10-2018

Predicting recessions using trends in the yield spread

Steven E. Kozlowski
Fairfield University, skozlowski@fairfield.edu

Thaddeus Sim

Published Citation
Kozlowski, Steven E., and Thaddeus Sim. "Predicting Recessions Using Trends in the Yield Spread." Journal of Applied Statistics 46, no. 7 (June 2019): 1323–35. doi:10.1080/02664763.2018.1537364.

APPLICATION NOTE

Predicting Recessions Using Trends in the Yield Spread

Steven E. Kozlowski[a] and Thaddeus Sim[b]

[a]Fairfield University, Dolan School of Business, Fairfield, CT 06824, USA; [b]Bank of New York Mellon, Oriskany, NY 13424, USA

This Draft: July 21, 2018

ABSTRACT
The yield spread, measured as the difference between long- and short-term interest rates, is widely regarded as one of the strongest predictors of economic recessions. In this paper, we propose an enhanced recession prediction model that incorporates trends in the value of the yield spread. We expect our model to generate stronger recession signals because a steadily declining value of the yield spread typically indicates growing pessimism associated with reduced future business activity. We capture trends in the yield spread by considering both the level of the yield spread at a lag of twelve months as well as its value at each of the previous two quarters leading up to the forecast origin, and we evaluate its predictive abilities using both logit and artificial neural network models. Our results indicate that models incorporating information from the time series of the yield spread correctly predict future recession periods much better than models only considering the spread value as of the forecast origin. Furthermore, the results are strongest for our artificial neural network model and logistic regression model that includes interaction terms, which we confirm using both a blocked cross-validation technique as well as an expanding estimation window approach.

KEYWORDS
Forecasting; Classification; Logistic regression; Artificial neural networks; Cross-validation.

JEL CLASSIFICATION
Codes: C45, C52, C53, E43

Contact Author: Steven E. Kozlowski. Email: skozlowski@fairfield.edu

1. Introduction

The yield spread, which is typically defined as the difference between the interest rates of the 10-year Treasury bond and the 3-month Treasury bill, has been shown to be an excellent predictor of recessions [6, 7, 5, 15]. Predicting recessions is an important task, because an unanticipated recession can create significant hardships for society, such as unemployment, declining portfolio values, reduced corporate profits, and bankruptcy. Therefore, an accurate recession forecast allows businesses, governments, and consumers to make decisions to minimize its negative impact. Additionally, because these decisions often take time to impact the economy, it is best if a recession can be predicted far in advance. Consequently, our focus is on developing a model that will generate better forecasts of whether a recession will occur one year in the future.

The existing literature on the prediction of recessions in the United States using the yield spread can be divided into two groups based on the type of dependent variable used in the study. In the first group, continuous measures such as the gross domestic product [21] and real manufacturing and trade sales [14] are used as the dependent variable. The predicted growth or decline in economic activity would then indicate if a recession is on the horizon.

The second group, into which our work belongs, consists of classification models, where the dependent variable is binary. The values of the dependent variable usually reflect the recessionary periods identified by the National Bureau of Economic Research (NBER).[1] [9] provides an overview of the different considerations and issues pertinent to this stream of research, and our discussion in the remainder of this section focuses on studies where the dependent variable is a binary recession indicator.

Initial work on classifying recessions using the yield spread focused on simple probit models where the independent variable is the yield spread measured at a particular forecast horizon; for example, 1, 2, 3, or 4 quarters prior to the time period being forecasted [6]. Later studies examined whether the inclusion of a second independent variable in addition to the yield spread improves the predictive accuracy of the probit models [5, 3, 8, 9, 15, 22]. Examples of these independent variables include gross domestic product, manufacturing and trade sales, income, employment levels, and stock price indexes.

However, because recessions and business cycles in general tend to be asymmetric, several studies have noted that linear, constant-parameter models may not adequately explain and predict these business cycles [18, 1], and more recent work has tended to rely primarily on less restrictive forecasting methodologies. Artificial neural networks (ANN) in particular, which are non-linear, non-parametric models, may be better equipped for modeling business cycles and financial crises [4], and thus predicting recessions as a result of their flexibility.

[18] analyzes ANN models with forecast horizons ranging from 1 to 8 quarters, and 25 different input variables such as interest rates and spreads, stock price indexes, macro indicators, and indexes of leading indicators. While it is found that the yield spread generates the lowest out-of-sample mean squared forecast error (MSFE) among the predictors, and performs best at a forecast horizon of 4 quarters, adding a second predictor variable does, in some instances, improve the explanatory power of the model. However, [18] also notes that the predictors that help improve the forecasting accuracy in one subperiod often differ from those that perform best in another subperiod. As a result, while we also explore the use of ANN models in our study, our focus is on extracting as much information as possible from the yield spread, which has a direct impact on the level of future business activity and has been the most reliable predictor

of future recessions.

Our primary contribution to this stream of research on predicting recessions is to provide evidence that forecasts can be improved by incorporating trends in the yield spread. Recessions do not occur instantaneously; instead, there is often a gradual decline in the economy to a point when one would consider the country to be in a recession. Given that interest rates have a direct impact on economic activity, we expect that trends in the spread will amplify the strength of this signal. For instance, a steadily declining value of the spread can capture growing pessimism associated with reduced future business activity.

We incorporate information produced by trends in the yield spread by including either the twelve month lagged value of the spread and its change over each of the preceding two quarters, or in the case of our ANN models by simply inputting its values at lags of 12, 15, and 18 months.[2] This approach is different from the previously described works, which only consider the value of the yield spread as of the forecast origin and ignore whether its value has been increasing, decreasing, or relatively stable. Accounting for trends in the yield spread should help pick up on broader movements in overall economic activity that could help indicate if a recession in imminent.

To the best of our knowledge, [15] is the only other study to consider using a lagged value of the yield spread to improve recession forecasts; however, their baseline model only includes a single value of the yield spread. As a second predictor, they explored the addition of various financial and macroeconomic variables as well as a single lagged value of the yield spread while relying on probit models to generate recession forecasts.

Our work differs from that of [15] in several important ways. First, our study explicitly focuses on whether trends in the yield spread provide enhanced information for predicting recessions. We incorporate information regarding how the spread's value evolved over multiple points in time leading up to the forecast origin to evaluate its impact. Second, we utilize artificial neural network models as well as a logistic regression model with multiple interaction terms to better extract information from the time series of the yield spread. The greater flexibility and lack of imposed functional form of artificial neural networks in particular has been shown to result in greater predictive ability for a variety of financial applications [13, 23, 16]. Third, we use the accuracy and sensitivity metrics to evaluate model performance in addition to mean absolute error and Brier score, whereas [15] focus exclusively on the area under the receiver operating characteristic curve. We do not view the use of the AUROC as a weakness of [15]; however, all individual performance measures have certain drawbacks. For instance, AUROC provides little information regarding the size of model errors, and because all models achieve very high specificity values even at modest cutoff values, the vast majority of the area under the curve is derived based on performance at trivially low cutoff values.[3]

We examine the impact of yield spread trends within both logistic regression and artificial neural network models using monthly interest rate data from the Federal Reserve on U.S. Treasury securities and business cycle dates from the National Bureau of Economic Research (NBER) from January 1981 to August 2015, which we evaluate through two approaches: a blocked cross-validation approach and an expanding-window out-of-sample forecasting approach. Our findings indicate that the inclusion of yield spread trends results in better overall recession predictions than traditional models that only consider the value of the spread as of the forecast origin. Additionally, an ANN model with a single hidden layer and two hidden-layer nodes and a logistic regression model with interaction terms produce the strongest forecasting results while traditional simple logistic regression models appear to ignore valuable information and larger ANN models are exposed to the risk of potential overfitting.

The remainder of the paper is organized as follows: Section 2 outlines our methodological approach including our logistic regression and artificial neural network models. We describe the data and discuss the results of our empirical analyses in Section 3. Concluding remarks follow in Section 4.

2. Methodology

Our focus is on whether the time series of the yield spread can be used to generate improved recession forecasts; therefore, we construct models that incorporate information regarding how the yield spread has changed over time. We hypothesize that since the yield spread has good predictive power for the future state of the economy, changes in its value should provide additional information that could strengthen the predictive accuracy of existing models. Given findings in [18] that the yield spread has its greatest predictive power at a one-year forecast horizon coupled with the fact that one-year provides sufficient lead time for model users to react, we focus our analysis on models with this forecast horizon.

Let x_t represent the value of the yield spread in period t, and let Δ_{t_1,t_2} denote the difference in the yield spread between time periods t_1 and t_2; that is, $\Delta_{t_1,t_2} = x_{t_2} - x_{t_1}$. With x_{t-12}, x_{t-15}, and x_{t-18} representing the yield spread values at lags of 12, 15, and 18 months, respectively, and the binary variable y_t indicating if period t is a recessionary period ($y_t = 1$) or not ($y_t = 0$), we define model LM which accounts for trends in the yield spread as follows:

$$\begin{aligned}\text{LM:} \quad \ln\left(\frac{y_t}{1-y_t}\right) &= \beta_0 + \beta_1 x_{t-12} + \beta_2 \Delta_{t-15,t-12} + \beta_3 \Delta_{t-18,t-15}. \\ &= \delta_0 + \delta_1 x_{t-12} + \delta_2 x_{t-15} + \delta_3 x_{t-18}.\end{aligned}$$

The first independent variable, x_{t-12}, helps capture the level of the yield spread, which prior work has shown has a tendency to be small or negative prior to economic recessions. The variables, $\Delta_{t-15,t-12}$ and $\Delta_{t-18,t-15}$, provide additional information about how the yield spread has evolved leading up to the forecast origin at time $t-12$ and may contain important information regarding whether the future economic outlook is improving or declining. While the two representations of model LM generate identical forecasts, which is our primary focus, it is worth noting that we cannot include all pairwise differences of the three yield spread lags, as the addition of $\Delta_{t-18,t-12}$ would render the parameters unidentifiable due to perfect multicollinearity.

As a basis for comparison, we also define the following traditional simple logistic regression models with either x_{t-12}, x_{t-15}, or x_{t-18} as the independent variable, where we acknowledge that the latter two models face the disadvantage of a longer forecast horizon:

$$\text{L12:} \qquad \ln\left(\frac{y_t}{1-y_t}\right) = \beta_0 + \beta_1 x_{t-12}.$$
$$\text{L15:} \qquad \ln\left(\frac{y_t}{1-y_t}\right) = \beta_0 + \beta_1 x_{t-15}.$$
$$\text{L18:} \qquad \ln\left(\frac{y_t}{1-y_t}\right) = \beta_0 + \beta_1 x_{t-18}.$$

Because the relationship between the predictive variables may be nonlinear, we also utilize ANN models with each of the three yield spread lags, x_{t-12}, x_{t-15}, and x_{t-18}, as input variables. We expect the less restrictive assumptions required by this non-parametric approach to allow for better utilization of the information in yield spread trends. In general, a neural network model consists of three parts: the input layer, hidden layer, and output layer (Figure 1).

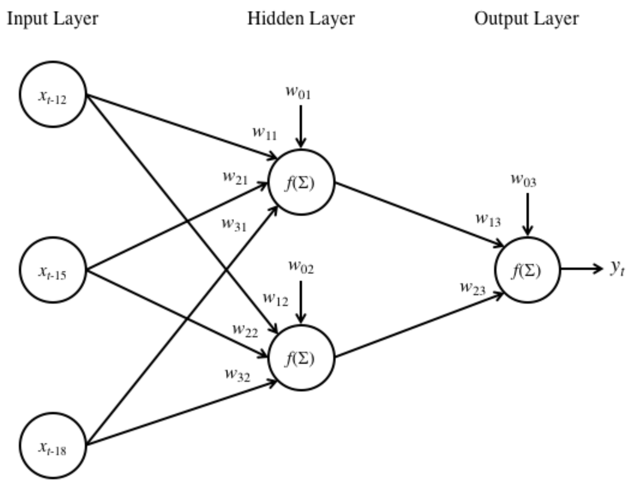


Figure 1. A three-layer artificial neural network with two hidden nodes in the hidden layer.

The input layer is comprised of nodes representing each of the predictor variables. In this study, the input layer has three nodes representing the three yield spread variables x_{t-12}, x_{t-15}, and x_{t-18}. The values of the input nodes are fed into the nodes in the hidden layer.

The hidden layer may contain multiple layers with each layer containing a pre-specified number of hidden nodes. In Figure 1, for example, there is only a single hidden layer, which contains two hidden nodes. Each hidden node combines information from all the input nodes as a weighted sum using weights $w_{1j}, w_{2j}, \ldots, w_{ij}$, and a bias value w_{0j} (where w_{ij} is the weight applied to the value passed from node i in a prior layer to node j in the subsequent layer). Thus, the weights for one hidden node can be different than those for another hidden node.

An activation function f then converts the weighted sum into a value that is fed forward to the next layer, which is either nodes in the next hidden layer or the output layer. We use the logistic sigmoid function as the activation function, which sends a signal with a value between 0 and 1 and is commonly used in binary classification problems.

$$f(\Sigma) = f(w_{0j} + w_{1j}x_{t-12} + w_{2j}x_{t-15} + w_{3j}x_{t-18}) = \frac{1}{1 + e^{-(w_{0j} + w_{1j}x_{t-12} + w_{2j}x_{t-15} + w_{3j}x_{t-18})}}$$

Finally, the output layer contains nodes representing the target variable. In this study, only a single output node is required because the dependent variable y_t is binary.

We explore ANN models with 1 to 10 hidden nodes in the single hidden layer. Each of these models will be denoted as NNh, where h is the number of hidden nodes in the model. We also explored ANN models with two hidden layers, but the results of these models were similar to those with only a single hidden layer. As such, we opted for the simpler single-hidden-layer model so as to minimize any potential issues with overfitting.

Although the multiple logistic regression model, LM, and the ANN models use an equal number of predictor variables, the ANN models possess the advantage of having more estimated parameters via their weights and biases that can be used to potentially generate better forecasts. While we acknowledge that not having to explicitly model a complex, unknown functional form is an advantage of the ANN models, we do explore logistic regression models that include interaction terms in order to provide a fairer comparison. We include the results for the interaction model, IM, shown below:[4]

$$\begin{aligned} \text{IM:} \quad \ln\left(\frac{y_t}{1-y_t}\right) &= \beta_0 + \beta_1 x_{t-12} + \beta_2 x_{t-15} + \beta_3 x_{t-18} + \beta_4 x_{t-12} * x_{t-15} \\ &+ \beta_5 x_{t-15} * x_{t-18} + \beta_6 x_{t-12} * x_{t-18}. \end{aligned}$$

3. Empirical Analyses

Our empirical tests use monthly recession data starting from January 1981 with the corresponding predictor variables available as of January 1980. We choose to generate forecasts beginning in 1980 because of dramatic changes in monetary policy announced by the Federal Reserve on October 6th, 1979 intended to combat high inflation rates. The policy changes involved a temporary shift of focus from controlling the level of the federal funds rate towards targeting monetary aggregates, and inflation has remained considerably more stable from 1980 until the present. As a result, including forecasts prior to 1980 may be inappropriate because economic conditions pre- and post-1980 were substantially different.

[17] lists the time periods for all recessions starting from June 1857. We use this information to code the monthly y_t variable to indicate if period t is a recessionary period or not. Monthly interest rates for the 10-year Treasury bond and 3-month Treasury bill were obtained from the *Federal Reserve Statistical Release H.15*.[5] The yield spread for each month is then calculated by taking the difference of the two interest rates. In total, our data set contains 416 records representing the months from January 1981 to August 2015.[6]

We first use a blocked cross-validation approach to evaluate the predictive performance of the logit and ANN models. We partition our sample into four equal-sized uninterrupted time series, or blocks, which each contain a significant recession and serve as validating sub-samples. This is different from previous work in this area in which a single hold-out data set is used to determine the best predictive model and is the method that [2] recommend become the standard for time series evaluation. We remove observations on each side of each validation set from its corresponding training set to ensure their independence as previously proposed in [19]. In prior studies, a predictive model is typically trained on data prior to a certain date, and the model is then tested on data beyond that date. An issue with this approach is that a particular model might perform better than another model in this single pair of training and test data sets, but it may not be better than the other model when applied to other instances. Cross-validation overcomes this overfitting issue, because it is a repeated hold-out method that evaluates the performance of a model using multiple holdout samples.

We also use an expanding-window out-of-sample forecasting approach to evaluate the performance of our models that incorporate trends in the yield spread as well as our baseline simple logistic regression models. This expanding-window approach is similar to those used by many previous studies in the literature and provides the benefit of mimicking how the predictive models would be utilized in practice.

The sensitivity and specificity metrics are commonly used to evaluate the performance of predictive models. For our study, sensitivity measures how well a model correctly predicts recessionary periods while specificity measures how well it predicts non-recessionary (i.e. expansionary) periods. Because the objective is to correctly predict recessionary periods, and there is likely far less harm done in producing a false warning signal before times of expansion than failing to identify an oncoming recession, the more relevant performance metric is sensitivity, which is also known as the true positive rate. However, it is important to ensure acceptable levels for both metrics are achieved. Additionally, for our expanding-window forecasting tests, we report each model's mean absolute error and Brier score with the latter being equivalent to the sole performance measure used in [18].

3.1. Blocked Cross-Validation Approach

Table 1 outlines the design of our blocked cross-validation tests. We divide our full sample period of 416 months evenly into four consecutive 104-month blocks, and although recessionary periods represent the minority in each block, all blocks contain a significant economic recession. This setup helps evaluate how the models might perform in predicting a new recession, as each validation set serves to evaluate each model's performance on new business cycle data it has never seen before, and it also provides a look at how the models perform within various subperiods.

Rather than using a single holdout period like traditional time series forecasting tests, each block is used as the validation sample once with the remaining blocks used to train each model. We ensure the independence between each training set and its validation set by ensuring the closest month in the training set is 12 months away. That is, we remove 11 months on each side of the validation sets from their training sets.[7] For example, Block 1 contains the first 104 records in our dataset, and when it serves as the validation dataset time periods 116 through 416 are used to train the models. The removal of periods 105 through 115 from the training set ensures there is no information leakage.

In our cross-validation as well as in subsequent tests, we use a cutoff threshold of

Table 1. Business cycle data and blocked cross-validation design.

	Total Periods	Recession	Expansion	Train	Validate
Block 1	104	16	88	116:416	1:104
Block 2	104	8	96	1:93; 220:416	105:208
Block 3	104	8	96	1:197; 324:416	209:312
Block 4	104	18	86	1:301	313:416
Total	416	50	366		

0.5 to indicate if the model predicts a recession or not; that is, if the predicted probability of a recession is at least 0.5, we classify that particular month's prediction as a recessionary period. This requires that models produce a sufficiently strong warning signal that a recession is coming to increase their sensitivity, so that its users would think a recession is more likely to occur than not and would be more likely to take corrective measures.[8]

All of the training and testing of the predictive models were performed using the statistical software R. We used the glm function for the logit models and the neuralnet package for the ANN models [12, 10]. For the neuralnet function, other than setting the maximum number of iterations to 5 million, we used the function's default settings, which uses the logistic function as the activation function and finds the best parameter values for the ANN using an algorithm that is based on resilient back-propagation with weight backtracking [20].

Table 2 reports the sensitivity and specificity values across all four blocks for each of the predictive models. Based on the sensitivity metric, model L12 records the highest sensitivity value among the three simple logistic regression models by correctly identifying an average of 29.3% of recession months. This is not surprising as the other two models generate predictions with a longer forecast horizon and merely serve as benchmarks. Logistic regression models LM and IM, which take into account trends in the yield spread, achieve slightly higher average sensitivity values of 31.8% and 37.8%, respectively, while models NN2 and NN3 correctly classify the highest average percentage of recession months among all models with sensitivity values of 43.4% and 43.8%, respectively. Although model NN3 achieves the highest average sensitivity value it fails to correctly classify any of the recession months in Block 2, which contains the 1990–1991 recession – a recession that has been noted in prior studies as being triggered by unusual events and particularly challenging to predict. For this reason, based on sensitivity alone we prefer model NN2, as it correctly identifies part of the recession in each block. In fact, its eight highest predicted recession probabilities within Block 2 are for the eight recession months.

While we want models to correctly identify the recession periods, for the models to be useful the number of false positives must also be low. The specificity columns reveal that the simple logistic regression models correctly classify the highest percentage of non-recessionary months, but this is attributable to their reduced tendency to generate recession predictions overall. Among the neural network models, model NN2 has a higher specificity than NN3 at 93.1% compared to 89.4%, so this also contributes to our considering NN2 the best performing ANN model. While we acknowledge that determining the optimal neural network size is a difficult task – one that we do not

Table 2. Sensitivity and specificity values for all blocks for each model in our blocked cross-validation process. The models include three simple logistic regression models (L12, L15, and L18), a multiple logistic regression model (LM), a multiple logistic regression model with interaction terms (IM), and ten neural network models denoted by NN# where the number denotes the number of nodes in the single hidden layer.

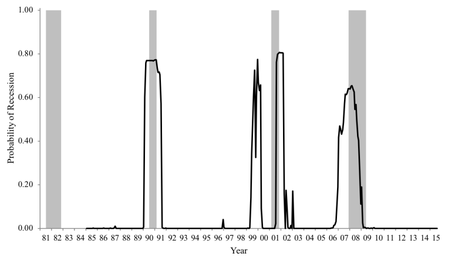
	Sensitivity					Specificity				
Block:	1	2	3	4	Avg	1	2	3	4	Avg
L12	56%	0%	50%	11%	29.3%	95%	100%	94%	99%	97.0%
L15	63%	0%	13%	0%	18.8%	92%	100%	94%	100%	96.4%
L18	63%	0%	0%	0%	15.6%	89%	100%	95%	100%	95.9%
LM	56%	0%	38%	33%	31.8%	91%	100%	93%	99%	95.6%
IM	50%	0%	63%	39%	37.8%	94%	100%	79%	94%	91.9%
NN1	50%	0%	50%	61%	40.2%	92%	100%	86%	94%	93.2%
NN2	50%	25%	38%	61%	43.4%	88%	100%	91%	94%	93.1%
NN3	75%	0%	50%	50%	43.8%	95%	93%	81%	88%	89.4%
NN4	44%	0%	50%	61%	38.7%	91%	99%	85%	93%	92.1%
NN5	38%	0%	50%	17%	26.0%	86%	88%	88%	91%	88.0%
NN6	63%	0%	38%	17%	29.2%	85%	90%	88%	91%	88.3%
NN7	44%	13%	38%	44%	34.5%	91%	88%	84%	91%	88.4%
NN8	44%	0%	25%	61%	32.5%	97%	93%	83%	91%	90.8%
NN9	19%	13%	38%	0%	17.2%	90%	85%	88%	94%	89.2%
NN10	31%	25%	38%	22%	29.0%	97%	88%	85%	91%	90.1%

definitively tackle – the cross-validation results suggest a smaller sized hidden layer is sufficient given our number of input variables and the complexity of our problem and that including many hidden layer nodes could lead to overfitting. As a result, we focus on model NN2 in our subsequent tests as we expand our analysis of model performance while acknowledging that similar sized neural network models generally yield comparable performance.

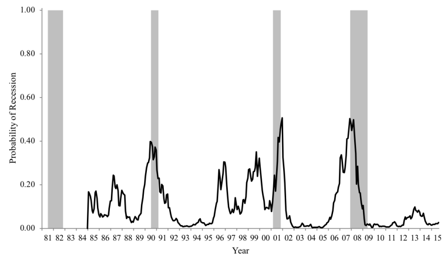
3.2. Expanding-Window Out-of-Sample Forecasting Approach

Figure 2 shows the probability of recession over time as predicted by the artificial neural network model, NN2, as well as the logistic regression models L12, LM, and IM, from using an expanding-window out-of-sample forecasting approach, which attempts to illustrate how the models would have performed in real-time.[9] The grey shaded time periods are the recessionary periods identified by NBER. To obtain the predicted probability values, we first trained each model using the first three years of monthly data (January 1981 to December 1983), and then generated the 12-month-ahead predicted probability of recession for December 1984 using the fitted model. Next, we retrained each model using the 37 months of data from January 1981 to January 1984 and generated the recession probability forecast for January 1985. We continued this expanding-window out-of-sample forecasting process until we obtained the predicted probabilities of recession for the final month in our sample, August 2015.

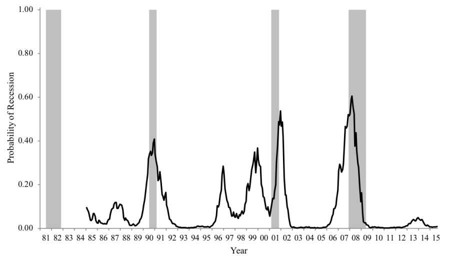
Figure 2a displays the results for the neural network model, NN2, and illustrates several interesting attributes. First, we observe that the predicted recession probabilities in most periods are either large or close to zero. This suggests that the model produces strong warning signals when a recession is probable and that the results are



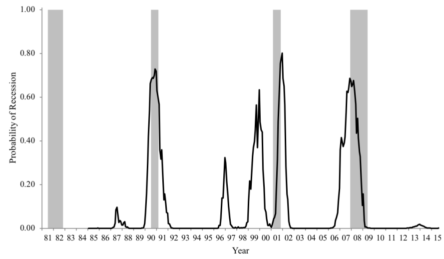
(a) Model NN2: Artificial neural network model with two hidden-layer nodes.



(b) Model L12: Logistic regression with yield spread at 12 month lag.



(c) Model LM: Logistic regression incorporating yield spread trends.



(d) Model IM: Logistic regression incorporating yield spread trends and interaction terms.

Figure 2. Recession probability forecasts generated by the artificial neural network model, NN2, and logistic regression models L12, LM, IM, from the expanding-window out-of-sample forecasting process. The grey shaded regions represent the actual recession periods identified by NBER.

relatively insensitive to the choice of the cutoff threshold.

Second, the model correctly classifies all or part of each recession. It succeeds in identifying all months in the 1990–1991 recession, the later part of the 2001 recession, and the first eight months in the 2007–2009 financial crisis. Although the model generates a few false positives towards the end of 1999 and beginning of the year 2000, a recession did occur approximately one year later, and the predicted probability of recession is near zero throughout most of the sample period.

The recession forecasts from the traditional simple logistic regression model, L12 (Figure 2b), also exhibit a spike near each of the recession periods in our sample, illustrating the yield spread's well-documented predictive ability; however, the predicted probability never exceeds 0.5 during any of the recession periods, and there are multiple instances where the probability increases to between 0.2 and 0.5 during non-recession periods. Overall, while the yield spread is a strong recession predictor as noted in prior studies, the model appears to produce a much noisier signal of oncoming recessions. The corresponding figure for model LM (Figure 2c) shows that while model LM generates slightly stronger recession signals at the beginning of the 2007–2009 financial crisis compared to model L12 and predicts a slightly lower probability of recession during the period of expansion during the mid to late 1980s, the overall pattern is fairly similar. Model IM (Figure 2d), however, produces noticeably stronger recession signals during each of the recession periods with relatively few spikes during non-recessionary periods.

Table 3 summarizes the sensitivity and specificity as well as the mean absolute error

Table 3. Performance metrics of the different forecasting models from the expanding-window out-of-sample forecasting process.

Model	Sensitivity	Specificity	MAE	Brier Score
L12	0.00%	99.40%	0.146	0.067
LM	17.65%	99.40%	0.125	0.061
IM	50.00%	95.52%	0.107	0.060
NN2	55.88%	91.94%	0.101	0.068

(MAE) and Brier score of the expanding-window forecast probabilities generated by the models shown in Figure 2.[10]

The simple logistic regression model achieves a sensitivity of 0% with a specificity of 99.40%. The model only generates a predicted probability of recession above 0.5 in two months, neither of which were during recessions; thus, the model fails to correctly classify any of the recession periods but classifies all but two of the non-recession periods correctly. Model LM achieves a higher sensitivity of 17.65% while correctly classifying the same percentage of non-recessionary periods as indicated by its specificity. Additionally, while models IM and NN2 achieve slightly lower specificity, they both identify at least half of the recession periods with specificity values of 50.00% and 55.88%, respectively. Given the greater costs associated with failing to foresee and plan for a period of recession than expansion, the performance of models IM and NN2 appear to be the most favorable.

The MAE and Brier score are also used to evaluate the performance of our various models, and while it is ideal that our preferred model is able to correctly classify recessions at a natural cutoff of 0.5, these measures offer an assessment that is independent of the choice of cutoff value. Based on the MAE measure, the models that incorporate yield spread trends, LM, IM, and NN2; outperform model L12 with mean absolute errors of 0.125, 0.107, and 0.101, respectively, compared to 0.146. Looking at Brier score, models LM and IM again show improvement over model L12 with mean squared errors of 0.061 and 0.060 compared to 0.067, but by this metric model NN2 actually fares slightly worse. However, this is largely caused by its large recession forecasts immediately preceding the start or following the end of the official recession periods. As a robustness test we removed the 6 months before and after each recession from the calculation to avoid penalizing models for producing early warning signals or predicting that a recession will persist for a while longer than it officially does given that a post-recession recovery is generally not felt immediately. In unreported results, we find this adjustment leads to a slight increase in the Brier score of model L12 to 0.068 while the Brier score for models LM, IM, and NN2, fall to 0.060, 0.053, and 0.055, respectively. Altogether, the evidence suggests model performance can be improved by accounting for trends in the yield spread, and this information is best incorporated using parametric models with interaction terms or non-parametric models such as artificial neural networks. Such evidence is expected to be of particular interest to both businesses and policymakers.

4. Conclusions

Prior research has consistently found the yield spread to be among the strongest predictors of future economic activity. In this paper, we expand on this stream of research by constructing predictive models that take into account how the yield spread has changed over time. The basic intuition is that since the yield spread contains a significant amount of predictive power regarding the future health of the economy, the time series evolution of the spread is likely to provide a stronger signal of recessions than its value at a single point in time and may better capture the overall trends in market sentiment and economic activity.

We use both logistic regression models (i.e., parametric models) and artificial neural network models (i.e., non-parametric models) to evaluate whether yield spread trends, measured based on the twelve, fifteen, and eighteen month lagged values of the spread, can be utilized to enhance recession forecasts. The empirical results show that our augmented models predict recessionary periods better than those that only consider the level of the spread as of the forecast origin. The new models achieve much higher sensitivity values while still maintaining high values for specificity and generally producing much lower forecast errors. In particular, the artificial neural network model with a single hidden layer of two nodes correctly classifies part of each recession in our blocked cross-validation tests and successfully predicts the highest percentage of recessionary periods in our expanding-window out-of-sample forecasting exercise. A logistic regression model with multiple interaction terms achieves comparable overall performance; however, the artificial neural network offers model users the benefit of not having to explicitly model the interactions.

Notes

[1]The NBER business cycle dates offer the benefit of not relying on a specific indicator of economic activity and instead focusing on economic activity as a whole. The business cycle dating committee states that it "examines and compares the behavior of various measures of broad activity: real GDP measured on the product and income sides, economy-wide employment, and real income", and may also consider additional indicators.

[2]The two approaches naturally yield identical probability forecasts in the case of the logistic regression models and have minimal effect on our ANN model results. For instance, we computed out-of-sample ANN probability forecasts using both sets of inputs and found their correlation exceeded 0.98.

[3]In unreported results we compute AUROC for our different models and find this metric generally ranks the models similarly to our other measures.

[4]As with the logit model, LM, an equivalent representation in terms of yield spread changes can be made for our interaction model. Given our focus on forecast performance rather than interpreting the individual coefficients, we opt for this straightforward model representation.

[5]Business cycle data available at http://www.nber.org/cycles/cyclesmain.html and U.S. Treasury interest rate data available at https://www.federalreserve.gov/releases/h15/.

[6][11] notes that official NBER business cycle dates are announced with a considerable lag. For instance, the start date of the recession which officially began in December 2007 was not announced until December 2008. Our models attempt to identify these periods of expansion or contraction in economic activity one year before they occur, and it is even longer before they are officially defined.

[7]The residual autocorrelation is small and statistically insignificant for all models well before 12 months; however, removing fewer months has minimal effect on the results. As noted in previous work in this area, the forecasting error is not usually a white noise process and may have temporal relationships; however, this does not create problems in our study since we do not make any statistical inference based on the forecasting errors.

[8]We report additional statistics that do not involve a choice of cutoff for our expanding-window tests but only report sensitivity and specificity here in the interest of space.

[9]We omit the figures for models L15 and L18 in the interest of space and to focus solely on the models with a one-year forecast horizon.

[10]Most prior studies rely on a single performance measure. We choose to report a variety of measures, because each provides additional information and, as noted in [2], all individual measures have certain shortcomings.

References

[1] E. Andersson, *Monitoring cyclical processes. a non-parametric approach*, Journal of Applied Statistics 29 (2002), pp. 973–990.

[2] C. Bergmeir and J.M. Benítez, *On the use of cross-validation for time series predictor evaluation*, Information Sciences 191 (2012), pp. 192–213.

[3] H. Bernard and S. Gerlach, *Does the term structure predict recessions? The international evidence*, International Journal of Finance & Economics 3 (1998), pp. 195–215.

[4] Y. Demyanyk and I. Hasan, *Financial crises and bank failures: A review of prediction methods*, Omega 38 (2010), pp. 315–324.

[5] M.J. Dueker, *Strengthening the case for the yield curve as a predictor of U.S. recessions*, Federal Reserve Bank of St. Louis Review 79 (1997), pp. 41–51.

[6] A. Estrella and G.A. Hardouvelis, *The term structure as a predictor of real economic activity*, The Journal of Finance 46 (1991), pp. 555–576.

[7] A. Estrella and F.S. Mishkin, *The yield curve as a predictor of U.S. recessions*, Current Issues in Economics and Finance 2 (1996), pp. 1–6.

[8] A. Estrella and F.S. Mishkin, *Predicting U.S. recessions: Financial variables as leading indicators*, Review of Economics and Statistics 80 (1998), pp. 45–61.

[9] A. Estrella and M.R. Trubin, *The yield curve as a leading indicator: Some practical issues*, Current Issues in Economics and Finance 12 (2006), pp. 1–7.

[10] S. Fritsch and F. Güenther, *neuralnet: Training of neural networks*, 2012. Available at https://CRAN.R-project.org/package=neuralnet, R package version 1.32.

[11] V. Golosnoy and J. Hogrefe, *Signaling nber turning points: a sequential approach*, Journal of Applied Statistics 40 (2013), pp. 438–448.

[12] F. Güenther and S. Fritsch, *neuralnet: Training of neural networks*, The R journal 2 (2010), pp. 30–38.

[13] D. Indro, C. Jiang, B. Patuwo, and G. Zhang, *Predicting mutual fund performance using artificial neural networks*, Omega 27 (1999), pp. 373–380.

[14] A. Joseph, M. Larrain, and E. Singh, *Predictive ability of the interest rate spread using neural networks*, Procedia Computer Science 6 (2011), pp. 207–212.

[15] W. Liu and E. Moench, *What predicts U.S. recessions?*, International Journal of Forecasting 32 (2016), pp. 1138–1150.

[16] R. Malhotra and D.K. Malhotra, *Evaluating consumer loans using neural networks*, Omega 31 (2003), pp. 83–96.

[17] National Bureau of Economic Research, *U.S. business cycle expansions and contractions*, 2015. http://www.nber.org/cycles/cyclesmain.html.

[18] M. Qi, *Predicting U.S. recessions with leading indicators via neural network models*, International Journal of Forecasting 17 (2001), pp. 383–401.

[19] J. Racine, *Consistent cross-validatory model-selection for dependent data: hv-block cross-validation*, Journal of econometrics 99 (2000), pp. 39–61.

[20] M. Riedmiller, *Advanced supervised learning in multi-layer perceptrons – from backpropagation to adaptive learning algorithms*, International Journal of Computer Standards and Interfaces 16 (1994), pp. 265–278.

[21] M. Shaaf, *Predicting recession using the yield curve: An artificial intelligence and econometric comparison*, Eastern Economic Journal 26 (2000), pp. 171–190.

[22] J.H. Wright, *The yield curve and predicting recessions*, 2006. Finance and Economics Discussion Series, Federal Reserve Board of Governors.

[23] J. Yao, Y. Li, and C.L. Tan, *Option price forecasting using neural networks*, Omega 28 (2000), pp. 455–466.