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

Norman R. Swanson
The 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

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. In order to compare our predictions to professionally available survey predictions, we implement an *ex ante* (or real-time) forecasting procedure. One dimension of this approach is that we construct a real-time economic data set which has the characteristic that data available at time *t* do not contain any information which has been allowed to "leak" in from future time periods, as often happens with fully revised macroeconomic data. We also investigate the issue of appropriate window sizes for rolling-window-based prediction methods. Results indicate that adaptive models often outperform a variety of nonadaptive models, as well as professional forecasters, when used to predict levels as well as the direction of change in various macroeconomic variables. Further, model selection based on an in-sample Schwarz Information Criterion (SIC) does not appear to be a reliable guide to out-of-sample performance, in the case of the variables considered here. Thus, the in-sample SIC apparently fails to offer a convenient shortcut to true out-of-sample performance measures.

March 1995

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, The Penn State University, University Park, PA 16802.

* Many thanks to Dina Nunez-Rocha for excellent research assistance. An early version of this paper was presented at the 1994 International Symposium of Forecastors in Stockholm, Sweden. We are indebted to H.O. Stekler and other conference participants for valuable comments. Research was supported by NSF Grant SES #9209023, and by the Research and Graduate Studies Office at Penn State University.

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. In a real-time analysis of forecasts of industrial production (IP), Diebold and Rudebusch (1991) use 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.

In this paper, we use linear models as well as a novel class of adaptive models capable of capturing nonlinearity called "artificial neural networks" to model 9 different macroeconomic series. We use only unrevised or "first reported" data, allowing construction of real-time 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 also made in real time. Most previous comparisons of professionally made forecasts with econometric models differ from ours in at least two respects: (1) Econometric models are generally constructed using "fully revised" data, which are available at the time that the models are constructed. This does not guard against future information creeping into the econometric specification, and thus forecasts. Two important examples are revised seasonal factors and revised benchmark figures. (2) Many forecasting models are linear and non-adaptive, in the sense that the same variables and lags are always used. By also estimating a class of neural network models we are able to answer the following question, "Given an array of nonadaptive linear and adaptive, possibly nonlinear, models for forecasting 9 macroeconomic variables, 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, 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: (1) Model selection allows us to focus directly on the issue at hand: out-of-sample forecasting performance; (2) Model selection does not require the specification of a correct model for its valid application, as does the traditional hypothesis testing approach; (3) 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)). To be sure, 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. Also, in certain cases, such as when an in-sample model selection criterion (the Schwarz Information Criterion (SIC)) is used to distinguish correctly specified models, the distinction between the model selection approach and the traditional hypothesis testing approaches is blurred, and the two approaches can be interpreted as different versions of the same technique.

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. One dimension of this contribution is that we construct a real-time economic data-set which has the characteristic that data available at time *t* do not contain any information which has been allowed to "leak" in from future time periods, as often happens with fully revised macroeconomic data. Further, we consider a variety of model selection criteria including the SIC, various forecast error measures, and forecast direction accuracy. Contributions are also attempted in a number of other related interesting directions. Specifically, we examine the usefulness for macroeconomic forecasting of a class of adaptive prediction models called artificial neural networks, and we examine the issue of "stability" by estimating our prediction models using fixed-length rolling windows of data.

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 "first available" (which we call "unrevised") U.S. data for 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. The choice of variables was dictated somewhat by the availability of "professional" forecasts (see below) which we compare with forecasts made using various linear and nonlinear econometric models. The data are all quarterly, and are published monthly in the Survey of Current Business (SCB). In order to collect the data, each monthly issue of the SCB 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.

Our data collection strategy allows us to formulate and estimate econometric models at time period *t*-1, for instance, using only data which were available prior to period *t*. This allows us to guard against future information creeping into our econometric specifications, and thus our forecasts. Specifically, we avoid using data available after period *t-h*, where *h* is the horizon of our forecasts. We thus address the data revision problem pointed out by Fair and Shiller (1990) and are able to construct truly *ex ante* forecasts.

As one measure of the usefulness of our real-time econometric forecasts, we compare our forecasts to professional forecasts of the variables in Table 1. The professional forecasts were provided by Dean Croushore, and are collected in the Survey of Professional Forecasters (SPF) data set. 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. They are asked to provide forecasts for each of the series listed in Table 1 (among others, see Croushore (1993)). We have simplified the available SPF panel by constructing a data set that consists of median 1-quarter and 4-quarter ahead forecasts for the period 1982:3-1993:3.

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. Because forecasts for *t* are collected at period *t-h*, *h*=1,4, the SPF data are also true *ex ante* forecasts. Thus, it is natural to ask whether our real-time forecasts based on adaptive, possibly nonlinear, econometric models compare favorably to forecasts based on non-adaptive linear econometric models as well to SPF forecasts over a similar period.

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 in part on the SPF data set 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. (Various other approaches to this problem are also available, such as basing linear econometric forecasts on information at period $t-1$, rather than t. However, further analysis of this and other approaches is left to future work.) 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 as we discuss below, 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 ("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.

Diebold and Rudebusch (1991) and Swanson and White (1995) examine real-time econometric forecasting models that use all available data (here called partially revised data) at each given point in time. To illustrate the amount of data collection necessary in order to conduct real-time experiments using partially revised data, it is worth noting that Diebold and Rudebusch (1991) use a partially revised data set for one variable, the composite leading index, for the period 1948:1-1988:11. The data matrix for this single series contains about 90,000 nontrivial entries. We leave for further research use of partially revised data in the current context.

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 $\alpha = 0$, $\beta_1 = 1$, β_i, $i=2,...,K1 = 0$, and $\gamma_i = \delta_i = 0$, $i=1,...,5$. The next five models are AR(K1) processes, where K1=1,...,5 and $\gamma_i = \delta_i = 0$, for all i. VAR(2) models are also considered, where alternately: (i) K1=K2=1,...,5 and $\delta_i = 0$, $i=1,...,K2$, and (ii) K1=K3=1,...,5 and $\gamma_i = 0$, $i=1,...,K3$. The final five models considered are VAR(3) specifications, where K1=K2=K3=1,...,5.

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 LINEAR VAR MODELS in Tables 2-12. Overall, the models differ primarily by the number of lags of the dependent and independent variables, and by the number of independent variables included. As all of the variables are measured in levels, the interpretation of the VAR models which we consider is of some interest. We leave this to Section 4, which examines estimation and model selection procedures.

3.2 Adaptive Models

We examine a class of flexible adaptive 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 (1993) 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). Other references containing relevant discussions about neural network models are Granger (1993) and Granger and Teräsvirta (1993). In this paper we closely follow the modeling strategies used in Swanson and White (1995), and the reader is referred to that paper for further references.

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:

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

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

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

$\theta = (\kappa', \lambda', \pi')'$, $\lambda = (\lambda_1, \ldots, \lambda_q)'$, $\pi = (\pi', \ldots, \pi_q')'$, 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 $w_0 (=1), w_1, \ldots, w_r$ over "connections" that amplify or attenuate the signals by a factor ("weight") π_{ji}, $i=0, \ldots, r$, $j=1, \ldots, q$. 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."

A number of authors including Hornik, Stinchcombe and White (1989, 1990) (see also Cybenko (1989), Caroll and Dickinson (1989) and Funahashi (1989)) have shown that functions of the form (2) are capable of approximating arbitrary functions of w arbitrarily well given q sufficiently large and a suitable choice of θ. This "universal approximation" property is one reason for the successful application of ANNs. White (1990) establishes that ANN models can be used to perform nonparametric regression, consistently estimating any unknown square integrable conditional expectation function.

Our approach is to 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 the number of "hidden units", $q = 5$. Note that when $\lambda_i = 0$, $i = 1,...,5$, we have the linear VAR(3) models as a special case. Using this framework, we attempt to determine whether inclusion of the nonlinear terms, $G(w' \pi_j)$, enhances forecasting ability, assuming that overfitting is properly avoided.

The ANN models we use are *adaptive* in the sense that in each time period and for each forecast horizon the parameters are re-estimated and the number of hidden units (and thus the appropriate degree of nonlinearity) is determined anew using the SIC. Further, which lags of the predictor variables are used is also determined anew from the same set used by the linear models via the SIC. Such models are designated ANN MODEL I in Tables 2-12. We also consider a final adaptive *linear* model, designated ANN MODEL II, in which no hidden units are included (ruling out nonlinearity), but for which a new group of linear regressors is chosen at each point in time, and for each forecast horizen. As in the ANN MODEL I case, in this final case the regressors are also chosen from among the same set of regressors used by the linear models.

4. Estimation and Model Selection Procedures

The parameters of the linear models are estimated by the method of least squares. Because all variables are in levels (and fail to reject the null of I(1) when subjected to augmented Dickey-Fuller tests), the VAR models that we consider can be interpreted as inefficiently estimated vector error correction (VEC) equations, as long as the variables considered are cointegrated. In our real-time scenario, this procedure allows us to avoid re-estimating the rank of any potential cointegration between the variables at each point in time, while still accounting for any cointegration that may be present, albeit in an inefficient way. As we are interested only in out-of-sample performance and model selection, we do not conduct any inference on the regression coefficients from our in-sample estimations, so that even if the distribution of the estimated coefficients were non-standard, our analysis is not affected in any way. Nevertheless, 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-sample, in terms of both cointegrating properties and in-sample fit. As the same set of predictor variables was used for both our adaptive and non-adaptive models, one interpretation of our approach is that we consider a very narrow class of econometric models, thereby increasing the probability that the SPF forecasts (which are based on the entire information set) will outperform our forecasts, all else fixed.

Another feature of our approach is that we assume the underlying relation between the economic variables considered may be evolving through time. Thus, we estimate the parameters of all our models using only a finite *window* of past data rather than all previously available data. 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 Fair and Shiller (1990), Leitch and Tanner (1991) and Peseran and Timmerman (1994a)).

To evaluate the nonadaptive regression models and the various window widths, sequences of out-of-sample 1-quarter and 4-quarter ahead forecast errors are generated by performing the regressions over a given window terminating at observation $t-h$, say, and then computing the error in forecasting 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 $t-h$. 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 by this process. In particular, because we use only unrevised data, we ensure that our forecasts are truly *ex ante* as no *future* information is allowed to creep into our data (see Section 2 for more details). For our study, the smallest value for $t-h$ corresponds to the second quarter of 1983 for $h=1$ (and the third quarter of 1982 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 1983:3-1993:3 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.

By simulating forecasts in real-time, we obtain measures of forecasting performance analogous to those recently discussed by Diebold and Rudebusch (1991). However, our procedure differs from theirs in three respects. First, Diebold and Rudebusch (1991) use a growing data window with fixed first observation, as they are not concerned with tracking a possibly evolving system. Second, Diebold and Rudebusch (1991) focus on the effects of using "partially revised" composite leading indicators (CPI) and "fully" revised industrial production (IP) data in real-time simulations, for predicting economic upturns and downturns. We also focus on macroeconomic data, but use strictly unrevised data, and consider forecasts of 9 different macroeconomic variables. Lastly, we focus on issues related to forecast model selection and evaluation, while Diebold and Rudebusch (1991) focus primarily on prediction, using squared error loss and absolute error loss to determine the usefulness of the CPI for predicting future IP.

We now turn to the issue of estimating our ANN models. As discussed in Swanson and White (1995), when the ANN I 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. 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. For example, an ANN model with no hidden units may be "chosen" based on data available at time $t-1$, while an ANN model with 5 hidden units may be chosen at time t. 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. In Tables 2-12 we report the frequency of SIC-best models which included hidden units in the ANN estimations. We thus simulate a fairly sophisticated real-time ANN forecasting implementation. On average, we should expect the ANN models to have SIC values superior to (i.e smaller than) those of the linear models, as the ANN model can choose any of the linear models as a special case. The ANN II models were fit in an analogous way, except that the process was terminated with $q=0$, enforcing linearity in the predictor variables.

Interestingly, in Swanson and White (1995) even this fairly conservative procedure did not entirely eliminate the tendency for the neural network model to overfit, as evidenced there by occasional totally wild 1-step forecasts of interest rate differentials from network models that fit very nicely in-sample. In forecasting the 9 macroeconomic series, however, we have come across no cases where resulting forecasts are unreasonable, so that the "insanity filter" placed on the network forecasts in Swanson and White (1995) is not used in our current analysis, and we thus have some evidence that the network models have not been overfit.

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.

As pointed out by Leitch and Tanner (1991) as well as Diebold and Mariano (1994), squared (or any other particular) error loss measures may not be closely related to some chosen profit measure. For this reason, Swanson and White (1995) also consider the ability of various models to predict the direction of change in interest rates. They also consider the "profitability" of the competing models when used to trade forward contracts based on interest rate variability. Here 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 (see below).

Using the MSE, we also calculate the out-of-sample forecast R^2, where

$$R^2 = 1 - MSE/S_y^2, \tag{3}$$

and S_y^2 is the sample variance of the dependent variable in the out-of-sample period. Of note is that (3) can take negative values because the MSE can exceed S_y^2, out-of-sample.

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 (1994). Their test considers a sample path $\{d_t\}_{t=1}^{T}$ of a loss differential series, and points out that

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

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 (1994) 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. We define our loss differential series as

$d_t \;=\; \hat{u}^f_{SPF,t} - \hat{u}^f_{ECO,t}$, for the MSE test;

$d_t \;=\; |\,\hat{u}^f_{SPF,t}\,| - |\,\hat{u}^f_{ECO,t}\,|$, for the MAD test; and

$d_t \;=\; |\,(\hat{y}^f_{SPF,t}/y_t) - 1\,| - |\,(\hat{y}^f_{ECO,t}/y_t) - 1\,|$, 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, in accord with our ex ante forecast period. Other similar tests of forecasting accuracy 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. One such test, discussed in Granger and Newbold (1977) makes use of the fact that $\varepsilon_t^1 = \hat{u}^f_{SPF,t} - \hat{u}^f_{ECO,t}$ is contemporaneously uncorrelated with $\varepsilon_t^2 = \hat{u}^f_{SPF,t} + \hat{u}^f_{ECO,t}$ so that the null hypothesis of equal forecast accuracy is the same as zero correlation between ε_t^1 and ε_t^2. Meese and Rogoff (1988) modify the Granger-Newbold test by allowing for serial correlation, while Mizrach (1991) further relaxes a Gaussianity assumption used in the Granger-Newbold and Meese-Rogoff tests.

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

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

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 Peseran and Timmerman (1994b) and Stekler (1994)).

Methods for testing the independence hypothesis in the context of forecasting the direction of asset price movements have been given by Henriksson and Merton (HM, 1981). Based on the hypergeometric distribution, the *p*-values delivered by HM's method require for their validity the independence of the directional forecast from the magnitude of the asset price change. Peseran and Timmerman (1994) show that the test proposed by HM is asymptotically equivalent to the standard χ^2-test of independence in a 2×2 contingency table, when the column and row sums are not *a priori* fixed (as in our case). When the column and row sums are fixed, Peseran and Timmerman (1994) further show that the HM-test of market timing is better interpreted as an exact test of independence within the framework of a 2×2 contingency table. We present confusion matrices, confusion rates, and both the HM *p*-values and the standard χ^2-test of independence *p*-values. As a final measure of independence (noting that the value of our χ^2 statistic is directly proportional to the sample size, T=45 in our case), we also report the standard ϕ coefficient, where $\phi = \sqrt{\chi^2/T}$. For our contingency tables, in which the number of rows and columns are each two, 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.

Our fifth and final in-sample model/window performance measure is Theil's U statistic. 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. The statistic takes the value 0 when the prediction is perfect, and takes the value unity when the MSE of the predicted change equals the MSE of the no change prediction.

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, Sawa 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. Two questions are of interest: First, what sort of guide is the in-sample SIC to out-of-sample performance? Second, are our SIC-selected adaptive ANN models comparable to our nonadaptive linear forecasting models with respect to their performance based on various out-of-sample performance measures? The first 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. The second question concerns the estimation of ANN models. The use of the SIC for choosing ANN models is very straightforward computationally, and if ANN models based on the SIC perform well based on out-of-sample performance measures, then our simple procedure for ANN estimation will be useful for the construction of further macroeconomic forecasting models.

For a model with p parameters estimated on a window of size n, the SIC is

$$\text{SIC} = \log s^2 + p(\log n)/n \ , \tag{6}$$

where s^2 is the regression mean-squared-error. The first term is a goodness of fit measure, and the second is the complexity penalty. We report the *mean* of the 45 values for the SIC, called MSIC, for given model/window combinations.

5. The Results For Statistical Performance Measures

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

SPF	Survey of Professional Forecasters.
NO CHANGE	No change or random walk model.
LINEAR VAR MODEL	Linear model: AR, VAR(2), or VAR(3)
ANN MODEL I	Adaptive Artificial Neural Network Model with up to 5 hidden units.
ANN MODEL II	Adaptive Linear Model.
SIC	Schwarz Information Criterion: $SIC = \log s^2 + p(\log n)/n$, n=window size.
MSIC	Mean Schwarz Information Criterion.
MSE	Forecast Mean Squared Error: Average of 45 1 or 4-quarter forecast errors.
R^2	Out-of-sample measure: $R^2 = 1 - MSE/S_y^2$.
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.
Confusion Matrix	2×2 Contingency Table.
Confusion Rate	Sum of off diagonal elements of the Confusion Matrix divided by the sum of all elements.
HM p-Value	p-Value for Henriksson-Merton Market Timing Test.
χ^2 p-Value	p-Value from χ^2-test of independence.
ϕ Coefficient	$\phi = \sqrt{\chi^2/T}$, T=45.
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 results for the 9 macroeconomic variables are contained in Tables 2-12. 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 1-quarter (h=1) and 1-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 in-sample SIC. Complete results for all models and all cases are available upon request from the authors. Table 1 contains variable acronyms which are used 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 all other models for IP and Net X, and MSE-dominate the no change model for R, NGNP, Π, and RGNP. Alternately, both the adaptive and nonadaptive models MSE-dominate the SPF forecasts for NGNP and PCE. Interestingly, comparisons using the MAD and MAPE statistics differ somewhat. Based on either MAD, MAPE, or both, the nonadaptive linear model dominates the SPF for U, R, IP, and PCE; the no change model dominates the SPF for U, R, PCE, and Net

X; and the adaptive models dominate for U, RGNP, and PCE. 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 window size, for $h=1$ the preferred window is 76 quarters, although windows of 58 quarters are also preferred for a small number of variables. Because the largest window widths are usually preferred to shorter ones, we have evidence of relative time stability for the variables. Results for $h=4$ are less straightforward. The SPF seems to dominate for R and IP, while the SPF loses for PCE. Based on MSE, MAD, and MAPE it is difficult to say more at this stage, since most 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 2-10 from a different perspective, two summary measures have been calculated. The first, shown in Table 11 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 5 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. However, as expected, the adaptive models "win" in all 9 cases based in MSIC. A surprising result in Table 11 is for the forecast horizon, $h=4$, where the adaptive models dominate all other models combined, for all selection criteria. Thus, while the Diebold-Mariano test suggests that there may be little to choose between the models when forecasting 1-year ahead, the point estimates overall suggest that the adaptive models hold some promise.

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, 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 12 lists the results of the sign test for the MSE, MAD, and MAPE criteria. For $h=1$, the adaptive and nonadaptive models both appear to outperform the SFP model, using a significance level of 5%. Interestingly, for $h=4$, only the adaptive models outperform the SFP 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.

Our fourth model selection criterion is the confusion rate. Not surprisingly, the MSE, MAD, and MAPE-best models are *not* generally the least confused (based 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 U, Π, RGNP, PCE, ΔBI, and Net X, 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 least confused for *PCE* ($h=1$)

and Net X (h =1); (ii) the nonadaptive models are least confused for U (h =1) and Π (h =4); (iii) the adaptive models are least confused for U (h =4), RGNP (h =4), and Net X (h =4); (iv) None of the models are confused for ΔBI (h =1,4). Thus, using a 10% LOS, each of the three competitors is seen to win twice, with the exception of the adaptive models which win 3 times. Perhaps not surprisingly, we conclude that while the adaptive model "wins" more often, the least confused models change on a case by case basis. Interestingly, based on point estimates alone, the least confused models are adaptive models in 6 of 9 cases for h =4, and in 3 of 9 cases for h =1.

Our final out-of-sample measure is Theil's U statistic. In passing, we note that based on Theil's U, the no change model beats all other models for R (h =1,4), Π (h =1), and Net X (h =1,4). In all other cases, the evidence is mixed. For instance, adaptive models win in 2 of 9 cases for h =1 and 5 of 9 cases for h =4. Of course, since Theil's U values are based on each variables' 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.

As our in-sample statistical performance measure, we consider the relation between the models identified as *best* in Tables 2-10 using the various ex ante model selection criteria with the MSIC-best models. As should be expected, the MSIC-best model is in each case ANN MODEL I, as these models are arrived at by minimizing the SIC in each window. However, the adaptive models deliver best out-of-sample MSE performance in 2 of 9 cases for h =1 and 5 of 9 cases for h =4 (see Table 11). Furthermore, as mentioned above, the adaptive models deliver least confused directional prediction 50% of the time. The adaptive models also correspond to the MAD and MAPE-best models around 50% 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 always be used as a reliable shortcut to identifying models that will perform optimally out-of-sample. 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. This result is similar to that found by Swanson and White (1995), who predict the spot differential in the interest rate at various forecast horizons.

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, mean absolute forecast error, and mean absolute deviation measures. However, when comparing 1-quarter ahead forecasts, SPF predictions of the direction of change outperform both linear and nonlinear models two thirds of the time. At a 1-year forecast horizon, though, this result is reversed, and adaptive models dominate two-thirds of the time. 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 prediction models 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, and for configuring adaptive 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 using the latest fully revised data, which often uses future data to help revise earlier data (such as when revised seasonal factors and revised benchmark figures are used). Finally, adaptive models appear to be promising for use in this context although we find little evidence that explicit nonlinearity is helpful in the present context. 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
U	Civilian Unemployment Rate: SA, %, Averaged from monthly.
R	Aaa Corporate Bond Yield: Moody's, %, Averaged from monthly.
IP	Industrial Production Index: SA, index, 1987=100, Averaged from monthly.
NGNP	Gross National Product: SA, $billions, Quarterly.
Π	Corporate Profits After Taxes: SA, $billions, Quarterly.
RGNP	Gross National Product: SA, $billions 1987, Quarterly.
PCE	Personal Consumption Expenditures: SA, $billions 1987, Quarterly.
ΔBI	Change in Business Inventories: SA, $billions 1987, Quarterly.
Net X	Net Exports of Goods and Services: SA, $billions 1987, Quarterly.

[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* as they became available in the Survey of Current Business, so that the GDP forecasts were *roughly* transformed to GNP forecasts. All real data are in 1987 dollars, and the IP index has 1987=100. When necessary, data were re-based using a simple calculation based on a comparison of overlapping quarters of data in the two different base years.

Table 2: Unemployment: Best Linear, SPF, and Neural Net Models by Selection Criterion[1]

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

Table 2a: $h=1,\ dep=U,\ ind1=PCE,\ ind2=R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(2,0,2) Window=76	*ANN MODEL I* Hidden Units in 2% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	-1.531	-1.857	-2.225	-2.224
MSE	0.133	0.102	0.072	0.095	0.095
(R^2)	0.931	0.947	0.963	0.951	0.951
MAD	0.287	0.241*	0.210**	0.238	0.238
MAPE	4.083	3.323**	2.999**	3.348*	3.341*
Confusion Matrix	10 14 6 12	-	10 9 7 17	10 11 7 15	10 11 7 15
Conf. Rate	0.476	-	0.372	0.419	0.419
(HM $p-Value$)	0.411	-	0.106	0.228	0.228
$(\chi^2 p-Value)$	0.818	-	0.212	0.455	0.455
ϕ Coefficient	0.035	-	0.191	0.114	0.114
Theil's U	1.142	1.000	0.841	0.963	0.963

Table 2b: $h=4,\ dep=U,\ ind1=PCE,\ ind2=R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(1,1,1) Window=76	*ANN MODEL I* Hidden Units in 29% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	0.517	-0.255	-1.125	-1.091
MSE	0.974	1.206	1.079	0.619	0.615
(R^2)	0.496	0.376	0.442	0.680	0.682
MAD	0.750	0.831	0.873	0.674	0.659
MAPE	10.29	11.39	12.47	9.593	9.510
Confusion Matrix	9 8 7 20	-	5 9 11 19	10 8 6 20	10 8 6 20
Conf. Rate	0.341	-	0.455	0.318	0.318
(HM $p-Value$)	0.068	-	0.650	0.030	0.030
$(\chi^2 p-Value)$	0.136	-	0.783	0.060	0.060
ϕ Coefficient	0.225	-	0.042	0.284	0.284
Theil's U	0.899	1.000	0.946	0.716	0.714

[1] The regression model which is shown is for the linear models. Values of K1, K2, and K3 are used to differentiate between the various specifications. SPF stands for Survey of Professional Forecasters. The Artificial Neural Network (ANN) Model is the flexible non-linear form, which may or may not include hidden units. All statistics are calculated using the *true* ex-post observation period from 1982:3-1993:3. In-sample models are calculated using *immediately* available *windows* of data. Available window sizes are 40, 58, and 76 quarters of data. The R^2 value is calculated as $R^2 = 1 - MSE/S_Y^2$, where MSE is the forecast mean squared error of the 45 out-of-sample, 1(4)-step-ahead forecasts, and S_Y^2 is the sample variance of the dependent variable in the out-of-sample period. Similarly, MAD is the mean absolute deviation, and MAPE is the mean absolute percentage error for the forecast sequence. Theil's U statistic values are also given. The 2x2 confusion matrices reported have diagonal cells corresponding to correct directional predictions, while off-diagonal cells correspond to incorrect predictions. The HM (Henriksson and Merton (1981)) p-values are based on the null hypothesis that a given model is of no value in predicting the direction of changes in the dependent variable. The χ^2 p-values have the same null hypothesis, and correspond to the standard χ^2-test of independence in a 2x2 contingency table. The ϕ coefficient is $(\chi^2/N)^{1/2}$, and The Yates correction is applied to the χ^2 calculations, and $\phi = (\chi^2/N)^{1/2}$, where N=45. Starred entries (**) denote significant difference between SPF and the starred entries at a 95% level of confidence. Similarly, * corresponds to a 90% level of confidence. Also, ** and * indicate "superior" performance of the econometric models, while @@ indicates "superior performance of the SPF model at a 95% level of confidence. The tests used are discussed above.

Table 3: Interest Rates: Best Linear, SPF, and Neural Net Models by Selection Criterion[1]

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

Table 3a: $h=1,\ dep=R,\ ind1=PCE,\ ind2=U$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(1,0,0) Window=76	*ANN MODEL I* Hidden Units in 0% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	-1.268	-1.251	-1.366	-1.366
MSE	0.483	0.284@@	0.321	0.388	0.388
(R^2)	0.829	0.900	0.887	0.863	0.863
MAD	0.518	0.396**	0.430**	0.501	0.501
MAPE	5.078	3.919**	4.131**	4.936	4.936
Confusion Matrix	8 20 4 2	-	13 32 0 0	8 20 5 11	8 20 5 11
Conf. Rate	0.546	-	0.711	0.568	0.568
(HM $p-Value$)	0.544	-	1.000	0.705	0.705
($\chi^2 p-Value$)	0.924	-	0.870	0.877	0.877
ϕ Coefficient	0.015	-	0.025	0.023	0.023
Theil's U	1.305	1.000	1.064	1.169	1.169

Table 3b: $h=4,\ dep=R,\ ind1=PCE,\ ind2=U$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(4,4,4) Window=76	*ANN MODEL I* Hidden Units in 53% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	0.565	0.714	-0.670	-0.105
MSE	2.143	1.809	3.384	3.222	3.128
(R^2)	0.244	0.361	<0	<0	<0
MAD	1.171	1.052	1.518@@	1.416@@	1.438@@
MAPE	11.88	10.58	15.13@@	14.22@@	14.54@@
Confusion Matrix	8 21 5 11	-	7 15 6 17	8 18 5 14	5 17 8 15
Conf. Rate	0.578	-	0.467	0.511	0.556
(HM $p-Value$)	0.729	-	0.462	0.506	0.889
($\chi^2 p-Value$)	0.933	-	0.422	1.000	0.573
ϕ Coefficient	0.013	-	0.115	0.001	0.084
Theil's U	1.088	1.000	1.368	1.335	1.315

[1] See notes to Table 2.

Table 4: Industrial Production: Best Linear, SPF, and Neural Net Models by Selection Criterion[1]

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

Table 4a: $h=1$, $dep=IP$, $ind1=\mathbf{\Delta}BI$, $ind2=Net\ X$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(2,2,0) Window=76	*ANN MODEL I* Hidden Units in 0% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	1.250	0.967	0.677	0.677
MSE	2.855	2.191$^{@@}$	1.544$^{@@}$	2.011$^{@@}$	2.011$^{@@}$
(R^2)	0.968	0.976	0.983	0.978	0.978
MAD	1.358	1.183	0.987**	1.186	1.186
MAPE	1.417	1.223	1.238**	1.172	1.172
Confusion Matrix	28 9 6 2	-	23 6 11 5	19 4 15 7	19 4 15 7
Conf. Rate	0.333	-	0.378	0.422	0.422
(HM $p-Value$)	0.641	-	0.330	0.219	0.219
($\chi^2 p-Value$)	0.679	-	0.670	0.436	0.436
ϕ Coefficient	0.062	-	0.064	0.116	0.116
Theil's U	1.142	1.000	0.840	0.958	0.958

Table 4b: $h=4$, $dep=IP$, $ind1=\mathbf{\Delta}BI$, $ind2=Net\ X$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(4,4,4) Window=76	*ANN MODEL I* Hidden Units in 24% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	3.336	3.132	2.456	2.504
MSE	17.97	22.43$^{@@}$	28.64	57.37	38.41
(R^2)	0.799	0.749	0.680	0.359	0.571
MAD	3.242	3.621	4.112	5.861$^{@@}$	5.124$^{@@}$
MAPE	3.478	3.784	4.284	6.040$^{@@}$	5.307$^{@@}$
Confusion Matrix	38 7 0 0	-	28 6 10 1	23 2 15 5	23 2 15 5
Conf. Rate	0.156	-	0.356	0.378	0.378
(HM $p-Value$)	1.000	-	0.356	0.378	0.378
($\chi^2 p-Value$)	0.679	-	0.840	0.250	0.250
ϕ Coefficient	0.031	-	0.030	0.171	0.171
Theil's U	0.895	1.000	1.130	1.599	1.309

[1] See notes to Table 2.

Table 5: Nominal GNP: Best Linear, SPF, and Neural Net Models by Selection Criterion[1]

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

Table 5a: $h=1$, $dep=NGNP$, $ind1=PCE$, $ind2=R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(4,0,4) Window=76	*ANN MODEL I* Hidden Units in 0% Window=76	*ANN MODEL II* No Hidden Units Window=76
MSIC	-	8.167	6.782	6.543	6.543
MSE	2086	6664$^{@@}$	1789**	1652**	1652**
(R^2)	0.998	0.993	0.998	0.998	0.998
MAD	37.06	74.44$^{@@}$	33.25	30.27	30.27
MAPE	0.814	1.631$^{@@}$	0.762	0.685	0.685
Confusion Matrix	45 0 0 0	-	45 0 0 0	45 0 0 0	45 0 0 0
Conf. Rate	0.000	-	0.000	0.000	0.000
(HM $p-Value$)	1.000	-	1.000	1.000	1.000
($\chi^2 p-Value$)	0.717	-	0.938	0.938	0.938
ϕ Coefficient	0.054	-	0.011	0.011	0.011
Theil's U	0.560	1.000	0.518	0.498	0.498

Table 5b: $h=4$, $dep=NGNP$, $ind1=PCE$, $ind2=R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(4,0,4) Window=76	*ANN MODEL I* Hidden Units in 2% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	10.76	8.541	8.110	8.104
MSE	10588	90232$^{@@}$	12839**	9941	10043
(R^2)	0.989	0.904	0.986	0.989	0.989
MAD	78.15	288.9$^{@@}$	89.73	79.19	80.49
MAPE	1.731	6.305$^{@@}$	2.052	1.766	1.796
Confusion Matrix	45 0 0 0	-	45 0 0 0	45 0 0 0	45 0 0 0
Conf. Rate	0.000	-	0.000	0.000	0.000
(HM $p-Value$)	1.000	-	1.000	1.000	1.000
($\chi^2 p-Value$)	0.938	-	0.938	0.938	0.938
ϕ Coefficient	0.011	-	0.011	0.011	0.011
Theil's U	0.343	1.000	0.377	0.332	0.334

[1] See notes to Table 2.

Table 6: Corporate Profits: Best Linear, SPF, and Neural Net Models by Selection Criterion[1]

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

Table 6a: $h=1$, $dep=\mathbf{\Pi}$, $ind1=\mathbf{\Delta}BI$, $ind2=R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(5,0,5) Window=76	*ANN MODEL I* Hidden Units in 0% Window=76	*ANN MODEL II* No Hidden Units Window=76
MSIC	-	4.446	4.693	4.411	4.411
MSE	170.4	146.6$^{@@}$	177.6	149.1	149.1
(R^2)	0.889	0.905	0.884	0.903	0.903
MAD	9.774	9.373	10.33	9.154	9.154
MAPE	6.229	5.779	6.572	5.712	5.712
Confusion Matrix	21 7 10 7	-	23 9 8 5	20 11 11 3	20 11 11 3
Conf. Rate	0.378	-	0.378	0.489	0.489
(HM $p-Value$)	0.210	-	0.367	0.904	0.904
($\chi^2 p-Value$)	0.421	-	0.746	0.552	0.552
ϕ Coefficient	0.120	-	0.048	0.089	0.089
Theil's U	1.078	1.000	1.100	1.008	1.008

Table 6b: $h=4$, $dep=\mathbf{\Pi}$, $ind1=\mathbf{\Delta}BI$, $ind2=R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(4,0,4) Window=76	*ANN MODEL I* Hidden Units in 2% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	5.683	5.447	5.164	5.165
MSE	650.1	617.0$^{@@}$	488.3	611.3	545.14
(R^2)	0.576	0.600	0.682	0.602	0.645
MAD	20.14	18.24	17.73	20.31	19.30
MAPE	13.28	10.32	10.73	12.20	11.74
Confusion Matrix	21 16 5 2	-	23 11 3 7	19 10 7 8	19 10 7 8
Conf. Rate	0.477	-	0.318	0.386	0.386
(HM $p-Value$)	0.875	-	0.040	0.189	0.189
($\chi^2 p-Value$)	0.760	-	0.078	0.378	0.378
ϕ Coefficient	0.046	-	0.266	0.133	0.133
Theil's U	1.026	1.000	0.890	0.995	0.940

[1] See notes to Table 2.

Table 7: Real GNP: Best Linear, SPF, and Neural Net Models by Selection Criterion[1]

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

Table 7a: $h=1$, $dep=RGNP$, $ind1=PCE$, $ind2=R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(4,0,4) Window=58	*ANN MODEL I* Hidden Units in 20% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	7.831	7.719	7.126	7.139
MSE	1839	$2033^{@@}$	1805	1050	994.1
(R^2)	0.989	0.988	0.990	0.994	0.994
MAD	35.49	38.04	32.27	26.66^{**}	25.98^{**}
MAPE	0.797	0.857	0.734	0.589^{**}	0.576^{**}
Confusion Matrix	32 4 8 1	-	35 5 5 0	35 4 5 1	35 4 5 1
Conf. Rate	0.267	-	0.222	0.200	0.200
(HM $p-Value$)	0.691	-	1.000	0.529	0.529
($\chi^2 p-Value$)	0.553	-	0.933	0.816	0.816
ϕ Coefficient	0.088	-	0.013	0.035	0.035
Theil's U	0.951	1.000	0.942	0.719	0.699

Table 7b: $h=4$, $dep=RGNP$, $ind1=PCE$, $ind2=R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(2,2,0) Window=58	*ANN MODEL I* Hidden Units in 0% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	10.00	8.682	8.464	8.464
MSE	9246	$23248^{@@}$	5914	6379	6379
(R^2)	0.948	0.868	0.966	0.964	0.964
MAD	79.14	$132.5^{@@}$	64.08	64.71	64.71
MAPE	1.786	$2.979^{@@}$	1.435	1.421	1.421
Confusion Matrix	39 6 0 0	-	37 4 2 2	38 4 1 2	38 4 1 2
Conf. Rate	0.133	-	0.133	0.111	0.111
(HM $p-Value$)	1.000	-	0.080	0.043	0.043
($\chi^2 p-Value$)	0.827	-	0.136	0.053	0.053
ϕ Coefficient	0.033	-	0.222	0.288	0.288
Theil's U	0.631	1.000	0.504	0.524	0.524

[1] See notes to Table 2.

Table 8: Consumption: Best Linear, SPF, and Neural Net Models by Selection Criterion[1]

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

Table 8a: $h=1$, $dep=PCE$, $ind1=RGNP$, $ind2=U$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(1,0,1) Window=58	*ANN MODEL I* Hidden Units in 0% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	6.946	6.626	6.453	6.453
MSE	6438	1295**	827.3**	967.8**	967.8**
(R^2)	0.923	0.985	0.990	0.989	0.989
MAD	68.14	30.51**	22.71**	24.82**	24.82**
MAPE	2.246	1.019**	0.776**	0.823**	0.823**
Confusion Matrix	10 0 25 10	-	35 10 0 0	35 10 0 0	35 10 0 0
Conf. Rate	0.556	-	0.222	0.222	0.222
(HM $p-Value$)	0.058	-	1.000	1.000	1.000
($\chi^2 p-Value$)	0.138	-	0.858	0.858	0.858
ϕ Coefficient	0.221	-	0.027	0.027	0.027
Theil's U	2.230	1.000	0.799	0.865	0.865

Table 8b: $h=4$, $dep=PCE$, $ind1=RGNP$, $ind2=U$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(1,1,1) Window=58	*ANN MODEL I* Hidden Units in 4% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	9.063	7.586	7.370	7.386
MSE	9103	3382	2832**	2784**	2819**
(R^2)	0.892	0.877	0.966	0.967	0.966
MAD	84.11	89.86	45.80**	44.30**	44.51**
MAPE	2.800	3.042	1.542**	1.468**	1.474**
Confusion Matrix	22 1 19 3	-	41 4 0 0	41 4 0 0	41 4 0 0
Conf. Rate	0.444	-	0.089	0.089	0.089
(HM $p-Value$)	0.287	-	1.000	1.000	1.000
($\chi^2 p-Value$)	0.569	-	0.793	0.793	0.793
ϕ Coefficient	0.085	-	0.039	0.039	0.039
Theil's U	2.941	1.000	0.525	0.520	0.523

[1] See notes to Table 2.

Table 9: Δ Business Inventories: Best Linear, SPF, and Neural Net Models by Selection Criterion[1]

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

Table 9a: $h=1$, $dep = \boldsymbol{\Delta} BI$, $ind1 = IP$, $ind2 = R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(2,2,0) Window=76	*ANN MODEL I* Hidden Units in 0% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	6.273	5.869	5.742	5.742
MSE	607.7	649.0	486.9	557.3	557.3
(R^2)	0.320	0.276	0.457	0.378	0.378
MAD	19.55	21.24	18.68	19.68	19.68
MAPE	688.7	447.2	322.0	266.1	266.1
Confusion Matrix	18 7 4 16	-	18 9 4 14	15 7 9 14	15 7 9 14
Conf. Rate	0.244	-	0.289	0.356	0.356
(HM $p-Value$)	0.001	-	0.004	0.049	0.049
($\chi^2 p-Value$)	0.002	-	0.009	0.098	0.098
ϕ Coefficient	0.472	-	0.390	0.247	0.247
Theil's U	0.969	1.000	0.866	0.927	0.927

Table 9b: $h=4$, $dep = \boldsymbol{\Delta} BI$, $ind1 = IP$, $ind2 = R$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(1,0,1) Window=76	*ANN MODEL I* Hidden Units in 0% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	7.046	6.491	6.162	6.162
MSE	983.6	1528@@	1009	753.2	753.2
(R^2)	<0	<0	<0	<0	<0
MAD	25.50	29.34	26.12	22.60	22.60
MAPE	1196	976.2	362.3	508.7	508.7
Confusion Matrix	20 8 4 13	-	13 6 11 15	18 6 6 15	18 6 6 15
Conf. Rate	0.267	-	0.378	0.267	0.267
(HM $p-Value$)	0.002	-	0.076	0.002	0.002
($\chi^2 p-Value$)	0.005	-	0.152	0.005	0.005
ϕ Coefficient	0.420	-	0.213	0.420	0.420
Theil's U	0.802	1.000	0.812	0.702	0.702

[1] See notes to Table 2.

Table 10: Net Exports: Best Linear, SPF, and Neural Net Models by Selection Criterion[1]

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

Table 10a: $h=1$, $dep = Net\ X$, $ind1 = \mathbf{\Delta} BI$, $ind2 = IP$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(1,0,0) Window=76	*ANN MODEL I* Hidden Units in 9% Window=76	*ANN MODEL II* No Hidden Units Window=76
MSIC	-	5.543	5.578	5.520	5.526
MSE	749.3	529.3@@	557.6@@	654.7@@	615.1@@
(R^2)	0.839	0.886	0.880	0.860	0.868
MAD	22.42	18.03*	18.73	20.47	19.86
MAPE	91.10	68.81	69.64	72.52	70.51
Confusion Matrix	14 19 1 10	-	6 13 9 16	6 18 9 11	6 17 9 12
Conf. Rate	0.455	-	0.568	0.614	0.591
(HM $p-Value$)	0.044	-	0.733	0.957	0.932
($\chi^2 p-Value$)	0.098	-	1.000	0.393	0.390
ϕ Coefficient	0.249	-	0.002	0.162	0.129
Theil's U	1.190	1.000	1.026	1.112	1.078

Table 10b: $h=4$, $dep = Net\ X$, $ind1 = \mathbf{\Delta} BI$, $ind2 = IP$

Selection Criterion	*SPF*	*NO CHANGE* Order=(1,0,0), $\beta_1=1$ $\beta_1=1$, $\alpha=0$	*LINEAR VAR MODEL* Order=(1,0,0) Window=58	*ANN MODEL I* Hidden Units in 31% Window=Variable	*ANN MODEL II* No Hidden Units Window=Variable
MSIC	-	7.161	7.343	6.667	6.706
MSE	2911	2539@@	3028	3322	2485
(R^2)	0.376	0.456	0.351	0.288	0.467
MAD	44.67	43.79	47.50	41.03	38.70
MAPE	157.1	141.9	124.9	103.8	100.2
Confusion Matrix	16 18 3 8	-	9 10 10 16	16 7 3 19	16 8 3 18
Conf. Rate	0.467	-	0.444	0.222	0.244
(HM $p-Value$)	0.212	-	0.385	0.000	0.000
($\chi^2 p-Value$)	0.420	-	1.000	0.000	0.001
ϕ Coefficient	0.120	-	0.001	0.521	0.484
Theil's U	1.071	1.000	1.092	1.144	0.989

[1] See notes to Table 2.

Table 11: Winners and Losers Among SPF, Nonadaptive Linear, and Adaptive Network Models by Selection Criterion[1]

Summary of Results by Number of Wins Using Point Estimates : *Tables* 2–10

Table 11a: *h* =1

Selection Criterion	*SPF*	*No Change*	*Nonadaptive Linear VAR Models*	*Adaptive Network Models*
MSIC	-	0	0	9
MSE	0	3	4	2
MAD	0	2	4	3
MAPE	0	2	2	5
Con. Rate	6	-	4	3
Theil's U	0	3	4	2

Table 11b: *h* =4

Selection Criterion	*SPF*	*No Change*	*Nonadaptive Linear VAR Models*	*Adaptive Network Models*
MSIC	-	0	0	9
MSE	1	1	2	5
MAD	2	1	2	4
MAPE	2	2	1	4
Con. Rate	3	-	4	6
Theil's U	1	1	2	5

[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 Network Model summarizes results from ANN Model I and ANN Model II in Tables 2-10. 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".

Table 12: Overall Performance Results Using the Sign Test[1]

Comparison of SPF With Linear and *Adaptive Network Models*

Table 12a: $h = 1$

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

Table 12b: $h = 4$

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

[1] The table summarizes the results of sign tests on the MSEs, MADs and MAPEs listed in Tables 2-10 for each of the variables and for forecast horizons: $h = 1$ (Table 12a) and $h = 4$ (Table 12b). Reported statistics are the number of positive differences, $S = \sum_{i=1}^{9} (MSS_{SPF}(i) - MSS_{ECO}(i))$, where ECO corresponds to the no change, linear, and 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 2-10), 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. All statistics are calculated using the *true* ex-post observation period from 1982:3-1993:3. 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. Theil's U statistic "wins" are also tabulated.

References

Akaike, H. (1973), "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, pp. 267-281.

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

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

Carroll, S.M. and Dickinson, B.W. (1989), "Construction of Neural Nets Using the Radon Transform," in *Proceedings of the International Joint Conference on Neural Networks*, Washington DC. New York: IEEE Press, pp. 607-611.

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

Cybenko, G. (1989), "Approximation by Superpositions of a Sigmoid Function," *Mathematics of Control Signals and Systems*, 2, pp. 303-14.

Diebold, F.X. and Mariano, R.S. (1994), "Comparing Predictive Accuracy," NBER Technical Working Paper Number 169.

Diebold, F.X. and Rudebusch, G.D. (1991), "Forecasting Output With the Composite Leading Index: A Real Time Analysis," *Journal of the American Statistical Association*, pp. 603-10.

Dorsey, R.E., Johnson, J.D. and van Boening, M.V. (1994), "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, pp. 19-40.

Engle, R.F. and Brown, S.J. (1986), "Model Selection for Forecasting," *Applied Mathematics and Computation*, 20.

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

Fair, R.C. and Shiller, R.J. (1990), "Comparing Information in Forecasts from Econometric Models," *American Economic Review*, 80, pp. 375-89.

Funahashi, K. (1989), "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks*, 2, pp. 183-92.

Granger, C.W.J. (1993), "Strategies for Modelling Nonlinear Time Series Relationships," *The Economic Record*, 69, pp. 233-8.

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

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

Henriksson, R.D. and Merton, R.C. (1981), "On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business*, 54, pp. 513-33.

Hornik, K., Stinchcombe, M. and White, H. (1989), "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, 2, pp. 359-66.

Hornik, K., Stinchcombe, M. and White, H. (1990), "Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks," *Neural Networks*, 3, pp. 551-60.

Keane, M.P. and Runkle, D.E. (1990), "Testing the Rationality of Price Forecasts," *American Economic Review*, 80, pp. 714-35.

Kuan, C.-M. and White, H. (1993), "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews*, 13, pp. 1-91.

Leitch, G. and Tanner, J.E. (1991), "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures," *American Economic Review*, 81, pp. 580-90.

Meese, R.A. and Rogoff, K. (1988), "Was it Real? The Exchange Rate - Interest Differential Relation over the Modern Floating-Rate Period," *Journal of Finance*, 43, 933-48.

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

Moody, J. and Utans, J. (1991), "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, pp. 683-690.

Peseran, M.H. and Timmerman, A.G. (1994a), "The Use of Recursive Model Selection Strategies in Forecasting Stock Returns," mimeo.

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

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

Sawa, T. (1978), "Information Criteria for Discriminating Among Alternative Regression Models," *Econometrica*, 46, pp. 1273-92.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, pp. 461-4.

Stekler, H.O. (1991), "Macroeconomic Forecast Evaluation Techniques," *International Journal of Forecasting*, 7, pp. 375-84.

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

Swanson, N.R. and White, H. (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*, forthcoming.

White, H. (1988), "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, pp. 451-458.

White, H. (1989), "Learning in Artificial Neural Networks: A Statistical Perspective," *Neural Computation*, 1, pp. 425-64.

White, H. (1990), "Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings," *Neural Networks*, 3, pp. 535-49.

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