BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM

DIVISION OF MONETARY AFFAIRS

Date: September 11, 2019

To: Distribution

From: Tyler Pike and Francisco Vazquez-Grande

Subject: Out-of-sample performance of recession probability models[1]

In recent months, the inversion of the yield curve —long-term bond yields below short-term interest rates—has sparked fears of an imminent recession risk for the US economy, and has reignited the debate on how to best use financial variables to predict recession risks. The workhorse statistical model used to estimate the probability of a recession is typically a probit model, with independent variables constructed using data from the Treasury and corporate bond markets (see for example Engstrom and Sharpe, 2019; Johansson and Meldrum, 2018; and Favara, Gilchrist, Lewis, and Zakrajsek, 2016). To assess the adequateness of these models, it is customary to rely on the in-sample coefficient of determination (R^2) to determine the goodness of fit. As policy makers take their decisions in real-time, however, in-sample R^2can be misleading in gauging the usefulness of a given model.

This memo discusses the *out-of-sample* (OOS) performance of a few probit models used by Board staff to assess the likelihood that the U.S. economy will be in a recession within the following year. We find that bi-variate probit models that use a measure of term spread and the excess bond premium (EBP) of Gilchrist and Zakrajsek (2012) performs well OOS across estimation and evaluation subsamples, while accounting for measures of the term premiums does not lead to an improvement of OOS performance. We also find that when estimating the probability of transitioning into a recession—and therefore not evaluating model performance when the economy is already in an recession—the OOS performance of models using measures of the term spread, the level of rates, and the principal components of the yield curve increase significantly, while the

[1] We thank Min Wei and Giovanni Favara for helpful comments.

OOS performance using EBP alone decreases significantly. As of Sep 2019, models with a spread and the EBP as independent variables estimate that the probability of being in recession at some point over the next 12 months is about 40 percent, while the probability of transitioning into a recession over the next 12 months implied by these models is between 60 and 65 percent.

Models and data

We estimate nine probit regression models using monthly data. The dependent variable is an indicator variable that is equal to 1 if the US economy is in a recession at any time over the next 12 months, and is 0 otherwise.[2] The independent variables include eight commonly used recession predictors from the Treasury and the nonfinancial corporate bond markets: two measures of the slope of the yield curve (the "long-term spreads" and the "near-term forward spread"), one measure of risk sentiment in the corporate bond market, the first three principal components of the yield curve, the yield on the three-month Treasury bill, and the 10-year term-premium estimated using the Treasury yield curve.

The long-term spread (TS)—the difference between the 10−year and 3−month Treasury yield—is commonly regarded as the best signal of impending recessions (Estrella and Mishkin, 1998; Ang, Piazzesi and Wei, 2006; Rudebusch and Williams, 2009). The near-term forward spread (NTFS) of Engstrom and Sharpe (2019)—the difference between the 3-month forward rate six quarters ahead and yield on the three-month T-bill—is a measure of market expectation for the near term direction of monetary policy. The EBP of Gilchrist and Zakarajsek (2012)—the component of corporate bond spreads in excess of an estimate of the compensation for expected losses from corporate defaults—is a measure of investor sentiment in the corporate bond market. The first three principal components of the term structure of Treasury yields as in Johansson and Meldrum (2018) and the level of the three month Treasury bill as alternative measure of the level factor. The term premium (TP) for Treasury securities with 10-year maturity—

[2] The dating of recessions follows the NBER convention. In particular, we use the "NBER based Recession Indicators for the United States from the Peak through the Trough" from FRED. For the quarterly models, we define a quarter to be in a recession if the monthly NBER indicator is equal to 1 for at least two months in the quarter.

estimated by Adrian et al. (2013)—is the compensation that investors require for bearing the risk that short-term Treasury yields do not evolve as expected.[3] The sample period goes from February 1973 until today.

Our OOS methodology consists of reestimating the models every month adding one monthly observation at the time.[4] We do so to simulate the prediction that would have been available to policy makers at each point in time[5]. We estimate seven models, three univariate models with the TS, NTFS and EBP; three bivariate models with TS and TP, TS and EBP and NTFS and EBP; and one model with three dependent variables TS, TP, and EBP.

Performance evaluation: the receiver operating characteristic curve

We use the model estimates of recession probabilities to forecast the occurrence of a recession in the coming year. We therefore evaluate the performance of the models as binary classifiers. For this purpose, we use the receiver operating characteristic (ROC) curve introduced in the 1950's in the field of radar signal detection theory, and applied in economics to recession dating and forecasting by Berger and Jordà (2012), Christiansen et al. (2013), Liu and Moench (2016).[6]

The ROC curve is constructed as follows. Let $Y_t \in \{1,0\}$ be a binary variable indicating whether the U.S. economy will be in recession at some point in the next 12 months. Let p_t be the OOS probability of recession implied by a model we are interested in evaluating (for example a probit model). Probability p_t together with the threshold $c \in [0, 1]$ define a binary prediction of a future recession whenever $p_t \geq c$, and expansion whenever $p_t < c$.

We estimate the following conditional probabilities:

$$TP(c) = P[p_t \geq c | Y_t = 1]$$
$$FP(c) = P[p_t \geq c | Y_t = 0]$$

[3] Using Kim-Wright term-premium estimates (starting in 1980) yields similar results.
[4] The sample used for the first estimation runs from February 1973 to February 1978.
[5] We present results from exercises using NBER recession dates as if known in real-time, however, we note that this is not true in reality. We find that the results tend to be generally robust when this informational lag is also accounted for. Also, for the TP we use the latest available estimates, instead of reestimating the TP model every month.
[6] Other measures of forecasting accuracy, such as the root mean square error, the mean absolute error or the log probability score, are available for binary classification. However, they rely on the specification of an underlying forecast loss function. The main advantage of the ROC curve is that it is a direct plot of the entire space of trade-offs for a given classification problem and it is not tied to a specific loss function.

For threshold c, $TP(c)$ is the "true positive rate" and $FP(c)$ is known as the "false positive rate".

The ROC curve plots the entire set of possible combinations of $TP(c)$ (on the y-axis) and $FP(c)$ (on the x-axis) for all values of c. By definition, for $c = 1, TP(1) = FP(1) = 0$, and conversely, when $c = 0, TP(0) = FP(0) = 1$, so that the ROC curve is an increasing function in [0,1] × [0,1] space. If p_t is unrelated to the variable of interest Y_t, and is an entirely uninformative classifier—for example, tossing a fair coin to make the forecast—$TP(c) = FP(c)$, for all c, and the ROC curve would be the 45° line, a natural benchmark with which to compare classifiers. On the other hand, if p_t is a very good classifier, then the ROC curve would display values of $TP(c)$ close to 1 and $FP(c)$ close to 0 for all $c \in (0, 1)$.

A commonly used measure of the trade-offs contained in the ROC curve is the area under the ROC curve (AUROC) where a perfect classifier has AUROC = 1, whereas a coin-toss classifier has AUROC = 0.5. AUROC is equal to the probability that the value of p_t conditioned on $Y_t = 1$, is greater than the value of p_t conditioned on $Y_t = 0$. This implies that the higher the value of AUROC the better probability p implied by the model of interest (for example a probit model) is able to discriminate the two categories in Y.[7]

Results

Table 1 provides the OOS performance comparison of the seven models, as measured by the AUROC, and provides confidence intervals for this summary statistic. We find that, using a bivariate model with both a measure of spreads and the EBP as explanatory variables displays the highest OOS forecasting power. Models that use TS or NTFS have similar OOS performance. Models using principal components or adding the level of the short rate or TP do not improve the forecasting ability.

A univariate probit model with EBP significantly outperforms the two other univariate models using term spreads. As discussed further in the robustness section below, this is likely driven by the ability of EBP to provided forecasting power when in or immediately before recessions.

[7] The literature provides an easily implementable non-parametric statistic to estimate AUROC, with an asymptotic Gaussian distribution that allows us to compare the classification ability across models.

For comparison we also include AUROC values for the in-sample estimates of the models, estimated with data up to August 2019. We also include estimated probabilities of recession as of September 2019 for each model. Models with only spreads have probabilities of recession of around 60 percent while models with spreads and the EBP estimate the probability of recession at around 40 percent.

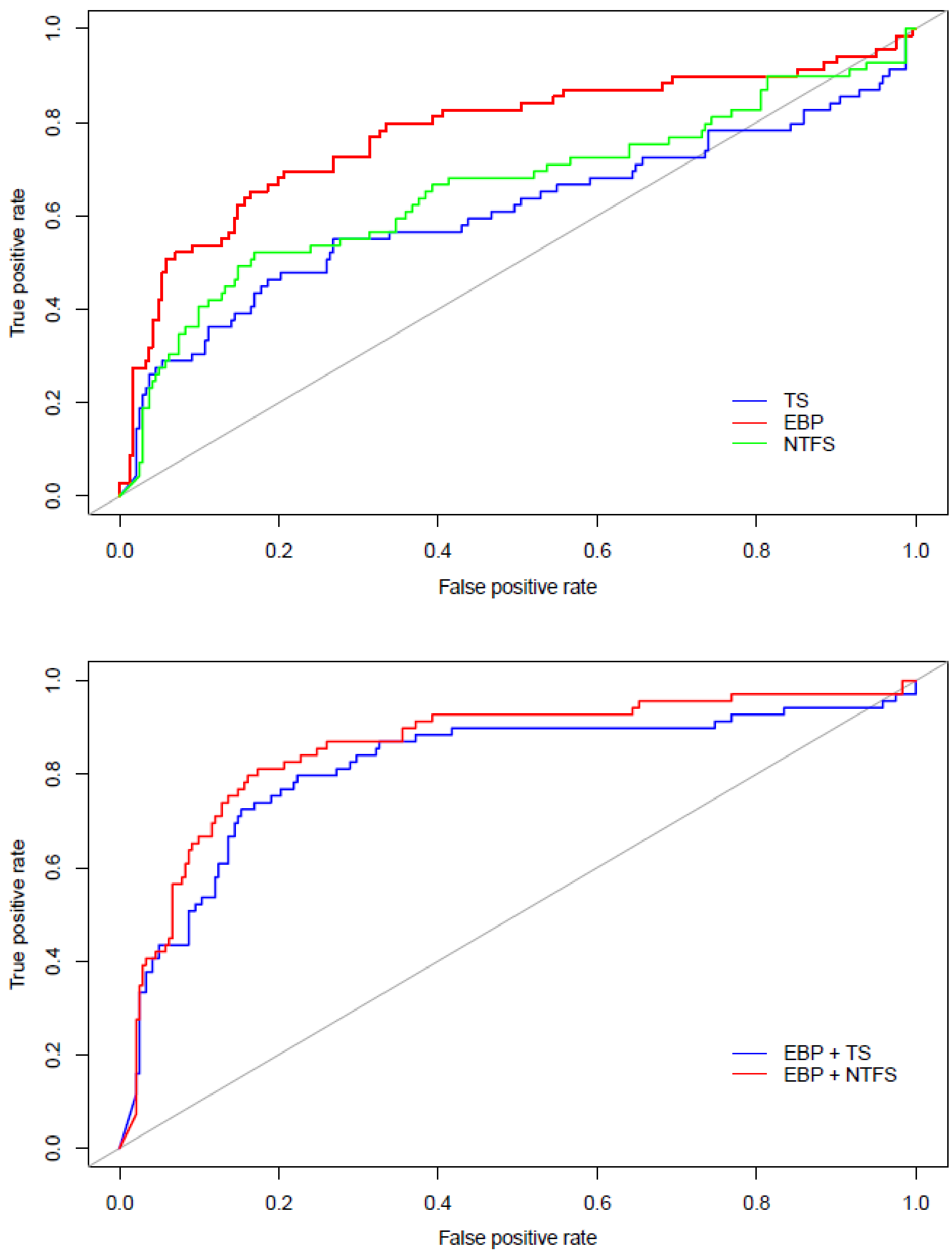
Table 1: OOS Performance of recession probability models for the U.S. economy being in a recession over the next 12 months

Model	Out-of-Sample AUROC	In-Sample AUROC	Prob. as of Sep. 2019
TS	0.61	0.74	0.62
	[0.52, 0.70]	[0.69, 0.79]	
NTFS	0.66	0.74	0.60
	[0.57, 0.74]	[0.68, 0.79]	
EBP	0.77	0.80	0.10
	[0.70, .85]	[0.75, .84]	
TS + TP	0.57	0.81	0.21
	[0.47, 0.66]	[0.76, 0.85]	
TS + EBP	0.82	0.90	0.42
	[0.75, 0.88]	[0.87, 0.93]	
NTFS + EBP	0.86	0.88	0.36
	[0.80, 0.91]	[0.85, 0.92]	
TS + TP + EBP	0.77	0.92	0.10
	[0.69, 0.85]	[0.89, 0.94]	
Tsy 3PC	0.51	0.82	0.25
	[0.44, 0.58]	[0.77, 0.86]	
TS + T. Bill	0.53	0.78	0.38
	[0.45, 0.60]	[0.73, 0.83]	

Notes: measured as AUROC. An AUROC of one denotes a perfect classifier. Brackets show 95-percent confidence intervals. Data starts 1973:02, forecasts start 1978:03 and run through 2018:08. Data is monthly. Predictors include the near-term-forward-spread (NFTS) of Engstrom and Sharpe (2019), excess bond premium (EBP), 3-month Treasury bill (T. Bill), 10-year minus 3-month Treasury Spread (TS), 10-year term premium implied by Adrian, Crump, and Moench (2013), and the first three principal components of the Treasury yield curve (Tsy 3PC) as in Johansson and Meldrum (2018).

Figure 1 plots the full ROC for a selection of models. These charts visualize the trade-offs between true and false positives implied by the models. The top panel plots the three univariate models, and it shows that the model with the EBP dominates the term spread models except for low value of thresholds. The bottom panel presents the two bivariate models of term spread and EBP, and shows that, as implied by AUROC, there is not a significant difference in terms of OOS performance between these two models.

Figure 1: Out-of-sample probit ROC curves



Notes: the top panel shows the ROC curve for the univariate probit models with TS, EBP, and NTFS as independent variables. The bottom panel shows ROC curves for the bivariate probit models with EBP and TS, and EBP and NTFS as independent variables.

Robustness: Alternative samples and forecasting horizons

One might be interested in knowing the probability of the economy transitioning into a recession, rather than simply being in a recession, in the next 12 months. For this purpose, we also consider the OOS performance of the probit models estimated and evaluated in a few different subsamples as well as for different forecasting horizons. Table 2 repeats Table 1 but with models estimated and evaluated over two different subsamples. The "Exclude Recessions" subsample excludes observations for which the economy was already in a recession during the previous month when estimating the model, while the "Exclude Recessions and ZLB" subsample excludes observations for which the economy was already in recession in the previous month as well as observations during the zero-lower-bound period (2008: m12–2015:m11). "Exclude Recessions and ZLB" is the monthly counterpart of the estimation sample used in Engstrom and Sharpe (2019).

By excluding recessionary periods from estimation and evaluation, these two subsamples estimate the probability of the U.S. economy transitioning into a recession over the next 12 months. The EBP model sees its OOS performance significantly decreased when the recession periods are excluded from estimation. This implies that much of the forecasting power of EBP comes from periods when the economy is already in recession. Meanwhile, models containing term spreads, the level of rates, or the three principal components of Treasury yields see their performance significantly improved. The bivariate models including either term spread and the EBP continue to work well out of sample, while adding TP continues to reduce the out-of-sample performance. Models with only spreads as predictors estimate a probability of transitioning into a recession of about 75 percent, while models with spreads and EBP estimate probabilities of between 60 and 65 percent.

Table 2: OOS Performance of recession probability models for the U.S. economy transitioning into a recession over the next 12 months.

Model	Exclude Recessions		Exclude Recessions and ZLB	
	AUROC	Prob. as of Sep. 2019	AUROC	Prob. as of Sep. 2019
TS	0.93	0.78	0.91	0.68
	[0.90, 0.95]		[0.88, 0.95]	
NTFS	0.93	0.73	0.93	0.62
	[0.90, 0.96]		[0.90, 0.96]	
EBP	0.59	0.09	0.46	0.08
	[0.52, 0.67]		[0.39, 0.54]	
TS + TP	0.75	0.45	0.73	0.15
	[0.67, 0.83]		[0.64, 0.81]	
TS + EBP	0.94	0.65	0.93	0.47
	[0.91, 0.94]		[0.89, 0.95]	
NTFS + EBP	0.94	0.60	0.93	0.46
	[0.91, 0.97]		[0.90, 0.96]	
TS + TP + EBP	0.71	0.47	0.69	0.12
	[0.61, 0.81]		[0.59, 0.79]	
Tsy 3PC	0.68	0.47	0.66	0.20
	[0.58, 0.78]		[0.56, 0.76]	
TS + T. Bill	0.71	0.54	0.69	0.20
	[0.62, 0.80]		[0.59, 0.78]	

Notes: measured as AUROC. An AUROC of one denotes a perfect classifier. Brackets show 95-percent confidence intervals. Data starts 1973:02, forecasts start 1978:03 and run through 2018:08. Data is monthly. "Exclude Recessions" denotes a subsample that drops observations for which the economy was already in a recession during the previous month. "Exclude Recessions and ZLB" denotes a subsample which drops observations for which the economy was already in a recession during the previous month and observations during the zero-lower-bound period (2008:m12 - 2015:m11). Predictors include the near-term-forward-spread (NFTS) of Engstrom and Sharpe (2019), excess bond premium (EBP), 3-month Treasury bill (T. Bill), 10-year minus 3-month Treasury Spread (TS), 10-year term premium implied by Adrian, Crump, and Moench (2013), and the first three principal components of the Treasury yield curve (Tsy 3PC) as in Johansson and Meldrum (2018).

Table 3 looks beyond the immediate horizon by looking at the probability of being in a recession at some point between 6 months and 12 months ahead, using either the full sample or a subsample dropping the ZLB period from the estimation. The models that contain a measure of the term spread and/or the EBP still perform well at this horizon. The EBP maintains OOS forecasting power, which when combined with the results of Table 2 implies that the EBP's forecasting power derives from improving the performance of the model when the economy is already in a recession. It is still the case that bivariate models including a term spread and the EBP work well out of sample and that adding TP reduces out of sample performance. Models with only spreads as predictors estimate a probability for the U.S. economy being in recession between 6 and 12 months ahead at about 60 percent, while models with spreads and EBP estimate probabilities of about 50 percent.

Table 3: OOS Performance of recession probability models for the U.S. economy being in a recession between 6 months and 12 months ahead.

Model	Full Sample		Excluding ZLB	
	AUROC	Prob as of Sep. 2019	AUROC	Prob as of Sep. 2019
TS	0.76	0.60	0.77	0.64
	[0.70, 0.83]		[0.70, 0.83]	
NTFS	0.80	0.61	0.81	0.65
	[0.74, 0.85]		[0.75, 0.87]	
EBP	0.75	0.10	0.70	0.13
	[0.69, 0.81]		[0.64, 0.76]	
TS + TP	0.60	0.29	0.61	0.26
	[0.52, 0.68]		[0.52, 0.69]	
TS + EBP	0.88	0.47	0.85	0.50
	[0.84, 0.91]		[0.81, 0.89]	
NTFS + EBP	0.89	0.46	0.88	0.52
	[0.85, 0.93]		[0.84, 0.92]	
TS + TP + EBP	0.72	0.23	0.69	0.22
	[0.64, 0.79]		[0.61, 0.76]	
Tsy 3PC	0.52	0.32	0.51	0.33
	[0.43, 0.60]		[0.42, 0.60]	
TS + T. Bill	0.51	0.41	0.49	0.35
	[0.43, 0.60]		[0.41, 0.58]	

Notes: measured as AUROC. An AUROC of one denotes a perfect classifier. Brackets show 95-percent confidence intervals. Data starts 1973:02, forecasts start 1978:03 and run through 2018:08. Data is monthly. "Excluding ZLB" denotes a subsample which drops observations during the zero-lower-bound period (2008:m12 - 2015:m11). Predictors include the near-term-forward-spread (NFTS) of Engstrom and Sharpe (2019), excess bond premium (EBP), 3-month Treasury bill (T. Bill), 10-year minus 3-month Treasury Spread (TS), 10-year term premium implied by Adrian, Crump, and Moench (2013), and the first three principal components of the Treasury yield curve (Tsy 3PC) as in Johansson and Meldrum (2018).

References

Adrian, Richard K. Crump, Emanuel Moench, (2013) "Pricing the term structure with linear regressions", Journal of Financial Economics, Volume 110, Issue 1, Pages 110-138.

Ang, Andrew, Piazzesi, Monika and Wei, Min, (2006), "What does the yield curve tell us about GDP growth? ", Journal of Econometrics, 131, issue 1-2, p. 359-403, https://EconPapers.repec.org/RePEc:eee:econom:v:131:y:2006:i:1-2:p:359-403.

Berge, Travis J., and Òscar Jordà. 2011. "Evaluating the Classification of Economic Activity into Recessions and Expansions." American Economic Journal: Macroeconomics, 3 (2): 246-77.

Christiansen, Charlotte , Jonas Nygaard Eriksen, and, Stig Vinther Møller (2014), "Forecasting US recessions: The role of sentiment," Journal of Banking & Finance, Volume 49, p. 459-468. https://doi.org/10.1016/j.jbankfin.2014.06.017.

Engstrom, Eric C. and Steven A. Sharpe (2019) "The Near-Term Forward Yield Spread as a Leading Indicator: A Less Distorted Mirror, " Financial Analysts Journal, https://doi.org/10.1080/0015198X.2019.1625617.

Estrella, Arturo, and Frederic S. Mishkin, (1998). "Predicting U.S. Recessions: Financial Variables As Leading Indicators," The Review of Economics and Statistics, MIT Press, vol. 80(1), pages 45-61, February.

Favara, Giovanni, Simon Gilchrist, Kurt F. Lewis, and Egon Zakrajsek (2016). "Recession Risk and the Excess Bond Premium," FEDS Notes. Washington: Board of Governors of the Federal Reserve System, April 8, 2016, http://dx.doi.org/10.17016/2380-7172.1739.

Gilchrist, Simon and Egon Zakrajsek (2012), "Credit Spreads and Business Cycle Fluctuations," American Economic Review, vol. 102 (June), pp. 1692–720.

Johansson, Peter, and Andrew Meldrum (2018). "Predicting Recession Probabilities Using the Slope of the Yield Curve," FEDS Notes. Washington: Board of Governors of the Federal Reserve System, March 1, 2018, https://doi.org/10.17016/2380-7172.2146.

Liu, Weiling, and, Emanuel Moench (2016), "What predicts US recessions?" International Journal of Forecasting, Volume 32, Issue 4, 2016, p. 1138-1150. https://doi.org/10.1016/j.ijforecast.2016.02.007.

Rudebusch, Glenn and Williams, John, (2009), Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve, Journal of Business & Economic Statistics, 27, issue 4, p. 492-503.