A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks*

Norman R. Swanson
Pennsylvania State University, Department of Economics,
University Park, PA, 16802

and

Halbert White
Research Group for Econometric Analysis,
University of California, San Diego, La Jolla, California 92093

March 1995
this revision: April 1996

ABSTRACT

We take a model selection approach to the question of whether a class of adaptive prediction models ("artificial neural networks") are useful for predicting future values of 9 macroeconomic variables. We use a variety of out-of-sample forecast-based model selection criteria including forecast error measures and forecast direction accuracy. *Ex ante* or real-time forecasting results based on rolling window prediction methods indicate that multivariate adaptive linear vector autoregression models often outperform a variety of: (i) adaptive and nonadaptive univariate models, (ii) nonadaptive multivariate models, (iii) adaptive nonlinear models, and (iv) professionally available survey predictions. Further, model selection based on the in-sample Schwarz Information Criterion apparently fails to offer a convenient shortcut to true out-of-sample performance measures.

JEL Classification: C22, C51, C53

Keywords: Adaptive Modeling, Artificial Neural Networks, Ex-ante Forecasting, Unrevised Data, Rolling Windows, Model Selection, Information Criteria.

Correspondence to: Norman R. Swanson, Department of Economics, Pennsylvania State University, 608 Kern graduate Building, University Park, PA 16802. Email: nswanson@psu.edu

* Many thanks to Dina Nunez-Rocha for excellent research assistance. An early version of this paper was presented at the 1994 International Symposium of Forecasters in Stockholm, Sweden. We are indebted to H.O. Stekler and other conference participants for valuable comments. Also, comments received from two anonymous referees and the editor have been instrumental in preparing this paper. Research was supported by NSF Grant SES #9209023, and by the Research and Graduate Studies Office at Penn State University. The usual disclaimer applies.

1. Introduction and Overview

An issue of continuing interest in the macroeconomics literature is the comparison of information in forecasts from econometric models. Fair and Shiller (1990) examine this issue by comparing forecasts for real GNP growth rates for different pairs of models using a regression of actual values on predicted values from the two models and find that the Fair (1976) model does very well relative to a variety of other models. Fair and Shiller also point out that information beyond *t-h* (where *t* is the time index and *h* is the forecast horizon) may have been used in the revisions of data for periods *t-h* and back, so that their forecasts are not truly *ex ante* (or real-time), in the sense that future information may creep into forecasts of current variables (e.g. through revised seasonal factors or revised benchmark figures). In an analysis of forecasts of industrial production (IP), Diebold and Rudebusch (1991) use a real time data set constructed using both preliminary and partially revised data (henceforth simply called "partially revised" data) on the composite leading index (CLI), which is constructed using data that were available only at *t-h* and back. In the context of linear forecasting models, they find that the performance of partially revised CLI data deteriorates substantially relative to revised data when used to predict IP. A number of other papers also address issues related to real-time forecasting. For example, Trivellato and Rettore (1986) discuss the decomposition of forecast errors into, among other things, the forecast error associated with preliminary data errors. A small sample of other related references includes Boschen (1982), Mariano and Tanizaki (1994), and Patterson (1995).

In this paper, we use a variety of adaptive and nonadaptive univariate and multivariate linear models, as well as a novel class of nonlinear models called "artificial neural networks", to model 9 different macroeconomic variables. Throughout the paper, by an adaptive model we mean that a new specification is chosen before each new rolling forecast is constructed. Our nonadaptive models are also re-estimated at each point in time, but the model specification remains fixed throughout the forecast horizon. We use only *unrevised* or *first reported* data, thus constructing *ex ante* forecasts. One advantage of this strategy is that we can compare our forecasts with forecasts from the Survey of Professional Forecasters (SPF: see Croushore (1993)), which are necessarily also made in real time. Most previous comparisons of professionally made forecasts with econometric models differ from ours in at least two respects: (i) Econometric models are generally constructed using "fully revised" data, which are available at the time that the models are constructed, so that forecasts are not *ex ante*. (ii) Many forecasting models are nonadaptive and linear. By estimating a class of adaptive linear and nonlinear models we are able to answer the following question, "Given an array of nonadaptive linear and adaptive, possibly nonlinear models, is there evidence that adaptive methods are useful?" If so, we have direct evidence of the usefulness of adaptive models for forecasting macroeconomic variables in a real-time setting. Swanson and White (1995) find that such models can be useful when the variable of interest is the spot-forward interest rate differential. We extend those results by considering a wider array of variables, financial and other.

We consider a number of different out-of-sample model selection criteria. In particular, for a 45 quarter out-of-sample evaluation period we calculate mean squared error, mean absolute deviation, mean absolute percentage error, and 2×2 contingency tables for directional accuracy, among others. To allow for the possibility that the economy is evolving over time, we use fixed-length rolling windows of 42, 58, and 76 quarters of data to estimate our models, calculating 1-quarter and 1-year ahead forecasts. This allows us to evaluate the stability of our econometric models throughout the sample period 1960 to 1993.

The model selection approach taken here is different from the more traditional hypothesis testing approach. As in Swanson and White (1995), we adopt this approach for a number of reasons: (i) Model selection allows us to focus directly on the issue at hand: out-of-sample forecasting performance. (ii) Model selection does not require the specification of a correct model for its valid application, as does the traditional hypothesis testing approach. (iii) Finally, if properly designed, the probability of selecting the truly best model approaches one as the sample size increases, in contrast to the traditional hypothesis testing approach (see Swanson and White (1995)). However, we note that it can sometimes be difficult to assess the Type I error associated with testing the implicit model selection hypothesis that two models under consideration truly perform equally well based on observed differences in realized model selection criteria. Nevertheless, this is a defect of the same order of magnitude as using a traditional test whose size is known only asymptotically. Further, if there exists a true model among the alternatives, then if the size of a traditional hypothesis test is appropriately made to approach zero as the sample size approaches infinity, model selection and hypothesis testing become comparable. A more detailed comparison of model selection and standard hypothesis testing approaches is given by Teräsvirta and Mellin (1986). Among other things, Teräsvirta and Mellin point out that model selection and hypothesis testing are comparable when the problem is to choose one of two nested models.

By adopting this model selection approach in a real-time forecasting scenario, we believe that we contribute not only to the discussion of the usefulness of econometric models for predicting macroeconomic variables, but also to the methodology of comparing econometric forecasts with professional forecasts. Two of our main findings are that: (1) *ex ante* forecasts based on rolling window prediction methods indicate that multivariate adaptive linear vector autoregression models often outperform a variety of: (i) adaptive and nonadaptive univariate models, (ii) nonadaptive multivariate models, (iii) adaptive nonlinear models, and (iv) professionally available survey predictions; and (2) model selection based on the in-sample Schwarz Information Criterion apparently fails to offer a convenient shortcut to true out-of-sample performance measures.

The rest of the paper is organized as follows. Section 2 discusses the data, while Section 3 outlines the prediction models considered in this study. Section 4 describes our estimation strategies, and outlines the model selection criteria used. Section 5 discusses the results for the statistical performance measures, and Section 6 provides a summary and concluding remarks.

2. The Data

For the period 1960:1 to 1993:3 we have collected unrevised quarterly U.S. data on unemployment, interest rates, industrial production, nominal gross national product, corporate profits, real gross national product, personal consumption expenditures, the change in business inventories, and net exports of goods and services. Table 1 expands on the series definitions. It should be noted that most of the variables are seasonally adjusted. Although the implications of this are not obvious, we nevertheless have ensured that future information due to the application of two-sided seasonal adjustment filters, for example, has not seeped into our data set. In order to collect the data, each monthly issue of the Survey of Current Business from 1960 to 1993 was examined. Each time a "new", or first available, observation for any of the series was reported, we added one more observation to our data set.

As one measure of the usefulness of our real-time econometric forecasts, we compare our forecasts to *median* professional forecasts. The professional forecasts were provided by Dean Croushore, and are collected in the Survey of Professional Forecasters (SPF) data set (Croushore (1993)). In the SPF (formerly known as the American Statistical Association/National Bureau of Economic Research Economic Outlook Survey) a number of professional forecasters from business, Wall Street, and certain universities are surveyed once each quarter. Zarnowitz and Braun (1992) provide a comprehensive study of the SPF. One of their findings is that taking the mean or median provides a consensus forecast with lower average errors than most individual forecasts. It should be pointed out, though, that using only median forecasts rather than the entire panel is a simplification that may lead to testing bias in certain cases. For instance, Keane and Runkle (1990) avoid aggregation bias by using the full panel of the SPF when testing the rationality of price forecasts, and find results different than when only mean forecasts are used.

Any comparison of forecast accuracy depends on the timing of the forecasts. While our comparison of adaptive and non-adaptive models is not affected by this issue, comparisons based on SPF forecasts depend crucially on the timing and availability of the data used to construct the competing forecasts. The real-time feature of our forecasts makes it relatively easy to pinpoint the timing used in our analysis. Generally, SPF surveys are mailed around the beginning of the current quarter, say period t. Responses are requested by shortly after the *middle* of the current quarter and consist of forecasts for periods t, $t+1$, ..., $t+5$. However, since respondents are asked to forecast current quarter values *during* the current quarter, their information sets may contain a large amount of the *same* information that is later used by the government to construct the actual data for period t. As might be expected in such a scenario, current quarter median SPF forecasts are extremely accurate and almost always outperform econometric models based on information available only during the previous quarter. Our approach to this issue is to compare SPF forecasts for $t+1$ and $t+4$ that are made during period t with corresponding econometric model forecasts made using data available at time t. Because SPF forecasts have the added advantage that they use *all* available information, while our econometric models use at most the unrevised and lagged information from at most two other macroeconomic variables, we feel that we will obtain, if anything, a conservative comparison. Also note that our econometric models are at a further disadvantage because we use only first available information; however, at any point in time t, not only are all lagged unrevised observations available, but all revisions (which show up in partially revised data) that have taken place prior to time t are also available. Thus, our unrevised data set is not as rich as the partially revised data available to SPF respondents at time t.

3. The Models

3.1 Linear Models

The linear models specified in this paper are all special cases of the following model:

$$y_{t+h-1} = \alpha_0 + \sum_{i=1}^{K1} \beta_i y_{t-i} + \sum_{i=1}^{K2} \gamma_i x_{t-i} + \sum_{i=1}^{K3} \delta_i z_{t-i} + u_{t+h-1}, \qquad (1)$$

where y_t is one of the nine macroeconomic variables, and h is the horizon of our forecast, in quarters. The independent variables, x_t and z_t, are two other variables chosen from our set of nine macroeconomic variables (see the discussion in Section 4).

In all, 21 versions of (1) are estimated. The first model corresponds to a random walk, where $\alpha_0 = 0$, $\beta_1 = 1$, β_i, $i=2,\ldots,K1 = 0$, and $\gamma_i = \delta_i = 0$, $i=1,\ldots,5$. The next five models are AR(K1) processes, where K1=1,...,5 and $\gamma_i = \delta_i = 0$, for all i. Two-variable VAR(5) models are also considered, where alternately: (i) K1=K2=1,...,5 and $\delta_i = 0$, $i=1,\ldots,K3$, and (ii) K1=K3=1,...,5 and $\gamma_i = 0$, $i=1,\ldots,K2$. The final five models considered are three-variable VAR(5) specifications, where K1=K2=K3=1,...,5. It should be noted, though, that as a new set of regressors is individually specified for each of the nine variables, all specifications based on (1) are actually AR or ARX and not VAR models. However, we call all of our models VARs, for convenience.

In this study, we consider these models as special cases of a fairly broad array of forecasting models, while realizing that various other linear models that we don't examine here are also available. The random walk model is called a "no change" model, while all other models are referred to by the ordered triplet (K1,K2,K3), and are called nonadaptive linear VAR models. We also examine "adaptive linear VAR" models, which have the additional feature that K1, K2, and K3 are selected anew as new observations become available. These models are discussed further in the next section.

3.2 Neural Network Models

We examine a class of flexible models that were first proposed by cognitive scientists. These so-called "artificial neural network" (ANN) models represent an attempt to emulate certain features of the way that the brain processes information (see Rumelhart and McClelland (1986) for further discussion). Because of their flexibility and simplicity, and because of demonstrated successes in a variety of empirical applications (see White (1989) and Kuan and White (1994) for some specifics), ANNs have become the focus of considerable attention as a possible vehicle for forecasting economic variables, and in particular, financial variables. A number of recent applications using financial data are contained in White (1988), Moody and Utans (1991), Dorsey, Johnson and van Boening (1991), and Swanson and White (1995). One possible reason for the success of ANNs is their ability to approximate arbitrary functions of variables arbitrarily well given a sufficiently large number of nonlinear terms and a suitable choice of parameters (see Hornik, Stinchcombe and White (1989, 1990), White (1990), and the references contained therein). In this paper we closely follow the modeling strategies used in Swanson and White (1995).

For present purposes, it suffices to treat these models as a potentially interesting black box, delivering a specific class of nonlinear regression models. In particular, the ANN regression models considered here have the form:

$$f(\tilde{w}, \theta) = \tilde{w}'\kappa + \sum_{j=1}^{q} G(\tilde{w}' \pi_j)\, \lambda_j \qquad (2)$$

where $\tilde{w} = (1, w')'$ is a (r+1)x1 vector of explanatory variables,

$$w = (y_{t-1}, \dots, y_{t-K1}, x_{t-1}, \dots, x_{t-K2}, z_{t-1}, \dots, z_{t-K3})',$$

$\theta = (\kappa', \lambda', \pi')'$, $\lambda = (\lambda_1, \dots, \lambda_q)'$, $\pi = (\pi_1', \dots, \pi_q')'$, q is a given integer, and G is a given nonlinear function, in our case, the logistic cumulative distribution function (c.d.f.) $G(z) = 1/(1+\exp(-z))$.

A network interpretation of (2) which is also given in Swanson and White (1995) is as follows. "Input units" send signals $\tilde{w} = (w_0(=1), w_1, \dots, w_r)$ over "connections" that amplify or attenuate the signals by a factor ("weight") π_{ji}, $i=0, \dots, r$, $j=1, \dots, q$. The signals arriving at "intermediate" or "hidden" units are first summed (resulting in $\tilde{w}'\pi_j$) and then converted to a "hidden unit activation" $G(\tilde{w}'\pi_j)$ by the operation of the "hidden unit activation function", G. The next layer operates similarly, with hidden activations sent over connections to the "output unit." As before, signals are attenuated or amplified by weights λ_j and summed. In addition, signals are sent directly from input to output over connections with weights κ. A nonlinear activation transformation at the output is also possible, but we avoid it here for simplicity. In network terminology, $f(w, \theta)$ is the "network output activation" of a "hidden layer feedforward network" with "inputs" w and "network weights" θ. The parameters π_j are called "input to hidden unit weights," while the parameters λ_j are called "hidden to output unit weights." The parameters κ are called "input to output unit weights."

Our approach is to first apply model (2) to the problem of forecasting y_{t+h-1} using explanatory variables w corresponding to the variables considered in the linear forecasting models described above, and with no "hidden units", $q = 0$ (i.e. we set $\lambda_i = 0$, $i=1,...,5$). This approach results in the specification of our linear adaptive VAR models, which are discussed in the previous section.

We then apply model (2) to the problem of forecasting y_{t+h-1} using explanatory variables w corresponding to the variables considered in the linear forecasting models described above, but with the highest number of "hidden units", $q = 5$ (note that q is chosen by the network, can take any value from 0 to 5, and can vary as we roll through the forecast horizon).

4. Estimation and Model Selection Procedures

4.1 Estimation

The parameters of the nonadaptive linear models are estimated by the method of least squares. The first 5 variables (see Table 1) are in growth rates, while the others are in levels. Although augmented Dickey-Fuller test evidence is mixed, it can be argued that the series are all I(0). However, even if some (or all) of the variables are I(1), the VAR models that we consider can be interpreted as a form of inefficiently estimated vector error correction (VEC) equations, as long as the I(1) variables are cointegrated. As we are interested only in out-of-sample performance and model selection, we do not concern ourselves overly with these and related issues. The variables chosen as predictor variables in each of our regression models were chosen by using a "training" set of data from 1960:1-1982:2 to determine which macroeconomic variables were most closely related, in terms of in-sample fit. One feature of our approach is that we assume the underlying relation between the economic variables considered may be evolving through time. As an approximate means by which we hope to capture this feature, we estimate the parameters of all our models using only a finite *window* of past data rather than all previously available data. In particular, window sizes of 40, 68, and 76 quarters are used for our regressions. Various other studies use "increasing" windows of data rather than the fixed windows used here (see for example Leitch and Tanner (1991), Pesaran and Timmerman (1994a), and Swanson (1995)).

To evaluate the nonadaptive regression models and the various window widths, we generate sequences of out-of-sample 1-quarter and 4-quarter ahead forecast errors by performing the regressions over a given window terminating at observation $t-h$, say, and then computing the error in forecasting the final revised value of y_t for $h=1,4$ using data available at time $t-h$ and the coefficients estimated using data in the window terminating at time $t-h$. (Although it is not clear whether data ever truly stops being revised, we obtained what we refer to as "finally revised" data for our 9 variables from CITIBASE in November of 1995 (see Swanson (1996) for further discussion)). Each time the window rolls forward one period, a new out-of-sample residual is generated, simulating true out-of-sample predictions and prediction errors made in real-time. For our study, the smallest value for $t-h$ corresponds to the second quarter of 1982 for $h=1$ (and the third quarter of 1981 for $h=4$) while the largest corresponds to the second quarter of 1993 for h=1 (and the third quarter of 1992 for $h=4$). We therefore have a sequence of 45 out-of-sample 1-quarter and 1-year ahead forecast errors based on forecasts for the period 1982:3-1993:3 with which to evaluate our models. The start date for this period coincides roughly with the last major shift in Federal Reserve monetary policy in late 1982, and the period includes recessionary as well as expansionary economic phases.

We now turn to the issue of estimating our adaptive linear and nonlinear (ANN) models. As discussed in Swanson and White (1995), when the ANN models are estimated it is inappropriate to simply fit the network parameters with (say) $q = 5$ hidden units by least squares, as the resulting network typically will have more parameters than observations, achieving a perfect or nearly perfect fit in-sample, with disastrous performance out-of-sample. To enforce a parsimonious fit, the ANN models were estimated by a process of forward stepwise (nonlinear) least squares regression, using an in-sample complexity penalized model selection criterion, the SIC (see below), to determine included regressors and the appropriate value for q. Specifically, a forward stepwise linear regression is performed first, with regressors added one at a time until no additional regressor can be added to improve the SIC. The linear regression coefficients are thereafter fixed. Next, a single hidden unit is added (i.e. q is set to 1), and regressors are selected one by one for connection to the first hidden unit, until the SIC can no longer be improved. Then a second hidden unit is added and the process repeated, until five hidden units have been tried, or the SIC for q hidden units exceeds that for $q-1$ hidden units. This ANN model selection procedure is begun anew each time the data window moves forward one period, as required for what we call our "adaptive" procedure. A different set of regressors and a different number of hidden units connected to different inputs may therefore be chosen at each point in time. Thus, in our ANN estimation strategy we allow for the eventuality that no hidden units may be preferable at any given point in time, suggesting that the SIC-best model may be linear for some periods, and perhaps nonlinear for others. We thus simulate a fairly sophisticated real-time ANN forecasting implementation. The adaptive linear models were fit in an analogous way, except that the process was terminated with $q=0$, enforcing linearity in the predictor variables.

4.2 Model Selection

In order to compare the various models, five measures of out-of-sample model/window performance are computed. The first is the *forecast mean squared error* (MSE) of the 45 forecast errors for each model and window, and for each horizon, $h = 1,4$. Using this measure, we can precisely address the question "Which model/window combination performs best in real-time macroeconomic forecasting based on an out-of-sample forecast error comparison?" Thus we have direct and specific evidence of the value of the various forecasting models. However, as pointed out by Leitch and Tanner (1991) as well as Diebold and Mariano (1995), squared (or any other particular) error loss measures may not be closely related to some chosen profit measure, for example. To address this concern, we examine a number of other model selection and performance measures, some of which are closely related to the MSE criterion, and some of which are not.

The second and third measures of forecast performance are closely related to the MSE. We calculate the *mean absolute deviation* (MAD) and the *mean absolute percentage error* (MAPE) of the 45 forecast errors for each model and window, and for each horizon, $h = 1,4$. For further discussion of these and other measures the reader is referred to Stekler (1991). In order to compare the MSE, MAD and MAPE statistics which derive from our adaptive and nonadaptive econometric models with those from the SPF forecasts over the same period, we use the asymptotic loss differential test proposed in Diebold and Mariano (1995). Their test considers a sample path $\{d_t\}_{t=1}^{T}$ of a loss differential series, and they point out that

$$S_1 = \bar{d} / [\, T^{-1} 2\pi \hat{f}_d(0) \,] \sim N(0,1) \;, \tag{3}$$

where $2\pi f_d(0)$ can be easily estimated in the usual way as a two-sided weighted sum of available sample autocovariances. Following Diebold and Mariano's (1995) suggestion, we use a uniform lag window, and assume $(h-1)$ dependence for our h-step ahead forecasts in order to choose the truncation lag, when constructing the statistic in (3). We define our loss differential series as (i) $d_t = \hat{u}^f_{SPF,t} - \hat{u}^f_{ECO,t}$, for the MSE test; (ii) $d_t = |\hat{u}^f_{SPF,t}| - |\hat{u}^f_{ECO,t}|$, for the MAD test; and (iii) $d_t = |(\hat{y}^f_{SPF,t}/y_t) - 1| - |(\hat{y}^f_{ECO,t}/y_t) - 1|$, for the MAPE test, where $\hat{u}^f$ is the prediction error, $\hat{y}^f$ is the predicted value, y_t is the actual value, SPF denotes forecasts from the Survey of Professional Forecasters, ECO denotes forecasts made using our adaptive or nonadaptive econometric models, and the index, t, runs from $t=1$ to 45. Other related tests of forecasting accuracy (e.g. Mizrach (1991) and Granger and Newbold (1986)) are also available, but are not examined here, partly because the Diebold-Mariano test is elegant, easy to construct, and assumes only that the loss differential series is covariance stationary and short memory.

Our fourth measure of forecast performance is how well a given forecasting procedure identifies the *direction* of change in the level of the variable being forecast, regardless of whether the *value* of the change is closely approximated. To examine this aspect of forecast performance, we calculate the "confusion matrix" of the model/window combination. A hypothetical confusion matrix is given as

$$\begin{array}{lcc} & & \text{actual} \\ & & \begin{array}{cc} \textit{up} & \textit{down} \end{array} \\ \text{predicted} & \begin{array}{c} \textit{up} \\ \textit{down} \end{array} & \begin{bmatrix} 23 & 3 \\ 12 & 7 \end{bmatrix} \end{array} \tag{4}$$

The columns in (4) correspond to *actual* moves, up or down, while the rows correspond to *predicted* moves. In this way, the diagonal cells correspond to correct directional predictions, while off-diagonal cells correspond to incorrect predictions. We measure overall performance in terms of the model’s "confusion rate," the sum of the off-diagonal elements, divided by the sum of all elements. As (4) is simply a 2×2 contingency table, the hypothesis that a given model/window combination is of no value in forecasting the direction of spot rate changes can be expressed as the hypothesis of independence between the actual and predicted directions (as discussed in Pesaran and Timmerman (1994b) and Stekler (1994)). Alternatively, a test which is asymptotically equivalent, in our context, has been given by Henriksson and

Merton (1981). We present confusion matrices, confusion rates, HM p-values, standard χ^2-test of independence p-values, and standard ϕ coefficients, where $\phi = \sqrt{\chi^2/T}$ (so that ϕ ranges from 0 when the variables are independent to 1 when the variables are perfectly related). Using confusion matrices should allow us to answer the question "Are the least confused models also the models which we would choose as best based on other out-of-sample forecast performance measures such as the MSE and MAD, in the context of real-time forecasting?" Also, a finding that a model rejects the null hypothesis of independence is direct evidence that the model is useful as a predictor of the sign of change in a particular macroeconomic variable.

As a final out-of-sample model/window performance measure we calculate Theil's U statistics. This statistic is well known, and can be viewed as the root MSE of a forecast divided by the root MSE of a naive no change forecast.

A drawback of the use of out-of-sample based model selection procedures is that they can be quite computationally intensive. Much less demanding procedures that use only in-sample information can be based on a variety of complexity-penalized likelihood measures. Among those most commonly used are the Akaike Information Criterion (AIC) (Akaike 1973, 1974) and the Schwarz Information Criterion (SIC) (Schwarz 1978, Rissanen 1978). These information criteria add a complexity *penalty* to the usual sample log-likelihood, and the model that optimizes this *penalized* log-likelihood is preferred. Because the SIC delivers the most conservative models (i.e. least complex), because more parsimonious models often outperform more complicated models when used to forecast macroeconomic variables, and because the SIC has been found to perform well in selecting forecasting models in other contexts (for example, see Engle and Brown (1986)), we examine its behavior in the present context as a final measure of forecast performance. One related question which we may ask is: "What sort of guide is the in-sample SIC to out-of-sample performance?" This question is of interest, for if the relatively straightforward SIC reliably identifies the model that performs best according to one of our out-of-sample criteria, then we may use SIC as a welcome computational shortcut.

5. The Results For Statistical Performance Measures

To aid in the discussion of the results, a list of the acronyms used is given.

Model 1: SPF	Survey of Professional Forecasters.
Model 2: No Change	Nonadaptive Linear Model: No change or random walk model.
Model 3: Linear VAR	Nonadaptive Linear VAR Model: AR or VAR. (Fixed Specification)
Model 4: Adaptive Linear	Adaptive Linear VAR Model: AR or VAR. (Flexible Specification)
Model 5: Adaptive Nonlinear	Adaptive Artificial Neural Network Model with up to 5 hidden units. (Flexible Specification)
SIC	Schwarz Information Criterion: $SIC = \log s^2 + p(\log n)/n$, n=window size, p=no. of parameters.
MSIC	Mean Schwarz Information Criterion.
MSE	Forecast Mean Squared Error: Average of 45 1 or 4-quarter forecast errors.
MAD	The Mean Absolute Deviation: Average of 45 1 or 4-quarter ahead forecast errors.
MAPE	The Mean Absolute Percentage Error: Average of 45 1 or 4-quarter ahead forecast errors.
CM	2×2 Contingency Table.
CR	Sum of off diagonal elements of the Confusion Matrix divided by the sum of all elements.
HM	p-Value for Henriksson-Merton Market Timing Test.
χ^2	p-Value from χ^2-test of independence.
ϕ	$\phi = \sqrt{\chi^2/T}$, T=45.
TU	Theil's U: The root MSE of the forecast divided by the root MSE of a naive no change forecast.
dep	The dependent variable for a particular forecasting model.
ind	The independent variables(s) (besides lags of the dependent variable) for a particular forecasting model.

The forecasting results for the 9 macroeconomic variables are contained in Tables 2-3. Because of space considerations, we include results for the best adaptive and nonadaptive models and window sizes, as well as results for the no change model for each variable, and for forecast horizons of one-quarter (h=1) and one-year (h=4). As discussed above, the best nonadaptive models are chosen based on a training set of observations from 1960:1-1982:2 (for h=1) and 1960:1-1981:3 (for h=4), while the best adaptive models are chosen based on the in-sample SIC. Complete results for all models and all cases are available upon request from the authors. Table 1 contains abbreviated names of variables which are used throughout. It should be reiterated that in our terminology, adaptive models include both adaptive linear VAR models as well as adaptive nonlinear network models. (Also, for these models, a simpler AR specification may also be "chosen".) For the adaptive models, a new specification is chosen before each forecast is made. This is in contrast to the nonadaptive linear VAR models which retain the same model specification throughout.

A number of fairly clear-cut conclusions emerge. First, it appears that the various models are all variously preferred, based on MSE, MAD, and MAPE measures. Using the Diebold-Mariano test for *h*=1, the SPF forecasts MSE-dominate the nonadaptive linear VAR model for NGNP. Alternately, the adaptive models dominate the SPF model for R, ΔBI, and Net X. Interestingly, comparisons using the MAD and MAPE statistics differ somewhat. Based on either MAD, MAPE, or both, the no change model dominates the SPF for RGNP, IP, R, and U; the nonadaptive linear VAR model dominates the SPF for PCE, Π, and U; and the adaptive models dominate for RGNP, PCE, and U. Thus, based on these three related model selection criteria, the results are mixed. This supports the well known observation that the choice of particular model selection criteria plays an important role in the final specification of forecasting models. Also, model selection varies depending on which macro variable is being forecast. Overall, it seems clear the *all* models have something to offer, depending on the context. With respect to our adaptive network models, forecasts of various series are seen to improve when flexibly adaptive models are fitted to the data. Thus, we have obtained some evidence for the usefulness of forecasting macroeconomic series with such models. Note, however, that explicit nonlinearity plays quite a *minor* role in these improvements; adaptivity without nonlinearity suffices, for the most part. With respect to linear VAR window size, for $h = 1$ the preferred window is 76 quarters, although a window of 58 quarters is preferred for PCE, at both forecast horizons. Because the largest window widths are almost always preferred to shorter ones, we have evidence of relative time stability for the relationships involved.

Results for $h = 4$ are less straightforward. The SPF seems to dominate for ΔBI and U, while the SPF loses for NGNP, RGNP, PCE, and Π. Based on MSE, MAD, and MAPE, however, it is difficult to say more at this stage, since many of the models cannot be distinguished from one another using the Diebold-Mariano test.

In order to examine the MSE, MAD, and MAPE results of Tables 2-3 from a different perspective, two summary measures have been calculated. The first, shown in Table 4 is a simple sum of the number of times that each model dominates all other models based on model selection criteria point estimates and for each of our six model selection statistics. For $h = 1$, the results show that the SPF "wins" based on confusion rate in 3 of 9 cases, based on MAPE in 2 of 9 cases, and based on MAD in 1 of 9 cases, but "loses" in all nine cases based on any of the other model selection criteria. Overall, the adaptive and nonadaptive models seem to dominate about equally, with linear model "winning" somewhat more frequently than nonlinear models. However, as expected, the adaptive models "win" in all 9 cases based in MSIC. A surprising result in Table 4 is for the forecast horizon, $h=4$, where the adaptive nonlinear models dominate all other models combined, based on the confusion rate. Further, the total number of wins (excluding MSIC) for the adaptive models is greater than that for either the SPF, no change, or nonadaptive linear VAR models, suggesting that the adaptive models hold some promise. Finally, note that for $h=4$ the incidence of "wins" for the no change model drops substantially, relative to the $h=1$ case.

The second summary measure is less crude than the first. Overall performance results are compared using the sign test (see for instance Bickel and Doksum (1977)). We consider the following version of the sign test. Assume that we wish to compare the performance of the SPF forecasts to each alternative econometric model. We assume that we have 9 independent pairs (i.e one pair is the MSE for the SPF model and the MSE for the No Change model for a single variable over the entire forecast period). Then we construct the difference between our "control" (the SPF values), and our "treated" model selection criterion (the "treated" model may be the adaptive, nonadaptive linear VAR, or No Change model). The hypothesis of no "treatment effect" corresponds to the assertion that the differences are symmetrically distributed about 0. The sign statistic, S, is simply the number of differences that are positive, and has a binomial distribution, $B(9,½)$, under the null hypothesis. This statistic is somewhat different from our other, *individual variable* model selection criteria, as it measures the *overall* performance of each model relative to the SPF model. Table 5 lists the results of the sign test for the MSE, MAD, and MAPE criteria. For $h=1$, the adaptive and nonadaptive linear VAR models both appear to outperform the SPF model, using a significance level of 5%, based on the MSE criterion. However, at a 10% significance level, the no change model is seen to dominate the SPF for MSE, MAD, and MAPE. Interestingly, for $h=4$, only the adaptive models outperform the SPF model, and this only for the MSE criterion at a 10% level of significance. Thus, while the adaptive and nonadaptive models are overall winners for $h=1$, it is much harder to choose among the competing models at the 1-year forecast horizon, with the adaptive models being marginally preferred.

Not surprisingly, the MSE, MAD, and MAPE-best models are *not* generally the least confused (based on confusion rates as well as on the HM and χ^2 *p*-values), as forecast errors for individual observations can simultaneously be small in magnitude and associated with a prediction of the wrong sign. This is especially likely in prediction of small changes. For all 9 variables, one or more of the models are not confused, based on rejections of the null hypothesis of independence at a 10% level of significance (10% LOS), using either the HM or the χ^2 p-values. Based on the same p-values and a 10% LOS we conclude: (i) the SPF model is not confused for NGNP, RGNP, PCE, Π, and Net X ($h=1$) and for RGNP, PCE, ΔBI, and U ($h=4$); (ii) the nonadaptive linear VAR and no change models are not confused for PCE, Π, IP, and Net X ($h=1$) and for RGNP, PCE, Π, IP, R, and ΔBI ($h=4$); (iii) the adaptive models are not confused for RGNP, PCE, Π, ΔBI, and U ($h=1$), and for PCE, Π, IP, ΔBI, and U ($h=4$). Thus, perhaps not surprisingly, we conclude that the least confused models change on a case by case basis. Interestingly, based on point estimates alone, the least confused models are the adaptive linear and nonlinear models in 5 of 9 cases for $h=1$, and in 5 of 9 cases for $h=4$.

Our final out-of-sample measure is Theil's U (TU) statistic. In passing, we note that based on TU, the no change model beats all other models for RGNP, IP, and R ($h=1$), and for NGNP and R ($h=4$). In all other cases, the evidence is mixed. For instance, adaptive models win in 4 of 9 cases for $h=1$ and 2 of 9 cases for $h=4$. Of course, since Theil's U values are based on each variable's root MSE (which varies by model), and no change root sum squared error (which is constant by model), then Theil's U statistic values yield exactly the same information as the MSE point estimates. Also, note that TU is not equal to unity, even for the no change models. This is because our forecasts (which are constructed using unrevised data) are compared with CITIBASE data, which have been revised.

To examine the behavior of our in-sample statistical performance measure, we consider the relation between the models identified as *best* in Tables 2-3 using the various ex ante model selection criteria with the MSIC-best models. The MSIC-best model is in each case the adaptive artificial neural network model, as should be expected, since these models are the most flexible of the models considered, and are arrived at by minimizing the SIC in each window. However, the nonlinear adaptive models deliver best out-of-sample MSE performance only for Net X ($h=1,4$). Furthermore, the adaptive nonlinear models deliver least confused directional prediction in 3 of 18 cases (all for $h=4$), and correspond to the MAD and MAPE-best models around 8% of the time (when point estimates are compared). This is interesting, as it suggests that (at least in the present macroeconomic forecasting context) the MSIC cannot reliably be used as a shortcut to identifying models that will perform optimally out-of-sample. Also, although computationally straightforward, the use of SIC for choosing ANN models may not be optimal, in the sense that a number of other predictive loss functions are clearly not minimized when SIC is minimized. However, it remains an open question whether or not the MSIC performs better than other in-sample measures, such as a mean R^2 statistic, for example. In network jargon, the MSIC-best model is not necessarily the model that "generalizes" best when presented with data not included in the "training set". Instead, it is necessary to do the appropriate out-of-sample analysis to find the best model, when using adaptive methods.

6. Summary and Concluding Remarks

We have used a model selection approach to compare real-time forecasts of 9 macroeconomic variables using various adaptive and nonadaptive models, linear and potentially nonlinear, and the Survey of Professional Forecasters (SPF) forecasts. We offer the following conclusions. First, even when we constrain our econometric models to include information available only on a real-time basis, our predictions still outperform SPF predictions for many of the variables, based on mean squared forecast error and mean absolute deviation measures. In particular, adaptive multivariate linear (and to a lesser extent nonlinear) models tend to outperform SPF, no change, and nonadaptive univariate and multivariate linear models. It should be noted, though, that the SPF forecasts appear to perform more or less as well as a number of econometric models when comparing predictions of the direction of change in a variable, and when minimizing the mean absolute percentage error model selection criterion. Nevertheless, based on predictions of the direction of change, the multivariate adaptive linear and nonlinear models (when grouped together) dominate all other models combined, providing the least confused forecasts for the majority of variables examined, for forecast horizons of both 1 quarter and 1 year. Overall, our results, which include Diebold-Mariano loss differential tests, χ^2 tests of independence, and sign tests, indicate that model selection should proceed on a case by case basis, with adaptive, nonadaptive, and SPF forecasts alternately dominating depending on which variable is being examined. Second, windows of observations less than the maximal size rarely appear in prediction-best models, suggesting relative stability in the relationships of interest. Third, the in-sample Schwarz Information Criterion does not appear to offer a convenient shortcut to true out-of-sample performance measures for selecting models, or for configuring adaptive neural network models, when forecasting macroeconomic variables. Fourth, the use of unrevised data in real-time forecasting appears to offer a valid guide for comparing real-time professionally available forecasts with econometric predictions. This is contrary to the common practice of building econometric models using the latest fully revised data, which is a mixture of unrevised, partially revised, and fully revised data, and has the feature that future data may have been used (perhaps inadvertantly) to revise earlier data (such as when revised seasonal factors and revised benchmark figures are used). Finally, multivariate adaptive models appear to be promising for use in this context, although we find little evidence that explicit nonlinearity is helpful. Further refinement and application of adaptive methods for modeling macroeconomic variables thus appears warranted, particularly in the context of truly *ex ante* or real-time forecasts.

The work here is meant as a starting point. From both a theoretical and an empirical perspective, a wide variety of further questions present themselves for subsequent research. On the theoretical side, it is of interest to establish the statistical properties of the model selection procedures followed here. On the empirical side, it is of interest to construct more refined prediction models using "partially revised" data rather than the unrevised data used here. Also, issues of timing and data availability when comparing competing predictions from different sources are of interest. Finally, while the in-sample SIC does not provide a convenient shortcut to out-of-sample predictive performance, other in-sample statistics may be more useful, and deserve examination in the current context. These and related issues are left to future work.

Table 1: Variable Definitions and Mnemonics[1]

Variable	Description
NGNP	Gross National Product: SA, $billions, nominal, Quarterly.
RGNP	Gross National Product: SA, $billions, real, various base years, Quarterly.
PCE	Personal Consumption Expenditures: SA, $billions, real, various base years, Quarterly.
Π	Corporate Profits After Taxes: SA, $billions, nominal, Quarterly.
IP	Industrial Production Index: SA, index, various base years, Averaged from monthly.
R	Aaa Corporate Bond Yield: Moody's, %, Averaged from monthly.
ΔBI	Change in Business Inventories: SA, $billions 1987, Quarterly.
Net X	Net Exports of Goods and Services: SA, $billions 1987, Quarterly.
U	Civilian Unemployment Rate: SA, %, Averaged from monthly.

[1] All data are collected from various issues of the Survey of Current Business. SA stands for seasonally adjusted. The full sample is 1960:1-1993:3. All ex-post model selection uses the sample 1982:3-1993:3. Linear and adaptive network models are compared to median forecasts from the Survey of Professional Forecasters (SPF). In 1992:1 participants in the SPF were asked to switch from forecasting GNP to GDP. In order to continue the ex-post sample through 1993:3, GDP median forecasts from the SPF for 1992:1-1993:3 were modified by adding the actual *rest of world* figures as they became available in the Survey of Current Business, so that the GDP forecasts were *roughly* transformed to GNP forecasts (This approximation does not have any notable effect on our estimation results, partly because $\Delta logs$ of the GNP variables are taken.) In the forecasting experiment, $\Delta logs$ are used for the first 5 variables in the table. Of the rest, ΔBI and Net X are re-based to 1987 dollars using a simple calculation based on a comparison of overlapping quarters of data in the two different base years. For all of the real national income and product accounts variables, base year changes occurred in 1965:3 (from 1954 to 1958 dollars), in 1975:4 (from 1958 to 1972 dollars), in 1985:4 (from 1972 to 1982 dollars), and in 1991:3 (from 1982 to 1987 dollars). For a more complete discussion of the data the reader is referred to Swanson (1996).

Table 2: Summary Model Selection Statistics: Forecast Horizon, h=1[1]

$$dep_{t+h-1} = \alpha + \sum_{i=1}^{K1} \beta_i \, dep_{t-i} + \sum_{i=1}^{K2} \delta_i \, ind1_{t-i} + \sum_{i=1}^{K3} \gamma_i \, ind2_{t-i} + u_{t+h-1}$$

Var	Mod	MSIC	MSE	MAD	MAPE	CM	CR	HM	χ^2	ϕ	TU
NGNP	1	--	0.000214	0.011	258.3	22,11,1,11	0.266	0.000	0.002	0.465	2.584
(2,0,2)	2	-8.690	0.000182	0.010	355.4	21,17,2,5	0.422	0.188	0.377	0.132	2.384
[OW=76]	3	-9.019	0.000304@@	0.015	549.1@@	23,22,0,0	0.488	1.000	0.888	0.022	3.081
HU:0%	4	-9.319	0.000181	0.011	449.6	21,20,2,2	0.488	0.679	0.639	0.071	2.376
	5	-9.319	0.000181	0.011	449.6	21,20,2,2	0.488	0.679	0.639	0.071	2.376
RGNP	1	--	0.000096	0.007	249.6	14,7,10,14	0.377	0.083	0.109	0.205	1.522
(3,3,0)	2	-8.605	0.000052	0.005*	138.7	13,9,11,12	0.444	0.323	0.647	0.068	1.118
[OW=76]	3	-8.849	0.000069	0.006	246.9	17,13,7,8	0.444	0.375	0.752	0.047	1.296
HU:9%	4	-9.312	0.000052	0.005	196.3*	17,9,7,12	0.355	0.055	0.112	0.237	1.124
	5	-9.315	0.000076	0.005**	176.1	17,10,7,11	0.377	0.100	0.200	0.190	1.354
PCE	1	--	0.000112	0.008	433.2	18,9,6,12	0.333	0.028	0.059	0.281	1.367
(2,2,0)	2	-8.366	0.000110	0.008	318.9	10,12,14,9	0.577	0.909	0.462	0.109	1.356
[OW=58]	3	-8.841	0.000065	0.006**	360.7	20,10,4,11	0.311	0.012	0.027	0.330	1.040
HU:7%	4	-9.267	0.000063	0.005**	306.1	18,7,6,14	0.288	0.005	0.012	0.373	1.027
	5	-9.270	0.000065	0.005**	282.9	18,7,6,14	0.288	0.005	0.012	0.373	1.043
Π	1	--	0.007867	0.074	415.2	14,9,8,14	0.377	0.088	0.180	0.200	0.821
(1,0,0)	2	-4.282	0.012665	0.087	329.7	16,7,6,16	0.288	0.005	0.011	0.378	1.042
[OW=76]	3	-4.871	0.005423	0.057**	140.9*	18,6,4,17	0.222	0.000	0.001	0.513	0.682
HU:0%	4	-5.076	0.006228	0.063	181.7	18,8,4,15	0.266	0.001	0.004	0.431	0.731
	5	-5.076	0.006228	0.063	181.7	18,8,4,15	0.266	0.001	0.004	0.431	0.731
IP	1	--	0.000227	0.012	203.6	14,14,8,9	0.488	0.546	0.920	0.017	1.386
(3,3,3)	2	-7.467	0.000156	0.009**	179.9**	14,11,8,12	0.422	0.221	0.446	0.114	1.149
[OW=76]	3	-7.708	0.000226	0.012	295.2	16,10,6,13	0.355	0.045	0.093	0.251	1.383
HU:0%	4	-8.217	0.000218	0.011	236.3	16,12,6,11	0.400	0.132	0.266	0.166	1.358
	5	-8.217	0.000218	0.011	236.3	16,12,6,11	0.400	0.132	0.266	0.166	1.358
R	1	--	0.483410	0.517	5.096	8,20,4,12	0.545	0.544	1.000	0.014	1.304
(3,3,3)	2	-1.267	0.284057**	0.395**	3.918**	5,13,3,11	0.500	0.503	1.000	0.000	0.999
[OW=76]	3	-1.064	0.414741**	0.494	4.844	5,14,8,18	0.488	0.742	1.000	0.001	1.208
HU:0%	4	-1.373	0.418057**	0.491	4.953	7,19,6,13	0.555	0.750	1.000	0.001	1.213
	5	-1.373	0.418057**	0.491	4.953	7,19,6,13	0.555	0.750	1.000	0.001	1.213
ΔBI	1	--	463.8686	15.33	89.50	17,9,8,11	0.377	0.105	0.213	0.186	0.967
(2,0,2)	2	6.272	805.3925	24.15@@	215.8@@	13,12,12,8	0.533	0.798	0.823	0.035	1.275
[OW=76]	3	5.742	421.6172	16.66	176.5	15,8,10,12	0.400	0.150	0.303	0.154	0.922
HU:0%	4	5.696	409.4494**	15.79	168.3	16,8,9,12	0.377	0.096	0.194	0.194	0.909
	5	5.696	409.4494**	15.79	168.3	16,8,9,12	0.377	0.096	0.194	0.194	0.909
Net X	1	--	3497.731	43.00	80.49	7,20,12,6	0.711	0.998	0.016	0.358	4.602
(RWD)	2	5.542	3126.692**	41.99	76.20	5,18,14,8	0.711	0.999	0.011	0.379	4.351
[OW=76]	3	5.534	3154.477**	42.41	76.89	4,17,15,9	0.711	0.999	0.008	0.393	4.371
HU:2%	4	5.461	3126.848**	43.59	76.32	5,18,14,8	0.711	0.999	0.011	0.379	4.351
	5	5.460	3123.217**	43.57	76.30	5,18,14,8	0.711	0.999	0.011	0.379	4.349
U	1	--	0.133551	0.292	4.169	7,13,6,10	0.527	0.693	0.863	0.032	1.173
(1,1,1)	2	-1.531	0.101581	0.238**	3.274**	6,12,6,10	0.529	0.730	0.920	0.018	1.023
[OW=76]	3	-1.822	0.087804	0.236**	3.402**	11,13,2,12	0.394	0.049	0.105	0.263	0.951
HU:0%	4	-2.258	0.097198	0.240**	3.336**	10,10,3,15	0.342	0.033	0.069	0.295	1.000
	5	-2.258	0.097198	0.240**	3.336**	10,10,3,15	0.342	0.033	0.069	0.295	1.000

[1] The equation shown above is the general specification for the linear models, where K1,K2,K3 can take values from 1 to 5. Model 1 is the SPF, Model 2 is the no change model, Model 3 is the nonadaptive linear VAR model (The selected lag structure is shown underneath the variable name in the first column of the table, with notation (K1,K2,K3), and the optimal window (OW) chosen for the nonadaptive linear VAR model is given in square brackets underneath the lag specification. RWD refers to the Random Walk with Drift Model.), Model 4 is the adaptive linear model, and Model 5 is the adaptive nonlinear (artificial neural network) model. (The percentage of forecasts which are made using specifications with hidden units is given underneath the variable name in the first column of the table, with notation HU: %.) The model selection statistic acronyms which are used are listed above. All statistics are calculated using the *true* ex-post observation period from 1982:3-1993:3. The 2x2 confusion matrices reported in the CM column of the table have diagonal cells (a11 and a22) corresponding to correct directional predictions, while off-diagonal cells (a12 and a21) correspond to incorrect predictions. The matrix is reported as a vector in the following order: a11, a12, a21, a22. The HM (Henriksson and Merton (1981)) and χ^2 confusion matrix tests of independence *p*-values are based on the null hypothesis that a given model is of no value in predicting the direction of change in the dependent variable. The Yates correction is applied to the χ^2 calculations. The Diebold-Mariano predictive accuracy test is applied to the MSE, MAD, and MAPE statistics (see above), where * and ** denote that the econometric models (Models 2-5) outperform the SPF (Model 1) at a 5 or 10% level respectively. Alternately, when the SPF "wins", the entries are marked with a @ or a @@.

Table 3: Summary Model Selection Statistics: Forecast Horizon, h=4[1]

$$dep_{t+h-1} = \alpha + \sum_{i=1}^{K1} \beta_i \, dep_{t-i} + \sum_{i=1}^{K2} \delta_i \, ind1_{t-i} + \sum_{i=1}^{K3} \gamma_i \, ind2_{t-i} + u_{t+h-1}$$

Var	Mod	MSIC	MSE	MAD	MAPE	CM	CR	HM	χ^2	ϕ	TU
NGNP	1	--	0.000334	0.014	346.3	18,16,4,7	0.444	0.272	0.543	0.090	2.253
(1,0,0)	2	-8.356	0.000191	0.011*	419.2	19,19,3,4	0.488	0.526	1.000	0.009	1.707
[OW=76]	3	-8.923	0.000229**	0.013	558.5	22,22,0,1	0.488	0.511	1.000	0.003	1.866
HU:0%	4	-9.162	0.000242**	0.014	532.9	22,22,0,1	0.488	0.511	1.000	0.003	1.919
	5	-9.162	0.000242**	0.014	532.9	22,22,0,1	0.488	0.511	1.000	0.003	1.919
RGNP	1	--	0.000221	0.010	306.6	15,9,6,15	0.333	0.023	0.043	0.294	1.546
(1,1,1)	2	-8.068	0.000100	0.007*	199.7**	13,10,8,14	0.400	0.145	0.292	0.157	1.042
[OW=76]	3	-8.706	0.000065	0.006*	291.5	18,12,3,12	0.333	0.012	0.027	0.330	0.842
HU:0%	4	-9.065	0.000086	0.006**	301.9	20,12,1,12	0.288	0.000	0.003	0.448	0.965
	5	-9.065	0.000085	0.006**	299.5	19,12,2,12	0.311	0.003	0.009	0.388	0.961
PCE	1	--	0.000147	0.009	277.0	16,10,5,14	0.333	0.020	0.042	0.303	1.388
(4,0,4)	2	-8.436	0.000131	0.009	371.8	8,15,13,9	0.622	0.973	0.182	0.199	1.314
[OW=58]	3	-8.757	0.000090	0.006*	252.2	18,6,3,18	0.200	0.000	0.000	0.562	1.085
HU:0%	4	-9.103	0.000034	0.004**	286.4	16,7,5,17	0.266	0.001	0.004	0.424	0.672
	5	-9.103	0.000034	0.004**	286.4	16,7,5,17	0.266	0.001	0.004	0.424	0.672
Π	1	--	0.013995	0.085	459.9	13,9,9,14	0.400	0.149	0.299	0.155	1.153
(1,1,1)	2	-4.239	0.013012	0.091	432.7	17,7,5,16	0.266	0.001	0.004	0.424	1.112
[OW=76]	3	-4.792	0.006550	0.065	228.4	19,9,3,14	0.266	0.001	0.003	0.441	0.789
HU:0%	4	-5.055	0.006698	0.062*	176.5**	18,6,4,17	0.222	0.000	0.001	0.513	0.797
	5	-5.055	0.006698	0.062*	176.5**	18,6,4,17	0.222	0.000	0.001	0.513	0.797
IP	1	--	0.000958	0.017	233.5	12,9,12,12	0.466	0.429	0.863	0.026	1.629
(1,1,1)	2	-6.731	0.000438	0.016	263.8	12,12,12,9	0.533	0.781	0.863	0.026	1.102
[OW=76]	3	-7.510	0.000257	0.011	212.3	22,8,2,13	0.222	0.000	0.001	0.519	0.845
HU:4%	4	-7.921	0.000430@	0.014	211.2	20,7,4,14	0.244	0.000	0.001	0.463	1.091
	5	-7.922	0.000432	0.014	215.0	20,7,4,14	0.244	0.000	0.001	0.463	1.095
R	1	--	2.142654	1.171	11.87	8,21,5,11	0.577	0.729	1.000	0.012	1.088
(1,1,1)	2	0.565	1.809012	1.052	10.58	7,11,4,8	0.500	0.533	1.000	0.014	1.000
[OW=76]	3	0.425	2.512500	1.239	12.19	2,15,11,17	0.577	0.992	0.102	0.243	1.178
HU:47%	4	-0.070	2.749688	1.421	14.46	5,16,8,16	0.533	0.849	0.718	0.055	1.232
	5	-0.621	2.610467	1.280	12.99	7,16,6,16	0.488	0.538	1.000	0.014	1.201
ΔBI	1	--	751.9552	20.56	173.0	20,16,0,9	0.355	0.002	0.009	0.391	0.728
(5,5,5)	2	7.045	1541.073	29.60@	247.3@	13,11,7,14	0.400	0.135	0.271	0.164	1.042
[OW=76]	3	6.606	1053.116	24.49	191.6	16,11,4,14	0.333	0.014	0.032	0.319	0.861
HU:2%	4	6.173	1007.060	24.19	191.6	17,12,3,13	0.333	0.010	0.024	0.337	0.842
	5	6.173	1001.895	24.00	189.6	18,12,2,13	0.311	0.003	0.008	0.395	0.840
Net X	1	--	6987.120	64.68	110.0	14,24,6,1	0.666	0.998	0.048	0.294	2.342
(1,1,0)	2	7.160	6786.074*	67.77@	116.3	5,20,15,5	0.777	0.999	0.001	0.505	2.308
[OW=76]	3	6.967	7254.314	72.32	122.3	7,23,13,2	0.800	0.999	0.000	0.553	2.386
HU:62%	4	6.529	5565.694	59.59	101.0	11,20,9,5	0.644	0.983	0.141	0.220	2.090
	5	6.438	4927.861	52.56	92.30	16,16,4,9	0.444	0.199	0.399	0.126	1.967
U	1	--	0.967537	0.751	10.33	10,7,7,20	0.318	0.031	0.062	0.281	0.898
(1,1,1)	2	0.517	1.206285@	0.834	11.43	8,14,9,10	0.560	0.848	0.699	0.061	1.002
[OW=76]	3	-0.274	0.710603	0.733	11.18	15,15,2,13	0.377	0.016	0.039	0.307	0.769
HU:22%	4	-0.834	0.726776	0.739	11.16	13,12,4,16	0.355	0.028	0.059	0.281	0.778
	5	-0.844	0.789363	0.746	11.34	13,12,4,16	0.355	0.028	0.059	0.281	0.811

[1] See notes to Table 2.

Table 4: Winners and Losers Among Adaptive and Nonadaptive Models by Selection Criterion[1]

Summary of Results by Number of Wins Using Point Estimates: Tables 2–3

Table 4a: $h=1$

Selection Criterion	*SPF*	*NO CHANGE*	*NONADAPTIVE LINEAR VAR MODELS*	*ADAPTIVE MODELS LINEAR*	*ADAPTIVE MODELS NONLINEAR (ANN)*
MSIC	-	0	0	6	3
MSE	0	3	2	4	0
MAD	1	4	2	1	1
MAPE	2	5	1	0	1
CR	3	1	4	5	2
TU	0	3	2	3	1

Table 4b: $h=4$

Selection Criterion	*SPF*	*NO CHANGE*	*NONADAPTIVE LINEAR VAR MODELS*	*ADAPTIVE MODELS LINEAR*	*ADAPTIVE MODELS NONLINEAR (ANN)*
MSIC	-	0	0	4	5
MSE	1	2	4	1	1
MAD	1	2	3	2	1
MAPE	3	2	1	2	1
CR	2	0	2	1	5
TU	1	2	4	1	1

[1] The table summarizes the winners and losers for all series by forecast horizon (h), and for the out-of-sample model selection criteria as given. SPF stands for Survey of Professional Forecasters. The Adaptive Models columns summarize results from the adaptive linear and adaptive nonlinear (artificial neural network: ANN) models. All statistics are calculated using the *true* ex-post observation period from 1982:3-1993:3. In the case of ties, each model was awarded with a "win", except when the final choice was between the adaptive linear and nonlinear models. In cases where the nonlinear model chooses no hidden units, the ANN model reported on in Tables 2 and 3 is the same as the adaptive linear model, and as such, ties between the two are reported as adaptive linear model "wins". The model selection statistic acronyms which are used are listed above. For example, the MSE is the forecast mean squared error of the 45 out-of-sample, 1-step-ahead (h=1) or 4-step ahead (h=4) forecasts. Similarly, MAD is the mean absolute deviation, and MAPE is the mean absolute percentage error.

Table 5: Overall Performance Results Using the Sign Test[1]

Comparison of SPF With Linear and *Adaptive Models*

Table 5a: *h* =1

Selection Criterion	*NO CHANGE*	*NONADAPTIVE LINEAR VAR MODELS*	*ADAPTIVE MODELS LINEAR*	*ADAPTIVE MODELS NONLINEAR (ANN)*
MSE	7 (0.090)	8 (0.020)	9 (0.002)	9 (0.002)
MAD	7 (0.090)	6 (0.254)	6 (0.254)	6 (0.254)
MAPE	7 (0.090)	6 (0.254)	6 (0.254)	6 (0.254)

Table 5b: *h* =4

Selection Criterion	*NO CHANGE*	*NONADAPTIVE LINEAR VAR MODELS*	*ADAPTIVE MODELS LINEAR*	*ADAPTIVE MODELS NONLINEAR (ANN)*
MSE	7 (0.090)	6 (0.254)	7 (0.090)	7 (0.090)
MAD	5 (0.500)	6 (0.254)	6 (0.254)	6 (0.254)
MAPE	3 (0.254)	4 (0.500)	4 (0.500)	4 (0.500)

[1] See notes to Table 4. The table summarizes the results of sign tests on the MSEs, MADs and MAPEs listed in Tables 2-3 for each of the variables and for forecast horizons: $h=1$ (Table 5a) and $h=4$(Table 5b). Reported statistics are the number of positive differences, $S = \sum_i (MSS_{SPF}(i) - MSS_{ECO}(i))$, where ECO corresponds to the no change, linear, and adaptive linear and adaptive nonlinear artificial neural network models, MSS is the value of the particular model selection statistic being examined, and the index, i, runs from 1 to 9, corresponding to the 9 economic variables. SPF stands for Survey of Professional Forecasters. In this way, the SPF model can be thought of as the "control". Bracketed p-values correspond to the probability of observing the reported number of positive differences between the SPF model selection criterion values and the econometric model selection criterion values (from Tables 2-3), under the null that the differences are symmetrically distributed about 0. A low p-value (when S>4) indicates that the econometric model outperforms the "control" SPF model, across all variables.

References

Akaike, Hirotuku, "Information Theory and an Extension of the Maximum Likelihood Principle," in B. N. Petrov and F. Csaki, eds., *2nd International Symposium on Information Theory* (Budapest: Akademiai Kiado, 1973) 267-281.

Akaike, Hirotuku, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control* AC-19 (1974) 716-23.

Bickel, Peter J. and Kjell A. Doksum, *Mathematical Statistics* (Englewood Cliffs, New Jersey: Prentice Hall, 1977).

Boschen, John F. and Herschel I. Grossman, "Tests of Equilibrium Macroeconomics Using Contemporaneous Monetary Data," *Journal of Monetary Economics* 10 (1982) 309-333.

Croushore, Dean, "Introducing: The Survey of Professional Forecasters," *The Federal Reserve Bank of Philadelphia Business Review* November-December (1993) 3-15.

Diebold, Francis X. and Roberto S. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (1995) 253-263.

Diebold, Frank X. and Glenn D. Rudebusch, "Forecasting Output With the Composite Leading Index: A Real Time Analysis," *Journal of the American Statistical Association* (1991) 603-610.

Dorsey, Robert E., J.D. Johnson, and Mark V. van Boening, "The Use of Artificial Neural Networks for Estimation of Decision Surfaces in First Price Sealed Bid Auctions," in W.W. Cooper and A. Whinston eds., *New Directions in Computational Economics.* (Boston: Kluwer, 1994) 19-40.

Engle, Robert F. and Scott J. Brown, "Model Selection for Forecasting," *Applied Mathematics and Computation* 20 (1986) 313-327.

Fair, Ray C., *A Model of Macroeconomic Activity, Vol. 2, The Empirical Models* (Cambridge MA: Ballinger Publishing, 1976).

Fair, Ray C. and Robert J. Shiller, "Comparing Information in Forecasts from Econometric Models," *American Economic Review* 80 (1990) 375-389.

Granger, Clive W.J., "Strategies for Modeling Nonlinear Time Series Relationships," *The Economic Record* 69 (1993) 233-238.

Granger, Clive W.J. and Paul Newbold, *Forecasting Economic Time Series* (San Diego: Academic Press, 1986).

Granger, Clive W.J. and Timo Teräsvirta, *Modelling Nonlinear Economic Relationships* (New York: Oxford, 1993).

Henriksson, Roy D. and Robert C. Merton, "On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business* 54 (1981) 513-533.

Hornik, Kurt, Max Stinchcombe, and Halbert White, "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks* 2 (1989) 359-366.

Hornik, Kurt, Max Stinchcombe, and Halbert White, "Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks," *Neural Networks* 3 (1990) 551-560.

Keane, Micheal P. and David E. Runkle, "Testing the Rationality of Price Forecasts," *American Economic Review* 80 (1990) 714-735.

Kuan, Chung-Ming and Halbert White, "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews* 13 (1994) 1-91.

Leitch, Gordon and J. Ernest Tanner, "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures," *American Economic Review* 81 (1991) 580-590.

Mariano, Roberto S. and Hisashi Tanizaki, "Prediction of Final Data with Use of Preliminary and/or Revised Data," *Journal of Forecasting* (1994) forthcoming.

Meese, Richard A. and Kenneth Rogoff, "Was it Real? The Exchange Rate - Interest Differential Relation over the Modern Floating-Rate Period," *Journal of Finance* 43 (1988) 933-948.

Mizrach, Bruce, "Forecast Comparison in L_2," mimeo, Department of Finance, Wharton School, University of Pennsylvania (1991).

Moody, John and J. Utans, "Principled Architecture Selection for Neural Networks: Applications to Corporate Bond Rating Predictions," in J.E. Moody, S.J. Hanson and R.P. Lippmann, eds., *Advances in Neural Information Processing Systems* 4 (San Mateo: Morgan Kaufman, 1991) 683-690.

Patterson, K.D., "An Integrated Model of the Data Measurement and Data Generation Processes with an Application to Consumers' Expenditure," *The Economic Journal* 105 (1995) 54-76.

Pesaran, M. Hashem and Allan G. Timmerman, "The Use of Recursive Model Selection Strategies in Forecasting Stock Returns," mimeo (1994a).

Pesaran, M. Hashem and Allan G. Timmerman, "A Generalization of the Non-Parametric Henriksson-Merton Test of Market Timing," *Economic Letters* 44 (1994b) 1-7.

Rissanen, Jorma, "Modeling by Shortest Data Description," *Automatica* 14 (1978) 465-471.

Rumelhart, David E. and James L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, (Cambridge: MIT Press, 1986).

Schwarz, Gideon, "Estimating the Dimension of a Model," *The Annals of Statistics* 6 (1978) 461-464.

Stekler, Herman O., "Macroeconomic Forecast Evaluation Techniques," *International Journal of Forecasting* 7 (1991) 375-384.

Stekler, H.O., "Are Economic Forecasts Valuable?" *Journal of Forecasting* 13 (1994) 495-505.

Swanson, Norman R. and Halbert White, "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 (1995) 265-275.

Swanson, Norman R., "A Rolling Window Analysis of the Marginal Predictive Content of Money for Real Output: New Evidence of the Money-Income Causal Relation," Working Paper # 10-94-2, Department of Economics, Pennsylvania State University (1995).

Swanson, Norman R., "Forecasting Using First Available Versus Fully Revised Economic Time Series Data," forthcoming, *Studies in Nonlinear Dynamics and Econometrics* (1996).

Teräsvirta, Timo and I. Mellin, "Model Selection Criteria and Model Selection Tests in Linear Models," *Scandinavian Journal of Statistics* 13 (1986) 159-171.

Trivellato, Ugo and Enrico Rettore, "Preliminary Data Errors and Their Impact on the Forecast Error of Simultaneous-Equations Models," *Journal of Business and Economic Statistics* 4 (1986) 445-453.

White, Halbert, "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns," in *Proceedings of the IEEE International Conference on Neural Networks, San Diego* (New York: IEEE Press, 1988) I:451-458.

White, Halbert, "Learning in Artificial Neural Networks: A Statistical Perspective," *Neural Computation* 1 (1989) 425-464.

White, Halbert, "Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings," *Neural Networks* 3 (1990) 535-549.

Zarnowitz, Victor and Phillip Braun, "Twenty-Two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance," NBER Working Paper Number 3965 (1992).