International Journal of Forecasting 32 (2016) 1138–1150

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

What predicts US recessions?

Weiling Liu [a], Emanuel Moench [b,*]

[a] Harvard Business School, Wyss House, Boston, MA 02163, United States

[b] Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany

ARTICLE INFO

Keywords:
Recession predictability
ROC
Term spread
Leading indicators

ABSTRACT

We reassess the in- and out-of-sample predictability of US recessions at horizons of three months to two years ahead for a large number of previously proposed leading indicator variables, using the Treasury term spread as a benchmark. We estimate both univariate and multivariate probit models, and evaluate the relative model performance based on the receiver operating characteristic (ROC) curve. At the three- and six-month-ahead horizons, various alternative predictor variables increase the accuracy of recession forecasts significantly relative to the term spread, with the annual return on the S&P500 index providing the strongest improvement. While the Treasury term spread is more difficult to outperform systematically at longer horizons, manufacturers' new orders of capital goods and balances in Broker-Dealer margin accounts increase the precision of recession predictions significantly at horizons of more than one year.

1. Introduction

The accurate prediction of business cycle turning points, and impending economic recessions in particular, is of great importance to households, businesses, investors and policy makers alike. Prior research has shown that a variety of economic and financial variables contain predictive information about future recessions in the United States. Most prominently, Estrella and Hardouvelis (1991) and Estrella and Mishkin (1998) have documented that the slope of the term structure of Treasury yields has strong predictive power for US output growth and US recessions at horizons of up to eight quarters into the future. Other variables that have been considered as leading recession indicators include stock prices (Estrella & Mishkin, 1998), the index of leading economic indicators (Berge & Jordà, 2011; Stock & Watson, 1989), credit market activity (Levanon, Manini, Ozyildirim, Schaitkin, & Tanchua, 2015), and various employment and interest rate measures (Ng, 2014).

In this paper, we reassess the predictability of US recessions since 1959 using a wide variety of leading indicator variables that have been considered in the literature by both academics and practitioners. In doing so, we pay special attention to the question of whether or not these leading indicators contain predictive information about future recessions over and above the Treasury term spread. Consistent with most of the prior literature, we use the business cycle dating chronology provided by the National Bureau of Economic Research (NBER) as the benchmark series of business cycle turning points. While the NBER recession indicator is a binary variable, most leading indicators have continuous distributions. Thus, much of the empirical literature has used the nonlinear *probit* model to map the changes in predictor variables into recession forecasts, and we follow this tradition.

The probability of recession implied by the probit model is rarely exactly zero or one. Therefore, when assessing whether or not a given probit model predicts

* Corresponding author. Tel.: +49 69 9566 2312.
E-mail addresses: wliu@hbs.edu (W. Liu), emanuel.moench@bundesbank.de (E. Moench).

http://dx.doi.org/10.1016/j.ijforecast.2016.02.007
0169-2070/

recessions and expansions correctly, one has to define a cutoff value for the fitted probability above which one interprets the model forecast as implying a recession. Thus, an objective evaluation of a model's ability to categorize future time periods into recessions versus expansions over an entire spectrum of different cutoffs requires the probit model to be complemented with a classification scheme. A classification scheme that has been being used for a long time in the statistics literature but has only recently found its way into economic research is the receiver operating characteristic (ROC) curve (see for example Jordà & Taylor, 2011, 2012; Khandani, Kim, & Lo, 2010; Lahiri & Wang, 2013).

The computation of the ROC curve involves several steps. First, for each point in a given grid of cutoff values of the implied recession probability, one calculates the percentages of true positives and false positives from classifying all periods in the sample. Then, one plots the collection of true and false positives against one another across the entire grid, to create the receiver operating curve. Finally, a common method of summarizing the forecast performance implied by each ROC curve is to integrate the area under the curve, abbreviated as AUROC. For example, a model which delivers a perfect classification of all recessions and expansions would only have true positives, giving an AUROC of one. In contrast, a model which is the equivalent of a random guess would have equal numbers of true and false positives on average, corresponding to an AUROC of 0.5. As the residuals from probit regressions that use a recession indicator as the dependent variable are likely to be autocorrelated, inference on the classification ability using the AUROC should take serial correlation into account. Here, we do so using the block bootstrap approach of Politis and Romano (1994).

We assess the predictability of recessions both for the full sample from 1959 to 2011 and in an out-of-sample analysis which uses the estimated parameters from the 1959–1985 sample to predict recessions after 1985. Our main findings can be summarized as follows. The predictive power of the Treasury term spread for future recessions peaks at the one-year-ahead horizon. Thus, it may not be surprising that adding lagged observations of the term spread improves the predictability substantially at shorter horizons. That being said, though, the annual return on the S&P500 index, when considered jointly with the term spread, improves the predictability of recessions further at forecast horizons shorter than one year ahead.

Furthermore, our analysis shows that the predictive ability of the term spread can be improved upon significantly at horizons longer than one year ahead by two main variables: new orders of non-defense capital goods and margin debit at NYSE brokers and dealers. As a measure of leverage in the financial sector, the latter variable highlights the importance of financial intermediary balance sheet conditions in the transmission of economic shocks (see for example Adrian, Moench, & Shin, 2010; Adrian & Shin, 2010). While the importance of financial intermediary leverage for the pricing of risk has been documented empirically by Adrian, Etula, and Muir (2014) and Adrian, Moench, and Shin (2014), its usefulness as a recession predictor has not been discussed in the academic literature previously, to the best of our knowledge.

Our paper is related to a large body of literature on the use of financial and macroeconomic leading indicators for predicting real output growth and recessions. Estrella and Hardouvelis (1991) were the first to popularize the Treasury term spread as a predictor of future output growth and recessions. They found that it has greater predictive power than the leading indicator index and outperforms survey forecasts both in- and out-of-sample. Estrella and Mishkin (1996) and Estrella and Mishkin (1998) considered the out-of-sample performances of a range of macroeconomic and financial variables in both univariate and multivariate models. Their findings suggest that stock returns are a valuable leading indicator in the short run, but that the Treasury term spread is still the single best performing predictor at horizons of one year ahead or more. Dueker (1997) revisited the term spread as a leading indicator within the context of the probit model studied in our paper. Confirming earlier results, he found the term spread to be the single best recession predictor and showed that the findings are robust to the augmentation of the probit model with lagged dependent variables and Markov switching. Chauvet and Potter (2005) examined further extensions of the yield curve probit model, including a business cycle dependent model, a model with autocorrelated errors, and combinations of these extensions. They concluded that the more sophisticated models capture the predictive instability of the yield curve better because they allow for breakpoints.

While all of the papers cited above have studied the predictive power of the term spread for output growth and recessions within the US, other work has documented government bond yield spreads internationally as having similarly strong predictive power. For example, Duarte, Venetis, and Payà (2005) found yield spreads to predict recessions in the European Monetary Union. Moreover, examining both the US and Germany, Nyberg (2010) concluded that the domestic term spread remains the best recession predictor.

Recently, Rudebusch and Williams (2009) found that the term spread to consistently outperform even professional forecasters in predicting recessions. This is surprising because these forecasters have a wealth of information and many other indicators available to them. Croushore and Marsten (2014) confirm that Rudebusch and Williams' findings are robust across several dimensions, including the sample choice, the use of rolling regression windows, and various measures of real output. Moreover, Lahiri, Monokroussos, and Zhao (2013) report that the result remains valid even after augmenting the model further with factors extracted from a large macroeconomic dataset. These papers' findings highlight the singular importance of the Treasury term spread as a predictor of recessions, and justify our use of this indicator as the benchmark predictor variable.

Methodologically, our paper borrows from the work of Berge and Jordà (2011), who use the AUROC to both validate the NBER's business cycle chronology and investigate which economic indices work best as classification mechanisms for recessions. They find no support for the idea of other indicators achieving statistically significant improvements relative to the NBER dates. Hence, their results support our use of the NBER business cycle chronology for examining the classification abilities of the various variables that we consider. However, we use an updated and considerably wider set of leading indicators than Berge and Jordà (2011), including novel additions like broker dealer debit margins. Moreover, while they use a univariate and nonparametric ROC test, we combine sets of indicators using a parametric probit model.

Our paper is organized as follows. Section 2 discusses the empirical methodology that is used to predict recession probabilities and evaluate the classification of future recession and expansion periods. Section 3 provides a description of the various recession indicators used in our analysis. Section 4 summarizes the in-sample and out-of-sample recession prediction results. Finally, Section 5 provides a discussion of the empirical findings.

2. Methodology

In this section, we provide a brief description of the empirical methods used in the paper. We start by revisiting the standard probit model that we use to estimate the recession probabilities as functions of observable predictor variables. We then briefly discuss an extension which allows for the inclusion of partially unobserved predictor variables. Finally, we describe the AUROC measure and related statistical tests which we employ to discriminate between models.

2.1. Predicting recessions

The state of the business cycle is a binary variable, taking the value of one during recessions and zero during expansions. On the other hand, most leading indicators are continuous variables. In order to account for this, a common tool for predicting recessions is the probit model (see Estrella & Hardouvelis, 1991; Estrella & Mishkin, 1996; Estrella & Trubin, 2006; Wright, 2006), which allows a mapping from a set of continuous explanatory variables onto a binary dependent variable. While other methods for predicting binary response variables are available, we restrict ourselves to this popular class of models because of its simplicity and ease of use.[1]

Some of the predictor variables that we consider in our empirical analysis are not observed over the full sample period. We therefore need to adjust the probit model to allow for missing observations. One commonly-used method of handling missing data is to disregard all dates on which any variables are missing, but this method inefficiently discards potentially useful data. Instead, we employ the "probitmiss" estimator, proposed by Conniffe and O'Neill (2011), which allows us to incorporate all relevant data. We summarize this estimator and its implementation in the Appendix.

[1] As the probit model is a standard tool in time series econometrics, we do not review it here. Instead, we refer the interested reader to, e.g., Estrella and Hardouvelis (1991) for a brief account of the use of the model in the context of predicting recessions.

2.2. Model evaluation

Previous research has used various different metrics to evaluate the fits of recession prediction models. For example, Moore and Shiskin (1967) present an explicit scoring system for business cycle indicators, focusing on the length of the lead before business cycle turns, the smoothness of the series, the clarity of cyclical movements, and the relationship to general business activity, among other criteria. Estrella and Mishkin (1996) and Estrella and Mishkin (1998) use the pseudo R^2 to evaluate the fits of probit models. Finally, Wright (2006) employs the BIC to measure the fit of his in-sample model and root mean squared forecast errors to evaluate the fit of his out-of-sample forecasts.[2]

However, all of these evaluation measures focus on the model fit, not specifically the classification ability, which is the object of interest in our application. Berge and Jordà (2011) recently used the receiver operating characteristic (ROC) curve to assess the recession classification abilities of various leading indicators. One can use the area under the ROC curve (AUROC) to evaluate the categorization ability of the model of interest over an entire spectrum of different cutoff points for determining a recession, instead of evaluating the predictive power at only one arbitrary threshold.

The basic ROC methodology was first developed in a seminal paper by Peterson and Birdsall (1953). The procedure has since been used widely in statistics and other fields, but has only recently found its way into the economics literature (see for example Jordà & Taylor, 2011, 2012; Khandani et al., 2010). In the context of predicting recessions, it can be summarized as follows:

1. Let
$$Z_t = \begin{cases} 1, & \text{if in recession} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$
denote the true, observed state of the economy. Let P_t be the prediction of Z_t, or the probability of recession, given by the probit model, where $0 \leq P_t \leq 1$.
2. Define evenly spaced thresholds (denoted C^*) along the interval [0, 1]. Larger numbers of thresholds lead to a smoother ROC curve with more points. For example, a potential set with 20 thresholds would be: $C^* = \{0, 0.05, \ldots, 0.95, 1\}$.
3. For each given threshold, C_i^*, record the model's predicted categories. More specifically, define the predicted categorization of Z_t, or $\hat{Z}_t$, as follows:
$$\hat{Z}_t = \begin{cases} 1, & \text{if } P_t \geq C_i^* \\ 0, & \text{if } P_t < C_i^*. \end{cases} \quad (2)$$
4. Comparing the true Z_t with the predicted categorizations $\hat{Z}_t$, calculate the percentages of true positives (PTP)

[2] For a comprehensive review of forecasting models for binary outcomes, see Lahiri and Yang (2013).

and false positives (PFP). More specifically, they can be defined using the sum of two indicator variables:

$$PTP_i = \frac{1}{n_R}\sum_{t=1}^{T} I_t^{tp};$$
$$\text{where } I_t^{tp} = \begin{cases} 1, & \text{if } Z_t = 1 \text{ and } \hat{Z}_t = 1 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$$PFP_i = \frac{1}{n_E}\sum_{t=1}^{T} I_t^{fp};$$
$$\text{where } I_t^{fp} = \begin{cases} 1, & \text{if } Z_t = 0 \text{ and } \hat{Z}_t = 1 \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where n_R is the number of times the true Z_t was in a recession and n_E is the number of times the true Z_t was not in a recession, such that $n_R + n_E = T$, where T is the total number of observations in our sample.

5. For each C_i^*, create a set of coordinates: (PFP_i, PTP_i).
6. After a coordinate is created for each threshold, plot the coordinates across all thresholds, with the false positive rate on the x-axis and the true positive rate on the y-axis. Connect these coordinates in order to trace out the ROC curve.

In summary, the ROC curve pinpoints the percentage of false negatives that one would have to trade for an additional one percent of true positives. A model with 100% accuracy would have a ROC curve that hugged the top left corner. A model which is the equivalent of a random guess would follow a 45% diagonal running from the bottom-left to top-right corners. By construction, if we defined Z_t in terms of expansions (i.e., let Z_t equal one during expansions and zero otherwise) instead of recessions, the new curve would look symmetric to the old curve about a 45° line from the bottom-right to top-left corners. By geometry, the area under the curve would then remain exactly the same as before.

Due to its ease of application and intuitive visual interpretation, integrating the ROC curve to get the area under the ROC curve (AUROC) is a popular method of classifying a given model's predictive ability. In our empirical analysis in Section 4, we will therefore compare the recession classification abilities of various different probit models using their implied AUROC values.

In principle, the differences between models' AUROC values could be tested for statistical significance using the asymptotic t-statistic provided by Hanley and McNeil (1983). However, the error terms in our probit estimations with a business cycle indicator as the dependent variable probably feature some serial correlation (see e.g. Pesaran & Timmermann, 2009), which this t-statistic does not take into account. To account for the serial correlation of the estimation errors properly, we instead use the block-bootstrap approach proposed by Politis and Romano (1994). Specifically, we re-sample the dependent and independent variables 500 times for each forecast horizon, then re-estimate the probit models and the corresponding ROC curves for each bootstrap sample. This ultimately creates an empirical distribution of AUROCs for each model (and each horizon), which we can then compare. A similar approach has been used by Pierdzioch, Reid, and Gupta (2014), who employed ROC curves to study the directional accuracy of South African survey data for short-term and longer-term inflation forecasts. We use an average block length of eight years in order to retain the typical business cycle length. In addition, robustness tests using average block lengths of five years and eleven years yield very similar results.

3. Data

We use monthly US data for the sample period January 1959–December 2011. The dependent variable is a binary recession indicator which takes on the value of one during a recession and zero during an expansion, both as defined by the NBER business cycle dating committee. This committee meets periodically to judge whether a peak or trough in economic activity has occurred, taking into account a variety of economic activity indicators, including the real GDP measured on the product and income sides, economy-wide employment, real income, and indicators covering real parts of the economy, such as retail sales and industrial production.[3] The NBER's dating rules are considered widely to be the benchmark for US business cycles. Moreover, as was discussed in the introduction, Berge and Jordà (2011) find that a number of sophisticated parametric models show no improvement over the NBER in terms of recession classification ability.

Our explanatory variables are taken from Haver Analytics, with the exception of the real-time macroeconomic indicators, which come from the ArchivaL Federal Reserve Economic Database (ALFRED). Following Estrella and Hardouvelis (1991); Estrella and Mishkin (1998), and others, we use the term spread, precisely defined as the difference between the ten-year and three-month Treasury yields ("10y–3m spd"), as our benchmark predictor variable. We then assess whether additional lags of the term spread or other candidate predictor variables add predictive power to the benchmark model.

Based on prior research, we consider the following list of additional predictor variables. First, following Estrella and Mishkin (1998), we add several financial indicators, including returns on the S&P500 common stock price index ("S&P 500, 1y% change", "S&P 500, 3y% change") and the interest rates of the 3-month ("3m rate") and 10-year ("10y rate") Treasuries. Second, we include each component of the Conference Board's leading economic index (LEI; see Levanon et al., 2015). These indicators have been selected for their abilities to signal peaks and troughs in the business cycle, and the aggregate index has been shown to drop ahead of recessions and rise before expansions. The individual factors consist of: average weekly hours of manufacturing ("Avg wkly hrs (manufacturing)"); average weekly initial claims for unemployment insurance ("Avg initial claims"); manufacturers' new orders of goods and materials ("New orders, goods, materials"); the ISM index of new orders ("ISM new orders index"); manufacturers' new orders of non-defense capital goods

[3] See http://www.nber.org/cycles/recessions.html.

excluding aircraft ("New orders, non-defense"); building permits, new private housing units ("Building permits"); the yield of the 10-year Treasury note less the federal funds ("10yr-FF spread"); average consumer expectations ("Michigan consumer survey"); and the leading credit index.

The leading credit index was introduced by the Conference Board to supplement the leading economic index and reflect potential structural changes in the changing credit and financial markets. Its components were selected in the spirit of financial intermediation models, as laid out by Adrian and Shin (2010), and were tested for their ability to signal business cycle changes using a Markov switching model (see Levanon et al., 2015). In a recent paper, Lahiri and Yang (2015) found that the forecasting ability of the leading economic index is improved significantly when M2 is replaced with the leading credit index. We include each of the LCI's components that is offered at a monthly frequency or greater going back to at least 1985: the LIBOR 3-month less 3-month Treasury bill yield spread ("LIBOR 3 month"), balances in broker-dealer margin accounts ("Debit margins (BD)"), and the AAII Sentiment Survey's Market Survey of the spread between bearish and bullish sentiments ("Bear less bull").

Finally, we add a number of predictor variables following the recent findings of Ng (2014). In her paper, Ng examines a comprehensive list of 132 different real and financial indicators and assesses their relevance for predicting recessions. We include all 14 of the unique variables that were found to be most important using cross-validation boosting techniques, together with the four additional unique variables that were found through a rolling window exercise. This list includes the spreads between the yields of several constant maturity Treasuries and the federal funds (FF) rate ("3m/6m/1yr/2yr/5yr/30yr-FF spread"); employment hours for total non-farm, government, manufacturing, and mining ("Emp: total/govt/mfg/mining"); the NAPM index of consumer commodity prices ("NAPM com price"); NAPM vendor deliveries ("NAPM vendor del"); the NAPM inventories index ("NAPM invent"); and the USD-Yen exchange rate ("Ex rate: Japan").

Table 1 provides a list of all of the predictor variables that we consider, with their available time spans, the databases from which they originate, and the transformations that we use to render the series stationary. These transformations follow those used in the papers cited above directly. Since macroeconomic variables are often revised after their release dates and these revisions are not available to the economist for out-of-sample forecasting, we use real-time or first-release indicators whenever possible. When available, real-time data are collected from the ALFRED database. For several macroeconomic variables, real-time data before 1985 do not exist. These variables are marked with asterisks in Table 1, and are excluded from our out-of-sample analysis.

4. Results

In this section, we describe our results from comparing various probit model specifications at different horizons. We carry out both an in-sample and an out-of-sample analysis. For in-sample recession prediction, our sample runs from 1959 to 2011 and covers a total of seven recessions, which range in duration from six to 18 months. We conduct out-of-sample forecasts using the period of January 1959–August 1985 (which covers a total of four recessions) as our training sample, and the period from September 1985 through December 2011 (covering three recessions) as our forecasting sample.

In both exercises, we consider forecasts at the 3-, 6-, 12-, 18-, and 24-month horizons. At each horizon, we begin by estimating a baseline probit model using just the contemporaneous term spread as an explanatory variable (see Estrella & Hardouvelis, 1991). Then, we augment the baseline model with the six-month lagged term spread as an additional explanatory variable. These two spread models make up our first and second benchmark models, respectively. We do this in order to determine whether it is the level or the change in the spread that is important for predicting recessions, and to capture the effects of the term spread at multiple horizons. Later, this will allow us to test whether leading indicators contain predictive information over and above that contained in the term spread.

Finally, we add each of the variables shown in Table 1 to the spread and lagged spread models in turn. We use the AUROC to evaluate the performance of each model and also to compare the forecasting ability of each to those of the two baseline models. For each specification, we provide information on the bootstrapped distribution of AUROCs. This allows us to evaluate the uncertainty around the estimates and to assess whether adding predictor variables to the specifications using only the term spread increases the predictability of recessions significantly. Since the 90% probability bands of the resulting bootstrap distributions overlap to some extent, we also provide the bootstrapped probabilities of a given model outperforming the benchmark models. More specifically, for each of our 500 iterations, we use the same bootstrapped sample of dependent and independent variables to compute the various models' AUROCs. We then report the empirical probability that the AUROC of a given model will be smaller than that of either benchmark model, using data across all iterations.

4.1. In-sample analysis

The results of our in-sample probit regressions are summarized in Table 2. The table is broken down into panels A to E, corresponding to forecast horizons of 3, 6, 12, 18, and 24 months, respectively. Each panel reports results from the spread-only model, a spread and six-month lagged spread model, and the three best-performing specifications with an additional predictor. For each model, we report the 5th percentile, mean, and 95th percentile of the bootstrap distribution of AUROC values, along with the empirical probability that this model's AUROC will be smaller than that of either of the two benchmark models.

At the three-month-ahead horizon, we find that one can improve the spread-only model dramatically simply by adding a six-month lag of the term spread. In fact, doing so increases the average AUROC from 0.67, which is only slightly better than a random guess, to 0.86, which is

Table 1
Summary of key variables.

Series name	Code	Source	Time span
10y–3m spd	1	USECON	Jan 1959–Dec 2011
10y rate	1	USECON	Jan 1959–Dec 2011
3m rate	1	USECON	Jan 1959–Dec 2011
S&P 500, 1y % change	1	USECON	Jan 1959–Dec 2011
S&P 500, 3y % change	1	USECON	Jan 1959–Dec 2011
Leading credit index	1	BCI	Jan 1959–Dec 2011
Michigan consumer survey	1	BCI	Jan 1978–Dec 2011
BD debit margins	1	BCI	Jan 1960–Dec 2011
Bear less bull	4	BCI	Jul 1987–Dec 2011
LIBOR 3 month	1	USECON	Jan 1963–Dec 2011
Baa-Aaa spread	1	USECON	Jan 1959–Dec 2011
Fed Funds	1	USECON	Jan 1959–Dec 2011
Aaa-FF spread	1	USECON	Jan 1959–Dec 2011
Baa-FF spread	1	USECON	Jan 1959–Dec 2011
3mo-FF spread	1	USECON	Jan 1959–Dec 2011
6m-FF spread	1	USECON	Jan 1959–Dec 2011
1yr-FF spread	1	USECON	Jan 1959–Dec 2011
2yr-FF spread	1	USECON	Jun 1976–Dec 2011
5yr-FF spread	1	USECON	Jan 1959–Dec 2011
10yr-FF spread	1	USECON	Jan 1959–Dec 2011
30yr-FF spread	1	USECON	Mar 1977–Dec 2011
Ex rate: Japan	2	USECON	Jan 1959–Dec 2011
Avg wkly hrs (manufacturing)	1	ALFRED	Jan 1959–Dec 2011
Avg initial claims[a]	3	BCI	Jan 1959–Dec 2011
New orders, goods, materials[a]	3	BCI	Jan 1959–Dec 2011
New orders, non-defense[a]	3	BCI	Jan 1959–Dec 2011
ISM new order index[a]	1	BCI	Jan 1959–Dec 2011
Building permits[a]	3	BCI	Jan 1959–Dec 2011
Emp: total	5	ALFRED	Jan 1959–Sep 2010
Emp: govt	5	ALFRED	Jan 1959–Dec 2011
Emp: mfg	6	ALFRED	Jan 1959–Dec 2011
Emp: mining	5	ALFRED	Jan 1959–Dec 2011
NAPM com price	1	ALFRED	Jan 1959–Dec 2011
NAPM vendor del[a]	1	USECON	Jan 1959–Dec 2011
NAPM invent[a]	1	USECON	Jan 1959–Dec 2011

This table reports all of the predictor variables considered in our analysis. For each indicator, we provide the series name, the transformation, the data source, and the time span for which the series is available. Transformation codes 1–6 correspond to levels, monthly log differences, annual log differences, annual differences, 6-month moving average smoother, and 12-month moving average smoother, respectively. The data sources USECON, BCI, and ALFRED refer to the US Economics Statistics database in Haver Analytics, the Business Cycle Indicators database in Haver Analytics, and the online ArchivaL Federal Reserve Economic Database at the St. Louis Fed, respectively.

[a] Denotes macroeconomic indicators for which real-time data are not available prior to 1985, and which are therefore excluded from our out-of-sample analysis.

quite accurate. That being said, one can improve further upon the spread and lagged-spread model at the three-month horizon by including one of various additional indicators. The strongest improvement comes from the annual return on the S&P 500 index, which lifts the mean AUROC to a near-perfect 0.96. This model is also strongly statistically different from both the spread-only and spread and lagged spread models, as is shown by the fact that the 90% probability bands of the bootstrap distributions do not overlap. The next two best-performing additional indicators are the debit balances in margin accounts at broker dealers and total payroll employment, both of which have average AUROCs of around 0.94. Given the sharp differences in the bootstrapped AUROC distributions of the three models, relative to the benchmark models using only the term spread, it is not surprising that the empirical probabilities of finding an AUROC that is smaller than those of either of the two benchmark models are zero.

At the six-month-ahead horizon, we again find that adding a six-month lagged spread improves the recession classification ability of the probit model significantly, raising the average bootstrapped AUROC from 0.77 to 0.88. In addition, the S&P 500 index continues to be one of the top additional indicators. The other two are the 5-year Treasury yield—fed funds rate spread and building permits. Again, we find that one can improve upon the spread and lagged spread model significantly by adding any of these three additional indicators. Hence, these other variables contain significant predictive information beyond that captured by the Treasury term spread. This is underscored by the fact that the empirical probabilities of finding an AUROC that is smaller than that of one of the benchmark models continue to be essentially zero.

At the twelve-month-ahead horizon, the spread-only model performs remarkably better than at shorter horizons, with an average AUROC of 0.87. While adding the lagged spread improves the predictive ability somewhat, similarly to the six-month horizon, we find the model with the 5-year Treasury-FF spread to perform best, with an AUROC bootstrap mean of 0.91. The two next best models use the 1-year Treasury-FF spread and the 10-year Treasury-FF spread, respectively. While all three models appear to outperform the spread-only model systematically, their empirical probabilities of not beating the spread

Table 2
In-sample summary of AUROC values.

Model	5th	Mean	95th	$p(\theta < \theta_1)$	$p(\theta < \theta_2)$
Panel A: Three months ahead					
Spread(t) only	0.529	0.668	0.810	–	–
Spread(t) + spread($t-6$)	0.802	0.861	0.925	0.000	–
BD debit margins	0.896	0.936	0.970	0.000	0.000
Emp: total	0.910	0.939	0.965	0.000	0.000
S&P 500, 1y % chg	0.933	0.956	0.978	0.000	0.000
Panel B: 6 months ahead					
Spread(t) only	0.609	0.769	0.896	–	–
Spread(t) + spread($t-6$)	0.806	0.878	0.944	0.000	–
Building permits	0.885	0.922	0.954	0.000	0.002
S&P 500, 1y% chg	0.884	0.965	0.962	0.000	0.003
5yr-FF spread	0.895	0.933	0.969	0.000	0.002
Panel C: 12 months ahead					
Spread(t) only	0.788	0.869	0.920	–	–
Spread(t) + spread($t-6$)	0.823	0.886	0.941	0.040	–
10 yr-FF spread	0.869	0.907	0.950	0.010	0.114
1 yr-FF spread	0.875	0.909	0.952	0.004	0.070
5 yr-FF spread	0.878	0.912	0.952	0.004	0.092
Panel D: 18 months ahead					
Spread(t) only	0.731	0.814	0.896	–	–
Spread(t) + spread($t-6$)	0.754	0.825	0.906	0.136	–
ISM new order index	0.777	0.846	0.915	0.022	0.108
NAPM com price	0.775	0.849	0.935	0.032	0.0700
BD debit margins	0.815	0.878	0.948	0.000	0.008
Panel E: 24 months ahead					
Spread(t) only	0.551	0.695	0.835	–	–
Spread(t) + spread($t-6$)	0.575	0.712	0.849	0.218	–
ISM new order index	0.683	0.786	0.886	0.0120	0.038
New orders, non-defense	0.710	0.787	0.869	0.006	0.010
BD debit margins	0.714	0.808	0.890	0.026	0.066

This table shows the in-sample forecast results for the spread-only and spread-and-lagged-spread models, as well as the three top performing models with one additional variable over the spread and lagged spread model. For each model, we show the 5th and 95th percentiles, as well as the mean of the bootstrap distribution of AUROCs. The last two columns show the empirical probability that a given model has a lower AUROC than the spread only ($\theta < \theta_1$) and spread and lagged spread ($\theta < \theta_2$) models, respectively. The sample period is January 1959–December 2011.

and lagged spread model all exceed 5%. Hence, it is difficult to find additional variables at the twelve-month horizon that significantly beat models based only on the term spread.

Turning to the 18-month-ahead forecast horizon, we see that, with a mean AUROC of 0.81, the predictive ability of the spread-only model is slightly worse than at the twelve-month-ahead horizon, but better than at the three- and six-month-ahead horizons. While adding a six-month lagged spread shifts the distribution of bootstrapped AUROCs marginally to the right, the empirical probability of reducing the recession prediction ability by adding the lagged spread is almost 14%. The top three additional predictor variables at the 18-month-ahead horizon are the ISM new order index, the NAPM commodity price index, and debit margins at broker-dealers. The last pushes the mean bootstrapped AUROC to 0.88. In fact, as the probabilities in the last two columns of the table indicate, broker-dealer margin debit is the only predictor variable that strongly significantly increases the AUROC when added to the two baseline models. This is interesting, given that the previous literature has generally found that the spread-only model is difficult to beat at the twelve- and eighteen-month horizons.

Finally, at the 24-month-ahead horizon, we see that the predictive ability is generally much lower for all models. However, that being said, all models are still considerably better than a random guess model. The spread-only model has an average AUROC of 0.70, which makes it comparable to its counterpart at the three-month-ahead horizon (0.67). Adding the lagged spread improves the prediction ability marginally. On the other hand, we also find that the addition of margin debit at broker-dealers, new non-defense orders, or the NAPM commodity price index improves upon the two baseline models significantly. All three perform about equally well, with AUROCs ranging from 0.79 to 0.81. While these are not strong predictive abilities, they perform better than similar models estimated using only components of the LEI, which are considered the benchmark leading indicators (see Berge & Jordà, 2011).

Figs. 1–3 summarize the findings in Table 2 visually. In these figures, the plots are paired by forecast horizon. The top graph shows the predicted mean probability of recession over time for the spread only, the spread and lagged spread, and the best performing model adding a third regressor, with actual recessions shown as shaded grey areas. The second graph shows the corresponding bootstrapped ROC curves, calculated as described in Section 2. In all graphs, the blue lines represent the bootstrapped estimates from the spread-only model, the green lines represent those for the spread and lagged spread model, and the red lines represent those for the best performing model with a third predictor. Superimposing the three sets of bootstrapped AUROCs gives a clear visual impression of the differences between the different

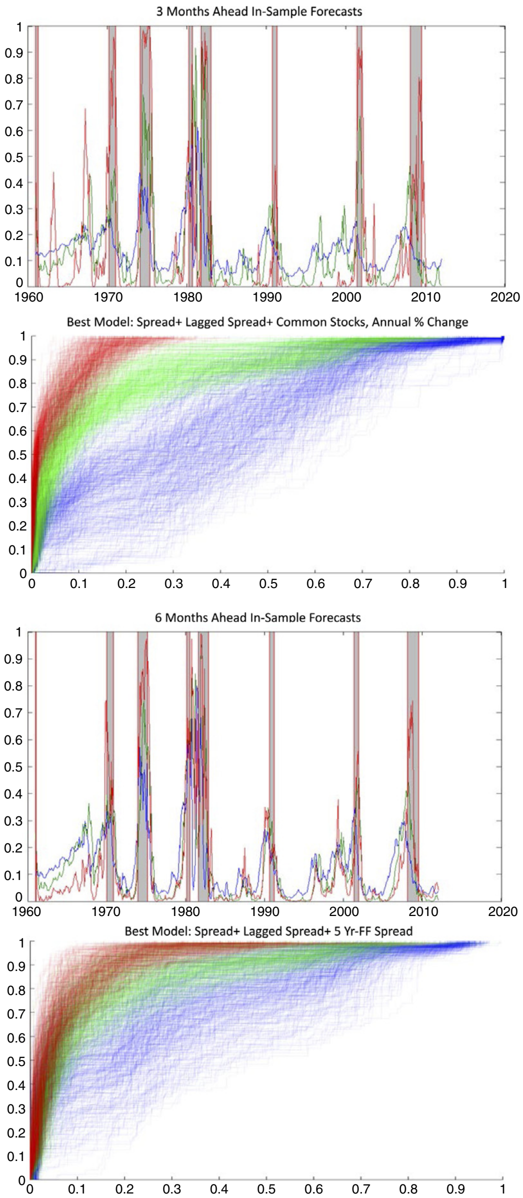

Fig. 1. 3m and 6m in-sample probabilities and ROC curves. For more information, see Fig. 3 (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

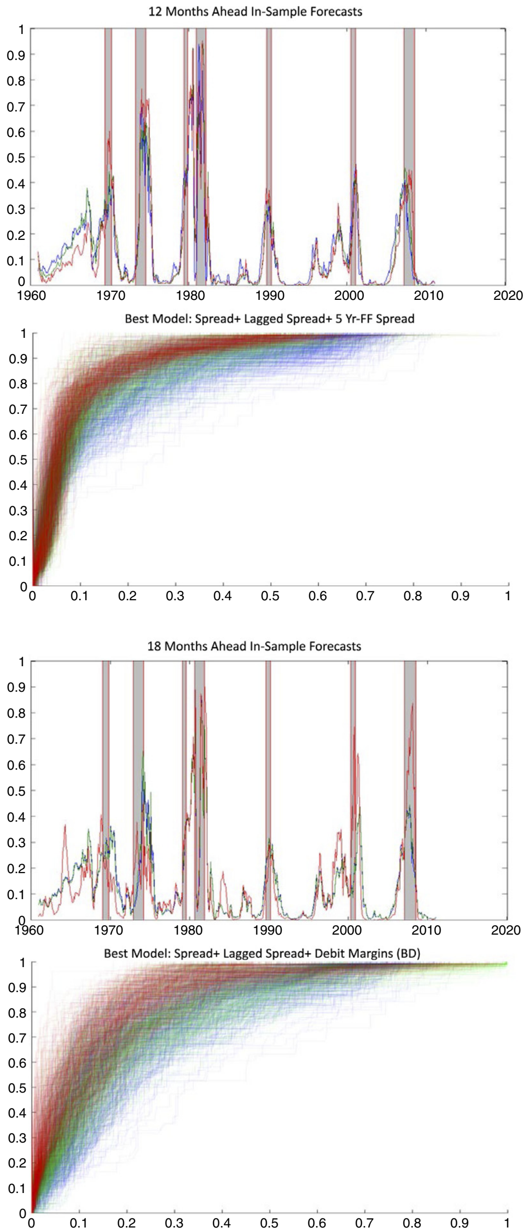

Fig. 2. 12m and 18m in-sample probabilities and ROC curves. For more details, see Fig. 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model specifications. In particular, the distinction between models is especially clear for the three-month-ahead horizon.

4.2. Out-of-sample analysis

To assess whether the forecasting performance documented for the full sample extends to an out-of-sample setting, we conduct the following exercise. We first estimate each probit model for the period January 1959–August 1985, covering a total of four recessions. We then use the estimated parameters to predict recessions over the period from September 1985 to December 2011, a period which covers three recessions.

The results of our baseline and best performing out-of-sample probit regressions are summarized in Table 3. As before, the table is organized in five panels that correspond to the five different forecast horizons. Each panel shows the AUROC for the spread-only model, the spread and lagged spread model, and the three best models obtained by adding a third predictor variable to the spread and lagged spread model. For each model, we again report the 5th percentile, mean and 95th percentile of the bootstrapped AUROC distribution, as well as the empirical probabilities that a given alternative model has an AUROC that is smaller than those of either of the two baseline models. The

Table 3
Out-of-sample summary of AUROC values.

Model	5th	Mean	95th	$p(\theta < \theta_1)$	$p(\theta < \theta_2)$
Panel A: 3 months ahead					
Spread(t) only	0.579	0.734	0.858	–	–
Spread(t) + spread($t-6$)	0.795	0.870	0.953	0.002	–
Avg initial claims	0.884	0.925	0.972	0.002	0.002
Emp: total	0.909	0.940	0.978	0.002	0.000
S&P 500, 1y % chg	0.908	0.946	0.980	0.002	0.002
Panel B: 6 months ahead					
Spread(t) only	0.731	0.842	0.937	–	–
Spread(t) + spread($t-6$)	0.863	0.910	0.955	0.000	–
6m-FF spread	0.905	0.947	0.983	0.002	0.002
3m-FF spread	0.911	0.949	0.983	0.002	0.002
Aaa-FF spread	0.908	0.950	0.985	0.002	0.002
Panel C: 12 months ahead					
Spread(t) only	0.830	0.868	0.908	–	–
Spread(t) + spread($t-6$)	0.838	0.891	0.946	0.100	–
NAPM com price	0.861	0.913	0.971	0.000	0.104
NAPM vendor del	0.865	0.913	0.962	0.010	0.180
Fed funds	0.862	0.915	0.970	0.024	0.038
Panel D: 18 months ahead					
Spread(t) only	0.620	0.716	0.826	–	–
Spread(t) + spread($t-6$)	0.685	0.768	0.874	0.128	–
5yr-FF spread	0.724	0.801	0.902	0.028	0.056
New orders, non-defense	0.721	0.803	0.909	0.012	0.082
BD debit margins	0.737	0.806	0.892	0.026	0.040
Panel E: 24 months ahead					
Spread(t) only	0.483	0.572	0.697	–	–
Spread(t) + spread($t-6$)	0.496	0.628	0.811	0.094	–
NAPM com price	0.607	0.744	0.894	0.002	0.032
New orders, non-defense	0.664	0.758	0.889	0.000	0.036
BD debit margins	0.762	0.836	0.919	0.000	0.034

This table shows the in-sample forecast results for the spread-only and spread-and-lagged-spread models, as well as the three top performing models with one additional variable over the spread and lagged spread model. For each model, we show the 5th and 95th percentiles, as well as the mean of the bootstrap distribution of AUROCs. The last two columns show the empirical probability that a given model has a lower AUROC than the spread only ($\theta < \theta_1$) or spread and lagged spread ($\theta < \theta_2$) models, respectively. The estimation sample is from January 1959 to August 1985 and the forecasting sample from September 1985 to December 2011.

recesssion predictions and corresponding AUROC curves are shown in Figs. 4–6.

At the three-month-ahead horizon, the spread-only model obtains an average AUROC of 0.73, which is mildly higher than a random guess. Adding a six-month lag of the term spread increases the AUROC substantially, to 0.87. In fact, the difference between the second benchmark model and the spread-only model is strongly significant, since the bootstrapped probability of the second benchmark model performing worse is essentially zero. In line with the in-sample analysis, we find that several key variables improve upon the spread and lagged spread benchmark model significantly. The three best performing variables are initial claims for unemployment insurance, total payroll employment, and the annual return on the S&P 500 index. All three improve upon the two baseline models significantly, as can be seen from the extremely low empirical probabilities of obtaining an AUROC that is smaller than those of the benchmark models. Similarly to the in-sample analysis, the best model, using the annual return on the S&P 500 index in addition to the spread and lagged spread, has a near-perfect predictive ability, as can be seen from its AUROC of 0.95.

Similarly to the in-sample analysis at the six-month horizon, we also find that the spread-only model performs better at the six-month horizon out-of-sample, reaching a mean bootstrapped AUROC of 0.84, with the 5th and 95th percentiles being 0.73 and 0.94 respectively. Adding the six-month lagged spread improves the model's predictive ability significantly, raising its average AUROC to 0.91. Next, when including additional predictor variables, we find evidence that various interest rate spreads over the fed funds rate contain predictive information about future recessions beyond that which is captured by the ten-year to three-month term spread. However, this set of best performing variables (short term interest spreads and a bond index spread) is different from the set of variables that was found to be most useful in-sample. This could suggest that the drivers of business cycle activity have varied within our sample period.

Turning to the one-year horizon, the spread-only model performs about equally well as in the in-sample analysis (with a bootstrap mean AUROC of 0.87). Moreover, adding the six-month lagged spread does not improve the predictability significantly. While the set of leading indicators that performs best out-of-sample at the twelve-month horizon differs from those that perform best in the in-sample analysis, only the federal funds rate improves the predictability compared to the spread and lagged spread model at the 5% significance level.

At the 18-month-ahead forecast horizon, similarly to the in-sample analysis, the recession predictive ability implied by the spread-only model is somewhat lower than for the twelve-month-ahead horizon. Hence, consistent with

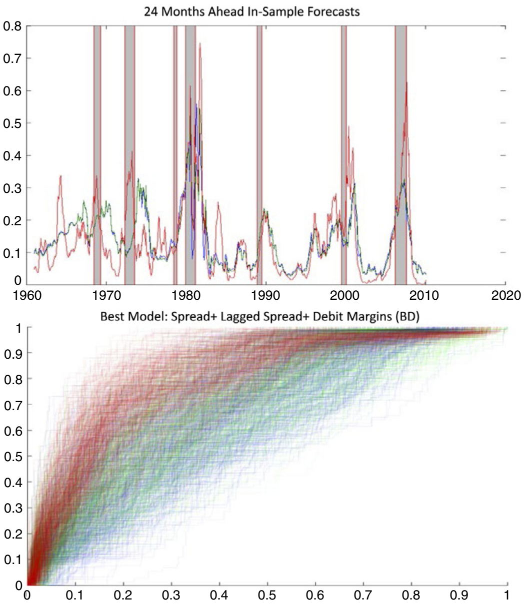

Fig. 3. 24m in-sample probabilities and ROC curves. The graphs show, at the 24-month forecast horizon, the probability of recession and the corresponding ROC curve for the spread-only model (blue lines), the spread and lagged spread model (green lines), and one additional model with best performance, as determined by the average bootstrapped AUROC (red lines). Each model is estimated using monthly data from January 1959 to December 2011. The first panel shows the probability of recession expressed as a decimal over our sample period, with the actual recession periods shaded in grey. Following this chart is a ROC curve that plots the trade-off between false positive (x-axis) and true positive (y-axis) rates for each model. The ROC curves correspond to 500 block bootstrap draws, each with a block length of eight years. The predicted probabilities are the averages of bootstraps across all 500 draws. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the in-sample analysis and with previous evidence (e.g. Estrella & Hardouvelis, 1991), we find that the Treasury term spread has its strongest predictive ability at a horizon of one year. While adding lagged observations of the spread does not improve the forecasting power significantly compared to the spread-only model, the addition of other variables such as new orders of non-defense goods and broker-dealer debit margins does. The latter is the only indicator which adds significant predictive information over and above the second benchmark model at the 18-month horizon.

Finally, at the 24-month-ahead horizon, the predictive ability of the spread-only model, as measured by its AUROC, drops sharply, to an average of 0.57. This is close to a random guess model, which has an expected AUROC of 0.50. While adding the six-month lag of the term spread does not increase the predictive power significantly, adding other indicators does improve the forecasts significantly. Most importantly, non-defense new orders and broker-dealer debit margins again show up as the top additional leading indicators at this horizon. Both raise the average AUROC substantially, and outperform the spread and lagged spread model significantly when comparing empirical probabilities.

Figs. 1–3 provide a visual summary of the findings in Table 3. Again, the plots are grouped by pairs and shown for each forecast horizon, with the first graph displaying the predicted probability of recession over time and the second graph showing the corresponding ROC graphs, calculated by comparing the predicted probability with the NBER business cycle chronology. In these figures, the plots are paired by forecast horizon. The top graph shows the predicted median probability of recession over time for the spread only, the spread and lagged spread, and the best performing model when adding a third regressor, with actual recessions shown as shaded grey areas. The second graph shows the corresponding bootstrapped ROC curves, calculated as described in Section 2. As before, the blue lines in each graph represent the bootstrapped estimates from the spread-only model, the green lines those for the spread and lagged spread model, and the red lines those for the best performing model with a third predictor. Superimposing the three sets of bootstrapped ROC curves gives a clear visual indication of the differences between the three model specifications. The charts underscore the findings from the table. While the bootstrapped ROC intervals based on the spread-only and spread and lagged spread models, and the best model adding a leading indicator, are clearly distinct at the three- and 24-month horizons, such is not the case at intermediate horizons. As the last two columns in Tables 2 and 3 show, however, this does not generally imply that the additional predictors are useless at these horizons. Rather, it shows that the incremental predictive ability (as measured by the AUROC) of the additional regressors over the term spread is rather small at intermediate and longer forecast horizons.

5. Summary of findings and concluding remarks

In summary, our results imply the following main messages. First, consistent with the past literature, we find that the ability of the Treasury term spread to predict recessions is strongest at the twelve-month-ahead horizon. In line with this result, adding lagged observations of the term spread improves the predictability of the term spread substantially at forecast horizons shorter than one year. However, this is not the case at longer forecast horizons, suggesting that it is the level, not the change in the term spread over time that matters most for its forecasting power.

Second, a few leading indicators improve the recession predictability relative to forecasts based only on the spread and its lags. At short forecast horizons, the annual return on the S&P500 index stands out. This leading indicator improves the ability to predict recessions both in-sample and out-of-sample significantly and substantially beyond the information contained in the term spread. While there is some evidence that other leading indicators can improve upon the term spread at the one-year or 18-month horizons, the evidence is less clear. That being said, new orders of non-defense capital goods and balances of margin debit at broker-dealers improve forecasts significantly at horizons of more than one year. This effect is most pronounced for the 24-month horizon and for out-of-sample forecasts, with the forecasting ability remaining

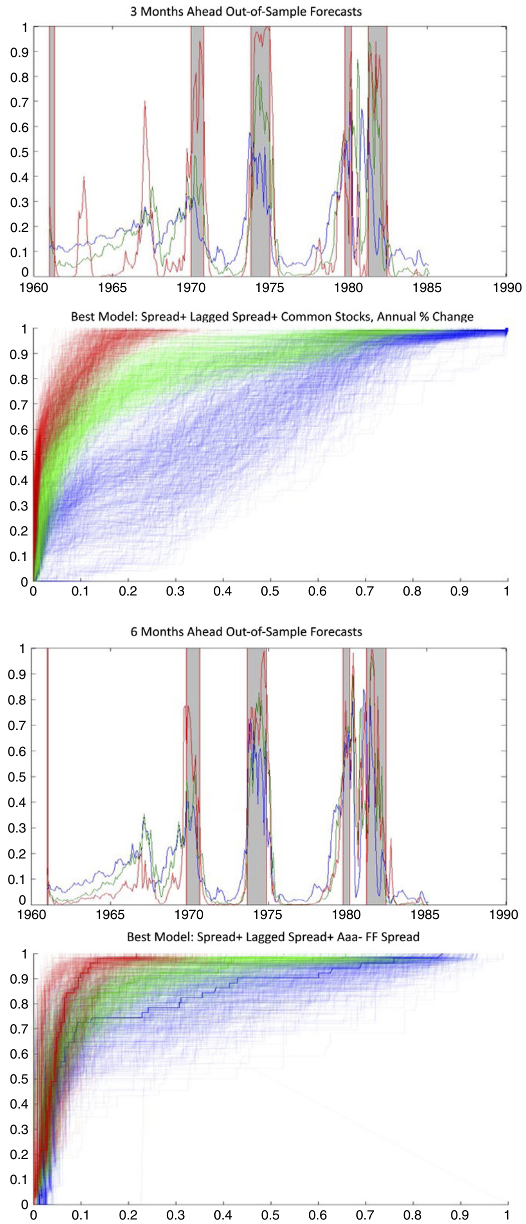

Fig. 4. 3m and 6m out-of-sample probabilities and ROC curves. For more information, see Fig. 6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

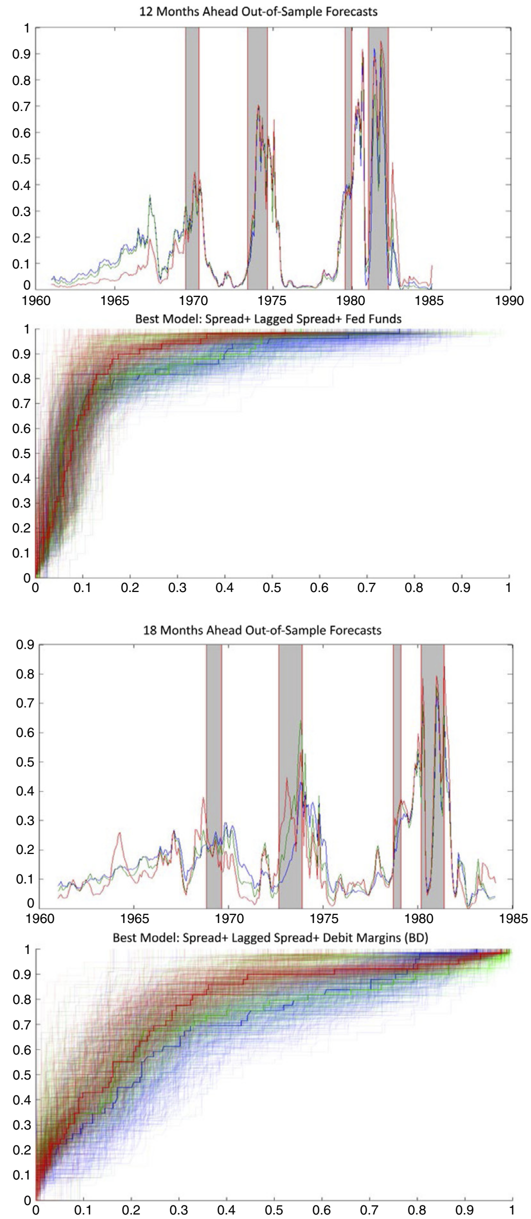

Fig. 5. 12m and 18m out-of-sample probabilities and ROC curves. For more information, see Fig. 6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

remarkably high (AUROC of roughly 0.80). In addition, the predictive content of broker-dealer margin balances for recessions is novel to the literature. As the margin debt at broker-dealers is typically considered to be a measure of the leverage in the financial system, its importance in predicting recessions highlights the role of financial intermediary balance sheet management in the transmission of economic shocks. A similar argument has been made by Adrian, Estrella, and Shin (2010), who document that a decline in the Treasury term spread has a causal impact on the net interest margin of banks and the growth rate of shadow bank balance sheets, which, in turn, have a direct effect on financial intermediaries' lending to the real economy. Both these authors' findings and our results suggest that the interaction between the term spread, financial intermediary balance sheets, and real economic activity is a promising area for future research.

Acknowledgments

Rui Yu provided excellent research assistance. The views expressed in this paper are those of the authors

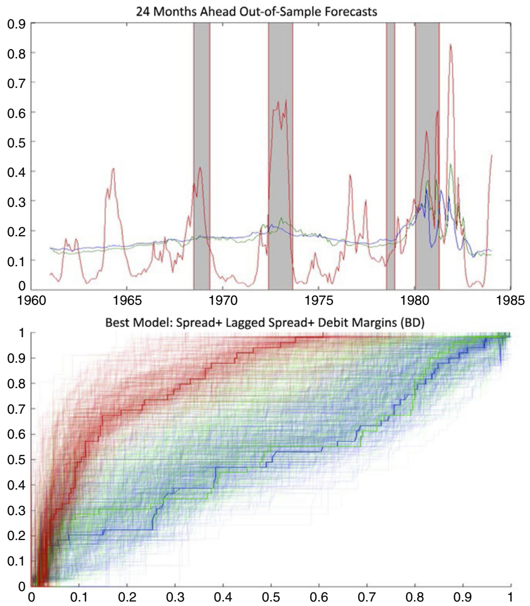

Fig. 6. 24m out-of-sample probabilities and ROC curves. The graphs show, at the 24-month forecast horizon, the probability of recession and the corresponding ROC curve for the spread-only model (blue lines), the spread and lagged spread model (green lines), and one additional model with the best performance, as determined by the average AUROC (red lines). The models are estimated over the period January 1959–August 1985, and the forecast accuracy calculations are estimated over the period September 1985–December 2011. The first panel shows the probability of a recession expressed as a decimal over our sample period, with the actual recession periods shaded in grey. Following this chart is a ROC curve that plots the trade-off between false positive (x-axis) and true positive (y-axis) rates for each model. The ROC curves correspond to 500 block bootstrap draws, each with a block length of eight years. The predicted probabilities are the averages across all 500 bootstrap draws. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and do not necessarily represent those of the Deutsche Bundesbank or the Eurosystem.

Appendix. The probit-miss estimator of Conniffe and O'Neill (2011)

Building on the work of Chesher (1984), Conniffe and O'Neill's model assumes that there exists one underlying unobservable, continuous latent variable Y_i, and an observed binary variable Z_i which follows the relationship:

$$Z_i = 1 \quad \text{if } Y_i > 0$$
$$Z_i = 0 \quad \text{if } Y_i < 0.$$

The regressors are grouped into two categories, denoted in vector form X_i (complete) and W_i (incomplete). The number of Xs is n, and that of Ws is l. We observe the complete sample of observations $\{X_i, W_i, Z_i\}$ for $i = 1, 2, \ldots, r$. This leaves $(T - r)$ observations on which $\{X_i, Z_i\}$ alone are measured. They follow the relationship:

$$Y_i = X_i' B_x + W_i' B_w + \varepsilon_i. \tag{5}$$

To make use of the $T - r$ non-missing observations, assume:

$$W_i' = X_i' C + u_i', \tag{6}$$

where C is a $(n \times l)$ matrix of parameters and $u_i \sim MVN(0, \Sigma)$. Then, combining Eqs. (5) and (6), one obtains:

$$Y_i = X_i'(B_x + CB_w) + e_{y_i}, \tag{7}$$

where (e_{y_i}, W_i) are multivariate normally distributed, conditional on X_i. The assumption of conditional joint normality is analytically convenient and allows for efficient estimation. Conniffe and O'Neill (2011) show that their proposed estimator is robust to various departures from the parametric assumptions in Eq. (6).

Here, rather than restating the estimator and its asymptotic variance derived by Conniffe and O'Neill (2011) explicitly, we simply summarize the various estimation steps[4]:

1. Run an OLS regression of X on W for the sample with r complete observations.
2. Run a standard probit of Y on X and W for the sample with r complete observations.
3. Run a probit of Y on X for just the sample with $T - r$ missing observations.
4. Calculate the coefficients and standard errors for the probit-miss estimator, using the estimation outputs from steps 1–3 as inputs.

It is important to point out that, in addition to the assumption of conditional multinormality, the efficient probit estimator also requires the missing data for W to be 'missing at random' (MAR). In other words, the reason for the data's absence should not be related to an omitted variable that is correlated with recessions, such as the state of the business cycle. Since the limitations of our database mean that we only have missing data at the beginning of the dataset, the MAR assumption is naturally satisfied.

References

Adrian, T., Estrella, A., & Shin, H. S. (2010). *Monetary cycles, financial cycles and the business cycle. Staff report 421*. Federal Reserve Bank of New York.

Adrian, T., Etula, E., & Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *The Journal of Finance*, *69*(6), 2557–2596.

Adrian, T., Moench, E., & Shin, H. S. (2010). Macro risk premium and intermediary balance sheet quantities. *IMF Economic Review*, *58*(1), 179–207.

Adrian, T., Moench, E., & Shin, H. S. (2014). *Dynamic leverage asset pricing. Staff report 625*. Federal Reserve Bank of New York.

Adrian, T., & Shin, H. S. (2010). Liquidity and leverage. *Journal of Financial Intermediation*, *19*(3), 418–437.

Berge, T. J., & Jordà, Ò. (2011). Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics*, *3*(2), 246–277.

Chauvet, M., & Potter, S. (2005). Forecasting recessions using the yield curve. *Journal of Forecasting*, *24*(2), 77–103.

Chesher, A. (1984). Improving the efficiency of probit estimators. *The Review of Economics and Statistics*, *66*, 523–527.

Conniffe, D., & O'Neill, D. (2011). Efficient probit estimation with partially missing covariates. *Advances in Econometrics*, *27*, 209–245.

Croushore, D., & Marsten, K. (2014). The continuing power of the yield spread in forecasting recessions. Working Paper 14-5. Federal Reserve Bank of Philadelphia. February.

Duarte, A., Venetis, I. A., & Payà, I. (2005). Predicting real growth and the probability of recession in the Euro area using the yield spread. *International Journal of Forecasting*, *21*(2), 261–277.

Dueker, M. J. (1997). Strengthening the case for the yield curve as a predictor of US recessions. *Federal Reserve Bank of St. Louis Review Mar*, 41–51.

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *The Journal of Finance*, *46*(2), 555–576.

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

[4] For further details, the interested reader is referred to the paper by Conniffe and O'Neill (2011).

Estrella, A., & Mishkin, F. S. (1998). Predicting US recessions: Financial variables as leading indicators. *The Review of Economics and Statistics*, *80*(1), 45–61.

Estrella, A., & Trubin, M. R. (2006). The yield curve as a leading indicator: Some practical issues. *Current Issues in Economics and Finance*, *12*(5), 1–7.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*(3), 839–843.

Jordà, Ò., & Taylor, A. M. (2011). Performance evaluation of zero net-investment strategies. Working Paper 17150, National Bureau of Economic Research.

Jordà, Ò., & Taylor, A. M. (2012). The carry trade and fundamentals: Nothing to fear but feer itself. *Journal of International Economics*, *88*(1), 74–90.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767–2787.

Lahiri, K., Monokroussos, G., & Zhao, Y. (2013). The yield spread puzzle and the information content of SPF forecasts. *Economics Letters*, *118*(1), 219–221.

Lahiri, K., & Wang, J. G. (2013). Evaluating probability forecasts for GDP declines using alternative methodologies. *International Journal of Forecasting*, *29*(1), 175–190.

Lahiri, K., & Yang, L. (2013). Forecasting binary outcomes. In G. Elliott, & A. Timmermann (Eds.), *Part B of handbook of economic forecasting*: *Vol. 2. Handbook of economic forecasting* (pp. 1025–1106). Elsevier, (Chapter 19).

Lahiri, K., & Yang, L. (2015). A further analysis of the conference board's new leading economic index. *International Journal of Forecasting*, *31*(2), 446–453.

Levanon, G., Manini, J.-C., Ozyildirim, A., Schaitkin, B., & Tanchua, J. (2015). Using a leading credit index to predict turning points in the US business cycle. *International Journal of Forecasting*, *31*(92), 392–398.

Moore, G. H., & Shiskin, J. (1967). *Indicators of business expansions and contractions*. NBER Books. National Bureau of Economic Research.

Ng, S. (2014). Viewpoint: Boosting recessions. *Canadian Journal of Economics*, *47*(1), 1–34.

Nyberg, H. (2010). Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting*, *29*(1–2), 215–230.

Pesaran, M. H., & Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, *104*(485), 325–337.

Peterson, W. W., & Birdsall, T. G. (1953). *The theory of signal detectability: Part I. The general theory. Technical report 13*. Electronic Defense Group.

Pierdzioch, C., Reid, M. B., & Gupta, R. (2014). *On the directional accuracy of inflation forecasts: Evidence from south african survey data. Technical Report 24/14, Stellenbosch Economic Working Papers*.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*(428), 1303–1313.

Rudebusch, G. D., & Williams, J. C. (2009). Forecasting recessions: The puzzle of the enduring power of the yield curve. *Journal of Business & Economic Statistics*, *27*(4), 492–503.

Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. In *NBER macroeconomics annual 1989. Volume 4* (pp. 351–409). National Bureau of Economic Research, Inc., NBER Chapters.

Wright, J. H. (2006). *The yield curve and predicting recessions. Technical report*. Board of Governors of the Federal Reserve System.