIC Economics Database (creation date 9/97). The format for each series is its DRI BASIC mnemonic; a brief description; and the first date used (in brackets). A series that was preliminarily transformed by taking its logarithm is denoted by "log" in parentheses; otherwise, the series was used without preliminary transformation.

Abbreviations: sa = seasonally adjusted; saar = seasonally adjusted at an annual rate; nsa = not seasonally adjusted.

IP	industrial production: total index (1992 = 100, sa) [1959:1] (log)
IPP	industrial production: products, total (1992 = 100, sa) [1959:1] (log)
IPF	industrial production: final products (1992 = 100, sa) [1959:1] (log)
IPC	industrial production: consumer goods (1992 = 100, sa) [1959:1] (log)
IPCD	industrial production: durable consumer goods (1992 = 100, sa) [1959:1] (log)
IPCN	industrial production: nondurable consumer goods (1992 = 100, sa) [1959:1] (log)
IPE	industrial production: business equipment (1992 = 100, sa) [1959:1] (log)
IPI	industrial production: intermediate products (1992 = 100, sa) [1959:1] (log)
IPM	industrial production: materials (1992 = 100, sa) [1959:1] (log)
IPMD	industrial production: durable goods materials (1992 = 100, sa) [1959:1] (log)
IPMND	industrial production: nondurable goods materials (1992 = 100, sa) [1959:1] (log)
IPMFG	industrial production: manufacturing (1992 = 100, sa) [1959:1] (log)
IPD	industrial production: durable manufacturing (1992 = 100, sa) [1959:1] (log)
IPN	industrial production: nondurable manufacturing (1992 = 100, sa) [1959:1] (log)
IPMIN	industrial production: mining (1992 = 100, sa) [1959:1] (log)
IPUT	industrial production: utilities (1992 = 100, sa) [1959:1] (log)
IPX	capacity util rate: total industry (% of capacity, sa)(frb) [1967:1]
IPXMCA	capacity util rate: manufacturing, total (% of capacity, sa)(frb) [1959:1]
IPXDCA	capacity util rate: durable mfg (% of capacity, sa)(frb) [1967:1]
IPXNCA	capacity util rate: nondurable mfg (% of capacity, sa)(frb) [1967:1]
IPXMIN	capacity util rate: mining (% of capacity, sa)(frb) [1967:1]
IPXUT	capacity util rate: utilities (% of capacity, sa)(frb) [1967:1]
LHEL	index of help-wanted advertising in newspapers (1967 = 100; sa) [1959:1]
LHELX	employment: ratio; help-wanted ads: no. unemployed clf [1959:1]

LHEM	civilian labor force: employed, total (thous., sa) [1959:1] (log)
LHNAG	civilian labor force: employed, nonagric.industries (thous., sa) [1959:1] (log)
LHUR	unemployment rate: all workers, 16 years & over (%, sa) [1959:1]
LHU680	unemploy. by duration: average (mean) duration in weeks (sa) [1959:1]
LHU5	unemploy. by duration: persons unempl. less than 5 wks (thous., sa) [1959:1] (log)
LHU14	unemploy. by duration: persons unempl. 5 to 14 wks (thous., sa) [1959:1] (log)
LHU15	unemploy. by duration: persons unempl. 15 wks + (thous., sa) [1959:1] (log)
LHU26	unemploy. by duration: persons unempl. 15 to 26 wks (thous., sa) [1959:1] (log)
LHU27	unemploy. by duration: persons unempl. 27 wks + (thous, sa) [1959:1] (log)
LHCH	average hours of work per week (household data)(sa) [1959:1]
LPNAG	employees on nonag. payrolls: total (thous., sa) [1959:1] (log)
LP	employees on nonag. payrolls: total, private (thous, sa) [1959:1] (log)
LPGD	employees on nonag. payrolls: goods-producing (thous., sa) [1959:1] (log)
LPMI	employees on nonag. payrolls: mining (thous., sa) [1959:1] (log)
LPCC	employees on nonag. payrolls: contract construction (thous., sa) [1959:1] (log)
LPEM	employees on nonag. payrolls: manufacturing (thous., sa) [1959:1] (log)
LPED	employees on nonag. payrolls: durable goods (thous., sa) [1959:1] (log)
LPEN	employees on nonag. payrolls: nondurable goods (thous., sa) [1959:1] (log)
LPSP	employees on nonag. payrolls: service-producing (thous., sa) [1959:1] (log)
LPTU	employees on nonag. payrolls: trans. & public utilities (thous., sa) [1959:1] (log)
LPT	employees on nonag. payrolls: wholesale & retail trade (thous., sa) [1959:1] (log)
LPFR	employees on nonag. payrolls: finance, insur.&real estate (thous., sa) [1959:1] (log)
LPS	employees on nonag. payrolls: services (thous., sa) [1959:1] (log)
LPGOV	employees on nonag. payrolls: government (thous., sa) [1959:1] (log)
LW	avg. weekly hrs. of prod. wkrs.: total private (sa) [1964:1]
LPHRM	avg. weekly hrs. of production wkrs.: manufacturing (sa) [1959:1]
LPMOSA	avg. weekly hrs. of production wkrs.: mfg., overtime hrs. (sa)[1959:1]
LEH	avg. hr earnings of prod wkrs: total private nonagric ($, sa) [1964:1] (log)
LEHCC	avg. hr earnings of constr wkrs: construction ($, sa) [1959:1] (log)
LEHM	avg. hr earnings of prod wkrs: manufacturing ($, sa) [1959:1] (log)
LEHTU	avg. hr earnings of nonsupv wkrs: trans & public util ($, sa) [1964:1] (log)
LEHTT	avg. hr earnings of prod wkrs: wholesale & retail trade (sa) [1964:1] (log)
LEHFR	avg. hr earnings of nonsupv wkrs: finance, insur., real est ($, sa) [1964:1] (log)
LEHS	avg. hr earnings of nonsupv wkrs: services ($, sa) [1964:1] (log)

HSFR	housing starts: nonfarm(1947–58); total farm&nonfarm(1959–)(thous., sa) [1959:1] (log)
HSNE	housing starts: northeast (thous.u.)s.a. [1959:1] (log)
HSMW	housing starts: midwest(thous.u.)s.a. [1959:1] (log)
HSSOU	housing starts: south (thous.u.)s.a. [1959:1] (log)
HSWST	housing starts: west (thous.u.)s.a. [1959:1] (log)
HSBR	housing authorized: total new priv. housing units (thous., saar) [1959:1] (log)
HSBNE	houses authorized by build. permits: northeast(thou.u.)s.a [1960:1] (log)
HSBMW	houses authorized by build. permits: midwest(thou.u.)s.a. [1960:1] (log)
HSBSOU	houses authorized by build. permits: south(thou.u.)s.a. [1960:1] (log)
HSBWST	houses authorized by build. permits: west(thou.u.)s.a. [1960:1] (log)
HNS	new one-family houses sold during month (thous, saar) [1963:1] (log)
HNSNE	one-family houses sold: northeast(thou.u., s.a.) [1973:1] (log)
HNSMW	one-family houses sold: midwest(thou.u., s.a.) [1973:1] (log)
HNSSOU	one-family houses sold: south(thou.u., s.a.) [1973:1] (log)
HNSWST	one-family houses sold: west(thou.u., s.a.) [1973:1] (log)
HNR	new one-family houses, month's supply @ current sales rate (ratio) [1963:1]
HMOB	mobile homes: manufacturers' shipments (thous.of units, saar) [1959:1] (log)
CONTC	construct.put in place: total priv & public 1987$ (mil$, saar) [1964:1] (log)
CONPC	construct.put in place: total private 1987$ (mil$, saar) [1964:1] (log)
CONQC	construct.put in place: public construction 87$ (mil$, saar) [1964:1] (log)
CONDO9	construct.contracts: comm'l & indus.bldgs (mil.sq.ft.floor sp.; sa) [1959:1] (log)
MSMTQ	manufacturing & trade: total (mil of chained 1992 dollars)(sa) [1959:1] (log)
MSMQ	manufacturing & trade: manufacturing; total (mil of chained 1992 dollars)(sa) [1959:1] (log)
MSDQ	manufacturing & trade: mfg; durable goods (mil of chained 1992 dollars)(sa) [1959:1] (log)
MSNQ	manufact. & trade: mfg; nondurable goods (mil of chained 1992 dollars)(sa) [1959:1] (log)
WTQ	merchant wholesalers: total (mil of chained 1992 dollars)(sa) [1959:1] (log)
WTDQ	merchant wholesalers: durable goods total (mil of chained 1992 dollars)(sa) [1959:1] (log)
WTNQ	merchant wholesalers: nondurable goods (mil of chained 1992 dollars)(sa) [1959:1] (log)
RTQ	retail trade: total (mil of chained 1992 dollars)(sa) [1959:1] (log)
RTDQ	retail trade: durable goods total (mil.87$)(s.a.) [1959:1] (log)
RTNQ	retail trade: nondurable goods (mil of 1992 dollars)(sa) [1959:1] (log)
IVMTQ	manufacturing & trade inventories: total (mil of chained 1992)(sa) [1959:1] (log)

IVMFGQ	inventories, business, mfg (mil of chained 1992 dollars, sa) [1959:1] (log)
IVMFDQ	inventories, business durables (mil of chained 1992 dollars, sa) [1959:1] (log)
IVMFNQ	inventories, business, nondurables (mil of chained 1992 dollars, sa) [1959:1] (log)
IVWRQ	manufacturing & trade inv: merchant wholesalers (mil of chained 1992 dollars)(sa) [1959:1] (log)
IVRRQ	manufacturing & trade inv: retail trade (mil of chained 1992 dollars)(sa) [1959:1] (log)
IVSRQ	ratio for mfg & trade: inventory/sales (chained 1992 dollars, sa) [1959:1]
IVSRMQ	ratio for mfg & trade: mfg; inventory/sales (87$)(s.a.) [1959:1]
IVSRWQ	ratio for mfg & trade: wholesaler; inventory/sales (87$)(s.a.) [1959:1]
IVSRRQ	ratio for mfg & trade: retail trade; inventory/sales (87$)(s.a.) [1959:1]
PMI	purchasing managers' index (sa) [1959:1]
PMP	napm production index (percent) [1959:1]
PMNO	napm new orders index (percent) [1959:1]
PMDEL	napm vendor deliveries index (percent) [1959:1]
PMNV	napm inventories index (percent) [1959:1]
PMEMP	napm employment index (percent) [1959:1]
PMCP	napm commodity prices index (percent) [1959:1]
MOCMQ	new orders (net)—consumer goods & materials, 1992 dollars (bci) [1959:1] (log)
MDOQ	new orders, durable goods industries, 1992 dollars (bci) [1959:1] (log)
MSONDQ	new orders, nondefense capital goods, in 1992 dollars (bci) [1959:1] (log)
MO	mfg new orders: all manufacturing industries, total (mil$, sa) [1959:1] (log)
MOWU	mfg new orders: mfg industries with unfilled orders (mil$, sa) [1959:1] (log)
MDO	mfg new orders: durable goods industries, total (mil$, sa) [1959:1] (log)
MDUWU	mfg new orders: durable goods industries with unfilled orders (mil$, sa) [1959:1] (log)
MNO	mfg new orders: nondurable goods industries, total (mil$, sa) [1959:1] (log)
MNOU	mfg new orders: nondurable gds ind.with unfilled orders (mil$, sa) [1959:1] (log)
MU	mfg unfilled orders: all manufacturing industries, total (mil$, sa) [1959:1] (log)
MDU	mfg unfilled orders: durable goods industries, total (mil$, sa) [1959:1] (log)
MNU	mfg unfilled orders: nondurable goods industries, total (mil$, sa) [1959:1] (log)
MPCON	contracts & orders for plant & equipment (bil$, sa) [1959:1] (log)
MPCONQ	contracts & orders for plant & equipment in 1992 dollars (bci) [1959:1] (log)
FM1	money stock: m1(curr, trav.cks, dem dep, other ck'able dep)(bil$, sa) [1959:1] (log)

FM2	money stock: m2(m1 + o'nite rps, euro$, g/p&b/d mmmfs&sav&sm time dep (bil$, sa) [1959:1] (log)
FM3	money stock: m3(m2 + lg time dep, term rp's&inst only mmmfs)(bil$, sa) [1959:1] (log)
FML	money stock: l(m3 + other liquid assets) (bil$, sa) [1959:1] (log)
FM2DQ	money supply—m2 in 1992 dollars (bci) [1959:1] (log)
FMFBA	monetary base, adj. for reserve requirement changes(mil$, sa) [1959:1] (log)
FMBASE	monetary base, adj. for reserve req chgs(frb of st.louis)(bil$, sa) [1959:1] (log)
FMRRA	depository inst reserves: total, adj for reserve req chgs(mil$, sa) [1959:1] (log)
FMRNBA	depository inst reserves: nonborrowed, adj res req chgs(mil$, sa) [1959:1] (log)
FMRNBC	depository inst reserves: nonborrow + ext cr, adj res req cgs(mil$, sa) [1959:1] (log)
FMFBA	monetary base, adj for reserve requirement changes(mil$, sa) [1959:1] (log)
FCLS	loans & sec @ all coml banks: total (bil$, sa) [1973:1] (log)
FCSGV	loans & sec @ all coml banks: US govt. securities (bil$, sa) [1973:1] (log)
FCLRE	loans & sec @ all coml banks: real estate loans (bil$, sa) [1973:1] (log)
FCLIN	loans & sec @ all coml banks: loans to individuals (bil$, sa) [1973:1] (log)
FCLNBF	loans & sec @ all coml banks: loans to nonbank fin. inst (bil$, sa) [1973:1] (log)
FCLNQ	commercial & industrial loans oustanding in 1992 dollars (bci) [1959:1] (log)
FCLBMC	wkly rp lg com'l banks:net change com'l & indus loans (bil$, saar) [1959:1]
CCI30M	consumer instal.loans: delinquency rate, 30 days & over, (%, sa) [1959:1]
CCINT	net change in consumer instal cr: total (mil$, sa) [1975:1]
CCINV	net change in consumer instal cr: automobile (mil$, sa) [1975:1]
FSNCOM	nyse common stock price index: composite (12/31/65 = 50) [1959:1] (log)
FSNIN	nyse common stock price index: industrial (12/31/65 = 50) [1966:1] (log)
FSNTR	nyse common stock price index: transportation (12/31/65 = 50) [1966:1] (log)
FSNUT	nyse common stock price index: utility (12/31/65 = 50) [1966:1] (log)
FSNFI	nyse common stock price index: finance (12/31/65 = 50) [1966:1] (log)
FSPCOM	s&p's common stock price index: composite (1941–43 = 10) [1959:1] (log)
FSPIN	s&p's common stock price index: industrials (1941–43 = 10) [1959:1] (log)
FSPCAP	s&p's common stock price index: capital goods (1941–43 = 10) [1959:1] (log)
FSPTR	s&p's common stock price index: transportation (1970 = 10) [1970:1] (log)

FSPUT	s&p's common stock price index: utilities (1941–43 = 10) [1959:1] (log)
FSPFI	s&p's common stock price index: financial (1970 = 10) [1970:1] (log)
FSDXP	s&p's composite common stock: dividend yield (% per annum) [1959:1] (log)
FSPXE	s&p's composite common stock: price–earnings ratio (%, nsa) [1959:1] (log)
FSNVV3	nyse mkt composition: reptd share vol by size, 5000 + shrs, % [1959:1] (log)
FYFF	interest rate: federal funds (effective) (% per annum, nsa) [1959:1]
FYCP	interest rate: commercial paper, 6-month (% per annum, nsa) [1959:1]
FYGM3	interest rate: US treasury bills, sec mkt, 3-mo.(% per ann, nsa) [1959:1]
FYGM6	interest rate: US treasury bills, sec mkt, 6-mo. (% per ann, nsa) [1959:1]
FYGT1	interest rate: US treasury const maturities, 1-yr. (% per ann, nsa) [1959:1]
FYGT5	interest rate: US treasury const maturities, 5-yr. (% per ann, nsa) [1959:1]
FYGT10	interest rate: US treasury const maturities, 10-yr. (% per ann, nsa) [1959:1]
FYAAAC	bond yield: moody's aaa corporate (% per annum) [1959:1]
FYBAAC	bond yield: moody's baa corporate (% per annum) [1959:1]
FWAFIT	weighted avg foreign interest rate(%, sa) [1959:1]
FYFHA	secondary market yields on fha mortgages (% per annum) [1959:1]
EXRUS	united states; effective exchange rate(merm)(index no.) [1973:1] (log)
EXRGER	foreign exchange rate: germany (deutsche mark per US$) [1973:1] (log)
EXRSW	foreign exchange rate: switzerland (swiss franc per US$) [1973:1] (log)
EXRJAN	foreign exchange rate: japan (yen per US$) [1973:1] (log)
EXRUK	foreign exchange rate: united kingdom (cents per pound) [1973:1] (log)
EXRCAN	foreign exchange rate: canada (canadian $ per US$) [1973:1] (log)
HHSNTN	u. of mich. index of consumer expectations (bcd–83) [1959:1]
F6EDM	US mdse exports: [1964:1] (log)
FTMC6	US mdse imports: crude materials & fuels (mil$, nsa) [1964:1] (log)
FTMM6	US mdse imports: manufactured goods (mil$, nsa) [1964:1] (log)
PWFSA	producer price index: finished goods (82 = 100, sa) [1959:1] (log)
PWFCSA	producer price index: finished consumer goods (82 = 100, sa) [1959:1] (log)
PWIMSA	producer price index: intermed mat. supplies & components (82 = 100, sa) [1959:1] (log)
PWCMSA	producer price index: crude materials (82 = 100, sa) [1959:1] (log)
PWFXSA	producer price index: finished goods, excl. foods (82 = 100, sa) [1967:1] (log)
PW160A	producer price index: crude materials less energy (82 = 100, sa) [1974:1] (log)
PW150A	producer price index: crude nonfood mat less energy (82 = 100, sa) [1974:1] (log)

PW561	producer price index: crude petroleum (82 = 100, nsa) [1959:1] (log)
PWCM	producer price index: construction materials (82 = 100, nsa) [1959:1] (log)
PWXFA	producer price index: all commodities ex. farm prod (82 = 100, nsa) [1959:1] (log)
PSM99Q	index of sensitive materials prices (1990 = 100)(bci–99a) [1959:1] (log)
PUNEW	cpi-u: all items (82–84 = 100, sa) [1959:1] (log)
PU81	cpi-u: food & beverages (82–84 = 100, sa) [1967:1] (log)
PUH	cpi-u: housing (82–84 = 100, sa) [1967:1] (log)
PU83	cpi-u: apparel & upkeep (82–84 = 100, sa) [1959:1] (log)
PU84	cpi-u: transportation (82–84 = 100, sa) [1959:1] (log)
PU85	cpi-u: medical care (82–84 = 100, sa) [1959:1] (log)
PUC	cpi-u: commodities (82–84 = 100, sa) [1959:1] (log)
PUCD	cpi-u: durables (82–84 = 100, sa) [1959:1] (log)
PUS	cpi-u: services (82–84 = 100, sa) [1959:1] (log)
PUXF	cpi-u: all items less food (82–84 = 100, sa) [1959:1] (log)
PUXHS	cpi-u: all items less shelter (82–84 = 100, sa) [1959:1] (log)
PUXM	cpi-u: all items less medical care (82–84 = 100, sa) [1959:1] (log)
PSCCOM	spot market price index:bls & crb: all commodities (67 = 100, nsa) [1959:1] (log)
PSCFOO	spot market price index:bls & crb: foodstuffs (67 = 100, nsa) [1959:1] (log)
PSCMAT	spot market price index:bls & crb: raw industrials (67 = 100, nsa) [1959:1] (log)
PZFR	prices received by farmers: all farm products (1977 = 100, nsa) [1975:1] (log)
PCGOLD	commodities price:gold, london noon fix, avg of daily rate, $ per oz [1975:1] (log)
GMDC	pce, impl pr defl: pce (1987 = 100) [1959:1] (log)
GMDCD	pce, impl pr defl: pce; durables (1987 = 100) [1959:1] (log)
GMDCN	pce, impl pr defl: pce; nondurables (1987 = 100) [1959:1] (log)
GMDCS	pce, impl pr defl: pce; services (1987 = 100) [1959:1] (log)
GMPYQ	personal income (chained) (series #52) (bil 92$, saar) [1959:1] (log)
GMYXPQ	personal income less transfer payments (chained) (#51) (bil 92$, saar) [1959:1] (log)
GMCQ	personal consumption expend (chained): total (bil 92$, saar) [1959:1] (log)
GMCDQ	personal consumption expend (chained): total durables (bil 92$, saar) [1959:1] (log)
GMCNQ	personal consumption expend (chained): nondurables (bil 92$, saar) [1959:1] (log)
GMCSQ	personal consumption expend (chained): services (bil 92$, saar) [1959:1] (log)
GMCANQ	personal cons expend (chained): new cars (bil 92$, saar) (log)

Notes

1. It should be emphasized that, like the experiment reported in this paper, these studies are simulated real-time exercises, not a comparison of true real-time forecasts. True real-time forecasts are based on preliminary data and often contain significant judgmental adjustments; see for example McNees (1986, 1990) and the surveys in Granger and Newbold (1977, ch. 8.4 and 1986, ch. 9.4). Although true out-of-sample MSEs would differ from those reported here, the simulated real-time nature of this experiment provides a controlled environment for comparing and ranking different forecasting methods.
2. A fixed lag length of six was used to compute the unit root test statistics. The unit root pretests were computed and applied recursively, that is, the forecast of y_{t+h} using data through time t were computed using the model selected at time t by the unit root pretest computed using data through time t. The critical values for the unit root tests were chosen so that the pretest constituted a consistent rule for selecting between the I(0) and I(1) specification. Specifically, for the DF-GLS$^\mu$ test, the critical value was $\ln(120/t) - 1.95$, and for the DF-GLS$^\tau$ test the critical value was $\ln(120/t) - 2.89$. When $t = 120$, these correspond to 5 percent significance level unit root pretests, with lower significance levels as the sample size increases.
3. See Swanson and White (1995, 1997) for discussion of ANN models in economics; for a monograph treatment, see Masters (1994).
4. See Granger and Teräsvirta (1993) for an exposition of the threshold autoregression and smooth transition autoregression family of models, including LSTAR models.
5. Other loss functions are possible, for example, the forecaster might have asymmetric loss, cf. Granger (1969) and Diebold and Christofferson (1997). Under nonquadratic loss, least squares forecasts are not optimal, but considering alternative estimation methods is beyond the scope of this paper.
6. The basic algorithm described above was augmented at each date t by 1,000 randomly selected trial values of the parameters. The four sets of parameters that yielded the smallest SSR together with the optimum at date $t - 1$ were each used as initial values for 250 Gauss–Newton iterations. Finally, the parameter values associated with the resulting smallest SSR was used to construct the forecasts and carried forward as initial condition for date $t + 1$. Experiments indicated that this algorithm yielded essentially identical function values in repeated trials, suggesting that it achieved a global optimum.

References

Bates, J. M. and C. W. J. Granger (1969), "The Combination of Forecasts," *Operations Research Quarterly*, 20: 451–68.

Brooks, C. (1997), "Linear and Non-linear (Non-)Forecastability of High-Frequency Exchange Rates," *Journal of Forecasting*, 16: 125–45.

Campbell, J. Y. and P. Perron (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," *NBER Macroeconomics Annual*, 20.

Chan, L., J. H. Stock, and M. W. Watson (1998), "A Dynamic Factor Model Framework for Forecast Combination," manuscript, Harvard University.

Diebold, F. X. and P. F. Christoffersen (1997), "Optimal Prediction under Asymmetric Loss," *Econometric Theory*, 13: 808–17.

Diebold, F. X. and L. Kilian (1997), "Unit Roots and Forecasting: To Difference or Not To Difference?" manuscript, University of Pennsylvania.

Diebold, F. X. and P. Pauly (1990), "The Use of Prior Information in Forecast Combination," *International Journal of Forecasting*, 6: 503–8.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64: 813–36.

Ghysels, E., C. W. J. Granger, and P. L. Siklos (1996), "Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process?" *Journal of Business and Economic Statistics*, 14: 374–86.

Granger, C. W. J. (1966), "The Typical Spectral Shape of an Economic Variable," *Econometrica*, 34: 150–61.

Granger, C. W. J. (1969), "Prediction with a Generalized Cost of Error Function," *Operational Research Quarterly*, 20: 199–207.

Granger, C. W. J. (1993), "Strategies for Modelling Nonlinear Time Series Relationships," *Economic Record*, 60: 233–8.

Granger, C. W. J. and J.-L. Lin (1994), "Forecasting from Non-Linear Models in Practice," *Journal of Forecasting*, 13: 1–10.

Granger, C. W. J. and P. Newbold (1977), *Forecasting Economic Time Series*. New York: Academic Press.

Granger, C. W. J. and P. Newbold (1986), *Forecasting Economic Time Series*, 2nd edn. New York: Academic Press.

Granger, C. W. J. and R. Ramanathan (1984), "Improved Methods of Combining Forecasting," *Journal of Forecasting*, 3: 197–204.

Granger, C. W. J. and T. Teräsvirta (1993), *Modelling Non-linear Economic Relationships*. Oxford: Oxford University Press.

Granger, C. W. J., T. Teräsvirta, and H. M. Anderson (1993), "Modeling Nonlinearity over the Business Cycle," ch. 8 in J. H. Stock and M. W. Watson (eds.), *Business Cycles, Indicators, and Forecasting*. Chicago: University of Chicago Press for the NBER, pp. 311–27.

McNees, S. K. (1986), "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts," *Journal of Business and Economic Statistics*, 4: 5–24.

McNees, S. K. (1990), "The Role of Judgment in Macroeconomic Forecasting Accuracy," *International Journal of Forecasting*, 6: 287–299.

Makridakis, S., A. Anderson, R. Carbonne, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler (1982), "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *Journal of Forecasting*, 1: 111–53.

Masters, T. (1994), *Signal and Image Processing with Neural Networks*. New York: John Wiley and Sons.

Meese, R. and J. Geweke (1984), "A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomic Time Series," *Journal of Business and Economic Statistics*, 2: 191–200.

Meese, R. and K. Rogoff (1983), "Empirical Exchange Rate Models of the Seventies: Do They Fit out of Sample?" *Journal of International Economics*, 14: 3–24.

Stock, J. H. (1996), "VAR, Error Correction and Pretest Forecasts at Long Horizons," *Oxford Bulletin of Economics and Statistics*, 58/4: 685–701; reprinted in A. Banerjee and D. F. Hendry (eds.), *The Econometrics of Economic Policy*. Oxford: Basil Blackwell, 1997, pp. 115–32.

Swanson, N. R. and H. White (1995), "A Model Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks," *Journal of Business and Economic Statistics*, 13: 265–75.

Swanson, N. R. and H. White (1997), "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks," *Review of Economics and Statistics*, 79: 540–50.

Teräsvirta, T., D. Tjostheim, and C. W. J. Granger (1994), "Aspects of Modelling Nonlinear Time Series," ch. 48 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, vol. 4. Amsterdam: Elsevier, pp. 2919–60.

Tiao, G. C. and D. Xu (1993), "Robustness of Maximum Likelihood Estimates for Multi-Step Predictions: The Exponential Smoothing Case," *Biometrika*, 80: 623–41.

Weigend, A. S. and N. A. Gershenfeld (1994), *Time Series Prediction: Forecasting the Future and Understanding the Past*. Reading, MA: Addison-Wesley for the Santa Fe Institute.