PREDICTING RECESSION
USING THE YIELD CURVE:
AN ARTIFICIAL INTELLIGENCE AND ECONOMETRIC COMPARISON

Mohammad Shaaf
University of Central Oklahoma

INTRODUCTION

Over the years, economists have provided evidence that both stock and bond market data contain information relevant for predicting future economic growth. Several authors investigated this subject empirically and found that the bond market predictions are more accurate. For example, Harvey [1980] showed that the bond market explained more than 30 percent of the variation in economic growth over 1953-89, while stock market variables explained only about 5 percent. This is due to variation in stock prices that reflect both changes in expected economic growth and changes in the expected rate of stock cash flows.

Other studies focused on the bond market or the yield curve as a predictor of recession. The yield curve most often used is the one representing the rates of return on Treasury bonds against their maturity dates. Normally, the yield curve has a positive slope and looks convex as is shown in Figure 1. However, when the curve gets flat or slopes downward, it is an indicator of economic recession in the future.

The theoretical basis of the yield curve goes back to Irving Fisher [1907]. There are several forward-looking hypotheses of the yield curve. One hypothesis suggests that tight monetary policy causes short-term interest rates to rise relative to the long-term rates. This action could reduce the spread between the long-term and the short-term interest rates, making the curve flat or downward sloping. Another hypothesis argues that tight monetary policy results in lower inflationary expectations and consequently reduces long-term interest rates, also reducing the spread and causing the curve to flatten.1

The empirical works of Harvey [1986, 1989, 1991, 1993], Mishkin [January 1990, August 1990, 1991], Estrella and Hardouvelis [1991], Hu [1993], Campbell [1995, 1996], and Benzoni [1996, 1997], and others confirm the predictive power of the yield curve. These investigators used different types of econometric models to show the relationship between the yield curve spread, with or without other variables, and real-GDP growth.2

Mohammad Shaaf. Economics Department, University of Central Oklahoma, Edmond, OK 73034. E-mail: mshaaf@ucok.edu


171
Fig 1. Treasury Yield Curve

Source: Federal Reserve of St. Louis, Monetary Trends, March 1998

Estrella and Mishkin used a probit model to estimate the probability of recession given the shape of the yield curve. Using pseudo-$R^2$ as a measure of fit and t-statistics in their model, they concluded that the four-quarter lags of the Treasury spread:

shows the best predictive performance across the range of horizons examined...it is substantially outperformed by a number of other indicators, including stock price indices, the Commerce and Stock-Watson leading indicators, and some of the Commerce indicator's components (1995, 20).

In recent years the artificial intelligence technique and its forecast have been used successfully in various economic studies, including investment and financial forecast (Trippi and Turban, 1992; Hsieh, 1993; Swales and Yoon, 1992), effectiveness of fiscal and monetary policy (Shaaf, 1999), housing markets (Shaaf and Erfani, 1996), exports growth as the source of economic growth (Shaaf and Ahmadi, 1999), and others. The purpose of this study is twofold: the first, using an artificial intelligence model called "neural networks," is to investigate the power of the yield curve in predicting economic activity and recession. The second is to compare the predictions of the artificial intelligence model with those of the econometric approach. This study uses the same model, variables, and lag structures as those determined to be the "best-fit" by Estrella and Mishkin (1995).

The results of the artificial intelligence model confirm those of the econometric findings: a flat or negative yield curve is the sign of recession in the near future. The findings of out-of-sample simulations also suggest that the artificial intelligence approach to predicting recession is more accurate than that of the traditional econometric method. This paper proceeds as follows. The first section describes the architecture of artificial intelligence. The next section presents the empirical results of the artificial neural network and regression models, and those of the in-sample and out-of-sample simulations. The last section concludes the paper by summarizing the neural networks and the findings.

ARTIFICIAL INTELLIGENCE MODEL

One of the branches of the study of artificial intelligence is the neural network. Neural networks are multiple-layer configurations similar to the structure of the brain, consisting of simple processing elements or nodes that interact with each other through weighted connections in the system. This study utilizes the most widely used neural network model termed 'back propagation,' which is a non-parametric algorithm for adjusting connection weights in a multiple-layer network. Back propagation is a learning design by which the multi-layer network is set for pattern recognition utilizing actual cross section or time series data as the external teacher. A typical artificial network consists of:

1. one input (independent-variable) layer
2. one or more intermediate or hidden layers
3. one output (dependent-variable) layer.

Similar to the simultaneous equations in econometric modeling, neural network architecture can have several independent variables and several dependent variables. A network similar to the one used in this study with two independent variables (inputs), a hidden layer, a bias, and one dependent variable (output) is shown in Figure 2. Each connection from the input layer to the hidden layer or from the hidden layer to the output layer has a weight. These weights represent the coefficients or the parameters of the model. The size of each weight represents the relative strength of the connection. Each node computes a weighted sum of the incoming values and passes this sum through a nonlinear or linear function as output. The output of one layer serves as the input to the next.

The hidden layer(s) is a processor of information and a bridge between the independent variable(s) and dependent variable(s). That layer(s) computes the information and does most of the work in the network. Since this layer is connected to independent variables (inputs) on one hand and dependent variables (output) on the other, and is hidden from the outside world, it is called the hidden layer. Besides its link to the hidden layer, the input layer can be directly connected to the output layer. Notice that the inputs of the network used in this study are also connected directly to the output layer (see Figure 2).

The network is "fully connected." That is, all the nodes are linked in adjacent layers. These links are the weights that can be strong or weak depending on their size. The weights are adjusted for the minimization of the mean square error as the
FIGURE 2
An Artificial Neural Network for the Yield Curve as a Predictor of Recession

![Diagram of an Artificial Neural Network]

objective function. Similar to the econometric method, the back propagation design is an optimization technique for finding optimum values of the weights as parameters. Furthermore, there is a bias-processing element that is similar to the intercept notion in equations of econometric models.

The three main phases in the operation of a network are learning, recall, and testing. In the learning phase, the neural network recognizes a pattern between independent variable(s) and dependent variable(s) and estimates the final value at the output layer. Subsequently, this estimated value is compared with the actual output, and errors are calculated. These errors become the factor used to adjust the weights and, subsequently, to reduce the errors and readjust the weights of the connections.

This process of weight adjustment continues further until the error declines to an acceptable level, when possible. Through this process, the neural network learns the rules and adapting patterns for processing the information. The resulting pattern of linking the independent variables to the dependent variables in the learning process can be used for testing the accuracy of the networks. In addition, the pattern can be used for prediction of the dependent variable(s), given the independent variable(s).

To explain the mathematical equations of the neural networks, assume there are \( I \) number of independent variables, \( j \) hidden nodes in the hidden layer, and \( k \) output nodes in the output layer. Accordingly, back propagation computes the summation of multiplication of independent variables, \( X_i \), by their corresponding weights (from the input layer to the hidden layer), and adds the bias weights (intercept) to it as follows:

\[
 w_j = w_{0j} + \sum_i w_{ij}X_i
\]

where \( w_{0j} \) is the bias weight \( j \), and \( w_{ij} \) are the weights from the input \( i \) to hidden node \( j \).

The weighted summations, \( u_j \), in equation (1) are then transformed by a transfer function in the hidden layer. It is called the "transfer function" because it acts to "transfer" the internally generated sum to a potential output value. Three common nonlinear transfer functions used in artificial neural networks are the sigmoid, hyperbolic tangent, and arctangent functions. The sigmoid transfer function, which is considered appropriate and is used in this study, is a continuous monotonic mapping of the input into a limited-range value between zero and one. The sigmoid function of \( u \) variable is a smooth version of \((0,1)\) step-function type, and is defined as:

\[
y_j = \sigma(u_j) = (1 + e^{-u_j})^{-1}
\]

where \( e \) is the base of natural logarithms. Thus, the output of the hidden node \( J \) is a sigmoid-shaped hypersurface of \( I \) dimensions in a space of \((I+1)\) dimensions. This output, \( y_j \), from the hidden layer enters the output node as input, and subsequently computes the output of the network with \( k \) output nodes, \( \pi_k \), as:

\[
z_k = g(y_j), \quad k = 1, 2, ..., K
\]

and \( g(u) \) is the sigmoid function of \( u \) as:

\[
y_k = w_{0k} + \sum_j w_{jk}y_j
\]

where \( w_{jk} \) are weights on the links from hidden node \( j \) to the output \( k \), and \( w_{0k} \) is the bias weight of output node \( k \). Notice that at the output node the process works the same way as in the hidden nodes, and the output in equation (4) enters the sigmoid function of equation (3) and transforms the information in the form of:

\[
z = f(u) = (1 + e^{-u})^{-1}
\]

If the neural network is directly connected from the input nodes to the output nodes, the \( y \) function in equation (4) is modified as follows:

\[
y_k = w_{0k} + \sum_j w_{jk}x_j + \sum_{j=1}^J w_{jk}y_j
\]

The third term in equation (6) represents the weighted summation of input nodes by the weights directly connecting them to the output nodes.

The next step in the neural network is the calculation of the error. Suppose the estimated final dependent variable at the output layer, which is produced by the network with \( N \) observations, is \( z_{e,n} \), and the actual dependent variable is \( z_{a,n} \), then the mean square error is computed as:

\[
E = \frac{1}{N} \sum_{n=1}^{N} (z_{e,n} - z_{a,n})^2
\]
Here, $E$ is the global error function which is differentiable of all the connection weights in the network.

The subsequent stage in the artificial neural network is to adjust the weights (parameters) of the model. The strength of the connections (weights) changes continuously for the minimization of the error as the objective function. The weights in the neural networks fall within two major categories. The first category connects the hidden nodes to the output nodes. The second connects the input nodes to the hidden nodes. The equations for these two types of weight adjustment are explained next.

**Hidden-to-Output-Node Weights**

When the error $E$ is determined, the network back propagates it for the increase or decrease of the weights. First, it selects some arbitrary point in weight space and computes the slope of the error surface at that point. Then, it computes the change or derivative of the error, $E$, with respect to the weight from the hidden node $j$ to the output $k$, $\Delta E/dw_{jk}$, through a chain-rule multiplication of three terms as follows:

$$ \Delta w_{jk} = (\Delta w_{jk})/(\Delta E/d\omega_{jk}) \cdot (\Delta E/d\omega_{jk}) \cdot (\Delta \omega_{jk}/\omega_{jk}). $$

(8)

The first term on the right side of the equation, $\Delta E/d\omega_{jk}$, is the derivative of the error with respect to the dependent variable $k$. The second term in the chain is the derivative of that node's output with respect to its weighted sum of inputs from the hidden layer to the output layer. The last term, $\Delta \omega_{jk}/\omega_{jk}$, is the derivative of that sum with respect to the weight connected to the output node $k$. According to equation (4), the last term in equation (8) depends on the output in the hidden node $j$ as:

$$ (\Delta \omega_{jk}/\omega_{jk}) = \gamma_j. $$

(9)

**Input-to-Hidden-Node Weights**

Similarly, the neural networks compute the relative change in the weights (from input nodes to the hidden nodes), $\Delta \omega_{ij}$ as a chain-rule multiplication of five terms as follows:

$$ \Delta \omega_{ij} = (\sum k (\Delta E/d\omega_{kj})/(\delta y_k/d\omega_{ijk})) \cdot (\delta y_k/d\omega_{ijk}) \cdot (\Delta \omega_{ijk}/\omega_{ijk}). $$

(10)

The first two terms on the right side of equation (10) are similar to those in equation (8). The third, $\delta y_k/d\omega_{ijk}$, term is the derivative of the input from the hidden layer to the output layer with respect to the output $k$ to the hidden node. The fourth term, $\Delta \omega_{ijk}/\omega_{ijk}$, is the derivative of the output in the hidden node with respect to the weight input sum of inputs in that node. The fifth term is the derivative of that sum with respect to the weight connecting inputs to the hidden node $j$.

The next task of the neural network is to find the direction of the weight adjustment in which the error surface declines the most. This approach is called the "gradi-

**Predicting Recession Using the Yield Curve**

Given a point in the weight space, including those connecting the input nodes to the hidden nodes and those connecting the hidden nodes to the output nodes, the network produces a certain mean square error. The network must determine the direction and the magnitude of the change in weight result which yields the greatest reduction of error. The direction and magnitude of the weight change is computed according to the following equations:

$$ w_p = w_p - \delta \sum_{s=1}^{N} (\Delta S/d\omega_p) $$

or

$$ \Delta \omega_p = -\delta \sum_{s=1}^{N} (\Delta S/d\omega_p) $$

(11)

(12)

Where $w_p$ is the new weight at process number $p$, $\Delta S$ is the weight change and $\delta$ is the learning coefficient as a parameter which can be set by the user. It actualizes the proportion and the speed of the weight change. Its best value is determined by trial and error.

The last term on the right side of equations (11) and (12) is the total derivative of error with respect to each weight in the weight surface. The negative signs in these equations change the direction of the error so that it declines. With the new weight values, the process continues until the magnitude of the error approaches the minimum point in the error surface.

The "explain" command in the neural network reveals the information about the extent of contributions of each independent variable on the dependent variable. In the literature of the artificial neural network, this measure is called "diluting," which is similar to the measure of elasticity in the econometric approach.

**Empirical Findings**

Quarterly data from the first quarter of 1959 to the first quarter of 1997 for real-GDP growth and the spread between one-year and three-month Treasury bill rates were used for two input mixes. The contemporaneous quarter of the real-GDP growth as the dependent variable (output) and four-quarter lag of each variable in the input mix were utilized (see Figure 1). One input mix has only one independent variable (four-quarter lag of yield spread), and the other has two independent variables (four-quarter lag of the yield spread and four-quarter lag of the real-GDP growth).

Table 1 presents the learning results of the neural networks for each of the two data mixes. The input value column represents results of the last output value of the node where the connection originates. Notice that the value for the bias is one, similar to the value of the intercept in the econometric method. The final estimated weights (parameters) of the connection between the input node and the output are also shown in the table.

The resulting signs of the weight (parameter) of the lag of the Treasury spreads, as expected, are positive for both mixes of data. This suggests that the magnitude of
TABLE 1
Results of the Artificial Intelligence Approach to the Yield Curve as a Predictor of Recession *

<table>
<thead>
<tr>
<th>Input (Independent Variables)</th>
<th>Input Weight Value</th>
<th>Elasticity AVE</th>
<th>(6% Dithering) SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without GDP, 4</td>
<td>+1.0000</td>
<td>-0.0570</td>
<td>9.33     0.053</td>
</tr>
<tr>
<td>Bias</td>
<td>+0.8618</td>
<td>-0.3751</td>
<td>9.33     0.053</td>
</tr>
<tr>
<td>(Yield Spread), 4</td>
<td>+0.8521</td>
<td>-0.0495</td>
<td>9.33     0.053</td>
</tr>
<tr>
<td>With GDP, 4</td>
<td>+1.0000</td>
<td>-0.1148</td>
<td>9.33     0.053</td>
</tr>
<tr>
<td>Bias</td>
<td>+0.6601</td>
<td>-0.4054</td>
<td>9.33     0.053</td>
</tr>
<tr>
<td>(Yield Spread), 4</td>
<td>+0.6908</td>
<td>-0.0719</td>
<td>9.33     0.053</td>
</tr>
<tr>
<td>Hidden</td>
<td>+0.4795</td>
<td>-0.0461</td>
<td>9.33     0.053</td>
</tr>
</tbody>
</table>

a. AVE = Average; SD = Standard Deviation; Number of runs = 20,000.

The yield spread between long- and short-term interest rates moves in the same direction as the growth of real-GDP. That is, a lower yield spread implies a slowdown in economic activity in the next four quarters. Similar conclusions were derived from the results of the traditional regression approach for each of the input mixes.

The neural networks also calculated the relative sensitivity of real-GDP growth (elasticity) for each observation and each data mix. It represents the percentage changes of real-GDP growth as a result of a 1 percent change in the Treasury spread. Table 1 presents the averages of this elasticity measure for all the observations. Accordingly, a 5 percent increase (decrease) in the yield spread resulted in 9.33 percent increase (decrease) in the growth of real GDP, almost double. When the lag of the real-GDP growth is added to the input mix, the elasticity slightly increased. In addition, Table 1 presents the average standard deviation (SD) of the elasticity of real-GDP growth to the changes in those input variables for each observation. The low SD of each of the input mixes implies that the elasticity measure did not vary across observations.

In conclusion, the results of the artificial neural network approach imply that a low or negative yield spread indicates a probable recession within a year.

SIMULATION AND COMPARISON

To test the accuracy and validity of the neural-network prediction, two types of simulations for each of the input mixes were performed: in-sample simulation, and out-of-sample simulation.

In-Sample Simulation

For the in-sample simulation, the whole data from the first quarter of 1959 to the first quarter of 1997 were used to test the predictive power of the yield curve. The
## TABLE 2 (Cont.)

<table>
<thead>
<tr>
<th>Forecasts</th>
<th>Actual</th>
<th>WIT</th>
<th>Regression</th>
<th>Actual</th>
<th>WIT</th>
<th>WIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Art Int</td>
<td>Art In 3</td>
<td>Art In 1</td>
<td>Art Int</td>
<td>Art In 3</td>
<td>Art In 1</td>
</tr>
<tr>
<td>1970.1</td>
<td>5.95</td>
<td>5.51</td>
<td>5.95</td>
<td>1.58</td>
<td>5.51</td>
<td>5.95</td>
</tr>
<tr>
<td>1970.2</td>
<td>5.37</td>
<td>5.51</td>
<td>5.37</td>
<td>1.58</td>
<td>5.51</td>
<td>5.37</td>
</tr>
<tr>
<td>1970.3</td>
<td>4.78</td>
<td>5.51</td>
<td>4.78</td>
<td>1.58</td>
<td>5.51</td>
<td>4.78</td>
</tr>
<tr>
<td>1970.4</td>
<td>4.22</td>
<td>5.51</td>
<td>4.22</td>
<td>1.58</td>
<td>5.51</td>
<td>4.22</td>
</tr>
<tr>
<td>1970.5</td>
<td>3.82</td>
<td>5.51</td>
<td>3.82</td>
<td>1.58</td>
<td>5.51</td>
<td>3.82</td>
</tr>
<tr>
<td>1970.6</td>
<td>3.46</td>
<td>5.51</td>
<td>3.46</td>
<td>1.58</td>
<td>5.51</td>
<td>3.46</td>
</tr>
<tr>
<td>1970.7</td>
<td>3.17</td>
<td>5.51</td>
<td>3.17</td>
<td>1.58</td>
<td>5.51</td>
<td>3.17</td>
</tr>
<tr>
<td>1970.8</td>
<td>2.90</td>
<td>5.51</td>
<td>2.90</td>
<td>1.58</td>
<td>5.51</td>
<td>2.90</td>
</tr>
<tr>
<td>1970.9</td>
<td>2.68</td>
<td>5.51</td>
<td>2.68</td>
<td>1.58</td>
<td>5.51</td>
<td>2.68</td>
</tr>
<tr>
<td>1970.10</td>
<td>2.50</td>
<td>5.51</td>
<td>2.50</td>
<td>1.58</td>
<td>5.51</td>
<td>2.50</td>
</tr>
<tr>
<td>1970.11</td>
<td>2.35</td>
<td>5.51</td>
<td>2.35</td>
<td>1.58</td>
<td>5.51</td>
<td>2.35</td>
</tr>
<tr>
<td>1970.12</td>
<td>2.20</td>
<td>5.51</td>
<td>2.20</td>
<td>1.58</td>
<td>5.51</td>
<td>2.20</td>
</tr>
<tr>
<td>1971.1</td>
<td>2.07</td>
<td>5.51</td>
<td>2.07</td>
<td>1.58</td>
<td>5.51</td>
<td>2.07</td>
</tr>
<tr>
<td>1971.2</td>
<td>1.95</td>
<td>5.51</td>
<td>1.95</td>
<td>1.58</td>
<td>5.51</td>
<td>1.95</td>
</tr>
<tr>
<td>1971.3</td>
<td>1.85</td>
<td>5.51</td>
<td>1.85</td>
<td>1.58</td>
<td>5.51</td>
<td>1.85</td>
</tr>
<tr>
<td>1971.4</td>
<td>1.77</td>
<td>5.51</td>
<td>1.77</td>
<td>1.58</td>
<td>5.51</td>
<td>1.77</td>
</tr>
<tr>
<td>1971.5</td>
<td>1.71</td>
<td>5.51</td>
<td>1.71</td>
<td>1.58</td>
<td>5.51</td>
<td>1.71</td>
</tr>
<tr>
<td>1971.6</td>
<td>1.67</td>
<td>5.51</td>
<td>1.67</td>
<td>1.58</td>
<td>5.51</td>
<td>1.67</td>
</tr>
<tr>
<td>1971.7</td>
<td>1.62</td>
<td>5.51</td>
<td>1.62</td>
<td>1.58</td>
<td>5.51</td>
<td>1.62</td>
</tr>
<tr>
<td>1971.8</td>
<td>1.58</td>
<td>5.51</td>
<td>1.58</td>
<td>1.58</td>
<td>5.51</td>
<td>1.58</td>
</tr>
<tr>
<td>1971.9</td>
<td>1.55</td>
<td>5.51</td>
<td>1.55</td>
<td>1.58</td>
<td>5.51</td>
<td>1.55</td>
</tr>
<tr>
<td>1971.10</td>
<td>1.51</td>
<td>5.51</td>
<td>1.51</td>
<td>1.58</td>
<td>5.51</td>
<td>1.51</td>
</tr>
<tr>
<td>1971.11</td>
<td>1.48</td>
<td>5.51</td>
<td>1.48</td>
<td>1.58</td>
<td>5.51</td>
<td>1.48</td>
</tr>
<tr>
<td>1971.12</td>
<td>1.45</td>
<td>5.51</td>
<td>1.45</td>
<td>1.58</td>
<td>5.51</td>
<td>1.45</td>
</tr>
</tbody>
</table>

Out-of-Sample Simulation

As a better test of the forecast accuracy of the artificial neural networks and the regression models, the out-of-sample simulation was performed. The data was divided into two equal parts—1959.1 to 1973.7 and 1974.1 to 1977.1. The first data set was used to estimate the parameters of each model (neural network, and regression). Then, the input data of the second part were injected into each of the estimated model to forecast the output (real-GDP growth).

The estimated forecast of the real-GDP growth of the neural networks is compared with that of actual real-GDP growth in Table 3 and Figures 7 and 8 for two input mixtures. Similarly, the estimated forecast of the real-GDP growth of the regression approach is compared with that of actual growth in Table 3 and Figures 9 and 10 for two input mixtures.

For comparison, the MSE and the SD of the error for both artificial neural networks and regression models are calculated and shown at the end of Table 3. Accordingly, the MSE for the artificial intelligence method for each of the input mixes (12.8 and 13.68) is far lower than those of the regression approach. 25.33 and 37.64. Thus, the findings of the out-of-sample simulation strongly suggest that the real-GDP growth prediction of the artificial intelligence approach is far more accurate than that of the traditional regression method.

Furthermore, the resulting MSE was higher for each of the input mixtures with the inclusion of the lag of growth for both the neural networks and the regression methods. This outcome suggests that adding a variable(s) in the input mix (over-fitting) does not necessarily increase the accuracy of the forecast of the out-of-sample simulation. In fact, over-fitting can result in less accuracy, as was the case here with a higher MSE.

In addition, the SD of the forecast error of the artificial intelligence model and that of the regression method were calculated and shown in the Table 3. The SD for the model without real-GDP lag is about equal for the artificial intelligence method compared to that of the regression method, while the SD for the model with real-GDP lag is significantly higher, suggesting that the model may be over-fitting the data.
FIGURE 3
Actual and In-Sample Forecast of Real-GDP Growth with Yield Spread as Input: Artificial Intelligence

FIGURE 4
Actual and In-Sample Forecast of Real-GDP Growth with Yield Spread and GDP as Inputs: Artificial Intelligence

FIGURE 5
Actual and In-Sample Forecast of Real-GDP Growth with Yield Spread as Input: Regression Method

FIGURE 6
Actual and In-Sample Forecast of Real-GDP Growth with Yield Spread GDP as Inputs: Regression Method
### TABLE 3

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual WOT</th>
<th>WIT</th>
<th>Regression WOT</th>
<th>WIT</th>
<th>Forecast WIT</th>
<th>Year</th>
<th>Actual WOT</th>
<th>WIT</th>
<th>Regression WOT</th>
<th>WIT</th>
<th>Forecast WIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>78.1</td>
<td>0.6145</td>
<td>4.0908</td>
<td>2.7326</td>
<td>6.3033</td>
<td>7.4146</td>
<td>78.2</td>
<td>0.3243</td>
<td>4.1024</td>
<td>2.5507</td>
<td>6.6847</td>
<td>3.5157</td>
</tr>
<tr>
<td>78.3</td>
<td>1.0465</td>
<td>4.2024</td>
<td>2.404</td>
<td>10.52</td>
<td>7.7464</td>
<td>78.4</td>
<td>1.1845</td>
<td>4.2501</td>
<td>2.4055</td>
<td>6.6857</td>
<td>3.5157</td>
</tr>
<tr>
<td>78.5</td>
<td>1.1455</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
<td>78.6</td>
<td>0.6256</td>
<td>4.2501</td>
<td>2.329</td>
<td>6.6847</td>
<td>3.5157</td>
</tr>
<tr>
<td>78.7</td>
<td>1.2356</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
<td>78.8</td>
<td>0.6256</td>
<td>4.2501</td>
<td>2.329</td>
<td>6.6847</td>
<td>3.5157</td>
</tr>
<tr>
<td>78.9</td>
<td>1.3256</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
<td>79.1</td>
<td>1.4156</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
</tr>
<tr>
<td>79.2</td>
<td>1.5056</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
<td>79.3</td>
<td>1.6956</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
</tr>
<tr>
<td>79.4</td>
<td>1.7856</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
<td>79.5</td>
<td>1.9656</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
</tr>
<tr>
<td>79.6</td>
<td>2.1456</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
<td>79.7</td>
<td>2.3256</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
</tr>
<tr>
<td>79.8</td>
<td>2.4956</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
<td>79.9</td>
<td>2.6756</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
</tr>
<tr>
<td>80.1</td>
<td>2.8456</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
<td>80.2</td>
<td>3.0256</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
</tr>
<tr>
<td>80.3</td>
<td>3.1956</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
<td>80.4</td>
<td>3.3656</td>
<td>4.2905</td>
<td>2.295</td>
<td>7.052</td>
<td>7.7476</td>
</tr>
</tbody>
</table>

### FIGURE 7
Actural and Out-of-Sample Forecast of Real-GDP Growth with Yield Spread as Input: Artificial Intelligence

### FIGURE 8
Actual and Out-of-Sample Forecast of Real-GDP Growth with Yield Spread and GDP as Inputs: Artificial Intelligence

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>1981</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Art INT</th>
<th>WOT</th>
<th>WIT</th>
<th>Regression</th>
<th>Year</th>
<th>Actual WOT</th>
<th>WIT</th>
<th>Regression WOT</th>
<th>WIT</th>
</tr>
</thead>
</table>

*Art INT = Artificial Intelligence; WOT = Without Lag of Real-GDP Growth; WIT = With Lag of Real-GDP Growth; MSE = Mean Square Error; SD = Standard Deviation Error.*
lag is lower for the artificial intelligence approach compared to that of the regression method. Thus, the findings of the out-of-sample forecasts suggest that the artificial intelligence method is more accurate with less error and less variation than those of the regression approach.

SUMMARY AND CONCLUSIONS

Economists typically utilize the shape of the yield curve to predict recessions. Using econometric models they concluded that a flat or a downward-sloping yield curve signals a slowdown in economic activity in the near future. This study used one branch of artificial intelligence model called neural networks as another method to confirm that conclusion and to compare the results with those of the econometric models.

The computation of the artificial network starts with loading a set of values of independent variables into the input layer of the network for each observation. Then, each hidden node calculates the weighted summation of the inputs and incorporates the threshold, or bias, weight. Next, each hidden node computes the sigmoid function of the summation, which represses the value of the summation to a range between one and zero. After that, each hidden node scales the results to the output nodes. In the output layer, computations similar to those in the hidden layer take place, and finally, in that layer, the value of the dependent variable(s) is estimated. The networks then compute the error as the difference between the estimated and the actual dependent variable(s). The output node propagates the amount of its error on each observation back through its links to the hidden nodes. This process continues and, based on this error minimization, the weights are determined.

Using quarterly data from the first quarter of 1969 to the first quarter of 1997, the findings confirm those of the earlier studies employing econometric methodology that a flat or negative-sloping yield curve is a reliable predictor of recession within a year. Furthermore, the findings of the out-of-sample simulation suggest that the forecast of the artificial neural networks is more accurate with less error and lower variation than those of the regression.

NOTES

The author gratefully acknowledges two anonymous referees whose comments helped to improve the content and style of this work. The usual disclaimer applies. This research was made possible by the Office of Sponsored Research and Grants of the University of Central Oklahoma.

1. For more discussion of the theoretical basis of the yield curve as a predictor of recession see Kimball (1997) and Dusten (1997).

2. The yield curve is also used both to gauge whether or not monetary policy is expansionary or restrictive to predict inflation. See for example, Frankel and Levin (1994), and Mathieson (January 1990, August 1990, 1991).

There are three basic questions regarding the topology of the artificial neural networks. The first is the number of hidden layers. When using a functional link, the first step is to use no hidden layer, and to connect the processing elements to both the inputs and the functional link. When accuracy is not achieved, a functional link layer should not be used, and one hidden layer can be added. Normally, most problems such as those with a small number of inputs and one output can be
solved with a single hidden layer, using conventional statistical methods. A theoretical upper boundary for the number of hidden units can be determined by the number of training observations for each weight.

3. The second question is the number of nodes in each hidden layer. Generally, the fewer the number of processing elements in the hidden layers, the better the network will generalize. Overgeneralization is the power to learn and interpolate between previously unseen examples generalized to the future prediction.

Finally, the third question relates to the learning rate and the learning rule. The learning rate for the last hidden layer must be twice that of the output layer. The delta rule measures error and changes the weights for each observation. Two extensions of the delta rule are the cumulative delta rule, which accumulates weight changes over several observations, and the normalized-cumulative delta rule. In general, the normalized-cumulative delta rule works well. Furthermore, the epoch, which is the number of presentations over which weight changes are accumulated, is set to 30, unless the data are very noisy. In this study, the normalized-cumulative delta rule, with epoch of 10 was considered appropriate.

4. The elasticity concept, called "disappearance" in the artificial neural network literature, is not necessarily present.

5. Theoretically, the real business cycle moves with real interest rates. Since real rates are not available, it is assumed that the spread between nominal rates is approximately equal to the spread between real rates. Furthermore, interest rates and spreads are measured on a quarterly average basis. While the precise starting date does not seem to be crucial, the data chosen maximize the availability of comparable data for all series.

6. The regression models used are as follows:

\[ (1) \quad GDP_{t} = a + b \cdot \text{Treasurer Spread}_{t-4} \]

\[ (2) \quad GFR_{t} - GDP_{t} = c + d \cdot \text{Treasurer Spread}_{t-4} + e \cdot GFR_{t} - GDP_{t} - 4 \]

where \( a, b, c, d, \) and \( e \) are the parameters of the equations, and \( t \) is the rate of growth. A similar structure is used for the artificial intelligence model.

7. The estimated results of equation (1N) for the whole period of the first quarter of 1959 to the first quarter of 1997 are as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.96515</td>
<td>0.36870</td>
<td>9.61782</td>
</tr>
<tr>
<td>(Treasurer Spread) (_t-4)</td>
<td>0.723436</td>
<td>0.214975</td>
<td>3.37128</td>
</tr>
</tbody>
</table>

and the estimated results of equation (2N) for the same period are as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.30895</td>
<td>0.44557</td>
<td>7.2958</td>
</tr>
<tr>
<td>(Treasurer Spread) (_t-4)</td>
<td>0.223355</td>
<td>0.160701</td>
<td>2.33415</td>
</tr>
<tr>
<td>GFR - GDP (_t-4)</td>
<td>0.303472</td>
<td>0.085049</td>
<td>3.60310</td>
</tr>
</tbody>
</table>

Accordingly, the \( t \) ratios for (Treasurer Spread) \(_t-4\) are statistically significant for both equations. However, the \( t \) ratio for GFR - GDP \(_t-4\) is not significant.

The estimated results of equation (1N) for the first quarter of 1959 to the third quarter of 1977 is as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.37356</td>
<td>0.566824</td>
<td>4.26956</td>
</tr>
<tr>
<td>(Treasurer Spread) (_t-4)</td>
<td>1.29830</td>
<td>0.460039</td>
<td>2.29100</td>
</tr>
</tbody>
</table>

The predicted results of equation (2N) for the same period is as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.43644</td>
<td>0.603621</td>
<td>3.49758</td>
</tr>
<tr>
<td>(Treasurer Spread) (_t-4)</td>
<td>0.417947</td>
<td>0.115109</td>
<td>2.95313</td>
</tr>
<tr>
<td>GFR - GDP (_t-4)</td>
<td>-0.387286</td>
<td>0.015890</td>
<td>-1.24500</td>
</tr>
</tbody>
</table>

Again, the results of the \( t \) ratios for (Treasurer Spread) \(_t-4\) are statistically significant for both equations. But, the \( t \) ratio for GFR - GDP \(_t-4\) is not significant.

8. Estrella and Mishkin (1990) came to the same conclusion from their regression equations.

REFERENCES


RATIONAL IRRATIONALITY:
A FRAMEWORK FOR THE NEOCLASSICAL-BEHAVIORAL DEBATE

Bryan Caplan
George Mason University

INTRODUCTION

Economists have characterized beliefs as “rational,” if agents satisfy Bayesian probability axioms; or, more strongly, if they also satisfy the rational expectations assumption (Shefrin, 1996; Wittman, 1996). A diverse body of experimental evidence shows that individuals’ beliefs deviate from these standards of rationality [Kahneman, Slovic, and Tversky, 1982; Camerer, 1995, Rabin, 1998]. Critics of these findings argue that the anomalies are suspect because financial incentives were absent or inadequate; people would be more rational—perhaps fully rational—if the monetary rewards were large enough [Harrison, 1992; 1999; 1989; Wittman, 1996; Friedman, 1998; Smith and Walker, 1999]. Defenders of the behavioral perspective reply that anomalies are generally robust to this criticism.

The purpose of the current article is not to resolve this controversy, but to provide a shared framework for debate. I present a model of “rational irrationality” and show that the main positions in the neoclassical-behavioral debate about beliefs are special cases of it. In this model, irrationality is a good like any other, and agents optimize by trading wealth for irrationality. Unless otherwise stated, “irrationality” is interpreted as “deviations from rational expectations,” of which deviation from Bayesian axioms is a subset. The central assumption of the model of rational irrationality is that agents perceive the price of irrationality without bias. On some level, they have rational expectations about the consequences of irrationality, even though they typically hold a positive quantity of irrationality in their consumption bundle.

The upshot is that it is not necessary to see the neoclassical and behavioral approaches as two irreconcilable paradigms. The neoclassical-behavioral dispute over beliefs can instead be seen as a disagreement within “normal science” about parameter values; even if researchers cannot agree about their conclusions, at least they are asking the same questions. At the same time, the rational irrationality model does not tautologically define genuine irrationality away; a key falsifiable implication of the model is that (compensated) demand for irrationality must be decreasing in price. Experimental findings that irrationality increases as monetary incentives rise thus differ in kind from other anomalies and merit special attention [Hogarth and Reder, 1987].

Bryan Caplan  Department of Economics, Center for Study of Public Choices, George Mason University, 4400 University Drive, Fairfax, Virginia 22030-4444. E-mail: brucapla@gmu.edu