JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 61, No. 3, May 2026, pp. 1216–1246
© THE AUTHOR(S), 2025. PUBLISHED BY CAMBRIDGE UNIVERSITY PRESS ON BEHALF OF THE MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON
doi:10.1017/S0022109025102093

Equity Premium Predictability over the Business Cycle

Emanuel Moench
Frankfurt School of Finance & Management, CEPR
e.moench@fs.de (corresponding author)

Tobias Stein
Independent
stein_to@web.de

Abstract

Equity returns follow a pronounced V-shape pattern around the onset of recessions. They sharply drop into negative territory just before business cycle peaks and then strongly recover as the recession unfolds. Recessions are typically preceded by a flat yield curve. Probit models relying on the term spread as a predictor therefore time the beginning of recessions well. We show that model-implied recession probabilities based on the term spread strongly improve equity premium prediction in- and out-of-sample and outperform several benchmark predictors. Correcting for a structural break in the mean of the term spread in 1982 further strengthens the forecast performance.

I. Introduction

Whether stock returns are predictable has been the subject of a long debate in finance. A large literature documents in-sample predictability using a host of financial and economic variables such as valuation ratios, the default spread, or the consumption-wealth ratio as predictors (see, e.g., Fama and French (1988), Campbell and Shiller (1988), and Lettau and Ludvigson (2001)). However, an influential paper by Welch and Goyal (2008) shows that none of the proposed predictors of the equity premium would have consistently outperformed a simple historical average return out-of-sample.[1] Since then, a growing literature has proposed alternative predictors and forecasting methods that appear to provide superior statistical predictability relative to the historical average benchmark; see Rapach and Zhou (2013) for an overview.[2]

We thank Jonah Schmidt for research assistance and Arturo Estrella, Roberto Gómez-Cram, Amit Goyal, Mete Kilic, and Stijn Van Nieuwerburgh as well as anonymous referees and Thierry Foucault (the editor) for helpful comments and suggestions.

[1]While the equity premium refers to the *expected* excess return on the stock market, much of the forecasting literature (including Welch and Goyal (2008)) uses the term "equity premium" interchangeably with the *realized* excess market return. Here, we follow this convention in the literature and do not consistently make the important distinction between expected and realized excess returns throughout the article.

[2]Among others, recently proposed predictors include the output gap (Cooper and Priestley (2009)), short interest (Rapach, Ringgenberg and Zhou (2016)), industrial electricity usage (Da, Huang, and Yun (2017)), gold-to-platinum ratio (Huang and Kilic (2019)), variance risk premium (Pyun (2019)), and investor attention (Chen, Tang, Yao, and Zhou (2020)). Methodological contributions include non-negativity constraints

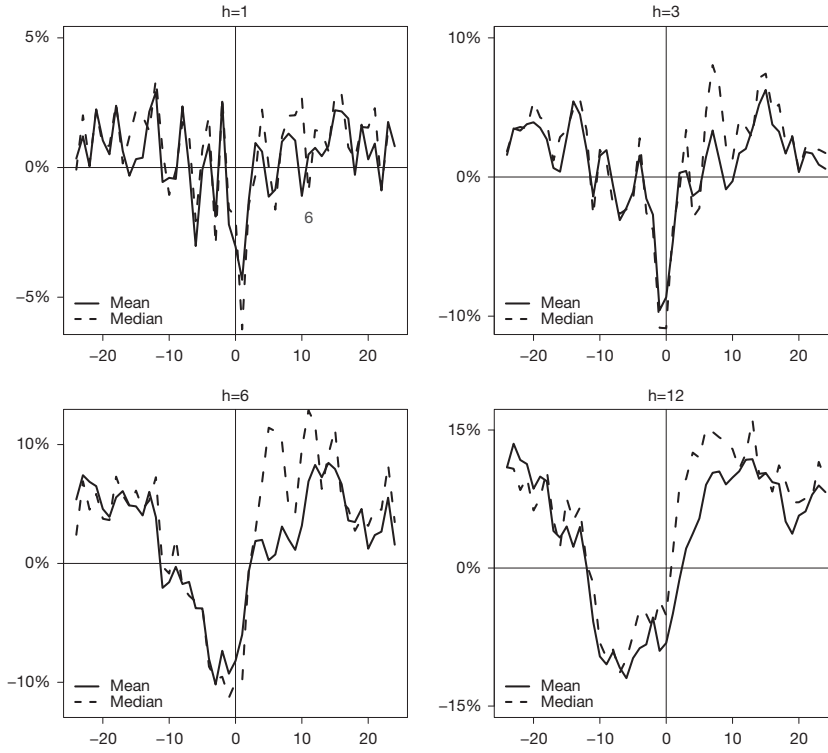
https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

A common finding in that literature is that predictability primarily arises around recessions. This is consistent with a related literature suggesting that expected equity returns vary over the business cycle (e.g., Fama and French (1989), Ferson and Harvey (1991), Cochrane (2007), and Campbell and Diebold (2009)). Lustig and Verdelhan (2012) document significant variation in realized excess returns around recessions, showing that they are negative at the business cycle peak and then sharply rise over the following quarters. This is confirmed by Figure 1 which depicts the forward-looking arithmetic mean and median of the U.S. log excess market return over different time windows around the 11 NBER recessions from March 1951 to December 2020. The realized market return is

FIGURE 1

Log Excess U.S. Equity Market Return Around Business Cycle Peaks

Figure 1 presents the arithmetic average and median of the (cumulative) log excess U.S. equity market return around the 11 U.S. recessions in the sample from 1951:3 to 2020:12. Specifically, the excess return is the difference between value-weighted returns on the S&P 500 index (including dividends) and the Treasury bill rate. The vertical axis depicts $\sum_{j=0}^{h-1} r_{t+j}$ for $t = -24, \ldots, -1, 0, 1, \ldots, 24$, where r_{t+j} is the realized excess return in the month $t+j$. The horizontal axis displays the 24 months before and after a business cycle peak, with $t = 0$ referring to the first month of a NBER-dated recession. Results are shown for $h = 1, 3, 6, 12$.



(Campbell and Thompson (2008)), combination forecasts (Rapach, Strauss, and Zhou (2010)), time-varying coefficient models (Dangl and Halling (2012)), principal component analysis (Neely, Rapach, Tu, and Zhou (2014)), economic constraints (Pettenuzzo, Timmermann, and Valkanov (2014)), and machine learning techniques (Gu, Kelly, and Xiu (2020)).

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

mostly negative but relatively volatile for the 1-month and 3-month windows around the beginning of recessions. However, a clear V-shape emerges for the cumulative 6- and 12-month-ahead horizons, highlighting that equity returns are sharply negative around business cycle peaks but strongly recover thereafter.

Table 1 presents moments of the annualized realized log excess market return over the business cycle for the same 70-year period. The total annual equity premium was 6.4% with a standard deviation of 14.6%, implying a Sharpe ratio of 0.4. Focusing only on NBER expansions, the excess market return amounted to 8.7% with an annualized Sharpe ratio of around 0.7. In recessions, it was negative at −7%. Zooming in around business cycle peaks, we see that the equity premium tended to be strongly negative during the 6 months before and after the business cycle peak, with annualized values of almost −8% and −17%, respectively. Hence, the stock market, on average, incurs large losses in the 1-year window around the beginning of recessions. While it tends to recover in the subsequent months, on average, it only gains an annualized 3.8% 6 to 11 months after the peak. The last 2 rows in Table 1 show the annualized log excess return for samples that exclude the 12 months and the 24 months around the beginning of recessions, respectively. When excluding 2 years (1 year) of observations around each peak, the average equity premium and Sharpe ratio rise to 10.9% and 0.8 (9.8% and 0.7), compared to 6.4% and 0.4 for the full sample. This evidence strongly suggests that, to the extent that one can predict the beginning of recessions, one should be able to time the market.

It is well documented that the term spread is a robust predictor of recessions for horizons of 1 year ahead and longer. In the post-war period, every single U.S. recession was preceded by an inverted yield curve. In a seminal paper, Estrella and Hardouvelis (1991) show that an inverted yield curve is indeed a strong predictor of recessions and future real economic activity. Estrella and Mishkin (1998) complement this finding by comparing the predictive power of the term

TABLE 1

Log Excess Stock Market Return over the Business Cycle

Table 1 reports the annualized mean, median, standard deviation, and Sharpe ratio of the monthly U.S. log equity premium. The equity premium is the difference between value-weighted returns on the S&P 500 index (including dividends) and the Treasury bill rate. $\rho(1)$ (N) denotes the first-order serial correlation (number of observations). The descriptive statistics are presented for the full sample from March 1951 to December 2020, as well as separately for recessions and expansions. Further statistics are provided for subsamples before and after business cycle peaks. Peak refers to the peak month of NBER-dated business cycle contractions (first month of a recession). The total number of recessions in the sample is 11.

	1951:3 to 2020:12							
Log Equity Premium	Mean	Median	Std. Dev.	Sharpe Ratio	Skewness	Kurtosis	$\rho(1)$	N
Full sample	6.38	10.96	14.59	0.44	−0.67	5.35	0.05	838
Recessions	−6.97	−3.76	20.58	−0.34	−0.25	2.92	0.16	124
Expansions	8.70	11.44	13.19	0.66	−0.68	6.12	−0.06	714
Before the peak:								
peak −12 to peak −7	7.77	11.22	12.48	0.62	0.22	2.56	−0.15	66
peak −6 to peak −1	−7.53	−4.85	12.81	−0.59	−0.37	2.9	−0.09	66
At/after the peak:								
peak to peak +5	−16.40	−11.14	17.93	−0.91	−0.26	2.62	0.04	66
peak +6 to peak +11	3.77	16.18	19.91	0.19	−0.61	3.85	0.26	65
Excl. peak −6 to peak +5	9.81	12.06	14.19	0.69	−0.72	6.11	0.01	706
Excl. peak −12 to peak +11	10.92	12.08	13.63	0.80	−0.74	6.54	−0.01	581

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

spread with financial variables such as stock prices and other spreads, as well as monetary aggregates. While some alternative predictors are useful over 1- to 3-quarter horizons, it is the term spread that predicts best over horizons of 1 year and longer. Moreover, the binary models for recessions are found to be more stable than continuous models for economic growth (Estrella, Rodrigues, and Schich (2003)), and the relation is also present in other countries (Bernard and Gerlach (1998), Chinn and Kucko (2015)).[3] More recently, Liu and Moench (2016) confirm that the term spread is a robust in-sample and out-of-sample predictor of U.S. recessions. They also document that the additional incorporation of lagged observations of the term spread further improves recession forecasts.

In this article, we make use of the recession-timing ability of the yield curve for predicting stock market excess returns. We thus exploit the above documented fact that equity premia follow a pronounced V-shaped pattern around the beginning of recessions, which in turn, can be reasonably well timed using information in the yield curve. Specifically, we estimate probit models in the spirit of Estrella and Hardouvelis (1991) to predict the likelihood of a recession starting in the next 12 months. We then use the estimated probabilities as inputs in linear predictive regressions to forecast the equity premium over horizons from 1 to 12 months ahead. We first show that recession probabilities based on term structure information are strong in-sample predictors of stock market excess returns in post-war U.S. data. We then document that the strong predictive power of recession probabilities carries over to an out-of-sample setting. Our key finding is that the V-shaped pattern of excess returns around business cycle peaks is well captured by real-time recession probability forecasts based on information in the yield curve and can be exploited in real time. While our evidence is mostly based on U.S. data, we also document this market timing ability for the United Kingdom, Germany, and France.

We start by documenting that a backward-looking 3-year moving average of the term spread substantially strengthens the recession classification ability of the term spread by reducing false positives and better timing the beginning of recessions. Several authors have argued that the probit model for forecasting recessions with the term spread suffers from a structural break (see, e.g., Chauvet and Potter (2002), (2010)). We indeed document strong evidence for a structural break in the mean of the term spread in 1982 and show that it would have been possible for investors to identify this break in real time a few years after it occurred. We follow Lettau and Van Nieuwerburgh (2008) and Pesaran and Timmermann (2007) and apply four different methods to correct for the break in the term spread. All further improve the out-of-sample R^2 for forecasting the equity premium using recession probabilities based on the term spread. This improvement is partly due to the fact that the real-time break-corrected recession probabilities better predict the beginning of the 2001 and 2008–2009 recessions.

In terms of predictive ability, our approach outperforms other recently proposed predictors including the variance risk premium of Bollerslev, Tauchen, and Zhou (2009), "short interest" of Rapach et al. (2016), and the "gold-to-platinum"

[3]Wheelock and Wohar (2009) provide a comprehensive survey of the ability of the term spread to predict recessions.

ratio of Huang and Kilic (2019). The out-of-sample R^2 is above 1% for monthly forecasts and often higher than 10% for cumulative 1-year-ahead forecasts. Moreover, we perform an asset allocation exercise for a mean–variance investor who invests in the equity market and the risk-free rate. This exercise reveals an excellent market timing ability of recession probability forecasts, which is even more pronounced for the break-correction methods. The models signal to run down equity exposure before the onset of recessions when the yield curve flattens and to re-enter the market toward the end of recessions when the yield curve steepens. An investor who forecasts with (break-corrected) recession probabilities increases the Sharpe ratio to around 0.85 compared to 0.50 for the buy-and-hold investor. Using a VAR-based decomposition in the spirit of Campbell (1991), we find that the predictability is driven by both higher anticipated discount rates and lower expected future dividends, consistent with countercyclical risk premia. We also show that our results are robust to taking into account transaction costs and that recession probabilities predict a wide range of portfolios sorted on various firm characteristics (e.g., industry portfolios).

In a related recent paper, Gómez-Cram (2022) studies 1-month-ahead equity premium predictability over the business cycle. Consistent with his results, we find that stock returns are negative at the beginning of recessions and that business cycle variables help to time these periods. Despite this broad similarity, there are a number of important differences between our article and his paper. First, while Gómez-Cram (2022) studies only 1-month-ahead forecasts, we predict equity returns also over longer horizons. Second and more importantly, we combine the recession and equity premium prediction literatures by directly using recession probability forecasts to forecast equity returns, while Gómez-Cram (2022) uses a state-space model to link expected excess equity returns to the business cycle. While he estimates a common growth component from real-time data of nine U.S. key macroeconomic aggregates, we instead confirm that the term spread is a robust leading indicator of recessions and strongly improves equity premium forecasts. In line with our results, Andreasen, Engsted, Møller, and Sander (2021) show that the yield spread better predicts bond risk premiums when conditioning on the business cycle. We compare our results with those implied by the common growth component of Gómez-Cram (2022) in the Supplementary Material.

More generally, our findings are in line with Rapach et al. (2010) and Dangl and Halling (2012) who find that the predictive power of combination forecasts and time-varying coefficient models primarily arises from business-cycle variation in the equity premium. Importantly, we show that equity premium forecasts based on recession probabilities outperform the historical average benchmark also in expansions. The reason is that by correctly anticipating low equity market returns heading into recessions, they also correctly predict higher returns in business cycle booms.

From an economic perspective, our results are consistent with Baron, Xiong, and Ye (2022) who document that a credit-based measure of disaster risk can predict severe output crashes while the implied disaster probabilities predict equity returns negatively, in line with slow information processing in the equity market. This is also consistent with recent work by Ghaderi, Kilic, and Seo (2024) who show theoretically that investors learning about the state of the economy are willing to

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

pay a premium for high future variance at times when the economy is performing poorly and future volatility is associated with better economic conditions.

II. Empirical Results

This section presents our empirical results. We describe our data in Section II.A. In Sections II.B and II.C, we first confirm the ability of probit models along the lines of Estrella and Mishkin (1998) to predict NBER recessions. We document that the forecast performance of the standard probit model using the term spread as the only explanatory variable strongly improves when lagged information about the term spread is added. We show in Section II.D that the recession probabilities implied by probit models have strong in-sample predictive power for the U.S. equity premium. In Section II.E, we then document that this predictive power carries over to a real-time out-of-sample forecast setting. We further provide evidence for a structural break in the mean of the term spread and show that adjusting for the break in real time improves recession and equity premium forecasts. In Section II.F, we perform an asset allocation exercise showing that the recession forecasts based on information in the yield curve significantly improve market timing. Finally, in Section II.G, we show that estimated recession probabilities forecast the equity premium by predicting higher discount rates and lower future dividends.

A. Data

We obtain data on the equity premium and term spread from Amit Goyal's homepage (https://sites.google.com/view/agoyal145). The equity premium is computed as the continuously compounded log return on the S&P 500 index, including dividends, minus the Treasury bill rate (Welch and Goyal (2008)). The term spread (TMS) is calculated as the difference between the long-term government bond yield and the Treasury bill rate. The yields are taken from Ibbotson's *Stocks, Bonds, Bills, and Inflation Yearbook* and have a maturity of approximately 20 years (Ibbotson and Sinquefield (1976)). Our data set consists of monthly observations from March 1951 to December 2020. We start our analysis in March 1951 after the Treasury-Federal Reserve Accord, which laid the foundation for an independent monetary policy (Lacker (2001)). During World War II and the 6 years afterward, the Fed was tasked to support the financing requirements of the Treasury by stabilizing long-term interest rates (Eichengreen and Garber (1991), Carlson and Wheelock (2014)). Hence, we begin our analysis after this extraordinary period of pegged interest rates. The business cycle chronology with classifications into expansions and recessions is taken from NBER. A business cycle peak is defined to be the first month of a recession. We start our pseudo out-of-sample forecasting exercise in 1980 when the Business Cycle Dating Committee of the NBER began to release timely announcements of its business cycle classifications.

B. Predicting Recessions

In this section, we are interested in predicting U.S. recessions as classified by the NBER Business Cycle Dating Committee. The literature typically

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

distinguishes between the probability of a recession in exactly h months and the probability of a recession within the next h months. Here, we focus on the latter, as we aim to forecast cumulative log equity premiums over the next h months in later sections; for similar definitions, see Wright (2006) and Johansson (2018). More precisely, let $Y_{t+1:t+h}=1$ if the NBER has classified at least 1 month between $t+1$ and $t+h$ as a recession. We follow common practice and assume that $Y_{t+1:t+h}$ is based on a latent variable $Y^*_{t+1:t+h}$ where $Y_{t+1:t+h}=1$ for $Y^*_{t+1:t+h}\geq 0$ and $Y_{t+1:t+h}=0$ for $Y^*_{t+1:t+h}<0$. The latent variable is assumed to follow a (multivariate) linear regression model:

$$Y^*_{t+1:t+h}=X_t'\beta+\epsilon_{t+1:t+h}, \tag{1}$$

$$\Pr\left[Y_{t+1:t+h}=1|X_t\right]=\Phi\left[X_t'\beta\right], \tag{2}$$

where $X_t'=\left(1,x_{1,t},\ldots,x_{p,t}\right)'$ is the $1\times(p+1)$ vector of predictor variables including the intercept, β is the $(p+1)\times 1$ vector of coefficients, and $\epsilon_{t+1:t+h}$ is the error term. Further, $\Phi[\cdot]$ is the cumulative distribution function of the standard normal distribution and Pr denotes probability. Let $\hat{p}_{t+1:t+h|t}$ be the out-of-sample forecast for $\Pr[Y_{t+1:t+h}=1]$ based on information contained in X_t. We follow Jacobsen, Marshall, and Visaltanachoti (2019) and account for the fact that the NBER typically publishes business cycle classifications with a substantial delay by estimating the β coefficients with information up to $t-24$. This is a conservative choice as other authors (see, e.g., Kauppi and Saikkonen (2008)) only account for a delay of 1 year.[4] The sample with T observations is split into an in-sample estimation period of M months and an out-of-sample period of $T-M$ months. We only use data that are available in real time mimicking as closely as possible the information an investor would have had.

In what follows, we focus on the simple probit model using (transformations of) the term spread as input. The first model only includes a constant and the term spread as predictors—this is the model of Estrella and Hardouvelis (1991). It is well known that this model performs well for forecast horizons of 1 to 2 years, with only weak predictability for shorter horizons. More recently, Liu and Moench (2016) show that the short-horizon forecasts substantially improve when adding lagged observations of the term spread. Building on this, the second model includes the term spread lagged by 6 months as an additional predictor. As we will see below, the recession probabilities implied by the two models closely track the dynamics of the term spread and thus tend to be quite volatile. This implies a number of false positive signals about impending recessions. To address this issue, we also consider a specification that includes a constant, the term spread, as well as a backward-looking 3-year moving average of the term spread (MA-TMS) which we construct as $\frac{1}{36}\sum_{j=0}^{35}\text{TMS}_{t-j}$.[5]

[4]The longest delay in our sample was 21 months: the NBER announced the November 2001 business cycle trough on July 17, 2003. Results are very similar when we use information up to $t-12$.

[5]We show in the Supplementary Material that our results are robust to the length of the moving average window.

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

FIGURE 2

Term Spread and Moving Average Term Spread

Figure 2 presents the term spread (dashed line) and the moving average term spread (solid line), whereby the latter is the moving average of past 3-year observations of the term spread. The time series are normalized to a mean of 0 and a standard deviation of 1. The sample is 1951:3 to 2020:12 and shaded areas indicate NBER-dated recession periods.

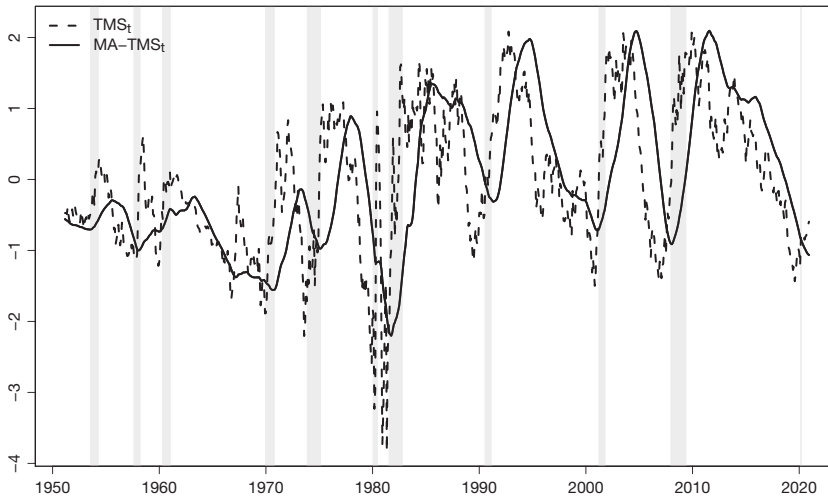


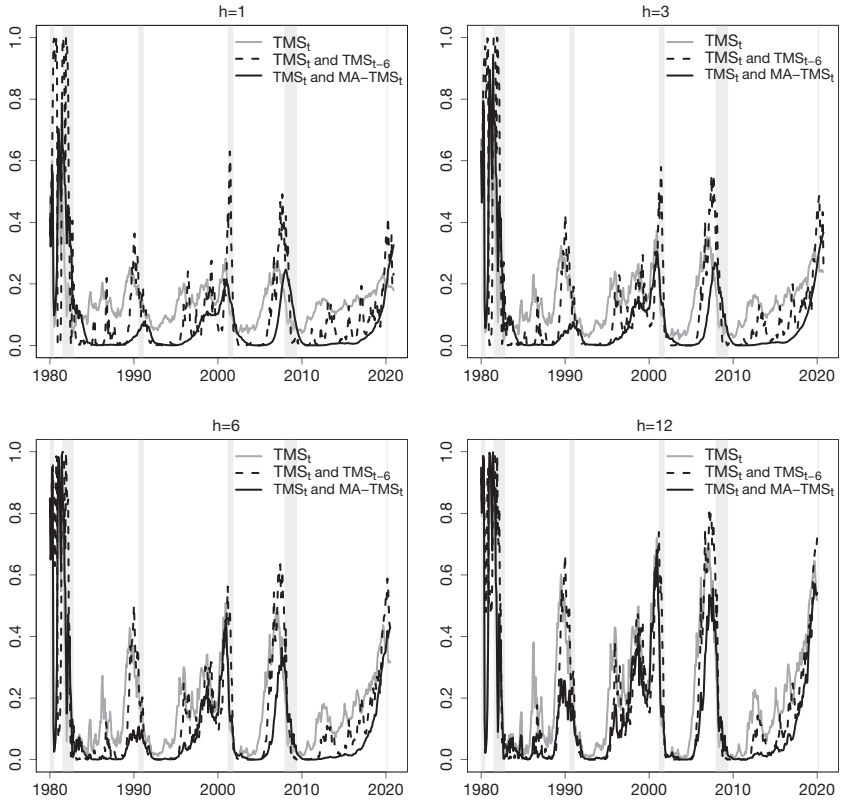
Figure 2 superimposes TMS and MA-TMS. While the local minima of the term spread usually lead the beginning of recessions by several months, the moving average term spread reaches its troughs often just before the onset of recessions. Moreover, the smoothed series averages out some local minima that are not followed by recessions. The smoothing thus emphasizes lower frequency components of the term spread which appear to be more relevant for signaling recessions. We will indeed show below that the incorporation of lagged and averaged term spread information into the probit model significantly enhances short horizon forecasts. Figure 3 shows the out-of-sample recession probability forecasts for the three models from 1980:1 to 2020:12. Several points are noteworthy. First, the model with only a constant and the term spread performs relatively poorly for $h = 1$ and $h = 3$, with several false positives and no pronounced differences between expansions and recessions since the mid-1980s. The performance of this model gradually increases in the forecast horizon. This is consistent with the prior literature which has documented an improved recession prediction with the term spread for horizons beyond 6 months. Second, the models adding lagged term spread information perform substantially better for short-horizon forecasts, where the model with the moving average term spread implies substantially smoother recession probabilities. This is in line with Rudebusch, Sack, and Swanson (2007), who show that the 1-year lagged term premium predicts future GDP growth, and that differences rather than levels of the expectations component and term premium matter more for forecasting real output growth. The finding that lagged observations of the term spread improve recession predictability is consistent with monetary policy affecting the economy with a delay of a few quarters (Rudebusch and Williams (2009)).

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

FIGURE 3

Out-of-Sample Recession Probability Forecasts

Figure 3 shows the out-of-sample recession probability forecasts at the 1-, 3-, 6-, and 12-month horizons. Results are presented for three different forecasting models: the solid gray line depicts forecasts from a model with the term spread as a predictor, whereas the solid (dashed) black line denotes a model with the term spread and the moving average of the past 3 years (6-month lagged value) of the term spread as predictors. Gray bars denote NBER-dated recession periods, and the out-of-sample period is 1980:1 to 2020:12.



Third, the recession probabilities in the 1990s and 2000s are less pronounced compared to the probabilities in the early 1980s, with values rarely exceeding 50% even in recessions. Similarly, Estrella et al. (2003) document that the signal of the models was weaker in the 1990–1991 recession compared to the early 1980s. Kauppi and Saikkonen (2008) also document—using probit models with lagged dependent variables—that the 1990–1991 and 2001 recessions were difficult to predict. This pattern is not unique to models using the term spread as predictor; the lack of predictability is also documented for models with larger sets of predictors (Hamilton (2011), Fornaro (2016)). We show in Supplementary Material Appendix C that the weaker recession signals result from a structural break in the mean of the term spread in the early 1980s. The probabilities are considerably stronger when this break is accounted for.

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

C. Forecast Evaluation

We follow the recession prediction literature and use the quadratic probability score (QPS), the logarithm score (LS), and the diagonal elementary score (DES) to formally evaluate the accuracy of recession probability forecasts.[6] Perfect classification ability results in values of 0 for all three scores; otherwise, they have positive values, with higher values indicating poorer forecast performance (see, e.g., Nyberg (2013), Christiansen, Eriksen, and Møller (2014), and Fornaro (2016) for recent applications). The loss function of LS penalizes large forecast errors more heavily than the loss function of QPS (Diebold and Rudebusch (1989)). Galvão and Owyang (2020) argue that the loss functions of LS and DES are more suitable in the context of rare events such as recessions. We further calculate the out-of-sample pseudo R^2 (Estrella (1998)) and the area under the receiver operating characteristic curve (AUROC). While the QPS, LS, and DES metrics evaluate the model accuracy, the AUROC is a metric to evaluate the classification ability—here into recessions and expansions—of a forecasting model. A perfect classifier has an AUROC of 1, and a coin-toss classifier has an AUROC of 0.5; for further details, see Berge and Jordà (2011).

Table 2 presents values of the QPS, LS, DES, pseudo R^2, and AUROC for the three different models for horizons $h = 1,3,6,12$ months ahead. The model with a constant and the term spread performs worst, with an AUROC close to 0.5 and a negative pseudo R^2 for $h = 1$ and $h = 3$. The performance improves with the forecast horizon, with an AUROC of 0.75 and a pseudo R^2 of 0.2 for $h = 12$, reflecting the well-known finding that the term spread has a lead time of about 4 to 6 quarters (Estrella and Mishkin (1998)). When adding the 6-month lagged term spread and the moving average term spread as predictors, the model accuracy and classification ability substantially increase, especially for short-horizon forecasts. Each of the evaluation statistics improves for these more sophisticated probit models. At the 1-month-ahead horizon, the AUROC jumps from 0.48 for the model with only the term spread to 0.81 for the model adding the lagged term spread and to 0.91 for the model adding MA-TMS. These differences in AUROC values, shown in the second to last column of the table, are highly statistically significant according to the test of Hanley and McNeil (1983). While the marginal improvement of predicting recessions by adding TMS_{t-6} or MA-TMS declines with the forecast horizon, it is statistically significant at the 1% level at all horizons. This highlights that adding lagged term spread information increases the recession classification precision of the probit models.[7]

The last column of Table 2 shows the correlation between the implied recession probabilities and the (cumulative) log equity premium over the next h months (ρ). The negative figures indicate that the equity premium tends to decrease when the recession probability rises, and that this correlation pattern is more pronounced

[6]Details on the estimation of all statistics in this section are provided in Supplementary Material Appendix A.

[7]We also formally test the null hypothesis that the AUROC value for the model with TMS and MA-TMS is equal to the AUROC value for the model with TMS and lagged TMS against the 1-sided alternative that the former is statistically significantly larger than the latter (Hanley and McNeil (1983)). While the null is rejected at the 5% level for $h = 1,3,6$, it cannot be rejected for $h = 12$.

TABLE 2

Out-of-Sample Recession Prediction Performance: Forecast Evaluation

Table 2 presents 5 forecast evaluation statistics for the out-of-sample performance of three different probit models, as well as the correlation between the probability forecasts and the (cumulative) log equity premium (ρ). The statistics are the quadratic probability score (QPS), logarithm score (LS), diagonal elementary score (DES), pseudo R^2, as well as the area under the receiver operating characteristic curve (AUROC). The predictor variables are the term spread (TMS) and lagged and averaged variants of the term spread. MA-TMS$_t$ refers to the backward-looking 3-year moving average of the term spread, and "Historical average" depicts forecasts from a probit model with only a constant. The recession probability forecasts refer to the probability that a recession occurs within the next h months. Results are shown for h = 1,3,6,12 and the out-of-sample period is 1980:1 to 2020:12. We test the null hypothesis that AUROC = 0.5 (random classification) against the 2-sided alternative (Hanley and McNeil (1982)); asterisks for this test are provided next to the AUROC value. ΔAUROC shows the gains relative to the probit model with TMS$_t$ only and asterisks denote that the AUROC of the respective bivariate model is significantly larger than the AUROC of the probit model with TMS$_t$ only based on the test of Hanley and McNeil (1983). *, **, and *** asterisks denote significance at the 10%, 5%, and 1% significance levels.

	1980:1 to 2020:12						
Variables in Probit Model	QPS	LS	DES	PseudoR^2	AUROC	ΔAUROC	ρ
Panel A. h = 1							
TMS$_t$	0.23	0.40	0.12	−0.01	0.48		−0.04
TMS$_t$, TMS$_{t-6}$	0.17	0.34	0.05	0.12	0.81***	0.33***	−0.07
TMS$_t$, MA-TMS$_t$	0.18	0.27	0.05	0.26	0.91***	0.43***	−0.12
Historical average	0.23	0.40	0.11	0.00	0.42**		0.04
Panel B. h = 3							
TMS$_t$	0.26	0.44	0.11	−0.00	0.56		−0.08
TMS$_t$, TMS$_{t-6}$	0.19	0.36	0.06	0.16	0.82***	0.26***	−0.14
TMS$_t$, MA-TMS$_t$	0.21	0.32	0.07	0.25	0.89***	0.33***	−0.20
Historical average	0.27	0.44	0.13	0.00	0.41***		0.08
Panel C. h = 6							
TMS$_t$	0.28	0.47	0.12	0.05	0.65***		−0.11
TMS$_t$, TMS$_{t-6}$	0.22	0.40	0.08	0.18	0.82***	0.17***	−0.17
TMS$_t$, MA-TMS$_t$	0.23	0.37	0.09	0.25	0.88***	0.23***	−0.23
Historical average	0.31	0.50	0.15	0.00	0.40***		0.10
Panel D. h = 12							
TMS$_t$	0.29	0.49	0.11	0.20	0.75***		−0.23
TMS$_t$, TMS$_{t-6}$	0.21	0.39	0.07	0.37	0.87***	0.12***	−0.32
TMS$_t$, MA-TMS$_t$	0.25	0.41	0.10	0.34	0.87***	0.12***	−0.36
Historical average	0.40	0.59	0.18	0.00	0.42**		0.12

for the models including lagged term spread information and for longer forecast horizons. At the 1-year-ahead horizon, the model using the backward-looking moving average term spread as additional regressor features a sizable negative 36% correlation of the implied recession probability with the cumulative equity premium over the next year.

The classification into expansions and recessions based on model-implied probabilities requires the definition of a threshold level. As a result, the proportion of correctly predicted recessions (percentage of true positives, PTP) and the proportion of falsely predicted recessions (percentage of false positives, PFP) are functions of this threshold. The receiver operating characteristic (ROC) curve traces all combinations of PTP and PFP for different thresholds in the unit box. The diagonal line represents uninformative forecasts (PTP = PFP). Curves above the diagonal line depict informative forecasts, and the ROC curve of a perfect classifier "will hug the north-west border of the positive unit quadrant" (Berge and Jordà (2011)).

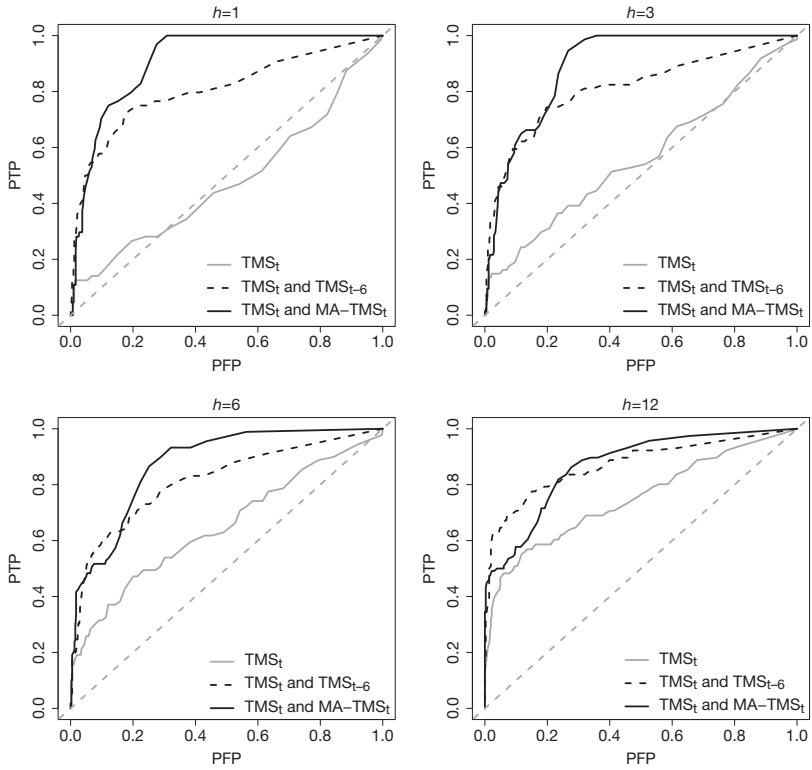
Figure 4 presents ROC curves for the three different forecasting models. The model with a constant and the term spread (solid gray line) is close to the diagonal line for h = 1 and h = 3 but shifts toward the north-west corner for h = 6 and h = 12. Hence,

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

FIGURE 4

Out-of-Sample Performance: ROC Curves

Figure 4 shows the receiver operating characteristic (ROC) curve for three different probit models. The solid gray line depicts a model with a constant and the term spread, and the solid (dashed) black line presents the performance of a model when the moving average term spread (6-month lagged term spread) is added as a predictor. The dashed gray line is the 45 degree line. Results are shown for the out-of-sample period from 1980:1 to 2020:12. The vertical axis depicts the percentage of true positives (PTP) and the horizontal axis depicts the percentage of false positives (PFP). Predictions are made for a recession starting within the next $h = 1,3,6,12$ months.



the model is relatively uninformative for short horizons but gains predictive power with increasing h. Adding the lagged term spread and the moving average component helps to predict recessions especially at these shorter horizons. The ROC curves substantially shift toward the north-west, and the improvements for the moving average component are highest for $h = 1,3,6$. This is consistent with Figure 3: the first and second models have relatively high recession probabilities during the 1990s and 2010s, thus generating some false positives for low threshold levels. Overall, the ROC curves of the models using lagged and averaged term spreads lie well above the corresponding curve of the spread-only model for all forecast horizons and threshold values. Thus, adding information on the lagged term spread strongly improves recession prediction, in line with Liu and Moench (2016).

D. Forecasting the Equity Premium with Recession Probability Forecasts

Having shown that equity premiums are particularly low around business cycle peaks and that the onset of recessions can be well predicted using yield curve

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

information, we next assess the usefulness of implied recession probabilities to forecast the equity premium. We use the standard linear predictive regression model:

$$r_{t+1:t+h} = \alpha_h + \beta_h \hat{\mathrm{p}}_{t+1:t+h|t} + \epsilon_{t+1:t+h},$$

where $r_{t+1:t+h} = \frac{1}{h}\sum_{j=1}^{h} r_{t+j}$ is the average of the cumulative log equity premium between $t+1$ and $t+h$, α_h and β_h are coefficients, and $\epsilon_{t+1:t+h}$ is the error term. The only predictor is the recession probability forecast $\hat{\mathrm{p}}_{t+1:t+h|t}$ which uses only information up to period t.

We start by documenting the in-sample predictive power of recession probabilities. Table 3 provides the estimated slope coefficients for the full sample from March 1951 through December 2020 (Panel A) and for a shorter sample starting in 1980 (Panel B). We consider this subsample in light of the evidence that the properties of the business cycle in the U.S. have changed in the early 1980s (McConnell and Perez-Quiros (2000)). We compare the results for recession probabilities from our three probit models with two benchmark recession probabilities: the recession probability provided by the Federal Reserve Bank of New York (https://www.newyorkfed.org/research/capital_markets/ycfaq#/) which is based on the model by Estrella and Hardouvelis (1991), and the 1-month-ahead probability from Chauvet and Piger (2008).[8] Importantly, while the New York Fed model also considers the term spread as the only predictor, it forecasts a recession in exactly 1 year instead of within the next 12 months as in our models.

Several remarks are in order. First, the implied recession probabilities for all three models significantly predict the equity premium at most horizons. Not surprisingly, the statistical significance of the estimated slope coefficients and the R^2 increase with the horizon. While the spread-only model performs marginally better in the full sample, the models using lagged term spread information have considerably stronger predictive ability in the post-1980 sample. The model using the moving average term spread as an additional predictor stands out, delivering a whopping 13% adjusted R^2 in this period. Second, and more importantly, the three models also strongly outperform the two benchmark models at forecast horizons of 3 months and beyond. The New York Fed recession probability and the one by Chauvet and Piger (2008) have better forecasting ability only at the 1-month-ahead horizon.

In the top panel of Table B.1 in the Supplementary Material, we provide in-sample results for a much longer time period, starting in 1926 and ending in 2022. This longer sample thus also covers the brief recession triggered by the COVID-19 pandemic. While the benchmark recession probabilities are not available going back to 1926, the results support our finding that recession probabilities derived from the term spread are useful predictors of equity market excess returns. Whether this information could indeed have been exploited by investors in real time will be assessed in the next section.

[8]The New York Fed recession probabilities are provided here: https://www.newyorkfed.org/research/capital_markets/ycfaq#/.

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

TABLE 3

In-Sample Equity Premium Prediction with Estimated Recession Probabilities

Table 3 reports in-sample results for the slope coefficient from regressions of log excess returns ($r_{t+1:t+h}$) on recession probability forecasts; $r_{t+1:t+h} = \alpha + \hat{p}_{t+1:t+h|t} + \epsilon_{t+1:t+h}$. $\hat{p}_{t+1:t+h|t}$ refers to recession probability forecasts from probit models with i) TMS_t, ii) TMS_t and TMS_{t-6}, and iii) TMS_t and MA-TMS_t, respectively. Additionally, results are shown for recession probability forecasts from the yield curve model provided by the New York Fed (NY FED rec. Prob.) as well as for the smoothed recession probabilities from Chauvet and Piger (2008). The *t*-statistics are given in parentheses and the adjusted R^2 is shown below in %. Standard errors are Newey–West adjusted with a lag length of 12 months. Panel A (Panel B) shows results for the sample from 1951:3 to 2020:12 (1980:1 to 2020:12). Results are reported for forecast horizons of $h = 1,3,6,12$ months. ∗, ∗∗, ∗∗∗ denote significance at the 10%, 5%, and 1% significance levels.

Variable	$h=1$	$h=3$	$h=6$	$h=12$
1	2	3	4	5
Panel A. In-Sample Results (1951:3 to 2020:12)				
$\hat{p}_{t+1:t+h\|t}$ (TMS_t)	−0.066**	−0.039**	−0.024**	−0.016***
t-stat.	(−2.113)	(−2.166)	(−2.300)	(−2.787)
R^2_{adj} (%)	0.47	1.55	2.96	6.14
$\hat{p}_{t+1:t+h\|t}$ (TMS_t, TMS_{t-6})	−0.010	−0.018**	−0.016**	−0.010***
t-stat.	(−0.879)	(−2.007)	(−2.318)	(−2.616)
R^2_{adj} (%)	0.04	1.52	3.00	4.21
$\hat{p}_{t+1:t+h\|t}$ (TMS_t, MA-TMS_t)	−0.012	−0.015*	−0.014**	−0.012***
t-stat.	(−1.163)	(−1.765)	(−2.277)	(−2.880)
R^2_{adj} (%)	0.08	0.98	2.56	5.90
NY FED rec. Prob. (1960:1–2020:12)	−0.016*	−0.011	−0.011	−0.011
t-stat.	(−1.700)	(−1.270)	(−1.183)	(−1.021)
R^2_{adj} (%)	0.21	0.37	0.87	1.64
Chauvet–Piger (1967:6–2020:12)	−0.018*			
t-stat.	(−1.855)			
R^2_{adj} (%)	0.85			
Panel B. In-Sample Results (1980:1 to 2020:12)				
$\hat{p}_{t+1:t+h\|t}$ (TMS_t)	−0.036	−0.024	−0.016	−0.016**
t-stat.	(−0.913)	(−1.141)	(−1.236)	(−2.350)
R^2_{adj} (%)	−0.03	0.44	1.10	5.84
$\hat{p}_{t+1:t+h\|t}$ (TMS_t, TMS_{t-6})	−0.014	−0.020*	−0.017*	−0.016***
t-stat.	(−0.882)	(−1.725)	(−1.927)	(−2.945)
R^2_{adj} (%)	0.12	1.88	3.18	9.38
$\hat{p}_{t+1:t+h\|t}$ (TMS_t, MA-TMS_t)	−0.023	−0.025*	−0.022**	−0.021***
t-stat.	(−1.383)	(−1.788)	(−2.090)	(−2.602)
R^2_{adj} (%)	0.50	2.52	5.21	12.69
NY FED rec. Prob.	−0.022**	−0.019*	−0.022*	−0.023
t-stat.	(−2.260)	(−1.873)	(−1.785)	(−1.527)
R^2_{adj} (%)	0.49	1.36	3.74	8.32
Chauvet–Piger	−0.021*			
t-stat.	(−1.888)			
R^2_{adj} (%)	1.14			

E. Out-of-Sample Equity Premium Prediction

Having documented the significant predictive power of the recession probabilities for the equity premium in-sample, we now turn to an out-of-sample assessment. Importantly, in this analysis, we only use information that would have been available to investors in real time. This is implemented in the following way. Suppose that we are interested in forecasting $r_{t+1:t+h}$ at time t. First, we estimate the coefficients of the probit model with information up to time $t-24$ to account for the fact that the NBER calls recessions typically with a few months delay. We then

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

combine these estimated coefficients with the values of the term spread and its backward-looking moving average up to month t into the implied recession probabilities $\hat{p}_{t+1:t+h|t}$. Second, we regress the log equity premium until time t on a constant and the estimated in-sample recession probabilities until period t. Third, we use the estimated coefficients $\hat{\alpha}_t$ and $\hat{\beta}_t$ and the out-of-sample recession probability forecast $\hat{p}_{t+1:t+h|t}$ to predict $r_{t+1:t+h}$. Thus, the log equity premium forecast is $\hat{r}_{t+1:t+h|t} = \hat{\alpha}_t + \hat{\beta}_t \hat{p}_{t+1:t+h|t}$. We recursively re-estimate the coefficients of the probit model and the linear predictive regression model and real-time forecasts for each month over the period from 1980:1 to 2020:12.

We follow the convention in the literature and evaluate the forecast performance based on the out-of-sample R^2 of Campbell and Thompson (2008). The R^2_{OS} statistic measures the proportional reduction in mean squared forecast error (MSFE) relative to the benchmark model with only a constant ($\beta_t = 0$); see, among others, Rapach et al. (2010), Jiang, Lee, Martin, and Zhou (2019):

$$R^2_{\text{OS}} = 1 - \frac{\sum_{t=M}^{T-h} \left(r_{t+1:t+h} - \hat{r}_{t+1:t+h|t}\right)^2}{\sum_{t=M}^{T-h} \left(r_{t+1:t+h} - \bar{r}_{t+1:t+h|t}\right)^2},$$

where $\bar{r}_{t+1:t+h|t}$ is the prevailing mean forecast using information up to period t. This benchmark assumes no predictability in the excess return series and therefore uses the conditional sample mean as the predicted value.[9] Welch and Goyal (2008) show that none of the theoretically motivated predictors such as the dividend–price ratio, term spread, or book-to-market ratio consistently outperform this naive benchmark over long samples. In fact, they find that the predictive power is mainly driven by the 1973–1975 oil shock, and that the period from 1975 to 2005 is characterized by "30 years of poor performance" (Welch and Goyal (2008), p. 1504). We test the null hypothesis of a lower or equal MSFE from forecasts of the historical average benchmark ($R^2_{\text{OS}} \leq 0$) against the alternative that forecasts from the models using recession probabilities as predictors have a lower MSFE ($R^2_{\text{OS}} > 0$) using the MSFE-adjusted statistic of Clark and West (2007), which corrects for the fact that the Diebold and Mariano (1995) statistic follows a non-standard distribution for nested models. We account for serial correlation in the residuals by estimating Newey and West (1987) standard errors with lag lengths of h months.

Panel A in Table 4 presents the R^2_{OS} statistics (in %) when using the same three probit models as in the previous sections to derive recession probability forecasts. The following results are worth noting. First, while the R^2_{OS} statistic for TMS is negative (−0.74%) and insignificant for $h = 1$, it is positive (3.88%) and significant at the 5% level for $h = 12$. Hence, the forecasts from the standard probit model with the term spread as explanatory variable are helpful in predicting cumulative log

[9]Specifically, for any horizon h, we use the sample mean of the h-period cumulative excess market return up to the point in time when the forecast is made as the predictor of the h-period cumulative excess market return, i.e., $\bar{r}_{t+1:t+h|t} = \frac{1}{t-h+1} \sum_{i=h}^{t} \left(\frac{1}{h} \sum_{j=0}^{h-1} r_{i-j}\right)$. In the case of 1-step-ahead predictions, for $h = 1$, this simply equals $\bar{r}_{t+1|t} = \frac{1}{t} \sum_{i=1}^{t} r_i$, the sample mean of the excess return up to period t.

TABLE 4

Out-of-Sample R^2 Statistics for Log Equity Premium Forecasts

Table 4 reports R^2_{OS} statistics in % for the out-of-sample predictability of (cumulative) log excess returns on the S&P 500 index at the h-month-ahead horizon relative to forecasts from the historical average. Forecasts are based on the linear predictive regression model with a constant and 1 predictor variable. Panel A shows results when forecasting with model-implied recession probabilities. The recession probability forecasts are derived by three different probit models: the first model only includes a constant and the term spread, whereas the second and third models add either the term spread lagged by 6-months (TMS_{t-6}) or the 3-year moving average of the term spread ($MA\text{-}TMS_t$) as additional predictors. Panel B shows results when the term spread variables are directly used as predictors in the OLS regression. CF-MEAN refers to an equal-weighted average of forecasts from univariate regressions with TMS_t, TMS_{t-6}, and $MA\text{-}TMS_t$, respectively. *, **, and *** indicate significance at the 10%, 5%, and 1% levels according to the Clark and West (2007) MSFE-adjusted statistic. The null hypothesis is equal MSFE and the alternative is that the more sophisticated model has a smaller MSFE than the historical average benchmark. The out-of-sample period runs from 1980:1 to 2020:12.

Variable	$h=1$	$h=3$	$h=6$	$h=12$
1	2	3	4	5
Panel A. Probit Model				
TMS_t	−0.74	−1.32	−1.69	3.88**
TMS_t, TMS_{t-6}	0.21	1.42**	1.54*	7.90***
TMS_t, $MA\text{-}TMS_t$	0.96**	2.76***	3.42**	8.99***
Panel B. OLS Model				
TMS_t	−0.91	−2.07	−3.19	1.85**
TMS_{t-6}	−0.06	0.89**	2.07**	4.61***
$MA\text{-}TMS_t$	0.29	1.02*	1.87**	4.16**
CF-MEAN	0.17	0.97*	1.71**	5.70***
TMS_t, $MA\text{-}TMS_t$	−1.15	−2.60	−4.30	0.02**

equity premiums over the next year, although the reduction in MSFE relative to the historical average is below 5%. Second, the models with lagged term spread information have consistently positive R^2_{OS} values. The improvements for $h=1$ are relatively small with values of 0.21% and 0.96%, respectively, but increase to almost 9% for $h=12$. While these gains may appear modest, Campbell and Thompson (2008) show that a monthly R^2_{OS} of only 0.50% can already translate into significant economic gains for an investor. Third, the best-performing model uses the term spread and the moving average term spread to predict recessions. The R^2_{OS} values are significant at the 5% level for each forecast horizon and monotonically increase in h. These results give a first indication that recession probabilities estimated from term structure information can be useful out-of-sample predictors of excess stock returns. In the next section, we document that the predictive power further improves once a structural break in the mean of the term spread is accounted for. We further compare the forecast performance with several recently proposed benchmark predictors of the equity premium.

1. Correcting for Breaks: Out-of-Sample Equity Premium Prediction

We have shown that the in-sample predictive power of recession probabilities is stronger in the post-1980 sample than in the full sample starting in 1951, and that this strong predictive ability carries over to an out-of-sample setup. At the same time, as shown by Figure 3, the implied recession probabilities are muted and rarely exceed 50% after the mid-1980s. This is consistent with, for example, Chauvet and Potter (2002) who find evidence for a structural break in the probit model based on the term spread but argue that the exact date of the

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

break is difficult to localize.[10] Other papers also document breaks in the dynamic relationship between the term spread and real growth in the United States and other countries (e.g., Schrimpf and Wang (2010)).

In sum, there is ample prior evidence for a structural break in the link between the term spread and output growth, as well as for a break in the estimated recession probabilities from the standard probit model. This suggests that accounting for such a break may improve the recession probability forecasts and hence equity premium predictability. In Supplementary Material Appendix C, we provide further evidence for a structural break. Instead of focusing on a break in the estimated relation between the term spread and future recessions, we focus our attention on a break in the mean of the term spread. The reason is that a break in the parameters of the probit model is generally difficult to identify and even more so difficult to narrow down to an exact point in time—not least because of the infrequent occurrence of recessions (Wright (2006)). In contrast, we show that the shift in the mean of the term spread can be identified precisely and in a more timely manner regardless of the state of the economy. We entertain several state-of-the-art methods to detect the break in the mean of the term spread in expanding samples. We then use various alternative approaches to forecasting in the presence of structural breaks following Pesaran and Timmermann (2007). These are labeled "Cross-validation," "Pooling," and "Post-break window." In addition, we follow Lettau and Van Nieuwerburgh (2008) and break-adjust the term spread by subtracting the pre- and post-break subsample means from the original series.

To assess the relative performance of our approach, we compare our results with several additional benchmark models. First, we generate forecasts from a simple two-state model that can perfectly foresee NBER recessions. We recursively estimate the following regression:

$$r_t = \alpha + \beta \times \mathbf{I}\{\text{NBER}_t = 1\} + \epsilon_t,$$

where $\mathbf{I}\{\text{NBER}_t = 1\}$ is an indicator function that equals 1 in recessions. If there is a recession in the next month, then the forecast equals the average log equity premium in past recessions and, vice versa, if the next month is in an expansion then the forecast equals the average during past expansions. We implement this benchmark to see if the inverse V-shape in estimated recession probabilities has any value above and beyond simply classifying periods into expansions and recessions. The second benchmark predictor is the short interest variable of Rapach et al. (2016), which they characterize as "*the* strongest known predictor of aggregate stock returns." This is calculated as the log of the equal-weighted mean of short interest across publicly listed stocks on U.S. exchanges. The series shows a strong linear upward trend and is therefore recursively detrended. Our third benchmark is the (log) gold-to-platinum ratio of Huang and Kilic (2019), constructed as the log ratio of gold to platinum prices, which the authors show to predict returns particularly

[10]When fixing the break to 1984, they show that the model is able to predict the recession in 2001 with probabilities as high as 90% for the 12-month-ahead forecast horizon. Galvão (2006) proposes a structural break threshold-VAR (SBTVAR) model that allows for non-linearities and breaks in the link between the term spread and U.S. output growth and identifies a break in 1985:2. She further shows that the SBTVAR model correctly anticipates the 2001 recession in real time.

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

well over longer forecast horizons. Due to data availability, we can only evaluate the latter two predictors over the sample from 1990:1 to 2013:12. As a final benchmark, we use the Variance Risk premium (VRP) which various authors have shown to be a good predictor of aggregate market returns (e.g., Bollerslev et al. (2009) and more recently Cheng (2019)).[11] This series is available starting in 1990, and we use it for OOS forecasting starting in 2000. Since Goyal, Welch, and Zafirov (2024) report that the VRP has performed particularly poorly in early 2020, we include the VRP for a sample ending 2019 as well as for a sample ending in 2020:12.

Table 5 compares the equity premium forecasts from recession probabilities of the probit model based on TMS and MA-TMS with forecasts from the same model based on different break-adjustments of the term spread and the benchmark predictors. Panels A through D provide the results for forecast horizons $h = 1,3,6,12$ months. Columns 1 through 4 report the R^2_{OS} for various out-of-sample prediction periods.

Several points are worth making. First, the term spread and moving average term spread are consistently outperformed by their break-adjusted counterparts. Among the break-adjustment methods, the approach using only post-break data performs best, closely followed by cross-validation. The break in the mean of the term spread is relatively sizable, and thus the break date can be estimated accurately in real time. Due to the large impact of pre-break information, the improvements from pooling are the smallest. Across horizons and prediction periods, cross-validation seems to be the most robust choice for improving recession and equity premium forecasts.[12]

Second, the gains in predictability relative to the historical average are economically and statistically significant for essentially all break-adjustment methods and sample periods. The R^2_{OS} statistics are consistently positive for the different subsamples and gradually increase in the forecast horizon h, with R^2_{OS} values above 10% for cumulative 1-year-ahead equity premiums. It is worth noting that the estimated predictive coefficients (not shown) are consistently negative, indicating that the superior performance relative to the historical average is driven by negative equity premium forecasts during recessions. This contrasts Campbell and Thompson (2008) who argue that imposing non-negativity constraints on the equity premium can improve performance.

Finally, the break-adjusted recession probabilities also outperform the four benchmark models, particularly at longer forecast horizons. The gains in terms of reduced MSFEs are substantial, with the R^2_{OS} statistic more than doubling. The R^2_{OS} of cross-validation is 2.94% for $h = 1$, compared to 1.74% for a perfect recession classifier, 1.67% for short interest, and 1.15% for the gold-to-platinum ratio.[13] While the VRP performs best at the 1-month forecast horizon with an

[11] We thank Dave Rapach for making the short interest and Amit Goyal the VRP series available on their respective homepages, and Darien Huang and Mete Kilic for kindly sharing their data with us.

[12] We show the recursively selected estimation windows from cross-validation in Supplementary Material Appendix C.

[13] The results differ slightly compared to those published in the papers because our historical average uses data from 1951:3 onward. We re-estimate the predictive coefficients recursively using expanding data available in real time.

TABLE 5

Out-of-Sample R^2 Statistics When Correcting for a Structural Break

Table 5 presents R^2_{OS} statistics (in %) for forecasts of the $h = 1,3,6,12$ months ahead (cumulative) log equity premium. This statistic measures the reduction in MSFE relative to forecasts from the historical average. Results are shown for a probit model with the term spread and the moving average component as predictors, and for four methods that correct for a structural break in the mean of the term spread. The first method forecasts with a break-adjusted term spread (TMS_t^{break}, $\frac{1}{36}\sum_{j=0}^{35} TMS_{t-j}^{break}$) following Lettau and Van Nieuwerburgh (2008). Cross-validation selects an optimal estimation window over a holdout period, whereas pooling combines forecasts from several models with a grid of different starting values. Post-break window refers to forecasts from a model that only uses post-break data. For further details, see Supplementary Material Appendix C and Pesaran and Timmermann (2007). Short interest (Rapach et al. (2016)), gold-to-platinum ratio (Huang and Kilic (2019)), and the Variance Risk Premium (VRP, Bollerslev et al. (2009)) are three prominent benchmark return predictors from the literature. Perfect classifier refers to a simple two-state model that can perfectly anticipate NBER-dated recessions. *, **, *** denote significance at the 10%, 5%, and 1% significance levels according to the Clark and West (2007) MSFE-adjusted statistic. Columns 1 to 4 report results for different subsamples, and Panels A to D present results for different forecasting horizons.

	1980:1–2020:12	1980:1–1999:12	2000:1–2020:12	1990:1–2013:12
Variable	1	2	3	4
Panel A. h = 1				
TMS_t, MA-TMS_t	0.96**	1.52**	0.43*	0.74**
TMS_t^{break}, MA-TMS_t^{break}	1.01**	0.03	1.93*	2.17**
Cross-validation	1.85**	1.30**	2.37**	2.94**
Pooling (average)	1.14**	1.33**	0.96*	1.40***
Post-break window	1.61**	1.64**	1.59*	4.57***
Perfect classifier	1.31**	0.10	2.44**	1.74*
Short interest				1.67**
Gold-to-platinum ratio				1.15**
VRP (ending 2019)			4.44**	
VRP (ending 2020)			−8.44	
Panel B. h = 3				
TMS_t, MA-TMS_t	2.76***	4.40**	1.35*	2.17**
TMS_t^{break}, MA-TMS_t^{break}	2.14**	−1.11	4.95**	4.56**
Cross-validation	4.08**	3.70**	4.45*	9.25**
Pooling (average)	3.41***	3.95**	2.93**	3.78***
Post-break window	5.14***	4.43**	5.84**	11.69***
Perfect classifier	2.95**	0.27	5.32**	3.74*
Short interest				5.46***
Gold-to-platinum ratio				4.54**
VRP (ending 2019)			6.75**	
VRP (ending 2020)			−15.66	
Panel C. h = 6				
TMS_t, MA-TMS_t	3.42**	5.04**	2.24	3.54**
TMS_t^{break}, MA-TMS_t^{break}	1.18**	−7.34	7.38**	5.91**
Cross-validation	9.32***	3.16*	13.68**	14.25**
Pooling (average)	4.49**	4.10*	4.76**	6.01***
Post-break window	6.14***	5.45**	6.61**	15.49**
Perfect classifier	3.75**	−0.13	6.48**	5.49*
Short interest				9.44**
Gold-to-platinum ratio				8.52**
VRP (ending 2019)			2.56**	
VRP (ending 2020)			−10.78	
Panel D. h = 12				
TMS_t, MA-TMS_t	8.99***	10.10**	8.41**	8.58***
TMS_t^{break}, MA-TMS_t^{break}	6.40**	−8.75	15.91***	8.17*
Cross-validation	11.03***	7.76*	13.21***	14.13***
Pooling (average)	10.88***	7.73*	12.89***	11.38***
Post-break window	12.43***	13.22**	12.17**	18.45**
Perfect classifier	9.92***	3.70*	13.97***	12.55**
Short interest				7.85*
Gold-to-platinum ratio				9.55**
VRP (ending 2019)			−3.46	
VRP (ending 2020)			−3.47	

$R^2_{OS} = 4.44\%$ when ending the sample in 2019, consistent with Goyal et al. (2024) this indicator entirely loses its predictive power in the COVID-19 pandemic and the R^2_{OS} becomes negative when including 2020. Looking beyond $h = 1$ in the subsequent panels, we see that the recession probabilities based on the break-adjusted

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

term spread information increasingly become the dominant forecasting variables relative to the benchmark predictors. At the 1-year-ahead horizon, cross-validation delivers an R^2_{OS} statistic of more than 14%, compared to 7.85% for short interest, 9.55% for the gold-to-platinum ratio, and −3.47% for the VRP.[14]

2. A Comparison to the Standard OLS Approach

So far, we have shown that recession probabilities derived from probit models using the term spread as a predictor significantly outperform the historical average benchmark. This contrasts with the common finding that forecasts from a linear regression of the equity premium on the term spread perform poorly for short horizons (Rapach and Zhou (2013)). In this section, we compare forecasts from model-implied recession probabilities to forecasts from linear predictive regressions with TMS and MA-TMS as predictors. This corresponds to 3 replacing $\hat{p}_{t+1:t+h|t}$ with TMS_t, TMS_{t-6}, and MA-TMS_t, respectively. Panel B in Table 4 provides the corresponding R^2_{OS} statistics. We further show results for a simple combination forecast that takes the average of the 3 individual OLS regression forecasts (CF-MEAN). While TMS only has mild predictive power over 1-year-ahead forecast horizons, TMS_{t-6} and MA-TMS_t also significantly predict the equity premium for $h = 3$ and $h = 6$. MA-TMS_t further outperforms the historical average for $h = 1$, although with only a R^2_{OS} of 0.29%. Among the linear forecasts, the combination performs the best with a R^2_{OS} of 5.7%.

Our finding that adding lagged information significantly improves equity premium forecasts is surprising in light of the efficient market hypothesis. However, it is consistent with Gómez-Cram (2022) who documents that analysts only sluggishly revise their expectations downward and that stock prices do not fully reflect publicly available information about turning points. The last row in Panel B shows the performance of a joint OLS regression with TMS_t and MA-TMS_t. The OLS regression analogue to our probit model performs worse, in line with the common finding that multivariate regression models with several parameters often underperform the historical average. Importantly, at all forecast horizons, even the best linear model is substantially outperformed by the equity premium forecasts based on recession probabilities.

Graph A of Figure 5 illustrates this result by superimposing 1-month-ahead equity premium forecasts from the recession probability based on TMS_t and MA-TMS_t (solid black line), the analogue OLS model with the two predictors (dashed black line), and the historical average benchmark (solid gray line). While the OLS model generates forecasts that are volatile both in recessions and expansions, the probit model forecasts are relatively stable in expansions and markedly higher than the historical average. As the implied recession probabilities are high just before the 1981–1982 recession, the implied equity premium forecasts are sharply negative around that time. This effect, although substantially less pronounced, is also visible around the business cycle peaks in 2001 and 2007.

[14]In unreported results, we find that the VRP is a good in-sample predictor of recessions at short forecasting horizons. This is broadly consistent with Cheng (2019) and Ghaderi et al. (2024). Moreover, the recession probabilities based on the VRP also predict the equity premium quite well.

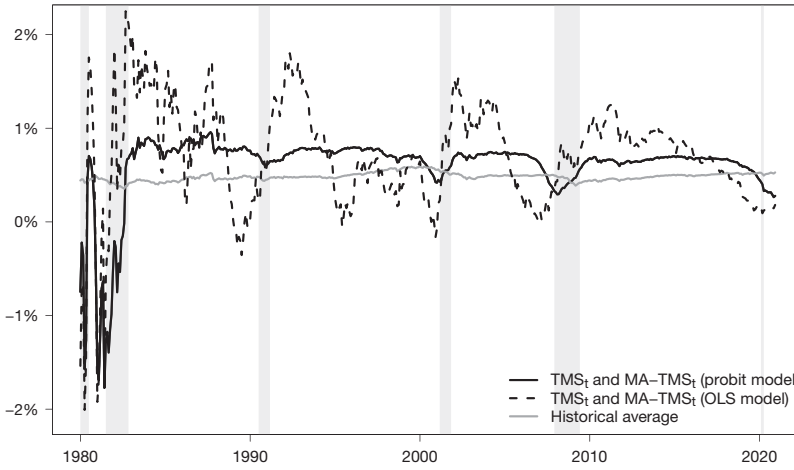
https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

FIGURE 5

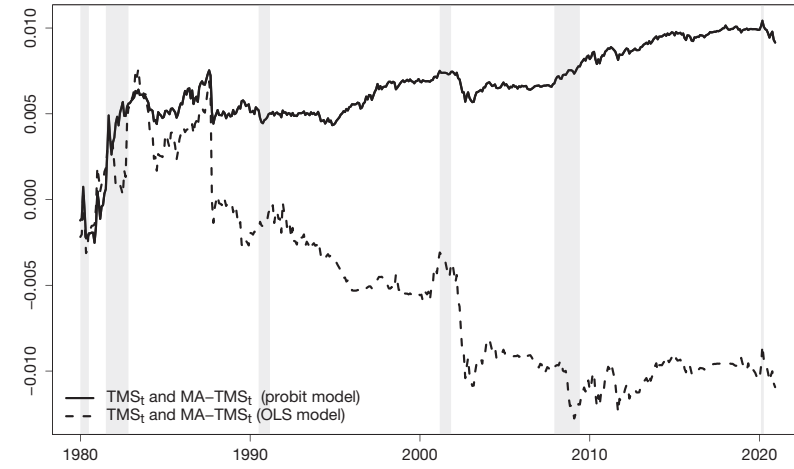
Out-of-Sample Forecasts and Performance over Time

Figure 5 shows 1-month-ahead forecasts of the log equity premium for three different models (Graph A). The solid black line depicts forecasts from model-implied recession probability forecasts of a probit model with the term spread (TMS) and the backward-looking 3-year moving average of the term spread (MA-TMS). The dashed black line presents forecasts from a standard linear predictive regression model with TMS and MA-TMS as predictors, and the solid gray line denotes the historical average. Graph B shows the difference between cumulative squared forecast errors (CSFE) of the historical average and the CSFE of the probit model forecasts (solid black line) and the OLS model forecasts (dashed black line). All forecasts are estimated with a recursively expanding information set that mimics the real-time situation of an investor. The out-of-sample period is 1980:1 to 2020:12.

Graph A. Comparison of 1-Month Ahead Equity Premium Forecasts



Graph B. Comparison of Cumulative Forecast Performance



Welch and Goyal (2008) have popularized an intuitive way to visualize the relative forecast performance of different prediction models over time. Graph B in Figure 5 follows their approach and plots the difference in cumulative squared forecast errors (CSFE) for the historical average and the CSFE for two different models: the solid black line depicts equity premium forecasts based on the implied

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

recession probability using TMS and MA-TMS, whereas the dashed black line shows the OLS model with both TMS and MA-TMS. An increasing curve indicates superior performance relative to the naive benchmark. We can see that the curve for the OLS model is decreasing over most of the sample, with reversed trends only around the 1981–1982 and 2001 recessions. In contrast, the curve for the probit model forecasts is rising over most of the sample, indicating that the recession probability forecast of the equity premium consistently outperforms the historical average.

This superior performance is driven by two distinct effects. First, and similar to the OLS model, the model significantly predicts the negative excess returns in the 1-year window around the peak in 1981. This is consistent with Table 1 and with the notion that term spread information anticipates recessions. Second, and more importantly, the model-implied recession probabilities outperform the naive benchmark also in expansions. This contrasts with the "no predictability in good times puzzle" that is often documented in related articles (Huang, Jiang, Tu, and Zhou (2017)). The explanation is simple: Table 1 shows that the annualized equity premium averages 6.38% in our sample, but is even higher at an annualized 8.70% in expansions. While the historical mean benchmark closely tracks the full-sample average, our recession probability-based forecast corrects for negative values around business cycle peaks and thus correctly predicts a higher equity premium in expansions.

In the bottom panel of Table B.1 in the Supplementary Material, we show out-of-sample results for a much longer time period, starting in 1946 and ending in 2022. While the forecasts start in 1946, we use data from 1871 onward to estimate the model parameters. The results show that the recession probabilities based on the probit model using TMS and MA-TMS have significantly predicted the equity premium out-of-sample even considering this much longer time span. Moreover, the recession probabilities based on term structure information consistently outperform their OLS counterpart, albeit both feature statistically and economically significant out-of-sample R^2.

To better understand the strong predictive power of recession probabilities for the equity premium, we provide separate R^2_{OS} statistics for expansions and recessions in Table 6. Strikingly, we see that for essentially all considered models and forecast horizons, there is an economically and statistically significant improvement over the historical average benchmark in both recessions *and* expansions. As discussed above, the reason is the following. Recession probabilities correctly predict low equity premiums in recessions. In addition, by adjusting for low equity premiums in recessions, they correctly predict higher equity premiums than the historical average in expansions.[15] That said, the break-correction methods improve the performance primarily in recessions, as the adjusted recession probabilities predict a highly negative equity premium around the peak in 2007. For

[15]Note that the R^2_{OS} statistic for the unadjusted probit model turns negative in recessions for 6- and 12-month-ahead forecasts and becomes highly significant in expansions. This reflects the fact that for longer horizons the recession forecasts substantially decrease prior to the peak and, therefore, have the strongest predictive power already before the beginning of the recession; see Figure 1.

TABLE 6

Out-of-Sample R^2 Statistics in Recessions and Expansions

Table 6 presents R^2_{OS} statistics (in %) for forecasts of the $h = 1,3,6,12$ months ahead (cumulative) log equity premium. Results are shown for a probit model with the term spread (TMS) and the moving average component (MA-TMS) as predictors, and for four methods that correct for a structural break in the steady state mean of the term spread. For further details, see Supplementary Material Appendix C. The R^2_{OS} statistics are displayed separately for recessions and expansions. *, **, *** denote significance at the 10%, 5%, and 1% significance levels according to the Clark and West (2007) MSFE-adjusted statistic. Column 1 reports the respective forecasting model, and columns 2 to 5 present results for different forecasting horizons. The out-of-sample period is 1980:1 to 2020:12.

Variable	$h=1$	$h=3$	$h=6$	$h=12$
1	2	3	4	5
Panel A. Recessions				
TMS_t, MA-TMS_t	1.60*	2.40*	−1.60	−1.20
TMS_t^{break}, MA-TMS_t^{break}	5.85**	7.86**	3.11*	9.44**
Cross-validation	5.16**	10.64**	14.13*	8.33***
Pooling (average)	3.07**	4.89**	1.20*	4.77***
Post-break window	8.71**	14.69**	11.51*	17.72**
Panel B. Expansions				
TMS_t, MA-TMS_t	0.71**	2.95***	6.67**	14.82***
TMS_t^{break}, MA-TMS_t^{break}	−0.91	−0.78*	−0.07**	4.66**
Cross-validation	0.54**	0.75**	6.21***	12.58***
Pooling (average)	0.37*	2.65**	6.62**	14.37***
Post-break window	−1.20	0.28**	2.66***	9.40***

example, the R^2_{OS} for the 1-month-ahead forecasts improves from 1.6% to 5.1% and 8.7% for cross-validation and post-break window.[16]

F. Implications for Asset Allocation

In this section, we analyze the economic value of the improved equity premium predictions. Cenesizoglu and Timmermann (2012) show that the correlation between statistical and economic measures of forecast performance is positive but typically of low magnitude. Specifically, many models produce negative R^2_{OS} values while still providing investors with improved Sharpe ratios and gains in the certainty equivalent return. Interestingly, the reverse—positive R^2_{OS} values and negative economic gains—is observed less often. We follow Dangl and Halling (2012) and Rapach et al. (2016) and others and consider a mean–variance investor who allocates funds across the equity market portfolio

[16]We perform encompassing tests to study whether the forecasting models provide distinct information. Specifically, we apply the ENC-T statistic of Harvey, Leybourne, and Newbold (1998) to test whether post-break window encompasses the information of the other four forecasting models. Formally, we test the null hypothesis that $\lambda = 0$ in a convex combination of forecasts, $\hat{r}^c_{t+1:t+h} = (1-\lambda)\hat{r}^{\text{post-break window}}_{t+1:t+h} + \lambda\hat{r}^j_{t+1:t+h}$ with j being equal to one of the four alternative models (TMS_t, MA-TMS_t; TMS_t^{break}, MA-TMS_t^{break}; Cross-validation; Pooling (average)). We cannot reject the null hypothesis that post-break window encompasses the other forecasts. It may, however, still be possible that the other models provide equally good forecasts. We thus test whether the other forecasts encompass the forecasts from post-break window ($\lambda = 1$). For $h = 1$ we reject the null hypothesis for TMS_t, MA-TMS_t and TMS_t^{break}, MA-TMS_t^{break}. For $h = 3$ we additionally reject the null for cross-validation. For $h = 6$ and $h = 12$ we cannot reject the null that TMS_t, MA-TMS_t, pooling (average), and cross-validation encompass the post-break window. In sum, we conclude that the post-break window encompasses some forecasts for $h = 1$ and $h = 3$ but that most models contain similar information for longer horizons. We therefore report results for all break-adjustment methods.

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

and the risk-free rate. At the end of period t, the investor optimally invests a share ω_t in the risky asset:

$$\omega_t = \frac{1}{\gamma}\frac{\hat{r}_{t+1|t}}{\hat{\sigma}^2_{t+1}}, \tag{3}$$

where γ is the coefficient of relative risk aversion and $\hat{r}_{t+1|t}$ is a forecast of the equity premium.[17] Similar to Rapach et al. (2016), we estimate the variance of excess returns, $\hat{\sigma}^2_{t+1}$, as a 10-year rolling window of past data. Thus, ω_t only differs because of the different equity premium forecasts implied by the various models, $\hat{r}_{t+1}$. The realized portfolio return, r^P_{t+1}, is as follows:

$$r^P_{t+1} = \omega_t r_{t+1} + r^f_{t+1}, \tag{4}$$

where r_{t+1} is the realized excess equity market return in period $t+1$ and r^f_{t+1} is the risk-free rate between periods t and $t+1$. Furthermore, the certainty equivalent return (CER) can be calculated as follows:

$$\text{CER}_P = \hat{\mu}_P - \frac{\gamma}{2}\hat{\sigma}^2_P, \tag{5}$$

where $\hat{\mu}_P$ and $\hat{\sigma}^2_P$ are the sample mean and variance of the portfolio over the out-of-sample period. We multiply the CER by 12 to interpret it as the annual risk-free rate that an investor would be willing to accept to not hold the risky portfolio (Chen et al. (2020)). The difference in CER of two models—also known as utility gain—can be interpreted as the annual portfolio management fee that an investor would be willing to pay to have access to the alternative forecasting model (Ferreira and Santa-Clara (2011)). Additionally, we calculate the annualized Sharpe ratio (SR) to evaluate the risk–return profile of the chosen portfolio allocations.

Following Welch and Goyal (2008) and others, we compare each forecasting model's asset allocation performance with that of the prevailing mean model; see the definition in Section II.E. Table 7 presents ΔCER and ΔSR for monthly re-balancing for the sample periods 1980–2020 (Panel A) and 1990–2013 (Panel B).[18] The CER and SR values for the prevailing mean model are shown in the first row of each panel. All subsequent rows then provide differences of CER and SR relative to this benchmark. We implement a bootstrap approach similar to DeMiguel, Plyakha, Uppal, and Vilkov (2013) to evaluate the statistical significance.[19] We consider different specifications for the coefficient of relative risk aversion (γ) and for leverage and short-selling constraints (range of ω). Over the baseline sample period from 1980 to 2020, the probit model with TMS and MA-TMS provides utility gains in the range of 1.2% to 2.6%. Hence, an investor would be willing to pay between 120 and 260 basis points annually – depending on risk preferences and constraints – to have access to the equity premium forecasts of this model. The gains are highest

[17]Forecasts in this section are based on excess returns rather than log excess returns.

[18]We show in Supplementary Material Appendix E that the results remain qualitatively the same when re-balancing in 3-, 6-, and 12-month intervals.

[19]We set the average block length to 3 months (Politis and Romano (1994)).

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

TABLE 7

Asset Allocation Exercise

Table 7 reports the annualized ΔCER and the annualized ΔSR for a mean–variance investor relative to forecasts from the historical average. The investor can invest in the S&P 500 index and the risk-free rate. Results are shown for 1-month-ahead forecasts of the equity premium and different values for the coefficient of relative risk aversion (γ), and different restrictions on the equity weights (ω). The "Prevailing mean" shows the CER and SR values, whereas all other values denote the improvements relative to this benchmark. Panel A (Panel B) shows results for the out-of-sample period from 1980:1 to 2020:12 (1990:1 to 2013:12). ∗, ∗∗, and ∗∗∗ indicate significantly improved performance relative to the prevailing mean benchmark at the 10%, 5%, and 1% significance levels. The *p*-values are obtained by using a bootstrap approach similar to DeMiguel et al. (2013) with the average block length set to 3 months (Politis and Romano (1994)).

	ΔCER				ΔSR			
γ	3	5	3	3	3	5	3	3
ω	[−0.5,1.5]	[−0.5,1.5]	[0,1.5]	[0,1]	[−0.5,1.5]	[−0.5,1.5]	[0,1.5]	[0,1]
	1	2	3	4	5	6	7	8
Panel A. 1980:1 to 2020:12								
Prevailing mean	8.03	6.37	8.03	8.62	0.48	0.47	0.48	0.50
Gains relative to prevailing mean:								
TMS_t, MA-TMS_t	2.58**	1.42**	2.27**	1.28***	0.14***	0.13**	0.13***	0.09***
TMS_t^{break}, MA-TMS_t^{break}	3.80*	2.33*	3.32**	1.73	0.28**	0.27*	0.26**	0.22**
Cross-validation	4.14***	2.73**	3.38***	2.24**	0.23***	0.23**	0.19***	0.18***
Pooling (average)	3.10***	1.91**	2.80**	1.62**	0.18**	0.18**	0.17**	0.13**
Post-break window	5.69***	3.87***	4.78***	2.80**	0.33***	0.33**	0.28***	0.25***
Buy-and-hold	1.02**	0.40	1.02**	0.44*	0.07**	0.08**	0.07**	0.02*
Panel B. 1990:1 to 2013:12								
Prevailing mean	5.80	4.61	5.80	6.56	0.41	0.39	0.41	0.45
Gains relative to prevailing mean:								
TMS_t, MA-TMS_t	2.05**	1.24**	2.05**	0.72**	0.12***	0.13***	0.12***	0.05**
TMS_t^{break}, MA-TMS_t^{break}	5.68**	3.71**	4.71**	2.68	0.45**	0.47**	0.38**	0.32**
Cross-validation	4.44**	3.29**	3.66**	2.23*	0.26**	0.29**	0.21**	0.17*
Pooling (average)	3.56***	2.16**	3.56***	2.04***	0.21***	0.21***	0.21***	0.15***
Post-break window	8.11***	5.69**	6.59***	3.89***	0.48***	0.49***	0.39***	0.32***
Short interest	3.58*	2.40*	3.24**	1.71	0.22*	0.22*	0.20**	0.16
Gold-to-platinum ratio	4.03*	2.93*	4.11**	2.67*	0.28*	0.31*	0.30**	0.27**
Buy-and-hold	1.23*	0.23	1.23*	0.48*	0.08**	0.09*	0.08**	0.03*

when $\gamma = 3$ and when leveraging and short-selling up to 50% is allowed. Nonetheless, even an investor with ω between 0 and 1 would be willing to pay a portfolio management fee above 100 basis points annually. These gains further increase when correcting for the structural break in the mean of the term spread, with the post-break window performing best. That said, any of these strategies outperforms a simple buy-and-hold strategy shown in the last row – often the gains are more than triple or quadruple. Similarly, the annualized Sharpe ratio rises from around 0.50 for the historical average to between 0.65 and 0.85 for the break-correction methods. Panel B shows that the gains are comparable with and often substantially better than those for short interest and the gold-to-platinum ratio.

G. Dissecting the Sources of Predictability

In this section, we aim to shed light on the question of whether the predictability of excess equity returns using recession probabilities derives from the cash flow or the discount rate channel. To this end, we follow Campbell (1991) and Campbell and Ammer (1993) and apply a VAR decomposition to U.S. equity market returns. According to Campbell and Shiller (1988), the log return r_{t+1} can be rewritten as follows:

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

$$r_{t+1} = E_t[r_{t+1}] + N^{CF}_{t+1} - N^{DR}_{t+1}, \tag{6}$$

where $E_t[r_{t+1}]$ is the conditional expectation of the log return at time t, and N^{CF}_{t+1} and N^{DR}_{t+1} represent cash flow and discount rate news, respectively. We use a first-order VAR model to estimate the three components:

$$z_{t+1} = \Gamma z_t + u_{t+1}. \tag{7}$$

Here, z_t is a k-dimensional vector of state variables, Γ is a $k \times k$ matrix of parameters, and u_{t+1} is k-dimensional vector of innovations. We demean the variables in the VAR model and, therefore, can omit an intercept. The cash flow news and discount rate news can be estimated as follows:

$$N^{CF}_{t+1} = (e1' + e1'\lambda)u_{t+1}, \tag{8}$$

$$N^{DR}_{t+1} = e1'\lambda u_{t+1}, \tag{9}$$

where $e1$ is a k-vector with a 1 in the first cell and a 0 in all remaining cells, and $\lambda \equiv \rho\Gamma(\mathbf{I} - \rho\Gamma)^{-1}$. We set $\rho = 0.95^{\frac{1}{12}}$, which corresponds to approximately 5% consumption of total wealth per annum (Campbell and Vuolteenaho (2004), Maio (2013)).

The state vector is assumed to capture the dynamics of log equity market returns. However, the estimated news components may be sensitive to the variables included in the VAR (Chen and Zhao (2009)). Therefore, we follow Rapach et al. (2016) and estimate a series of trivariate VAR models including the log return series, the dividend–price ratio, and one of the 14 commonly used Welch and Goyal (2008) predictors at a time.[20] Additionally, we show results for a VAR model, where we add the first three principal components of the predictors of Welch and Goyal (2008) to the log return and the log dividend–price ratio (Rapach et al. (2016)). We use the OLS estimates of Γ and u_{t+1} to calculate $\hat{E}_t[r_{t+1}]$, $\hat{N}^{CF}_{t+1}$, and $\hat{N}^{DR}_{t+1}$. Then, we run the following regressions to estimate the effect of our model-implied recession probabilities on the components:

$$\hat{E}_t[r_{t+1}] = \alpha^E + \beta^E \times \hat{p}_{t+1|t} + \epsilon^E_{t+1} \tag{10}$$

$$\hat{N}^{CF}_{t+1} = \alpha^{CF} + \beta^{CF} \times \hat{p}_{t+1|t} + \epsilon^{CF}_{t+1} \tag{11}$$

$$\hat{N}^{DR}_{t+1} = \alpha^{DR} + \beta^{DR} \times \hat{p}_{t+1|t} + \epsilon^{DR}_{t+1} \tag{12}$$

[20]The 14 predictors are the log dividend–price ratio (DP), log dividend yield (DY), log earnings–price ratio (EP), log dividend–payout ratio (DE), excess stock return volatility (RVOL), book-to-market ratio (BM), net equity expansion (NTIS), Treasury bill rate (TBL), long-term yield (LTY), long-term return (LTR), term spread (TMS), default yield spread (DFY), default return spread (DFR), and inflation (INFL).

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

TABLE 8

Predictive Regressions for Stock Market Return Components

Table 8 presents slope coefficients for regressions of log stock market return components on model-implied recession probabilities. The stock market return is decomposed into the conditional return expectation ($\hat{E}_t[r_{t+1}]$), a cash flow news component ($\hat{N}^{CF}_{t+1}$), and a discount rate news component ($\hat{N}^{DR}_{t+1}$). The decomposition is based on the VAR approach of Campbell (1991) and Campbell and Ammer (1993) and includes as states the variables in columns 1 and 5. The log return on the S&P 500 index (r) and the log dividend–price ratio (DP) are included in each of the VAR models. PC denotes the first three principal components of the 14 popular predictors of Welch and Goyal (2008). The three return components are separately regressed on a constant and model-implied recession probabilities from cross-validation. The probabilities are recursively estimated and identical to those used in the out-of-sample exercises in the previous sections. The regressions are based on data from 1980:1 to 2020:12. The *t*-statistics for the slope coefficients are shown in the brackets below, and the standard errors are HAC-robust (Andrews (1991)). ∗, ∗∗, and ∗∗∗ denote significance at the 10%, 5%, and 1% significance levels.

VAR Variables	$\hat{\beta}^{E}$	$\hat{\beta}^{CF}$	$\hat{\beta}^{DR}$	VAR Variables	$\hat{\beta}^{E}$	$\hat{\beta}^{CF}$	$\hat{\beta}^{DR}$
1	2	3	4	5	6	7	8
r, DP	0.03	−0.17***	0.24*	r, DP, LTY	0.00	−0.16***	0.21*
	[0.51]	[−2.84]	[1.86]		[0.08]	[−2.81]	[1.67]
r, DP, DY	0.03	−0.17***	0.24*	r, DP, LTR	0.05	−0.17***	0.25**
	[0.72]	[−2.85]	[1.92]		[1.56]	[−2.94]	[2.00]
r, DP, EP	0.03	−0.80**	−0.39	r, DP, TMS	−0.06	−0.14**	0.18
	[0.48]	[−2.19]	[−1.53]		[−1.19]	[−2.30]	[1.36]
r, DP, DE	0.03	−0.80**	−0.39	r, DP, DFY	0.07	−0.21***	0.23*
	[0.48]	[−2.19]	[−1.53]		[1.14]	[−2.96]	[1.83]
r, DP, RVOL	0.08	−0.21***	0.25**	r, DP, DFR	0.02	−0.16***	0.23*
	[1.29]	[−2.94]	[2.09]		[0.34]	[−2.83]	[1.76]
r, DP, BM	−0.02	−0.18***	0.17	r, DP, INFL	−0.01	−0.16***	0.20
	[−0.52]	[−2.75]	[1.42]		[−0.23]	[−2.86]	[1.50]
r, DP, NTIS	0.06	−0.15*	0.28**	r, DP, PC	0.00	−0.31***	0.06
	[1.47]	[−1.75]	[2.00]		[0.04]	[−3.49]	[0.52]
r, DP, TBL	−0.06**	−0.15***	0.16				
	[−2.11]	[−2.57]	[1.26]				

While the return decomposition is based on the VAR estimates over the full sample from 1951:3 to 2020:12, we estimate the recession probabilities $\hat{p}_{t+1|t}$ with recursively expanding information sets to rule out any look-ahead bias. Table 8 presents the coefficients $\hat{\beta}^{E}$, $\hat{\beta}^{CF}$, and $\hat{\beta}^{DR}$ estimated using data from 1980:1 to 2020:12 which corresponds to our out-of-sample forecasting period. As predictor we use the recession probabilities estimated by cross-validation, but the results are essentially unchanged for the alternative break-correction methods. The estimates show that the recession probability has significant predictive power for both the cash flow and the discount rate news component in most VAR specifications. Specifically, increased recession probabilities signal low future dividends and high discount rates. The returns recover as the recession unfolds and recession probabilities approach 0.

III. Robustness and Additional Results

In this section, we show the robustness of our results along several dimensions and briefly discuss additional results. Details are provided in the Supplementary Material.

Our baseline results rely on a moving average of 3 years for the term spread. In Supplementary Material Appendix A, we provide R^2_{OS} statistics for different

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

averaging windows to construct MA-TMS and show that the results are robust to the exact choice of moving average. The long-horizon forecasts for $h = 3,6,12$ are statistically significant also for moving averages between 1 to 5 years, whereas the 1-month-ahead forecasts perform best for averages between 3 and 5 years.

In Supplementary Material Appendix B, we present out-of-sample R^2_{OS} values for log raw returns (without subtracting the short rate). The previous results carry over to raw returns; however, the predictive power decreases slightly for cumulative 6- and 12-month-ahead forecasts.

In Supplementary Material Appendix C, we provide the certainty equivalent returns and Sharpe ratios when correcting for proportional transaction costs of 50 basis points per transaction (Balduzzi and Lynch (1999)). The gains relative to the historical average and the buy-and-hold strategy remain sizable.

In Supplementary Material Appendix D, we provide the decomposition of predictability into the cash-flow and discount rate channels based on a VAR (1) applied to non-overlapping annual data.

In Supplementary Material Appendix E, we compare our results with those implied by the common growth component extracted from a set of real-time macroeconomic indicators by Gómez-Cram (2022). We find that while the common growth component and its 12-month moving average predict recessions about equally well as the term spread and moving average term spread, the implied recession probabilities tend to spike toward the end and not the beginning of recessions. Accordingly, these recession probabilities do not predict excess stock market returns nearly as well as those implied by term structure information.

In Supplementary Material Appendix A, we extend our analysis to characteristic portfolios. We document that recession probabilities predict a wide range of industry portfolios. We find that industries more strongly exposed to business cycle variation such as durable goods, manufacturing, energy, and technology are most predictable.

Finally, in Supplementary Material Appendix B, we provide additional international evidence, showing that recession probabilities based on the term spread have significant predictive power also for excess stock market returns in Germany, France, Canada, and the United Kingdom.

IV. Conclusion

Excess equity market returns are negative around business cycle peaks and sharply recover during recessions. In this article, we have shown that recession probabilities derived from probit models using the term spread have strong predictive power for the U.S. equity premium. The gains are statistically and economically significant and further improve when adding a backward-looking moving average of the term spread to the probit model. Equity premium forecasts based on recession probabilities correctly anticipate negative market returns heading into recessions and positive returns in expansions. We provide evidence for a structural break in the mean of the term spread in 1982. When correcting for this structural break, both recession and equity premium forecasts further improve. Our article thus provides further evidence for the strong link between the business cycle and the equity

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

premium. More specifically, it shows that information in the yield curve can be used to time the equity market.

Supplementary Material

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022109025102093.

References

Andreasen, M. M.; T. Engsted; S. V. Møller; and M. Sander. "The Yield Spread and Bond Return Predictability in Expansions and Recessions." *Review of Financial Studies*, 34 (2021), 2773–2812.

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

Balduzzi, P., and A. W. Lynch. "Transaction Costs and Predictability: Some Utility Cost Calculations." *Journal of Financial Economics*, 52 (1999), 47–78.

Baron, M.; W. Xiong; and Z. Ye. "Measuring Time-Varying Disaster Risk: An Empirical Analysis of Dark Matter in Asset Prices." Tech. Rep., Mimeo, Princeton University (2022).

Berge, T. J., and O. Jordà. "Evaluating the Classification of Economic Activity into Recessions and Expansions." *American Economic Journal: Macroeconomics*, 3 (2011), 246–277.

Bernard, H., and S. Gerlach. "Does the Term Structure Predict Recessions? The International Evidence." *International Journal of Finance and Economics*, 3 (1998), 195–215.

Bollerslev, T.; G. Tauchen; and H. Zhou. "Expected Stock Returns and Variance Risk Premia." *Review of Financial Studies*, 22 (2009), 4463–4492.

Campbell, J. Y. "A Variance Decomposition for Stock Returns." *Economic Journal*, 101 (1991), 157–179.

Campbell, J. Y., and J. Ammer. "What Moves the Stock and Bond Market? A Variance Decomposition for Long-Term Asset Returns." *Journal of Finance*, 48 (1993), 3–37.

Campbell, J. Y., and R. J. Shiller. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, 1 (1988), 195–228.

Campbell, J. Y., and S. Thompson. "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, 21 (2008), 1509–1531.

Campbell, J. Y., and T. Vuolteenaho. "Bad Beta, Good Beta." *American Economic Review*, 94 (2004), 1249–1275.

Campbell, S. D., and F. X. Diebold. "Stock Returns and Expected Business Conditions: Half a Century of Direct Evidence." *Journal of Business and Economic Statistics*, 27 (2009), 266–278.

Carlson, M. A., and D. C. Wheelock. "Navigating Constraints: The Evolution of Federal Reserve Monetary Policy, 1935–1959." Federal Reserve Bank of St. Louis Working Paper No. 2014-13 (2014).

Cenesizoglu, T., and A. Timmermann. "Do Return Prediction Models Add Economic Value?" *Journal of Banking & Finance*, 36 (2012), 2974–2987.

Chauvet, M., and J. M. Piger. "A Comparison of the Real-Time Performance of Business Cycle Dating Methods." *Journal of Business and Economic Statistics*, 26 (2008), 42–49.

Chauvet, M., and S. Potter. "Predicting a Recession: Evidence from the Yield Curve in the Presence of Structural Breaks." *Economics Letters*, 77 (2002), 245–253.

Chauvet, M., and S. Potter. "Business Cycle Monitoring with Structural Changes." *International Journal of Forecasting*, 26 (2010), 777–793.

Chen, J.; G. Tang; J. Yao; and G. Zhou. "Investor Attention and Stock Returns." Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3194387 (2020).

Chen, L., and X. Zhao. "Return Decomposition." *Review of Financial Studies*, 22 (2009), 5213–5249.

Cheng, I.-H. "The VIX Premium." *Review of Financial Studies*, 32 (2019), 180–227.

Chinn, M., and K. Kucko. "The Predictive Power of the Yield Curve Across Countries and Time." *International Finance*, 18 (2015), 129–156.

Christiansen, C.; J. N. Eriksen; and S. V. Møller. "Forecasting US Recessions: The Role of Sentiment." *Journal of Banking and Finance*, 49 (2014), 459–468.

Clark, T. E., and K. D. West. "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics*, 138 (2007), 291–311.

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

Cochrane, J. H. "Financial Markets and the Real Economy." In *Handbook of the Equity Risk Premium*, chap. 7, R. Mehra, ed. Amsterdam, Netherlands: Elsevier (2007), 237–325.
Cooper, I., and R. Priestley. "Time-Varying Risk Premiums and the Output Gap." *Review of Financial Studies*, 22 (2009), 2801–2833.
Da, Z.; D. Huang; and H. Yun. "Industrial Electricity Usage and Stock Returns." *Journal of Financial and Quantitative Analysis*, 52 (2017), 37–69.
Dangl, T., and M. Halling. "Predictive Regressions with Time-Varying Coefficients." *Journal of Financial Economics*, 106 (2012), 157–181.
DeMiguel, V.; Y. Plyakha; R. Uppal; and G. Vilkov. "Improving Portfolio Selection Using Option-Implied Volatility and Skewness." *Journal of Financial and Quantitative Analysis*, 48 (2013), 1813–1845.
Diebold, F. X., and R. S. Mariano. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13 (1995), 253–263.
Diebold, F. X., and G. D. Rudebusch. "Scoring the Leading Indicators." *Journal of Business*, 62 (1989), 369–391.
Eichengreen, B., and P. Garber. "Before the Accord: U.S. Monetary-Financial Policy, 1945–1951." In *Financial Markets and Financial Crises*. Chicago: University of Chicago Press, National Bureau of Economic Research, Inc. (1991), 175–206.
Estrella, A. "A New Measure of Fit for Equations with Dichotomous Dependent Variables." *Journal of Business and Economic Statistics*, 16 (1998), 198–205.
Estrella, A., and G. A. Hardouvelis. "The Term Structure as a Predictor of Real Economic Activity." *Journal of Finance*, 46 (1991), 555–576.
Estrella, A., and F. S. Mishkin. "Predicting U.S. Recessions: Financial Variables as Leading Indicators." *Review of Economics and Statistics*, 80 (1998), 45–61.
Estrella, A.; A. P. Rodrigues; and S. Schich. "How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States." *Review of Economics and Statistics*, 85 (2003), 629–644.
Fama, E. F., and K. R. French. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics*, 22 (1988), 3–25.
Fama, E. F., and K. R. French. "Business Conditions and Expected Returns on Stocks and Bonds." *Journal of Financial Economics*, 25 (1989), 23–49.
Ferreira, M. A., and P. Santa-Clara. "Forecasting Stock Market Returns: The Sum of the Parts is More Than the Whole." *Journal of Financial Economics*, 100 (2011), 514–537.
Ferson, W. E., and C. R. Harvey. "The Variation of Economic Risk Premiums." *Journal of Political Economy*, 99 (1991), 385–415.
Fornaro, P. "Forecasting US Recessions with a Large Set of Predictors." *Journal of Forecasting*, 35 (2016), 477–492.
Galvão, A. B. "Structural Break Threshold VARs for Predicting US Recessions Using the Spread." *Journal of Applied Econometrics*, 21 (2006), 463–487.
Galvão, A. B., and M. T. Owyang. "Forecasting Low Frequency Macroeconomic Events with High Frequency Data." Federal Reserve Bank of St. Louis Working Paper No. 2020-028 (2020).
Ghaderi, M.; M. Kilic; and S. B. Seo. "Why Do Rational Investors Like Variance at the Peak of a Crisis? A Learning-Based Explanation." *Journal of Monetary Economics*, 142 (2024), 103513.
Gómez-Cram, R. "Late to Recessions: Stocks and the Business Cycle." *Journal of Finance*, 77 (2022), 923–966.
Goyal, A.; I. Welch; and A. Zafirov. "A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, 37 (2024), 3490–3557.
Gu, S.; B. Kelly; and D. Xiu. "Empirical Asset Pricing via Machine Learning." *Review of Financial Studies*, 33 (2020), 2223–2273.
Hamilton, J. D. "Calling Recessions in Real Time." *International Journal of Forecasting*, 27 (2011), 1006–1026.
Hanley, J. A., and B. J. McNeil. "The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve." *Radiology*, 143 (1982), 29–36.
Hanley, J. A., and B. J. McNeil. "A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases." *Radiology*, 148 (1983), 839–843.
Harvey, D. I.; S. J. Leybourne; and P. Newbold. "Tests for Forecast Encompassing." *Journal of Business and Economic Statistics*, 16 (1998), 254–259.
Huang, D.; F. Jiang; J. Tu; and G. Zhou. "Forecasting Stock Returns in Good and Bad Times: The Role of Market States." Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2188989 (2017).
Huang, D., and M. Kilic. "Gold, Platinum, and Expected Stock Returns." *Journal of Financial Economics*, 132 (2019), 50–75.

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press

Ibbotson, R. G., and R. A. Sinquefield. "Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926-1974)." *Journal of Business*, 49 (1976), 11–47.
Jacobsen, B.; B. R. Marshall; and N. Visaltanachoti. "Stock Market Predictability and Industrial Metal Returns." *Management Science*, 65 (2019), 2947–3448.
Jiang, F.; J. Lee; X. Martin; and G. Zhou. "Manager Sentiment and Stock Returns." *Journal of Financial Economics*, 132 (2019), 126–149.
Johansson, P. "Predicting Recession Probabilities Using the Slope of the Yield Curve." FEDS Notes. Washington, DC: Board of Governors of the Federal Reserve System, March 1 (2018).
Kauppi, H., and P. Saikkonen. "Predicting U.S. Recessions with Dynamic Binary Response Models." *Review of Economics and Statistics*, 90 (2008), 777–791.
Lacker, J. M. "Introduction to Special Issue on the 50th Anniversary of the Treasury-Federal Reserve Accord." *Federal Reserve Bank of Richmond Economic Quarterly*, 87 (2001).
Lettau, M., and S. Ludvigson. "Consumption, Aggregate Wealth, and Expected Stock Returns." *Journal of Finance*, 56 (2001), 815–849.
Lettau, M., and S. Van Nieuwerburgh. "Reconciling the Return Predictability Evidence." *Review of Financial Studies*, 21 (2008), 1607–1652.
Liu, W., and E. Moench. "What Predicts US Recessions?" *International Journal of Forecasting*, 32 (2016), 1138–1150.
Lustig, H., and A. Verdelhan. "Business Cycle Variation in the Risk-Return Trade-Off." *Journal of Monetary Economics*, 58 (2012), 535–549.
Maio, P. "Return Decomposition and the Intertemporal CAPM." *Journal of Banking and Finance*, 37 (2013), 4958–4972.
McConnell, M. M., and G. Perez-Quiros. "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *American Economic Review*, 90 (2000), 1464–1476.
Neely, C. J.; D. E. Rapach; J. Tu; and G. Zhou. "Forecasting the Equity Risk Premium: The Role of Technical Indicators." *Management Science*, 60 (2014), 1772–1791.
Newey, W. K., and K. D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55 (1987), 703–708.
Nyberg, H. "Predicting Bear and Bull Stock Markets with Dynamic Binary Time Series Models." *Journal of Banking and Finance*, 37 (2013), 3351–3363.
Pesaran, M. H., and A. Timmermann. "Selection of Estimation Window in the Presence of Breaks." *Journal of Econometrics*, (2007), 134–161.
Pettenuzzo, D.; A. Timmermann; and R. Valkanov. "Forecasting Stock Returns Under Economic Constraints." *Journal of Financial Economics*, 114 (2014), 517–553.
Politis, D. N., and J. P. Romano. "The Stationary Bootstrap." *Journal of the American Statistical Association*, 89 (1994), 1303–1313.
Pyun, S. "Variance Risk in Aggregate Stock Returns and Time-Varying Return Predictability." *Journal of Financial Economics*, 132 (2019), 150–174.
Rapach, D. E.; M. C. Ringgenberg; and G. Zhou. "Short Interest and Aggregate Stock Returns." *Journal of Financial Economics*, 121 (2016), 46–65.
Rapach, D. E.; J. K. Strauss; and G. Zhou. "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *Review of Financial Studies*, 23 (2010), 821–862.
Rapach, D. E., and G. Zhou. "Forecasting Stock Returns." In *Handbook of Economic Forecasting*, Vol. 2A, G. Elliott and A. Timmermann, eds. Amsterdam, Netherlands: Elsevier (2013), 329–383.
Rudebusch, G. D.; B. P. Sack; and E. T. Swanson. "Macroeconomic Implications of Changes in the Term Premium." Federal Reserve Bank of St. Louis Review, July/August (2007), 241–269.
Rudebusch, G. D., and J. C. Williams. "Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve." *Journal of Business and Economic Statistics*, 27 (2009), 492–503.
Schrimpf, A., and Q. Wang. "A Reappraisal of the Leading Indicator Properties of the Yield Curve under Structural Instability." *International Journal of Forecasting*, 26 (2010), 836–857.
Welch, I., and A. Goyal. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, 21 (2008), 1455–1508.
Wheelock, D. C., and M. E. Wohar. "Can the Term Spread Predict Output Growth and Recessions? A Survey of the Literature." Federal Reserve Bank of St. Louis Review, September/October, Part 1 (2009), 419–440.
Wright, J. H. "The Yield Curve and Predicting Recessions." FEDS Working Paper No. 2006–7 (2006).

https://doi.org/10.1017/S0022109025102093 Published online by Cambridge University Press