Nonlinear Predictability of Stock Returns Using Financial and Economic Variables

Min Qi

Journal of Business & Economic Statistics, Vol. 17, No. 4. (Oct., 1999), pp. 419-429.

Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199910%2917%3A4%3C419%3ANPOSRU%3E2.0.CO%3B2-8

Journal of Business & Economic Statistics is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://uk.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://uk.jstor.org/journals/astata.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://uk.jstor.org/
Mon May 1 09:51:10 2006

Nonlinear Predictability of Stock Returns Using Financial and Economic Variables

Min QI
Department of Economics, College of Business Administration, Kent State University, Kent, OH 44242-0001
(*mqi@bsa3.kent.edu*)

Inspired by the linear predictability and nonlinearity found in the finance literature, this article examines the nonlinear predictability of the excess returns. The relationship between the excess returns and the predicting variables is recursively modeled by a neural-network model, which is capable of performing flexible nonlinear functional approximation. The nonlinear neural-network model is found to have better in-sample fit and out-of-sample forecasts compared to its linear counterpart. Moreover, the switching portfolio based on the recursive neural-network forecasts generates higher profits with lower risks than both the buy-and-hold market portfolio and the switching portfolio based on linear recursive forecasts.

KEY WORDS: Ex ante forecasting; Neural networks; Recursive modeling; Stock-market prediction; Switching portfolio; Trading profits.

Many recent studies find that stock returns can be predicted—for example, the articles by Campbell (1987), French, Schwert, and Stambaugh (1987), Fama and French (1989), Balvers, Cosimano, and McDonald (1990), Breen, Glosten, and Jagannathan (1990), Cochrane (1991), Ferson and Harvey (1993), Glosten, Jagannathan, and Runkle (1993), and Pesaran and Timmermann (1995). Publicly available information, such as financial time series data and macroeconomic variables, can predict a significant portion of stock returns. Despite the difficulty in economic interpretation, the conclusion holds across international stock markets as well as over different time horizons.

Using a recursive linear regression modeling approach, Pesaran and Timmermann (1995) examined the robustness of the evidence on predictability of U.S. stock returns by simulating the decision process of an open-minded investor who, at each point in time, uses only historically available information and a predefined model-selection criterion to select a set of economic factors. The chosen set of variables is then used to make one-period-ahead prediction of excess returns, and the resulting recursive forecasts are employed to make investment decisions. They find that the predictive power of various economic factors over stock returns changes over time and tends to vary with the market volatility.

This research extends that of Pesaran and Timmerman by changing the investor's choice set. Instead of being open-minded in selecting economic factors, the investor is liberal in selecting the functional form through which the chosen economic factors predict stock-market returns. The investor is not confined to linear models; he or she is free to choose from a set of linear and nonlinear models.

Neural networks (NN's) are an ideal choice for flexible nonlinear modeling and are gaining attention in the area of stock-return prediction. As will be reviewed in Section 1, the existing literature focuses primarily on prediction using past returns or technical indicators that are generated from past returns, no fundamentals, or macroeconomic variables. Given the numerous empirical findings that stock returns are linearly predictable using some financial and economic variables, the current research is intended to gauge the usefulness of nonlinear models in stock-return prediction using these financial and economic variables. Many financial series have recently been found essentially nonlinear in nature. [See Abhyankar, Copeland, and Wong (1997) for a summary of published nonlinearity test results.] These findings provide strong motivation for assessing the predictability of stock returns using nonlinear models. The results of this study should shed some light on whether the well-documented nonlinearity could have been used to provide accurate forecasts of the stock-market returns.

The objectives are (a) to compare the goodness of fit of the linear regression (LR) and nonlinear NN models, (b) to compare the predictive performance of various forecasts, and (c) to test the profitability of the switching portfolios based on alternative forecasts.

1. METHODOLOGY

In this research, the investor is assumed to believe that stock returns can be predicted by means of a set of financial and macroeconomic indicators but does not know the "true" underlying specification, let alone the "true" parameter values. In this case, the best the investor can do is to search for a suitable model specification among the set of models that are believed a priori to be capable of predicting stock returns. As time goes on and the historical observations available to the investor increase, the investor updates the forecasting model.

Pesaran and Timmermann (1995) simulated the investor's search for a linear forecasting model by applying standard statistical and financial model-selection criteria to the set of 2^k different linear regression models spanned by all possible permutations of the k factors $\{x_1, x_2, \ldots, x_k\}$ using only information that is publicly available at a particular point in time. The best model is then chosen to make a

© 1999 American Statistical Association
Journal of Business & Economic Statistics
October 1999, Vol. 17, No. 4

one-period-ahead forecast. For their set of nine regressors, this means comparing $2^9 = 512$ models at each point in time and over the period 1959(12) to 1992(11). This gives a total of 202,752 regressions to be computed. Despite all these computations, the model with all nine regressors is viable among the various models based on various selection criteria in terms of market timing and economic significance. This article proposes a change in the dimension of the investor's search—from among combinations of regressors to alternative functional forms. Instead of being confined to a linear model and only allowed to search for the best subset of regressors at each point in time, the investor is free to choose whatever functional form best describes the underlying relationship between the excess returns and the nine financial and economic variables. A similar search procedure as given by Pesaran and Timmermann (1995) was conducted for the NN model at each particular point in time. The selected NN model almost always contains all nine financial and economic variables, and the results are robust no matter which model-selection criterion—such as R^2, adjusted R^2, direction accuracy, mean absolute percentage error, Akaike information criterion, and Bayesian information criterion—is used. Therefore, in this article I only focus on the search of the functional form rather than the choice of the subset of the nine financial and economic variables.

This is accomplished by artificial NN modeling. The benefit of an NN model is that it is a universal approximator, which can approximate any functional form arbitrarily well. An NN allows the investor to recursively approximate the best model specification and use it to make a recursive forecast, without having to know the "true" underlying specification and the "true" parameter values.

The recursive modeling approach is used to update the information set once the new data become available. The estimation and prediction were carried out in an ex ante (or real time) fashion so that data available at time t do not contain information that could have "leaked" in from future time periods. Let X_{t-1} be a vector of financial and economic variables available at time $t-1$ that are used to explain ρ_t, the excess return on the S&P 500 index at time t; a general model of excess returns can be written as

$$\rho_t = f(X_{t-1}, \theta_t) + e_t, \tag{1}$$

where f represents the functional relationship between X_{t-1} and ρ_t, θ_t is a vector of parameters, and e_t is an error term. f can be a linear function, or a set of linear and nonlinear functions, such as a three-layer feedforward NN. Let $\hat{\theta}_t$ be the estimated values of the parameter set θ_t based on sample information; then the conditional forecast for the excess return in time $t+1$ based on X_t is

$$\hat{\rho}_{t+1} = f(X_t, \hat{\theta}_t). \tag{2}$$

The estimation and prediction are carried out recursively once new information becomes available. This recursive procedure allows the investor to adapt to the changing financial and economic conditions.

The present research is intended to gauge the usefulness of alternative linear and nonlinear models in stock-return prediction using financial and economic variables, with particular attention to nonlinear NN's forecast in light of the profuse evidence of financial nonlinearity. The linear and NN models are specified in greater detail as follows.

1.1 The Linear Regression Model

Linear regression is by far the most popular model in studies of stock-return prediction using financial and economic variables. It is easy to estimate and interpret, and the statistical properties of its estimators are readily available for statistical inference and hypothesis testing. With relatively low computational cost, it produces reasonably good forecasts across a diverse set of series. Therefore, the LR model is included in the investor's choice set here. The LR model can be written as

$$\rho_t = \beta_t' X_{t-1} + e_t. \tag{3}$$

Let $\hat{\beta}_t$ be the vectors of regression coefficients estimated from the information available at time t; then the conditional forecasts for period $t+1$ will be

$$\hat{\rho}_{t+1} = \hat{\beta}_t' X_t. \tag{4}$$

1.2 Neural Networks

NN's are a class of generalized nonlinear nonparametric models inspired by studies of the brain and nerve system. The comparative advantage of NN's over more conventional econometric models is that they can approximate any nonlinear (or linear) function to an arbitrary degree of accuracy with a suitable number of hidden units through the composition of a network of relatively simple functions (Hornik, Stinchcombe, and White 1989, 1990; White 1990; White, Gallant, Hornik, Stinchcombe, and Wooldridge 1992). The recent development in NN theory even allows the construction of asymptotically valid prediction intervals (Hwang and Ding 1997). The interested reader is directed to Kuan and White (1994) for a detailed discussion of artificial NN's and their application in economics. NN's are appealing in financial applications in which there exists clear nonlinearity in financial variables yet very few testable models are able to account for the nonlinearity. Some representative studies are those of White (1988), Hutchinson, Lo, and Poggio (1994), Swanson and White (1995, 1997), Gencay (1998, 1999), and Qi and Maddala (1999). Although applications in option pricing and bankruptcy prediction have benefited from NN's, limited success has been achieved for stock-return prediction (Qi 1996).

For IBM daily stock returns, White (1988) found that the NN models wildly overfit in sample, with no ability to forecast out of sample. For monthly New York Stock Exchange stock-index returns, Chuah (1993) found no market timing ability, and the forecast errors of the NN are not significantly different from those of the benchmark linear model. Nevertheless, the network forecasts are able to generate a much larger profit than its linear counterpart and a buy-and-hold strategy. Gencay (1998) and Gencay and Stengos (1998) found strong evidence of predictability of daily Dow Jones Industrial Average Index returns using NN's with the past buy and sell signals of the moving average trading rules.

All of these applications of NN's use only delayed stock returns or technical indicators that are generated from past returns, no fundamentals, or macroeconomic indicators. Given the numerous empirical findings that stock returns are linearly predictable using some financial and economic variables, an NN with these input variables is an ideal choice for investors.

Despite the many desirable features of NN's, constructing a good network for a particular application is not a trivial task. Just like any other nonparametric model, it is subject to underfitting and overfitting. In addition, due to the relatively large number of parameters and the nonlinearity inherent in these specifications, the objective function is unlikely to be globally convex and can have many local minima. Constructing and estimating an NN model often involve the choices of an appropriate architecture (the number of layers, the number of units in each layer, and the connections among units), the selection of transfer functions of the middle and output units, the training algorithm, and the initial weights, and so forth. It has been widely accepted that a three-layer feedforward network with an identity transfer function in the output unit and logistic functions in the middle-layer units can approximate any continuous function arbitrarily well given sufficiently many middle-layer units. Thus, the network used in this research is a three-layer feedforward one. The inputs X (similar to regressors in a linear regression model) are connected to the output ρ (similar to the regressand) via a middle layer. The network model can be written as

$$\begin{aligned}\rho_t &= f(X_{t-1}, \alpha_t, \beta_t) + e_t \\ &= \alpha_{0t} + \sum_{j=1}^{n} \alpha_{jt} \text{logsig}\left(\sum_{i=1}^{k} \beta_{ijt} x_{it-1} + \beta_{0jt}\right) + e_t, \quad (5)\end{aligned}$$

where n is the number of units in the middle layer, k is the number of inputs, logsig is a logistic transfer function $\text{logsig}(a) = 1/(1+\exp(-a))$, α_t represents a vector of the coefficients (weights) from the middle to output layer units, and β_t represents a matrix of the coefficients from the input to middle-layer units at time t.

The initial values of α_t and β_t are generated with the Nguyen–Widrow method, which chooses initial values so that the active regions of the layer's units will be distributed roughly evenly over the range of input space. As a result, fewer units are wasted and training works faster compared to purely random initial values. Technique details are given in Appendix A.

The NN is trained by Levenberg–Marquardt algorithm, the fastest method for training moderate-sized feedforward NN's (up to several hundredweights). To prevent overfitting, Bayesian regularization (MacKay 1992) has been implemented in the training algorithm that provides a measure of how many network parameters are being effectively used by the network no matter how large the total number of parameters in the network becomes. This procedure prevents overfitting and produces networks that generalize well and thus eliminates the guesswork required in determining the optimum network size. See Appendix B for more information on Bayesian regularization. Based on both rule of thumb and experimentation, the number of middle units is set to be 8. The network is recursively estimated in the same fashion as the linear regression. The corresponding one-step-ahead conditional forecasts are

$$\begin{aligned}\hat{\rho}_{t+1} &= f(X_t, \hat{\alpha}_t, \hat{\beta}_t) \\ &= \hat{\alpha}_{0t} + \sum_{j=1}^{n} \hat{\alpha}_{jt} \text{logsig}\left(\sum_{i=1}^{k} \hat{\beta}_{ijt} x_{it} + \hat{\beta}_{0jt}\right). \quad (6)\end{aligned}$$

2. DATA

$X_{t-1} = \{\text{DY}_{t-1}, \text{EP}_{t-1}, \text{I1}_{t-1}, \text{I1}_{t-2}, \text{I12}_{t-1}, \text{I12}_{t-2}, \pi_{t-2}, \Delta\text{IP}_{t-2}, \Delta\text{M}_{t-2}\}$ are the nine financial and economic variables that are used to explain ρ_t, the excess returns on the S&P 500 index at time t. DY is the dividend yield, EP is the earnings-price ratio, I1 is the one-month Treasury-bill rate, I12 is the 12-month Treasury-bond rate, π is the year-on-year rate of inflation, ΔIP is the year-on-year rate of change in industrial output, and ΔM is the year-on-year growth rate of narrow money stock. The macroeconomic indicators such as ΔIP and ΔM are computed using 12-month averages to reduce the impact of historical data revision on results. As noted by an anonymous referee and by Swanson and White (1997), the historically revised data may contain useful information that investors do not have available at the time but that has been allowed to "leak" in from future time periods. It is difficult to weigh the impact of the data-revision problem and the effectiveness of the moving average correction procedure used in the present study. Because our focus is to show that nonlinear models can improve forecast accuracy over its linear counterpart, however, the comparison between the two models should not be largely affected because the same data are used in these two models. Both models are subject to the same bias (if there is any). Because the dividend and earning yields are based on 12-month moving averages, only a one-period lag of these variables was included in the base set. The delay in the publication of macroeconomic indicators means that the most recently available values must be included in the base set with a two-month time lag. To allow for the possibility that changes in interest rates rather than their absolute levels affect stock returns, a two-month, as well as a one-month, lagged value of the interest variables is included. Finally, following the standard practice in the stock-return literature, excess return (ρ_t) is calculated by capital gain plus dividend yield minus the one-month Treasury-bill rate. $\rho_t = [(\text{P}_t - \text{P}_{t-1} + \text{D}_t)/\text{P}_{t-1}] - \text{I1}_{t-1}$, where P_t is the stock price, D_t is dividends, and I1_{t-1} is the return from holding a one-month Treasury bill from the end of month $t-1$ to the end of month t. The share repurchases and takeover distributions considered in the broad definition of dividends (Ackert and Smith 1993) are captured by the capital-gain portion of the stock returns; thus, the broadly defined dividends they suggested are irrelevant in the present study. The dataset has the same definition and sources as those used by Pesaran and Timmermann (1995). Some summary statistics and results of nonlinearity tests are presented in Table 1.

Table 1. Summary Statistics and Nonlinearity Tests of the Data.

	Mean	Std.	Min. Max.	Skewness	Kurtosis	JB test	LB(5)	Tsay	RESET (k = 2)	RESET (k = 3)	BDS (m = 2, ε = σ)	BDS (m = 2, ε = σ/2)
ρ	.0059	.0424	−.2206	−.2895	4.9578	81.28	6.9593	1.7455	9.7644	5.1451	.0910	.447
			.1628			(.00)	(.22)	(.19)	(.00)	(.01)	(.93)	(.65)
DY	.0032	.0007	.0022	.8813	2.8480	61.04	2037.14	.7533	7.9311	7.7711	−3.6018	−2.089
			.0053			(.00)	(.00)	(.39)	(.01)	(.00)	(.00)	(.04)
π	.0356	.0369	−.0161	1.2372	3.8084	132.14	2265.65	1.9237	8.5871	11.6272	19.351	−8.253
			.1439			(.00)	(.00)	(.12)	(.00)	(.00)	(.00)	(.00)
ΔIP	.0325	.0491	−.1011	−.4418	2.6360	17.81	1847.79	3.1875	19.942	13.2802	−2.8235	−4.083
			.1276			(.00)	(.00)	(.01)	(.00)	(.00)	(.00)	(.00)
I1	5.314	2.919	.3550	.9419	4.0327	90.00	1996.10	6.1399	13.268	11.0284	−3.4637	.938
			16.146			(.00)	(.00)	(.01)	(.00)	(.00)	(.00)	(.35)
I12	6.159	3.030	.6330	.6984	3.3417	40.32	2124.20	6.0071	13.084	12.7673	−.1430	4.177
			15.812			(.00)	(.00)	(.01)	(.00)	(.00)	(.87)	(.00)
EP	.0064	.0021	.0032	.9157	2.9422	65.47	2154.24	1.7938	3.9586	8.1465	−8.9579	−5.046
			.0125			(.00)	(.00)	(.18)	(.05)	(.00)	(.00)	(.00)
ΔM	.0555	.0267	.0040	−.5597	2.0251	42.97	2292.72	5.3379	17.270	12.4799	10.213	−3.799
			.0958			(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)

NOTE: Numbers in parentheses are p values. JB test is the Jarque–Bera (1987) test for normality; LB(5) test is the Ljung–Box (1978) test for autocorrelation up to the order of 5; Tsay is Tsay's (1986) test for nonlinearity; RESET is Ramsey's (1969) regression specification error test for nonlinearity; BDS is the test of Brock, Dechert, Scheinkman, and Lebaron (1996). All tests for nonlinearity are performed on the residuals of a prewhitening autoregressive model (order chosen by AIC or BIC).

It is evident from Table 1 that all variables are asymmetrically distributed: Excess return (ρ), change in industrial production (ΔIP), and change in narrow money stock (ΔM) are right skewed, and the rest of the variables are left skewed. Although excess return, inflation rate (π), one-month Treasury-bill rate (I1), and twelve-month Treasury-bond rate (I12) are leptokurtic or slim-tailed, dividend yield (DY), growth rate of industrial production (ΔIP), earnings-price ratio (EP), and money growth rate (ΔM) are platykurtic or fat-tailed. The normality of all variables is strongly rejected by the Jarque-Bera test. Except for the excess returns, all variables exhibit strong autocorrelation according to the Ljung–Box test of order up to 5. As is well known, the significance of the Ljung–Box test could also be due to the autoregressive conditional heteroscedasticity (ARCH) effects, in which case the series are nonlinear in variance. According to Tsay's test for nonlinearity, half of the eight variables show strong evidence of nonlinearity. Ramsey's regression specification error tests (orders 2 and 3) show strong evidence of nonlinearity for all variables. All variables except ρ and I12 show significant nonlinearity by the BDS test with embedding dimension $m = 2$ and the neighborhood parameter $\varepsilon = \sigma$. With $\varepsilon = \sigma/2$ and $m = 2$, the BDS test statistics are highly significant for all variables except ρ and I1. The results confirm the evidence of nonlinearity in returns on the S&P 500 index of different horizons found by other studies summarized by Abhyankar et al. (1997).

3. RESULTS OF RECURSIVE IN-SAMPLE ESTIMATION

The sample period goes from January 1954 to December 1992 and contains a total of 468 observations. The recursive forecast starts in January 1960 and ends in December 1992; therefore, a total of 396 recursive estimation and one-step-ahead forecasts have been made by both the linear and the NN models.

The recursive model estimation and forecasting proceeds as follows: The linear and NN models are estimated with observations over the 1954(1) to 1959(12) period, and the parameter values are used to forecast excess returns for 1960(1). To forecast excess returns for 1960(2), the procedure is repeated for both models using monthly data over the period 1954(1) to 1960(1), and so on. The recursive procedure simulates the search process that the investor could have accomplished in real time to account for the possible structural change.

Five traditional measures are used to compare the fit and the forecasting accuracy of alternative models:

1. RMSE $= \sqrt{1/T \sum_{i=1}^{T} (\rho_i - \hat{\rho}_i)^2}$:

The root mean squared error between the actual and the fitted (in-sample) or predicted (out-of-sample) returns.

2. MAE $= 1/T \sum_{i=1}^{T} |\rho_i - \hat{\rho}_i|$:

The mean absolute error between the actual and the fitted or predicted returns.

3. MAPE $= 1/T \sum_{i=1}^{T} |(\rho_i - \hat{\rho}_i)/\rho_i|$:

The mean absolute percentage error between the actual and fitted or predicted returns.

4.

$$\text{CORR} = \frac{\sum(\rho_i - \bar{\rho})(\hat{\rho} - \bar{\hat{\rho}})}{\sqrt{\sum(\rho_i - \bar{\rho})^2}\sqrt{\sum(\hat{\rho}_i - \bar{\hat{\rho}})^2}} :$$

The Pearson correlation coefficient between the actual and the fitted or predicted returns.

5. Sign $= 1/T \sum z_i$, where

$$z_i = \begin{cases} 1 & \text{if } \rho_{i+1} \cdot \hat{\rho}_{i+1} > 0, \\ 0 & \text{otherwise.} \end{cases} :$$

The proportion of times the sign of excess returns is correctly forecasted.

Table 2. Average In-Sample Model Fit Measures Across 396 Recursive Estimations

Model	RMSE	MAE	MAPE	CORR	Sign
NN	.0344	.0274	1.8242	.4555	.6763
LR	.0360	.0284	1.8622	.3915	.6545

Because a total of 396 recursive estimations and one-period-ahead forecasts have been carried out, it is impossible to report all the details of the results. Table 2 reports the average of each of the five measures across all the recursive estimations by both the LR and NN models. By all measures, the NN model provides better fit to the data than the LR. The RMSE's, MAE's, and MAPE's are obviously smaller and the Pearson correlation coefficient and the percentage of correct signs are larger for the NN than for the LR model.

Some graphic displays of the main results are provided to see how the fit of these two models changes over time. Figure 1 shows the RMSE of different models. The RMSE of the NN model (in solid line) is much more volatile than that of the LR model (in dotted line), especially after the middle 1970s. One interesting observation is that, prior to the middle 1970s, the RMSE of the linear model is actually slightly better than the NN. The superior RMSE of the nonlinear NN model primarily occurs after the middle 1970s. Substantial increases in RMSE appear in 1962, 1974, and after the October 1987 crash. The time plot of the percentage of correct signs has a similar pattern (Fig. 2).

The better fit of the nonlinear NN model indicates that the NN model is able to capture substantial nonlinearity that cannot be fully captured by linear models. The NN in-sample explanatory power almost certainly must be better than that of the LR model, however, because there are many more free parameters in an NN model. The better fit of a neural net may therefore be due to overfit, in which case a model may have superior in-sample fit but mediocre out-of-sample performance.

4. RESULTS OF RECURSIVE OUT-OF-SAMPLE FORECASTS

Figure 3 shows the time plots of the actual excess returns, and the recursively predicted excess returns based on the NN and the linear regression models. The top panel of the figure shows the actual excess returns (in dotted line) and the recursive NN forecasts (solid line). The bottom panel shows the actual excess returns (dotted line) and the recursive linear predictions (solid line). Though both the recursive linear and nonlinear forecasts appear to be less volatile compared to the actual excess returns, they do capture some general patterns of the actual excess returns. All forecasts show relatively high volatility during the early 1980s. This may reflect the changes in the Federal Reserve's operating procedure between 1979 and 1982 that resulted in highly volatile nominal interest rates.

The predictive performance of the NN and the LR models in the whole forecast period and the three subperiods are summarized in Table 3. Panel A reports the overall performance measures in the whole forecast period from 1960(1) to 1992(12). Compared to the linear forecasts, the nonlinear NN forecasts have smaller RMSE, MAE, MAPE,

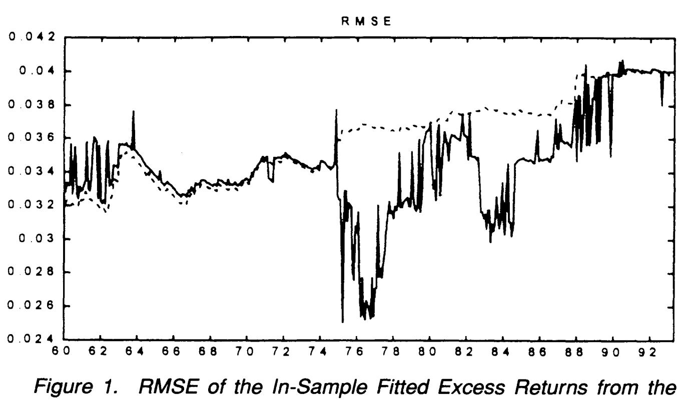

Figure 1. RMSE of the In-Sample Fitted Excess Returns from the Recursive Linear Regression (LR) and Neural Network (NN) Models, 1960(1) to 1992(12): ———, NN; — — —, LR.

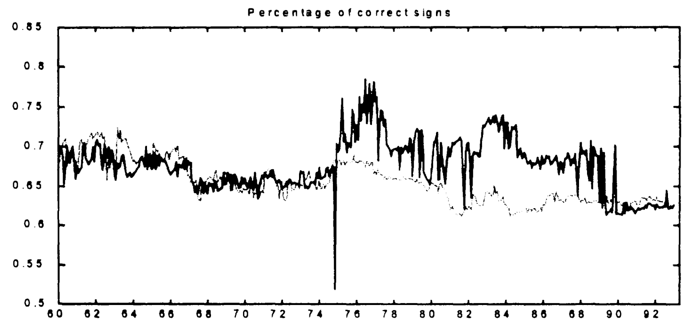

Figure 2. In-Sample Percentage of Correct Signs of the Alternative Recursive Models, 1960(1) to 1992(12): ———, NN; — — —, LR.

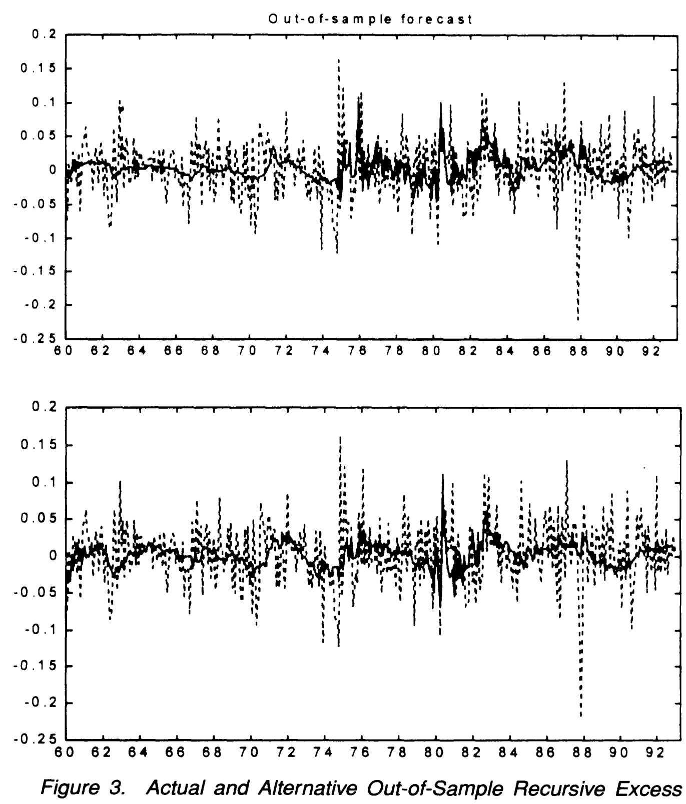

Figure 3. Actual and Alternative Out-of-Sample Recursive Excess Return Forecasts, 1960(1) to 1992(12): — — —, Actual; ———, NN.

Table 3. Out-of-Sample Forecast Performance of Alternative Models, 1960(1) to 1992(12), and the Three Subperiods

Model	RMSE	MAE	MAPE	CORR	Sign
		Panel A: 1960—1992			
NN	.0429	.0328	1.8289	.2292	.6237
LR	.0430	.0329	2.1095	.2081	.5960
		Panel B: 1960—1969			
NN	.0352	.0274	1.1150	.0661	.6083
LR	.0361	.0283	1.3726	.0656	.5667
		Panel C: 1970—1979			
NN	.0444	.0345	2.5604	.3067	.6750
LR	.0451	.0349	3.5491	.2458	.6250
		Panel D: 1980—1989			
NN	.0487	.0368	1.7134	.2070	.5750
LR	.0476	.0359	1.3894	.2287	.5750

and higher Pearson correlation coefficient and percentage of correct signs, though for both models the values of all out-of-sample performance measures are poorer than those in-sample. Panels B, C, and D of Table 3 report the forecast performance measures for the three subperiods—1960–1969, 1970–1979, and 1980–1989—respectively. As can be seen, in the first two subperiods, the NN forecasts outperform the linear forecasts by all five performance measures. In stark contrast, they underperform the linear forecasts in period 1980–1989 by four out of five measures. Nevertheless, both models generate the same percentage of correct signs in the last subperiod.

Table 3 also shows that, based on RMSE, MAE, and MAPE, both the linear and the nonlinear forecasts are more accurate in the 1960s than in the other two subperiods and the whole period of 1960–1992. Based on the Pearson correlation and the percentage of correct signs, however, both forecasts are more accurate in the 1970s than in the rest of the subperiods. Two conclusions thus can be drawn from the results. First, the predictive accuracy of all models changes over time. Second, the predictive accuracy varies with the performance measures. A model may perform better during certain periods and by certain measures but worse at other times or by other measures. The relative performance based on RMSE, MAE, and MAPE does not always agree with that based on CORR and Sign.

To assess the statistical significance of the difference between the LR and the nonlinear NN forecasts, the Diebold and Mariano (1995) test is used to test the null hypothesis that the two forecasts have the same expected squared errors. (I thank the associate editor and an anonymous referee for suggesting this test to me.) A consistent estimate of the spectral density at frequency 0 is obtained using the method of Newey and West (1987) with the truncation lag set by Andrews's (1991) AR(1) approximating rule. The results are given in Table 4. Though the NN model provides smaller squared forecast errors than linear regression during the 1960s, 1970s, and throughout the whole forecasting period, the Diebold and Mariano test shows that the improvement is not statistically significant.

Table 4. Diebold and Mariano (DM) Test of the Null of No Difference Between the Squared Forecast Errors LR and NN Forecasts

Models	DM statistics	p value
	Panel A: 1960 to 1992	
LR vs. NN	.0945	.4623
	Panel B: 1960 to 1969	
LR vs. NN	.7975	.2126
	Panel C: 1970 to 1979	
LR vs. NN	.5350	.2963
	Panel D: 1980 to 1989	
LR vs. NN	−.9990	.8411

According to Leitch and Tanner (1991), traditional measures of forecasting performance based on point forecast errors, such as RMSE, MAE, and MAPE, are not strongly correlated with the profits that may be generated from the forecast using certain trading strategies. In this research, the percentage of correct signs is used to measure the market timing ability of each forecast model. From Table 3, the percentage of correct signs of both the NN and the LR forecasts are all well above 50%, which indicates some market timing ability of both models.

The significance of the market timing ability is tested using the Pesaran and Timmermann (PT) test. The PT test is a nonparametric test that was generalized by Pesaran and Timmerman (1992, 1994) from the Henriksson–Merton (Henriksson and Merton 1981) test for independence between forecast and actual values. The PT test statistics and their p values are reported in Table 5. Panel A shows that, over the whole out-of-sample forecast period of 1960(1) to 1992(12), both the NN and linear models have significant market timing ability, and the p values are .0000 and .0003, respectively. Panels B, C, and D show the PT test results in the 1960s, 1970s, and 1980s. As can be seen, the NN model shows statistically significant market timing ability in both the 1960s and 1970s. The market timing ability of the NN model is less significant (p value $= .1597$) only during the 1980s. In contrast, the linear forecast model shows significant market timing ability only during the 1970s. Though

Table 5. PT Test of Out-of-Sample Market Timing Ability of Alternative Forecasts, 1960(1) to 1992(12) and the Three Subperiods

Model	Sign(%)	PT	p value
	Panel A: 1960 to 1992		
NN	62.37	4.5476	.0000
LR	59.60	3.4097	.0003
	Panel B: 1960 to 1969		
NN	60.83	1.8463	.0324
LR	56.67	1.1351	.1282
	Panel C: 1970 to 1979		
NN	67.50	3.8350	.0001
LR	62.50	2.8136	.0024
	Panel D: 1980 to 1989		
NN	57.50	.9956	.1597
LR	57.50	1.1144	.1326

the degree to which stock returns are linearly predictable seems quite low during the relatively calm markets in the 1960s (Pesaran and Timmermann 1995), the nonlinear NN still shows significant market timing ability during that period.

Finally, the relative predictive power of the LR and nonlinear NN forecasts is evaluated by a simple encompassing test [Fair and Shiller (1990), I thank an anonymous referee for suggesting this to me]. The actual realized value of excess returns (ρ) is regressed on a constant, and the nonlinear NN forecasts ($\hat{\rho}^{\mathrm{NN}}$) and the LR forecasts ($\hat{\rho}^{\mathrm{LR}}$). If the models all contain independent information that has power in predicting excess returns, then the coefficients on both $\hat{\rho}^{\mathrm{NN}}$ and $\hat{\rho}^{\mathrm{LR}}$ should be significantly different from 0. If, however, the information contained in one forecast is simply a subset of that contained in the other, then the coefficient on the former should be insignificant. The results from the encompassing test are

$$\hat{\rho} = \underset{\substack{(.7933)\\ [.4281]}}{.0017} + \underset{\substack{(2.2389)\\ [.0257]}}{.3947}\,\hat{\rho}^{\mathrm{NN}} + \underset{\substack{(1.0837)\\ [.2792]}}{.2113}\,\hat{\rho}^{\mathrm{LR}}$$

$$\mathrm{F} = 11.5160(p \text{ value } = .0000),$$

in which the numbers in parentheses are t ratios and the numbers in brackets are p values. Although the intercept and the coefficient on the LR forecasts are not significantly different from 0, the coefficient on the nonlinear NN forecasts are significant at 5% significance level. The results from the encompassing test show that the nonlinear NN model has significant explanatory power and the information contained in the LR forecasts is simply a subset of that contained in the flexible nonlinear NN forecasts.

5. PROFITABILITY

As is well known, predictability does not necessarily imply profitability. Whether the investor can make profit, how much profit the investor can make, and how much risk the investor has to bear to make so much profit depend also on what trading strategy one uses and the magnitude of transaction costs. Especially when the positions are evaluated monthly based on monthly recursive forecasts, the profits may be eroded by transaction costs. As such, an investment strategy that is based on recursive forecasts is likely to incur higher transaction costs and may not be as profitable as the buy-and-hold strategy. To determine whether the recursive predictions could have been used to make a higher profit than that from following a buy-and-hold strategy in the market portfolio, a switching strategy that has been extensively employed in the finance literature is adopted. The investor should hold equity in periods when the economic factors suggest that equity returns are going to outperform returns from bonds and otherwise hold bonds. The performance measures, such as mean return, standard deviation, Sharpe ratio, and final wealth of various switching portfolios relative to the market portfolio and Treasury bills are used to reveal the economic significance of alternative forecasting models. Assume no short selling or leverage during trading, and transaction costs are proportional to the value of the trade. Transaction costs are also assumed to be constant through time and symmetric with respect to whether the investor is buying or selling assets.

Table 6 reports the trading results, including the mean return, standard deviation, Sharpe ratio, and the final wealth, of the preceding switching strategy based on the forecasts of alternative models. The final wealth is calculated based on the assumption that the investor begins with $100 at the end of 1959(12). In the case of the market and bond portfolios, only the dividends or interests are reinvested on monthly bases, and in the case of the switching portfolio, funds may be reallocated between bonds and shares, depending on whether a change in the sign of the excess return is predicted. The results of 0, low, and high transaction costs are reported.

Consider first the results based on zero transaction costs. The mean annual return on the market index over the period 1960(1) to 1992(12) is 11.15%. The switching portfolios based on the recursive linear and NN forecasts outperform the market portfolio by 2.51% and 4.44%, respectively. These differences in mean returns are reflected in the end-of-period wealth accrued to the investment strategies based on reinvesting the funds in either bonds or stocks at the end of each month. Although the final wealth of the switching portfolio based on the linear forecasts ($7,458) is approximately three times as large as that of the market-index portfolio ($2,503), the final wealth generated from the switching portfolio based on the NN forecasts ($13,820) is more than five times that of the market portfolio. The standard deviation of the returns on the switching portfolios based on linear and NN forecasts are 10.08% and 10.61% (annual), respectively, which are substantially lower than that of the market portfolio (14.90%). Combining the mean return and standard deviation, the Sharpe ratio measures the excess return of a portfolio taking into consideration the risk of the portfolio. The switching portfolio based on the

Table 6. Risks and Profits of Market, Bond, and Switching Portfolios Based on the Out-of-Sample Forecasts of Alternative Models, 1960(1) to 1992(12)

Transaction costs	Mean return (%)	Std. of return	Sharpe ratio	Final wealth ($)
Panel A: Market portfolio				
Zero	11.15	14.90	.35	2,503
Low	11.13	14.90	.43	2,463
High	11.11	14.89	.429	2,424
Panel B: Bond portfolio				
Zero	5.93	2.74	—	700
Low	4.72	2.74	—	471
High	4.72	2.74	—	471
Panel C: Switching portfolio based on linear forecast				
Zero	13.66	10.08	.77	7,458
Low	12.21	10.18	.74	4,631
High	11.23	10.34	.63	3,346
Panel D: Switching portfolio based on NN forecast				
Zero	15.59	10.49	.92	13,820
Low	14.08	10.52	.89	8,420
High	13.03	10.61	.78	5,963

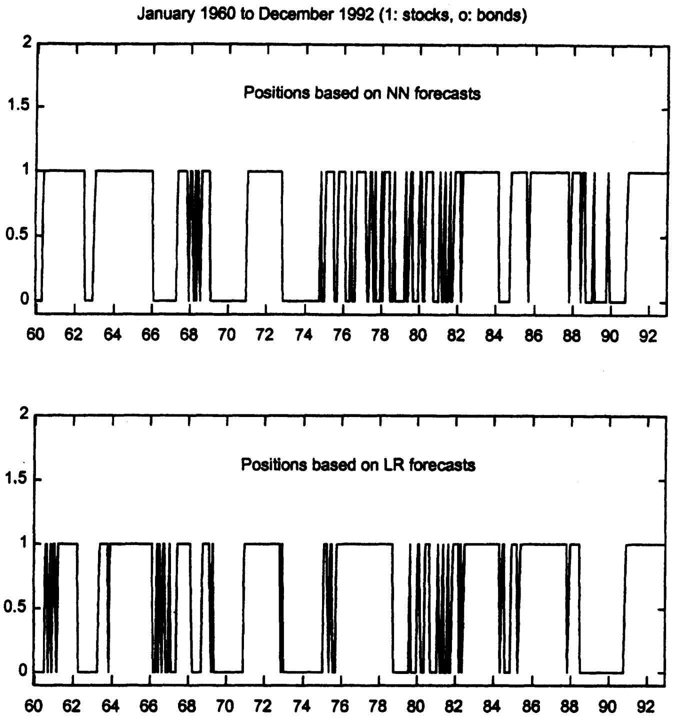

Figure 4. State of the Switching Portfolio: 1960(1) to 1992(12).

NN forecasts has the highest Sharpe ratio, .92. The Sharpe ratio of the switching portfolio based on the linear forecasts is .77, which is still higher than that of the market index buy-and-hold portfolio (.35).

With low transaction costs of .5% on trading in stocks and .1% in bonds, the mean returns on the switching portfolios based on the linear and NN forecasts decline by 1.45% and 1.51% per annum. In contrast, the transaction costs hardly affect the mean return on the market portfolio because it only involves dividend reinvestment. Yet, the mean returns, Sharpe ratio, and final wealth of the linear and NN switching portfolios are still much higher than those of the market portfolio, and the standard deviations are lower. Moreover, the switching portfolio based on the NN forecasts outperforms the switching portfolio based on the LR forecasts.

Now consider high transaction costs of 1% on shares and .1% on bonds. Although the switching portfolio based on the linear forecasts is marginally better than the market portfolio by measures of mean return (11.23% vs. 11.11%) and final wealth ($3,346 vs. $2,424), it has lower risk (10.34% vs. 14.90%) and thus a higher Sharpe ratio (.63 vs. .43). By all measures, the switching portfolio based on the NN forecasts still outperforms both the market portfolio and the switching portfolio based on the linear forecasts.

For switching portfolios, how often the portfolio is in bonds, and in particular when it is in bonds, provides additional information on the market timing ability of the forecasts that the switching decisions are based on (I thank an anonymous referee for suggesting this to me). Figure 4 plots over time the state of the portfolio; "1" indicates in stocks and "0" indicates in bonds. Although 61.11% of the time the portfolio is in stocks based on linear forecasts, it is in stocks 61.36% of the time based on nonlinear NN forecasts. Of particular interest, both portfolios based on the LR and nonlinear NN forecasts switched to bonds right before the market crash in October 1987, which further confirms the significant market timing ability shown by the PT nonparametric test results reported in Table 5.

The performance of various portfolios is also analyzed over the subperiods of 1960s, 1970s, and 1980s, and the results are reported in Table 7. Under the zero-transaction-costs scenario, the mean return on the switching portfolio based on the NN forecasts in all three subperiods and the

Table 7. Out-of-Sample Risk and Returns of Different Portfolios for Subperiods: 1960s, 1970s, and 1980s (Annual)

	Zero transaction costs			*Low transaction costs*			*High transaction costs*		
Portfolios	*1960s*	*1970s*	*1980s*	*1960s*	*1970s*	*1980s*	*1960s*	*1970s*	*1980s*
				Panel A: Mean return					
Market	8.99	7.68	17.04	8.97	7.66	17.02	8.96	7.64	17.00
Bond	3.71	6.02	8.23	2.51	4.81	7.02	2.51	4.81	7.02
LR	7.91	13.95	17.57	6.12	12.78	16.00	4.80	12.15	14.85
NN	9.55	16.47	19.83	8.45	14.47	18.18	7.73	13.12	16.92
				Panel B: Standard deviation					
Market	11.95	15.86	16.41	11.95	15.86	16.41	11.95	15.86	16.41
Bond	1.32	1.87	2.92	1.32	1.87	2.92	1.32	1.87	2.92
LR	7.50	9.93	11.99	7.62	9.77	12.30	7.86	9.65	12.65
NN	8.50	9.18	11.99	8.52	9.22	12.20	8.59	9.31	12.46
				Panel C: Sharpe ratio					
Market	.44	.10	.54	.54	.18	.61	.54	.18	.61
Bond	—	—	—	—	—	—	—	—	—
LR	.56	.80	.78	.47	.82	.73	.29	.76	.62
NN	.69	.99	.97	.70	.93	.91	.61	.81	.79
				Panel D: Final wealth					
Market	2,111	1,753	4,982	2,086	1,732	4,922	2,062	1,711	4,863
Bond	1,444	1,814	2,255	1,281	1,608	2,000	1,281	1,608	2,000
LR	2,132	3,800	5,593	1,783	3,390	4,756	1,563	3,189	4,210
NN	2,487	4,841	6,975	2,227	3,980	5,895	2,074	3,490	5,166

switching portfolio based on the linear recursive forecasts in the 1970s and 1980s outperform that of the market portfolio in the corresponding period. With low transaction costs, although the mean return on the switching portfolio based on the recursive linear forecast is higher than that of the market portfolio only during the 1970s, the mean return on the NN switching portfolio is better than that of the market portfolio during the 1970s and 1980s. When transaction costs are high, however, the mean return on both the linear and NN forecasts are better than that of the market only during the 1970s. Nevertheless, during the 1970s and 1980s, even with transaction costs the switching portfolios based on both the linear and NN forecasts have higher Sharpe ratios than the market portfolio because they have lower standard deviations in all subperiods and under all transaction-cost scenarios. The results confirm that "if ever there was a possibility that investors could improve their market timing based on a simple forecasting procedure ... , this was during the volatile periods in the 1970s where macroeconomic risk and volatility in nominal magnitudes, such as the rate of inflation and nominal interest rates, mattered the most" (Pesaran and Timmermann 1995). This suggests that an investor can use the nonlinearity captured by an NN to improve the forecasting accuracy for a larger risk-adjusted return.

6. SUMMARY AND DISCUSSION

In this research a recursive modeling approach has been applied to examine the predictability of S&P 500 index returns using the LR and nonlinear NN forecast models with monthly observations on nine financial and economic variables. The nonlinear NN approach accounts for the model-specification uncertainty in selecting appropriate functional forms faced by virtually all investors who try to forecast asset returns in real time. The recursive approach allows various models to evolve over time to account for possible structural changes in the underlying relationships.

The nonlinear NN model not only fits the data better than the linear model in sample, it also provides fairly accurate forecasts out of sample. The recursive NN model has smaller RMSE, MAE, and MAPE and higher Pearson correlation and percentage of correct signs than the LR model in the whole out-of-sample forecast period and in two out of three subperiods.

The profitability of the switching portfolios based on the linear and nonlinear forecasts has also been examined. With zero, low, and high transaction costs, the switching portfolio based on the recursive nonlinear NN forecasts earns higher risk-adjusted returns than the switching portfolio based on the recursive linear forecasts. Both switching portfolios outperform the buy-and-hold market-index portfolio during the whole forecast period from 1960 to 1992. When the three subperiods are studied separately, it is found that the outperformance occurred primarily during the volatile markets of the 1970s and 1980s.

This research provides clear evidence of nonlinear predictability of U.S. stock-market returns using financial and economic variables, in addition to the evidence of linear predictability documented in the existing literature. This is not surprising given the numerous findings of nonlinearity in financial and economic variables. The additional complexity of the NN model and the loss of interpretation and statistical inference seem to be compensated with more accurate forecasts and higher profitability. The nonlinear pattern represented by the recursively estimated NN to account for possible structure change can be used by investors to generate accurate forecast and higher risk-adjusted return.

Our finding of nonlinear predictability of stock returns using financial and economic variables in a nonparametric NN framework not only has practical value but also has theoretical importance. Even though at this stage it is still not clear at each point in time what is the "true" underlying model, the nonlinear predictability we found encourages an extensive search among a vast number of parametric nonlinear models. Future research can be carried out to compare alternative parametric nonlinear models, such as ARCH-M, multivariate polynomial, and piecewise linear, as well as Markov-switching models, and so forth.

ACKNOWLEDGMENTS

I thank Allan Timmermann for providing the dataset used in this research. I also acknowledge the helpful comments and suggestions of James Baker, Zhuanxin Ding, Clive Granger, Christine Jiang, G. S. Maddala, Allan Timmermann, Jeff Wooldridge (the editor), the associate editor, two anonymous referees, and the participants at the 1998 Midwest Finance Association Annual Meeting. I am grateful for the financial support from the College of Business Administration at Kent State University. The usual disclaimer applies. An earlier version of this article won the Best Paper in Investments Award (sponsored by the American Association of Individual Investors) at the 1998 Midwest Finance Association Annual Meeting in Chicago.

APPENDIX A: NGUYEN–WIDROW INITIALIZATION ALGORITHM

This section gives a brief introduction to the initialization algorithm suggested by Nguyen and Widrow (1990). Without loss of generality, consider the NN model

$$\begin{aligned} y &= f(X, \alpha, \beta) + e \\ &= \sum_{j=1}^{n} \alpha_j \text{logsig}\left(\sum_{i=1}^{k} \beta_{ij} x_i + \beta_{0j}\right) + e, \end{aligned} \quad \text{(A.1)}$$

where y is the network's output, X is the input vector, n is the number of middle layer units, $\beta_j = \{\beta_{ij}, i = 1, 2, \ldots, k\}$ is the weight vector of the jth unit of the middle layer, β_{0j} is the bias weight of the jth middle layer unit, α_j is the weight of the output layer that connects the jth hidden layer unit to the output, and logsig is a logistic function $\text{logsig}(x) = 1/(1 + \exp(-x))$ that is approximately linear with slope 1 for x between -1 and 1 but saturates to 0 or 1 as x becomes large in magnitude.

Define $y_j(X)$ to be the jth term of the sum in Equation (A.1),

$$y_j(X) = \alpha_j \text{logsig}\left(\sum_{i=1}^{k} \beta_{ij} x_i + \beta_{0j}\right), \qquad \text{(A.2)}$$

and let $Y_j(U)$ be the Fourier transform of $y_j(X)$. Because $Y_j(U)$ is a line impulse going through the origin of the transform space U and the orientation of the line impulse is dependent on the direction of the vector β_j, Y_j thus can be interpreted as a part of an approximation of a slice through the origin of the Fourier transform $F(U)$ of $f(x)$. The direction of β_j determines the direction of the jth slice of $F(U)$, and the magnitude of β_j determines the interval size in making piecewise linear approximation to the inverse transform of the jth slice of $F(U)$. The value of β_{0j} determines the location of the interval. Finally, α_j determines the slope of the linear approximation.

Assume that choosing initial values in such a way that the middle layer units are scattered in the input space can substantially improve the learning speed, and that after data normalization the input elements range from -1 to 1. The Nguyen–Widrow method proceeds as follows. First, the elements of β_j are assigned values from a uniform random distribution between -1 and 1 so that its direction is random. Next, the magnitude of the weight vectors β_j is adjusted so that each middle layer unit is linear over only a small interval. Then the n middle layer units will be used to form S slices, and I intervals per slice; that is, $n = S \cdot I$. The network weights are set so that $S = I^{k-1}$. Each element of the input vector X ranges from -1 to 1, which means the length of each interval is approximately $2/I$. The magnitude of β_j is then adjusted as

$$|\beta_j| = I = n^{1/k}. \qquad \text{(A.3)}$$

The magnitude of β_j is often set to $.7n^{1/k}$ to provide some overlap between the intervals. Finally, the center of the interval is located at a random location along the slice by setting β_{0j} to uniform random number between $-|\beta_j|$ and $|\beta_j|$.

APPENDIX B: BAYESIAN REGULARIZATION

Overfitting often occurs during NN training. The error on the training sample can be made very small, but the error is large when out-of-sample data are presented to the network. One way to improve the out-of-sample fit is to use a network that is just large enough to provide an adequate fit. When the network is small enough, it will not overfit the data. The problem is that it is difficult to know beforehand how large a network should be for a specific application. A recent development in the prevention of overfitting is Bayesian regularization suggested by MacKay (1992). This involves modifying the objective function by adding a term that consists of the mean of the sum of squares of the network coefficients. For the NN represented by Equation (A.1), the objective function with Bayesian regularization, MSEBR, becomes

$$\text{MSEBR} = \gamma\text{MSE} + (1-\gamma)\text{MSW}, \qquad \text{(B.1)}$$

where $\text{MSE} = 1/T \sum_{t=1}^{T} e_t^2$, the usual NN objective function, $\text{MSW} = 1/(n(k+1))(\sum_{j=1}^{n} \alpha_j^2 + \sum_{i=1}^{k} \sum_{j=1}^{n} \beta_{ij}^2)$, the mean of the sum of all the squared network coefficients, and γ is the performance ratio.

Using the performance function represented by (B.1) will cause the network to have smaller coefficients, and this will force the network response to be smoother and less likely to overfit. The problem with such regularization is that it is difficult to determine the optimum value for the performance ratio parameter. If γ is too large, the network may still overfit. If it is too small, the network will not adequately fit the training data.

One approach to determine the optimal regularization parameters γ is the Bayesian framework of MacKay (1992). In this framework the coefficients of the network are assumed to be random variables with specified distributions. The regularization parameters are related to the unknown variances associated with these distributions and thus can be estimated with statistical techniques. A detailed discussion of Bayesian regularization is beyond the scope of this article, Foresee and Hagan (1997) gave a detailed discussion of the use of Bayesian regularization in combination with the Levenberg–Marquardt training algorithm.

[Received January 1998. Revised November 1998.]

REFERENCES

Abhyankar, A., Copeland, L. S., and Wong, W. (1997), "Uncovering Nonlinear Structure in Real-Time Stock-Market Indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100," *Journal of Business & Economic Statistics*, 15, 1–14.

Ackert, L. F., and Smith, B. F. (1993), "Stock Price Volatility, Ordinary Dividends, and Other Cash Flows to Shareholders," *Journal of Finance*, 48, 1147–1160.

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrics*, 59, 817–858.

Balvers, R. J., Cosimano, T. F., and McDonald, B. (1990), "Predicting Stock Returns in an Efficient Market," *Journal of Finance*, 45, 1109–1128.

Breen, W., Glosten, L. R., and Jagannathan, R. (1990), "Predictable Variations in Stock Index Returns," *Journal of Finance*, 44, 1177–1189.

Brock, W. A., Dechert, W. D., Scheinkman, J. A., and Lebaron, B. (1996), "A Test for Independence Based on the Correlation Dimension," *Econometric Review*, 15, 197–235.

Chuah, K. L. (1993), "A Nonlinear Approach to Return Predictability in the Securities Markets Using Feedforward Neural Network," unpublished dissertation, Washington State University, Dept. of Economics.

Cochrane, J. H. (1991), "Production-Based Asset Pricing and the Link Between Stock Return and Economic Fluctuations," *Journal of Finance*, 46, 209–238.

Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

Fair, R. C., and Shiller, R. J. (1990), "Comparing Information in Forecasts From Econometric Models," *American Economic Review*, 80, 375–389.

Fama, E. F., and French, K. R. (1989), "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics*, 25, 23–49.

Ferson, W. E., and Harvey, C. R. (1993), "The Risk and Predictability of International Equity Returns," *Review of Financial Studies*, 6, 527–566.

Foresee, F. D., and Hagan, M. T. (1997), "Gauss-Newton Approximation to Bayesian Regularization," in *IEEE International Conference on Neural Networks 1997* (Vol. 3), New York: IEEE, pp. 1930–1935.

French, K. R., Schwert, G. S., and Stambaugh, R. F. (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–30.

Gencay, R. (1998), "The Predictability of Security Returns With Simple Technical Trading Rules," *Journal of Empirical Finance*, 5, 347–359.

——— (1999), "Linear, Nonlinear and Essential Foreign Exchange Rate Prediction With Simple Technical Trading Rules," *Journal of International Economics*, 47, 91–107.

Gencay, R., and Stengos, T. (1998), "Moving Average Rules, Volume and the Predictability of Security Returns With Feedforward Networks," *Journal of Forecasting*, 17, 401–414.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Returns on Stocks," *Journal of Finance*, 48, 663–680.

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, 513–533.

Hornik, K., Stinchcombe, M., and White, H. (1989), "Multilayer Feedforward Networks Are Universal Approximators," *Neural Networks*, 2, 359–366.

——— (1990), "Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks," *Neural Networks*, 3, 551–560.

Hutchinson, J., Lo, A., and Poggio, T. (1994), "A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks," *Journal of Finance*, 49, 851–889.

Hwang, J. T. G., and Ding, A. A. (1997), "Prediction Intervals for Artificial Neural Networks," *Journal of the American Statistical Association*, 92, 748–757.

Jarque, C. M., and Bera, A. K. (1987), "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55, 163–172.

Kuan, C., and White, H. (1994), "Artificial Neural Networks: An Econometric Perspective," *Econometric Review*, 13, 1–91.

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

Ljung, G. M., and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65, 297–303.

MacKay, D. J. C. (1992), "Bayesian Interpolation," *Neural Computation*, 4, 415–447.

Newey, W. K., and West, K. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Nguyen, D., and Widrow, B. (1990), "Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights," *Proceedings of the International Joint Conference of Neural Networks*, 3, 21–26.

Pesaran, M. H., and Timmermann, A. (1992), "A Simple Nonparametric Test of Predictive Performance," *Journal of Business & Economic Statistics*, 10, 461–465.

——— (1994), "A Generalization of the Nonparametric Henriksson–Merton Test of Market Timing," *Economics Letters*, 44, 1–7.

——— (1995), "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance*, 50, 1201–1228.

Qi, M. (1996), "Financial Applications of Artificial Neural Networks," in *Handbook of Statistics, 14: Statistical Methods in Finance*, eds. G. S. Maddala and C. R. Rao, Amsterdam: North-Holland, pp. 529–552.

Qi, M., and Maddala, G. S. (1999), "Economic Factors and the Stock Market: A New Perspective," *Journal of Forecasting*, 18, 151–166.

Ramsey, J. B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," *Journal of the Royal Statistical Society*, Ser. B, 31, 350–371.

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 & Economic Statistics*, 13, 265–275.

——— (1997), "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks," *The Review of Economics and Statistics*, 79, 540–550.

Tsay, R. S. (1986), "Nonlinearity Tests for Time Series," *Biometrika*, 73, 461–466.

White, H. (1988), "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns," *Proceedings of the IEEE International Conference on Neural Networks*, 2, 451–458.

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

White, H., Gallant, A. R., Hornik, K., Stinchcombe, M., and Wooldridge, J. (1992), *Artificial Neural Networks: Approximation and Learning Theory*, Cambridge, MA: Blackwell.

http://www.jstor.org

LINKED CITATIONS

You have printed the following article:

Nonlinear Predictability of Stock Returns Using Financial and Economic Variables
Min Qi
Journal of Business & Economic Statistics, Vol. 17, No. 4. (Oct., 1999), pp. 419-429.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199910%2917%3A4%3C419%3ANPOSRU%3E2.0.CO%3B2-8

This article references the following linked citations:

References

Uncovering Nonlinear Structure in Real-Time Stock-Market Indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100
A. Abhyankar; L. S. Copeland; W. Wong
Journal of Business & Economic Statistics, Vol. 15, No. 1. (Jan., 1997), pp. 1-14.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199701%2915%3A1%3C1%3AUNSIRS%3E2.0.CO%3B2-Q

Stock Price Volatility, Ordinary Dividends, and Other Cash Flows to Shareholders
Lucy F. Ackert; Brian F. Smith
The Journal of Finance, Vol. 48, No. 4. (Sep., 1993), pp. 1147-1160.
Stable URL:
http://links.jstor.org/sici?sici=0022-1082%28199309%2948%3A4%3C1147%3ASPVODA%3E2.0.CO%3B2-B

Predicting Stock Returns in an Efficient Market
Ronald J. Balvers; Thomas F. Cosimano; Bill McDonald
The Journal of Finance, Vol. 45, No. 4. (Sep., 1990), pp. 1109-1128.
Stable URL:
http://links.jstor.org/sici?sici=0022-1082%28199009%2945%3A4%3C1109%3APSRIAE%3E2.0.CO%3B2-3

Economic Significance of Predictable Variations in Stock Index Returns
William Breen; Lawrence R. Glosten; Ravi Jagannathan
The Journal of Finance, Vol. 44, No. 5. (Dec., 1989), pp. 1177-1189.
Stable URL:
http://links.jstor.org/sici?sici=0022-1082%28198912%2944%3A5%3C1177%3AESOPVI%3E2.0.CO%3B2-G

Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations
John H. Cochrane
The Journal of Finance, Vol. 46, No. 1. (Mar., 1991), pp. 209-237.
Stable URL:
http://links.jstor.org/sici?sici=0022-1082%28199103%2946%3A1%3C209%3APAPATL%3E2.0.CO%3B2-7

Comparing Predictive Accuracy
Francis X. Diebold; Roberto S. Mariano
Journal of Business & Economic Statistics, Vol. 13, No. 3. (Jul., 1995), pp. 253-263.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199507%2913%3A3%3C253%3ACPA%3E2.0.CO%3B2-T

Comparing Information in Forecasts from Econometric Models
Ray C. Fair; Robert J. Shiller
The American Economic Review, Vol. 80, No. 3. (Jun., 1990), pp. 375-389.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28199006%2980%3A3%3C375%3ACIIFFE%3E2.0.CO%3B2-7

The Risk and Predictability of International Equity Returns
Wayne E. Ferson; Campbell R. Harvey
The Review of Financial Studies, Vol. 6, No. 3. (1993), pp. 527-566.
Stable URL:
http://links.jstor.org/sici?sici=0893-9454%281993%296%3A3%3C527%3ATRAPOI%3E2.0.CO%3B2-K

On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills
Roy D. Henriksson; Robert C. Merton
The Journal of Business, Vol. 54, No. 4. (Oct., 1981), pp. 513-533.
Stable URL:
http://links.jstor.org/sici?sici=0021-9398%28198110%2954%3A4%3C513%3AOMTAIP%3E2.0.CO%3B2-D

LINKED CITATIONS

A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks

James M. Hutchinson; Andrew W. Lo; Tomaso Poggio

The Journal of Finance, Vol. 49, No. 3, Papers and Proceedings Fifty-Fourth Annual Meeting of the American Finance Association, Boston, Massachusetts, January 3-5, 1994. (Jul., 1994), pp. 851-889.

Stable URL:

http://links.jstor.org/sici?sici=0022-1082%28199407%2949%3A3%3C851%3AANATPA%3E2.0.CO%3B2-D

Prediction Intervals for Artificial Neural Networks

J. T. Gene Hwang; A. Adam Ding

Journal of the American Statistical Association, Vol. 92, No. 438. (Jun., 1997), pp. 748-757.

Stable URL:

http://links.jstor.org/sici?sici=0162-1459%28199706%2992%3A438%3C748%3APIFANN%3E2.0.CO%3B2-L

A Test for Normality of Observations and Regression Residuals

Carlos M. Jarque; Anil K. Bera

International Statistical Review / Revue Internationale de Statistique, Vol. 55, No. 2. (Aug., 1987), pp. 163-172.

Stable URL:

http://links.jstor.org/sici?sici=0306-7734%28198708%2955%3A2%3C163%3AATFNOO%3E2.0.CO%3B2-3

Economic Forecast Evaluation: Profits Versus the Conventional Error Measures

Gordon Leitch; J. Ernest Tanner

The American Economic Review, Vol. 81, No. 3. (Jun., 1991), pp. 580-590.

Stable URL:

http://links.jstor.org/sici?sici=0002-8282%28199106%2981%3A3%3C580%3AEFEPVT%3E2.0.CO%3B2-Y

On a Measure of Lack of Fit in Time Series Models

G. M. Ljung; G. E. P. Box

Biometrika, Vol. 65, No. 2. (Aug., 1978), pp. 297-303.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28197808%2965%3A2%3C297%3AOAMOLO%3E2.0.CO%3B2-4

LINKED CITATIONS

A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix

Whitney K. Newey; Kenneth D. West

Econometrica, Vol. 55, No. 3. (May, 1987), pp. 703-708.

Stable URL:

http://links.jstor.org/sici?sici=0012-9682%28198705%2955%3A3%3C703%3AASPSHA%3E2.0.CO%3B2-F

A Simple Nonparametric Test of Predictive Performance

M. Hashem Pesaran; Allan Timmermann

Journal of Business & Economic Statistics, Vol. 10, No. 4. (Oct., 1992), pp. 461-465.

Stable URL:

http://links.jstor.org/sici?sici=0735-0015%28199210%2910%3A4%3C461%3AASNTOP%3E2.0.CO%3B2-C

Predictability of Stock Returns: Robustness and Economic Significance

M. Hashem Pesaran; Allan Timmermann

The Journal of Finance, Vol. 50, No. 4. (Sep., 1995), pp. 1201-1228.

Stable URL:

http://links.jstor.org/sici?sici=0022-1082%28199509%2950%3A4%3C1201%3APOSRRA%3E2.0.CO%3B2-Z

Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis

J. B. Ramsey

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 31, No. 2. (1969), pp. 350-371.

Stable URL:

http://links.jstor.org/sici?sici=0035-9246%281969%2931%3A2%3C350%3ATFSEIC%3E2.0.CO%3B2-8

A Model-Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks

Norman R. Swanson; Halbert White

Journal of Business & Economic Statistics, Vol. 13, No. 3. (Jul., 1995), pp. 265-275.

Stable URL:

http://links.jstor.org/sici?sici=0735-0015%28199507%2913%3A3%3C265%3AAMATAT%3E2.0.CO%3B2-C

http://www.jstor.org

LINKED CITATIONS

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

Norman R. Swanson; Halbert White

The Review of Economics and Statistics, Vol. 79, No. 4. (Nov., 1997), pp. 540-550.

Stable URL:

http://links.jstor.org/sici?sici=0034-6535%28199711%2979%3A4%3C540%3AAMSATR%3E2.0.CO%3B2-R

Nonlinearity Tests for Time Series

Ruey S. Tsay

Biometrika, Vol. 73, No. 2. (Aug., 1986), pp. 461-466.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28198608%2973%3A2%3C461%3ANTFTS%3E2.0.CO%3B2-9