Equity premium predictability over the business cycle

Emanuel Moench and Tobias Stein*

July 2, 2025

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.

JEL classification: E32, E37, C53, G11, G17

Keywords: Recession predictability, return predictability, business cycle, probit model, term spread

*Moench (corresponding author), e.moench@fs.de, Frankfurt School of Finance & Management, CEPR; Stein, Independent, stein_to@web.de. We thank Jonah Schmidt for research assistance and Arturo Estrella, Roberto Gómez-Cram, Amit Goyal, Mete Kilic, Stijn Van Nieuwerburgh as well as anonymous referees and the editor, Thierry Foucault, for helpful comments and suggestions.

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), 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]

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

[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 paper.

[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 (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).

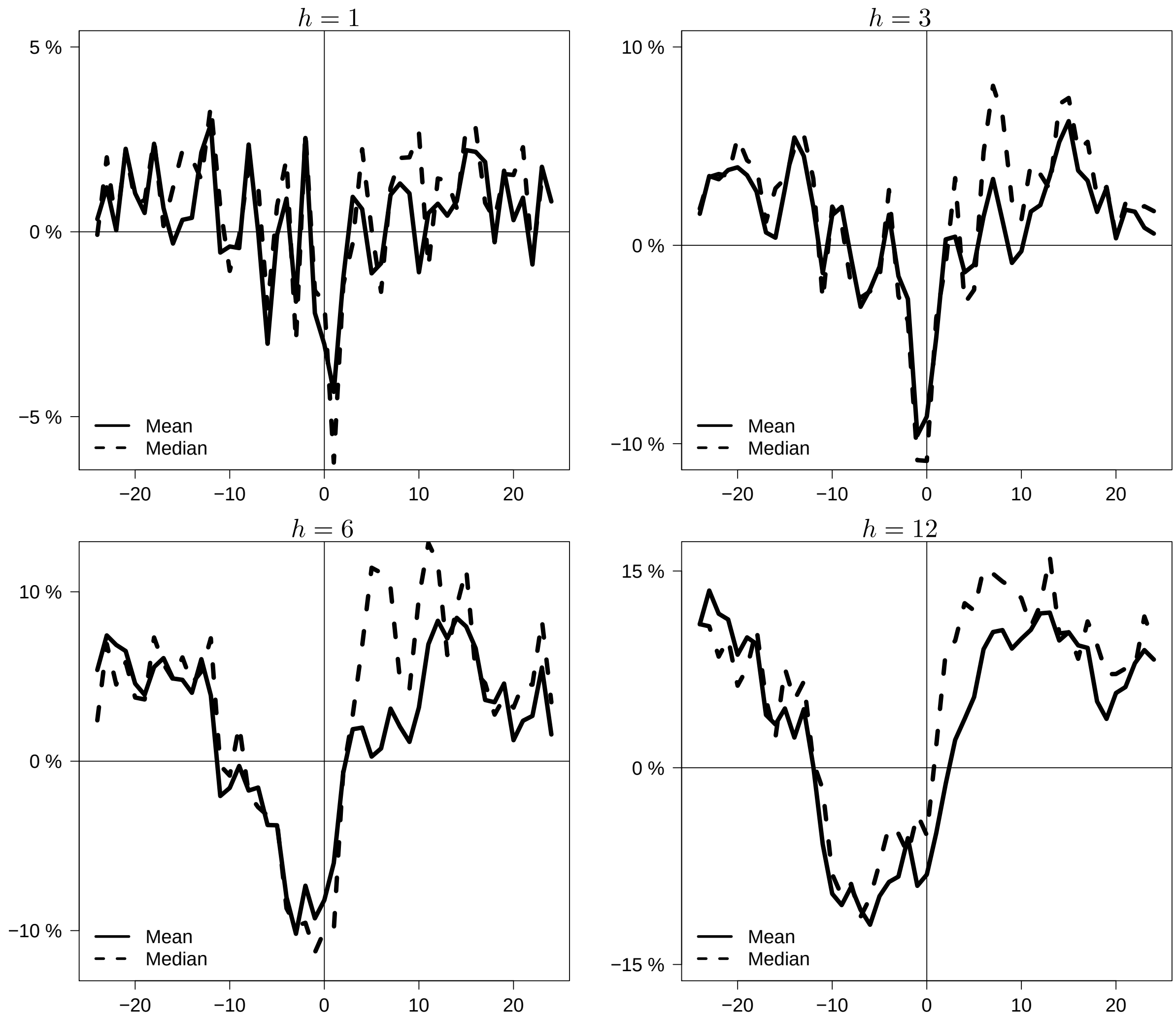
cycle (e.g. Fama and French (1989), Ferson and Harvey (1991), Cochrane (2007), Campbell and Diebold (2009)). Lustig and Verdelhan (2012) document significant variation in realized excess returns around recessions, showing that they are negative at the business cycle peak and then sharply rise over the 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 eleven NBER recessions from March 1951 to December 2020. The realized market return is mostly negative but relatively volatile for the one-month and three-month window around the beginning of recessions. However, a clear V-shape emerges for the cumulative six- and twelve-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 six 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 one-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% six to eleven months after the peak. The last two 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 two years (one year) of observations around each peak, the average equity premium and

FIGURE 1

Log excess U.S. equity market return around business cycle peaks

This figure 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 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 $h = 1, 3, 6, 12$.



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.

TABLE 1

Log excess stock market return over the business cycle

This table 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 sub-samples 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

It is well documented that the term spread is a robust predictor of recessions for horizons of one 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 spread with financial variables such as stock prices and other spreads, as well as monetary aggregates. While some alternative predictors are useful over one- to three-quarter horizons, it is the term spread that predicts best over horizons of one-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 paper, 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 twelve 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 UK, Germany, and France.

We start by documenting that a backward-looking three-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

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

model for forecasting recessions with the term spread suffers from a structural break (see, e.g., Chauvet and Potter (2002,0)). 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" 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 one-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, for example industry portfolios.

In a related recent paper, Gómez-Cram (2022) studies one-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 and his paper. First, while Gómez-Cram (2022) studies only one-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 Online Appendix.

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 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 A. In Sections B and 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 D that the recession probabilities implied by probit models have strong in-sample predictive power for the U.S. equity premium. In Section 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 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 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.[4] 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 six years afterwards 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 the National Bureau of Economic Research (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 distinguishes between the probability

[4]See `http://www.hec.unil.ch/agoyal/`

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 one 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:

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

$$\text{Pr}\big[Y_{t+1:t+h} = 1|X_t\big] = \Phi\big[X'_t\beta\big], \tag{2}$$

where $X'_t = (1, x_{1,t}, \ldots, x_{p,t})'$ 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, $\Phi\big[\,\cdot\,\big]$ is the cumulative distribution function of the standard normal distribution and Pr denotes probability. Let $\hat{\text{p}}_{t+1:t+h|t}$ be the out-of-sample forecast for $\text{Pr}\big[Y_{t+1:t+h} = 1\big]$ 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 one year.[5] 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

[5]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 $t - 12$.

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 one to two 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 six 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 three-year moving average of the term spread (MA-TMS) which we construct as $\frac{1}{36}\sum_{j=0}^{35}\text{TMS}_{t-j}$.[6]

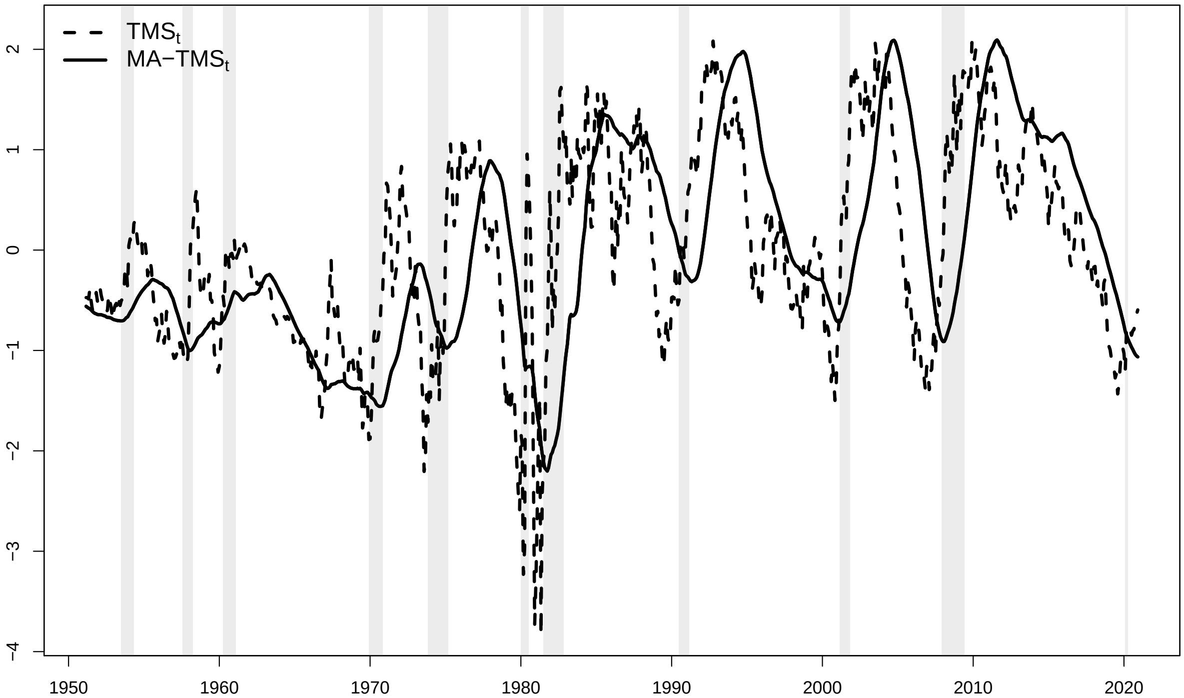
Figure 2 superimposes TMS and MA-TMS. While the local minima of the term spread usually lead the beginning of recessions by several months, the moving average term spread reaches its troughs often just before the onset of recessions. Moreover, the smoothed series averages out some local minima that are 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

[6]We show in the Online Appendix that our results are robust to the length of the moving average window.

FIGURE 2

Term spread and moving average term spread

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



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 for 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 six 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 one-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).

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 1990-91 recession compared to the early 1980s. Kauppi and Saikkonen (2008) also document - using probit models with lagged dependent variables - that the 1990-91 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 Appendix A 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.

C. Forecast evaluation

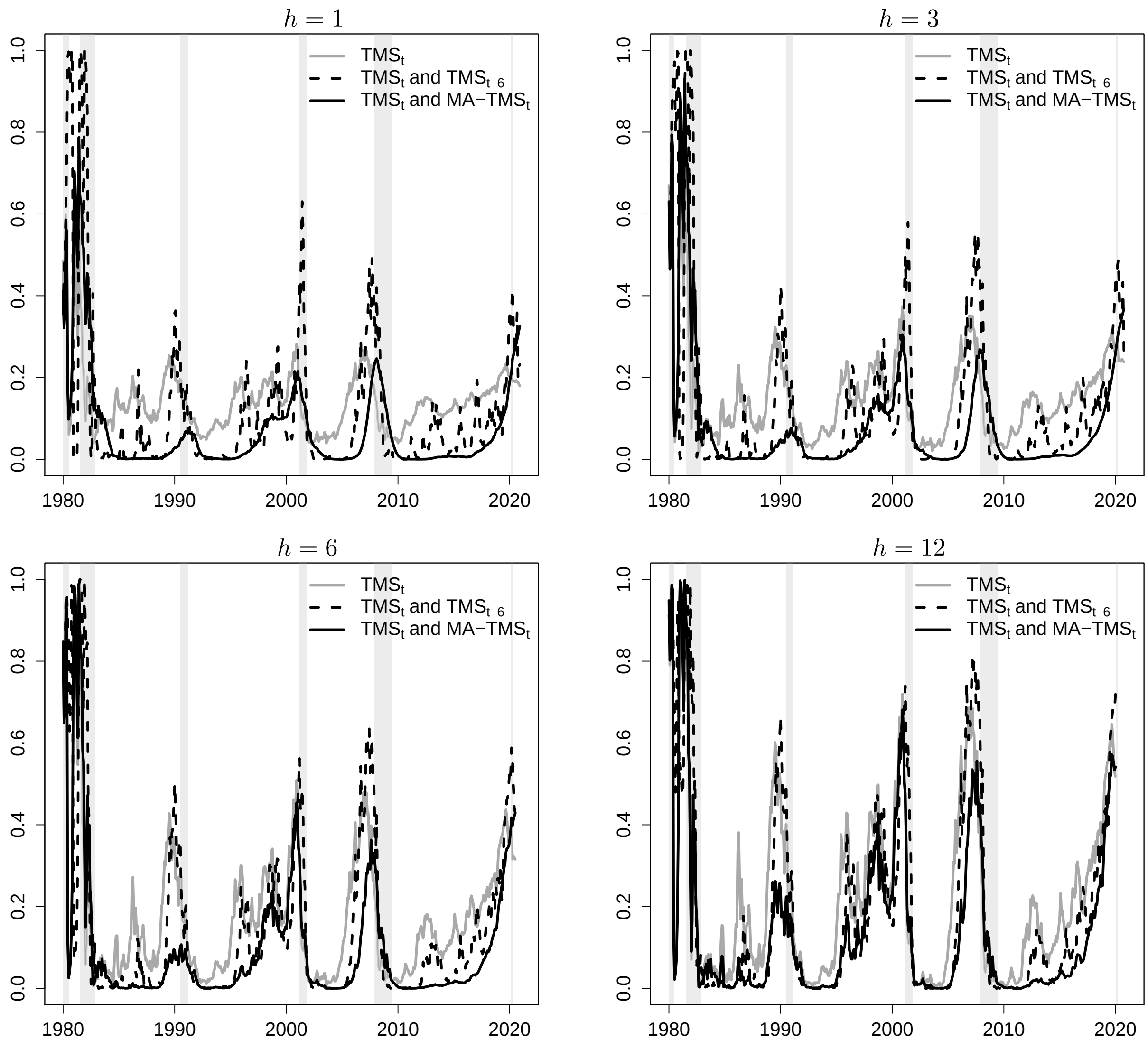
We follow the recession prediction literature and use the quadratic probability score (QPS), the logarithm score (LS) and the diagonal elementary score (DES) to formally evaluate the accuracy of recession probability forecasts.[7] Perfect classification ability results in values of zero for all three scores; otherwise they have positive values, with higher values indicating poorer forecast

[7]Details on the estimation of all statistics in this section are provided in Online Appendix A.

FIGURE 3

Out-of-sample recession probability forecasts

This figure shows the out-of-sample recession probability forecasts at the 1-, 3-, 6-, 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 three years (six 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.



performance, see, e.g., Nyberg (2013); Christiansen, Eriksen and Møller (2014); 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 one 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 four to six quarters (Estrella and Mishkin, 1998). When adding the six-months 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 one-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.[8]

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 for the models including lagged term spread information and for longer forecast horizons. At the one-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

[8] 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 one-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

This table presents five 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 three-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 two-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 $***$ denote significance at the 10%, 5%, and 1% significance levels.

	1980:1 to 2020:12						
Variables in probit model	QPS	LS	DES	pseudo R^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

toward the north-west corner for $h = 6$ and $h = 12$. Hence, 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 model 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 information, we next assess the usefulness of implied recession probabilities to forecast the equity premium. We use the standard linear predictive regression model:

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

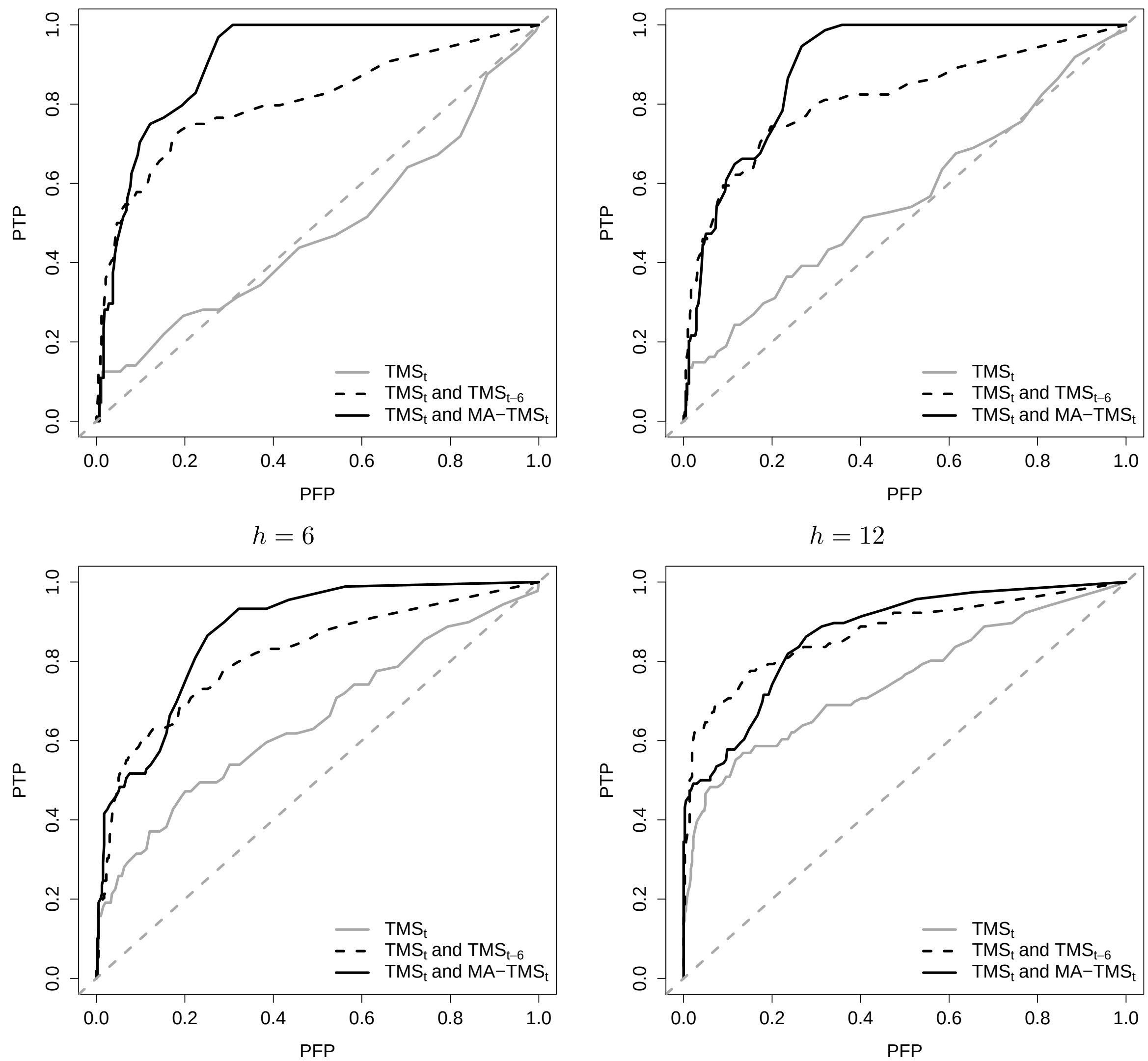
where $r_{t+1:t+h} = \frac{1}{h}\sum_{j=1}^{h} r_{t+j}$ is the average of the cumulative log equity premium between $t + 1$ and $t + h$, α_h and β_h are coefficients, and $\epsilon_{t+1:t+h}$ is the error term. The only predictor is the recession probability forecast $\hat{\mathrm{p}}_{t+1:t+h|t}$ which 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 (top panel) and for a shorter sample starting in 1980 (bottom panel). We consider this subsample in light of the evidence that the properties of the business cycle in the U.S. have

FIGURE 4

Out-of-sample performance: ROC curves

This figure 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 (six-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. $h = 1$
$h = 3$



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 which is based on the model by Estrella and Hardouvelis (1991), and the one-month ahead probability from Chauvet and Piger (2008).[9] Importantly, while the New York Fed model also considers the term spread as the only predictor, it forecasts a recession in exactly one year instead of within the next twelve 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-squared 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 additional predictor stands out, delivering a whopping 13 percent adjusted R-squared in this period. Second, and more importantly, the three models also strongly outperform the two benchmark models at forecast horizons of three months and beyond. The New York Fed recession probability and the one by Chauvet and Piger (2008) have better forecasting ability only at the one-month ahead horizon.

In the top panel of Table B.1 in the Online Appendix, 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

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

information could indeed have exploited by investors in real time will be assessed in the next section.

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 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 $\hat{\mathrm{p}}_{t+1:t+h|t}$. 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 $\hat{\mathrm{p}}_{t+1:t+h|t}$ to predict $r_{t+1:t+h}$. Thus, the log equity premium forecast is $\hat{r}_{t+1:t+h|t} = \hat{\alpha}_t + \hat{\beta}_t \hat{\mathrm{p}}_{t+1:t+h|t}$. 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

TABLE 3

In-sample equity premium prediction with estimated recession probabilities

This table reports in-sample results for the slope coefficient from regressions of log excess returns ($r_{t+1:t+h}$) on recession probability forecasts; $r_{t+1:t+h} = \alpha + \hat{\mathrm{p}}_{t+1:t+h|t} + \epsilon_{t+1:t+h}$. $\hat{\mathrm{p}}_{t+1:t+h|t}$ refers to recession probability forecasts from probit models with (1) TMS_t, (2) TMS_t and TMS_{t-6}, (3) 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 parenthesis 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.

(1) Variable	(2) $h = 1$	(3) $h = 3$	(4) $h = 6$	(5) $h = 12$
	Panel A: In-sample results (1951:3 to 2020:12)			
$\hat{\mathrm{p}}_{t+1:t+h\|t}$ (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
$\hat{\mathrm{p}}_{t+1:t+h\|t}$ (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
$\hat{\mathrm{p}}_{t+1:t+h\|t}$ (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)			
$\hat{\mathrm{p}}_{t+1:t+h\|t}$ (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
$\hat{\mathrm{p}}_{t+1:t+h\|t}$ (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
$\hat{\mathrm{p}}_{t+1:t+h\|t}$ (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			

constant ($\beta_t = 0$), see, among others, Rapach et al. (2010); Jiang, Lee, Martin and Zhou (2019):

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

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.[10] 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, page 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

[10] 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 one-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.

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 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 e.g. Chauvet and Potter (2002) who find evidence for a structural break in the probit model

TABLE 4

Out-of-sample R^2 statistics for log equity premium forecasts

This table 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 one 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 model add either the term spread lagged by six-months (TMS_{t-6}) or the three-year moving average of the term spread (MA-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 equally-weighted average of forecasts from univariate regressions with TMS_t, TMS_{t-6}, and MA-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 smaller MSFE than the historical average benchmark. The out-of-sample period runs from 1980:1 to 2020:12.

(1) Variable	(2) $h=1$	(3) $h=3$	(4) $h=6$	(5) $h=12$
		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-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-TMS_t	0.29	1.02*	1.87**	4.16**
CF-MEAN	0.17	0.97*	1.71**	5.70***
TMS_t, MA-TMS_t	-1.15	-2.60	-4.30	0.02**

based on the term spread but argue that the exact date of the break is difficult to localize.[11] Other papers also document breaks in the dynamic relationship between the term spread and real growth in the U.S. 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

[11] 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.

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 Online 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 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 sub-sample 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 one 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 equally-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 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)).[12] 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.

---

[12] 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.

TABLE 5

Out-of-sample R^2 statistics when correcting for a structural break

This table 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 ($\mathrm{TMS}_t^{\mathrm{break}}$, $\frac{1}{36}\sum_{j=0}^{35}\mathrm{TMS}_{t-j}^{\mathrm{break}}$) 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 Online 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 sub-samples, and panels A to D present results for different forecasting horizons.

Variable	(1) 1980:1-2020:12	(2) 1980:1-1999:12	(3) 2000:1-2020:12	(4) 1990:1-2013:12
Panel A: $\mathbf{h = 1}$				
TMS_t, MA-TMS_t	0.96**	1.52**	0.43*	0.74**
$\mathrm{TMS}_t^{\mathrm{break}}$, MA-$\mathrm{TMS}_t^{\mathrm{break}}$	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: $\mathbf{h = 3}$				
TMS_t, MA-TMS_t	2.76***	4.40**	1.35*	2.17**
$\mathrm{TMS}_t^{\mathrm{break}}$, MA-$\mathrm{TMS}_t^{\mathrm{break}}$	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: $\mathbf{h = 6}$				
TMS_t, MA-TMS_t	3.42**	5.04**	2.24	3.54**
$\mathrm{TMS}_t^{\mathrm{break}}$, MA-$\mathrm{TMS}_t^{\mathrm{break}}$	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: $\mathbf{h = 12}$				
TMS_t, MA-TMS_t	8.99***	10.10**	8.41**	8.58***
$\mathrm{TMS}_t^{\mathrm{break}}$, MA-$\mathrm{TMS}_t^{\mathrm{break}}$	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	

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 smallest. Across horizons and prediction periods, cross-validation seems to be the most robust choice for improving recession and equity premium forecasts.[13]

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 sub-samples and gradually increase in the forecast horizon h - with R^2_{OS} values above 10% for cumulative one 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 gold-to-platinum ratio.[14] While the VRP performs best at the one-month forecast horizon with an $R^2_{\text{OS}} = 4.44\%$ when ending the sample in 2019, consistent with Goyal et al. (2024) this indicator entirely

[13]We show the recursively selected estimation windows from cross-validation in Online Appendix C.

[14]The results differ slightly compared to those published in the papers because our historical average uses data

loses its predictive power in the Covid 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 term spread information increasingly become the dominant forecasting variables relative to the benchmark predictors. At the one-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.[15]

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 predictor significantly outperform the historical average benchmark. This contrasts 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 $\hat{\mathrm{p}}_{t+1:t+h|t}$ 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 three individual OLS regression forecasts (CF-MEAN). While TMS only has mild predictive power over one-year ahead forecast horizons, TMS_{t-6} and MA-TMS_t also significantly predict the equity premium for

from 1951:3 onwards. We re-estimate the predictive coefficients recursively using expanding data available in real-time.

[15]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.

$h = 3$ 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.

The upper panel of Figure 5 illustrates this result by superimposing one-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 1981-82 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.

Welch and Goyal (2008) have popularized an intuitive way to visualize the relative forecast performance of different prediction models over time. The lower panel 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 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 1981-82 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 one-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 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 Online Appendix, we show out-of-sample results for a much

TABLE 6

Out-of-sample R^2 statistics in recessions and expansions

This table 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 Online 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.

(1)	(2)	(3)	(4)	(5)
Variable	$h = 1$	$h = 3$	$h = 6$	$h = 12$
		Panel A: Recessions		
TMS_t, $\mathrm{MA\text{-}TMS}_t$	1.60*	2.40*	-1.60	-1.20
$\mathrm{TMS}_t^{\mathrm{break}}$, $\mathrm{MA\text{-}TMS}_t^{\mathrm{break}}$	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, $\mathrm{MA\text{-}TMS}_t$	0.71**	2.95***	6.67**	14.82***
$\mathrm{TMS}_t^{\mathrm{break}}$, $\mathrm{MA\text{-}TMS}_t^{\mathrm{break}}$	-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***

longer time period, starting in 1946 and ending in 2022. While the forecasts start in 1946, we use data from 1871 onwards 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-squared.

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.[16] 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 example, the R^2_{OS} for the one-month ahead forecasts improves from 1.6% to 5.1% and 8.7% for cross-validation and post-break window.[17]

[16]Note that the R^2_{OS} statistic for the unadjusted probit model turns negative in recessions for six- and twelve-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.

[17]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$_t^{\text{break}}$, MA-TMS$_t^{\text{break}}$; 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$_t^{\text{break}}$, MA-TMS$_t^{\text{break}}$. 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 post-break window. In sum, we conclude that 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.

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

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

where γ is the coefficient of relative risk aversion and $\hat{r}_{t+1|t}$ is a forecast of the equity premium.[18] 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:

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

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 period t and $t+1$. Furthermore, the certainty equivalent return (CER) can be calculated

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

as:

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

where $\hat{\mu}_P$ and $\hat{\sigma}_P^2$ 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 E. Table 7 presents ΔCER and ΔSR for monthly re-balancing for the sample periods 1980-2020 (Panel A) and 1990-2013 (Panel B).[19] 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.[20] 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 1980-2020, the probit model with TMS and MA-TMS

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

[20]We set the average block length to three months (Politis and Romano, 1994).

provides utility gains in the range of 1.2% to 2.6%. Hence, an investor would be willing to pay between 120 to 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 when $\gamma = 3$ and when leveraging and short-selling up to 50% is allowed. Nonetheless, even an investor with ω between zero and one 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 more than triple or quadruple. Similarly, the annualized Sharpe ratio rises from around 0.50 for the historical average to between 0.65 to 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 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:

(6) $$r_{t+1} = E_t[r_{t+1}] + N_{t+1}^{CF} - N_{t+1}^{DR},$$

TABLE 7

Asset allocation exercise

This table 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 one 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). $*$, $**$, $***$ indicate significantly improved performance relative to the prevailing mean benchmark at the 10%, 5%, and 1% significance level. The p-values are obtained by using a bootstrap approach similar to DeMiguel et al. (2013) with the average block length set to three months (Politis and Romano, 1994).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: 1980:1 to 2020:12							
	ΔCER				ΔSR			
γ	3	5	3	3	3	5	3	3
ω	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$
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***
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	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							
	ΔCER				ΔSR			
γ	3	5	3	3	3	5	3	3
ω	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$
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**
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	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*

where $E_t[r_{t+1}]$ is the conditional expectation of the log return at time t, and N_{t+1}^{CF} and N_{t+1}^{DR} 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:

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

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

where $e1$ is a k-vector with a one in the first cell and a zero 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.[21] Additionally, we show results for a VAR model,

[21]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).

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:

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

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

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

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 $\hat{\mathrm{p}}_{t+1|t}$ 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 zero.

TABLE 8

Predictive regressions for stock market return components

This table 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). $*$, $**$, $***$ denote significance at the 10%, 5%, and 1% significance levels.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VAR variables	$\hat{\beta}^E$	$\hat{\beta}^{CF}$	$\hat{\beta}^{DR}$	VAR variables	$\hat{\beta}^E$	$\hat{\beta}^{CF}$	$\hat{\beta}^{DR}$
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]				

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 Online Appendix.

Our baseline results rely on a moving average of three years for the term spread. In Online Appendix A, we provide R^2_{OS} statistics for different 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 one to five years, whereas the one-month ahead forecasts perform best for averages between three to five years.

In Online 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 six- and twelve-month ahead forecasts.

In Online 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 Online 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 Online 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 twelve-month moving average predict recessions about equally well as the term spread and moving average term spread, the implied recession probabilities tend to spike towards 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 Online Appendix A, we extend our analysis to characteristics 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 Online 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 UK.

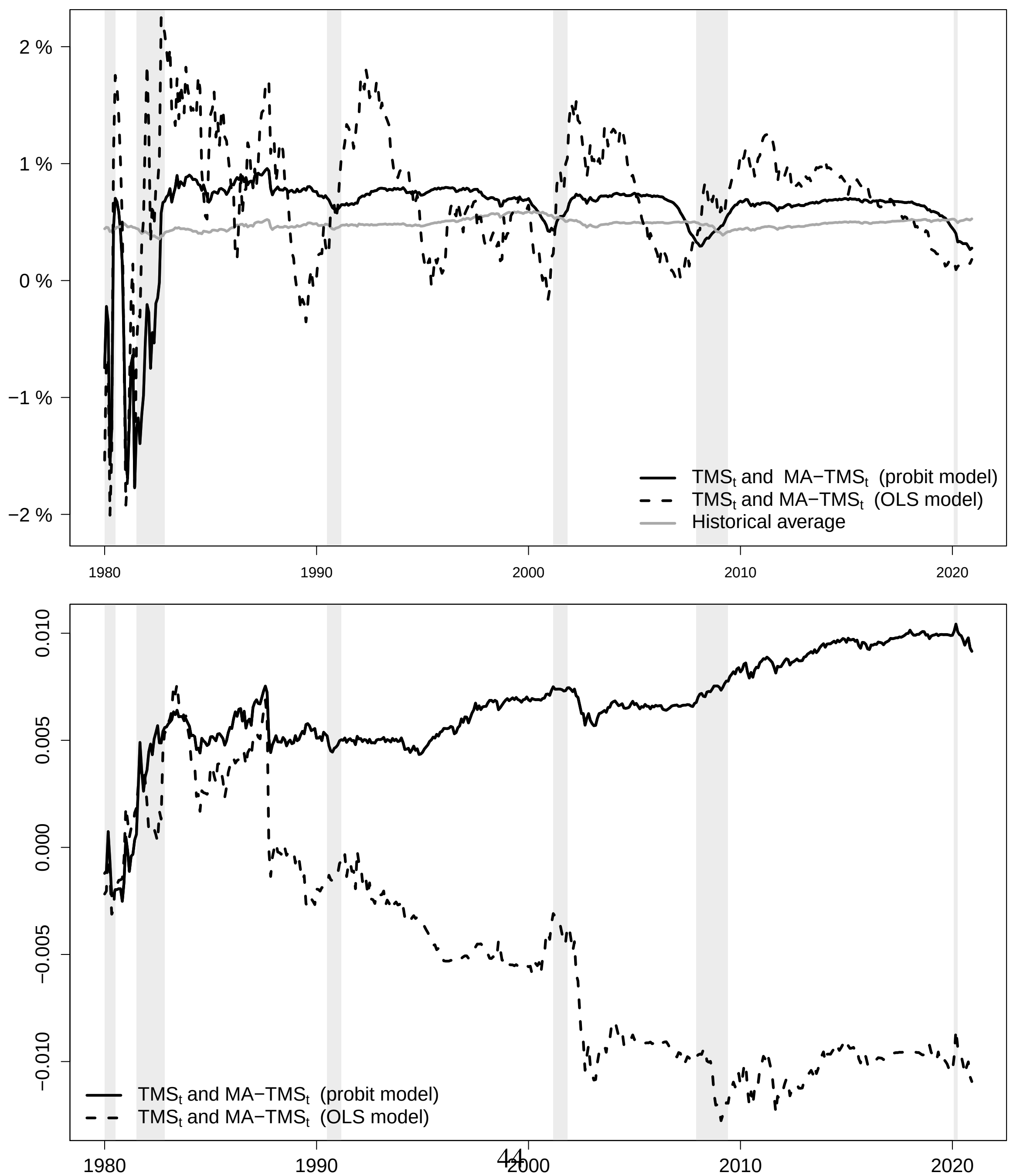
IV. Conclusion

Excess equity market returns are negative around business cycle peaks and sharply recover during recessions. In this paper, 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 paper thus provides further evidence for the strong link between the business cycle and the equity premium. More specifically, it shows that information in the yield curve can be used to time the equity market.

FIGURE 5

Out-of-sample forecasts and performance over time

This figure shows one-month ahead forecasts of the log equity premium for three different models (upper panel). 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 three-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. The lower panel 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.



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*, 2021, *34*, 2773–2812.

Andrews, D. W. K., "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 1991, *59* (3), 817–858.

__ , "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 1993, *61*, 821–856.

__ **and W. Ploberger**, "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 1994, *62*, 1383–1414.

Bai, J., "Estimation of a change point in multiple regression models," *Review of Economics and Statistics*, 1997, *79*, 551–563.

__ **and P. Perron**, "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 1998, *66* (1), 47–78.

__ **and** __ , "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 2003, *18* (1), 1–22.

Balduzzi, P. and A. W. Lynch, "Transaction costs and predictability: some utility cost calculations," *Journal of Financial Economics*, 1999, *52*, 47–78.

Baron, Matthew, Wei Xiong, and Zhijiang Ye, "Measuring time-varying disaster risk: An empirical analysis of dark matter in asset prices," Technical Report, Mimeo, Princeton University 2022.

Berge, T. J. and Ò. Jordà, "Evaluating the Classification of Economic Activity into Recessions and Expansions," *American Economic Journal: Macroeconomics*, 2011, *3* (2), 246–277.

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

Bollerslev, Tim, George Tauchen, and Hao Zhou, "Expected stock returns and variance risk premia," *The Review of Financial Studies*, 2009, *22* (11), 4463–4492.

Bouallègue, Z. B., L. Magnusson, T. Haiden, and D. S. Richardson, "Monitoring trends in ensemble forecast performance focusing on surface variables and high-impact events," *Quarterly Journal of the Royal Meteorological Society*, 2019, *145*, 1741–1755.

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

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

— **and R. J. Shiller**, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1988, *1* (3), 195–228.

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

— **and T. Vuolteenaho**, "Bad Beta, Good Beta," *American Economic Review*, 2004, *94* (5), 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*, 2009, *27*, 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 Papers 2014-13 2014.

Cenesizoglu, T. and A. Timmermann, "Do return prediction models add economic value?," *Journal of Banking & Finance*, 2012, *36*, 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*, 2008, *26* (1), 42–49.

— **and S. Potter**, "Predicting a recession: evidence from the yield curve in the presence of structural breaks," *Economics Letters*, 2002, *77*, 245–253.

— **and** —, "Business cycle monitoring with structural changes," *International Journal of Forecasting*, 2010, *26*, 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*, 2009, *22* (12), 5213–5249.

Chen, S.-S., Y.-H. Chou, and C.-Y. Yen, "Predicting US recessions with stock market illiquidity," *B.E. Journal of Macroeconomics*, 2016, *16*, 93–123.

Cheng, Ing-Haw, "The VIX premium," *The Review of Financial Studies*, 2019, *32* (1), 180–227.

Chinn, Menzie and Kavan Kucko, "The predictive power of the yield curve across countries and time," *International Finance*, 2015, *18* (2), 129–156.

Chow, G. C., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 1960, *28*, 591–605.

Christiansen, C., J. N. Eriksen, and S. V. Møller, "Forecasting US recessions: The role of sentiment," *Journal of Banking and Finance*, 2014, *49*, 459–468.

Clark, T. E. and K. D. West, "Approximately normal tests for equal predictive accuracy in nested models," *Journal of Econometrics*, 2007, *138*, 291–311.

Cochrane, J. H., "Financial Markets and the Real Economy," in R. Mehra, ed., *Handbook of the Equity Risk Premium*, Elsevier, 2007, chapter 7, pp. 237–325.

Cooper, I. and R. Priestley, "Time-Varying Risk Premiums and the Output Gap," *Review of Financial Studies*, 2009, *22*, 2801–2833.

Da, Z., D. Huang, and H. Yun, "Industrial Electricity Usage and Stock Returns," *Journal of Financial and Quantitative Analysis*, 2017, *52* (1), 37–69.

Dangl, T. and M. Halling, "Predictive regressions with time-varying coefficients," *Journal of Financial Economics*, 2012, *106*, 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*, 2013, *48* (6), 1813–1845.

Diebold, F. X. and G. D. Rudebusch, "Scoring the Leading Indicators," *Journal of Business*, 1989, *62*, 369–391.

_ **and R. S. Mariano**, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 1995, *13*, 253–263.

Eichengreen, B. and P. Garber, "Before the Accord: U.S. Monetary-Financial Policy, 1945-1951," in "Financial Markets and Financial Crises," National Bureau of Economic Research, Inc, 1991, pp. 175–206.

Estrella, A., "A New Measure of Fit for Equations With Dichotomous Dependent Variables," *Journal of Business and Economic Statistics*, 1998, *16*, 198–205.

_ **, 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*, 2003, *85*, 629–644.

_ **and F. S. Mishkin**, "Predicting U.S. recessions: financial variables as leading indicators," *Review of Economics and Statistics*, 1998, *80* (1), 45–61.

_ **and G. A. Hardouvelis**, "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance*, 1991, *46*, 555–576.

Fama, E. F. and K. R. French, "Dividend yields and expected stock returns," *Journal of Financial Economics*, 1988, *22*, 3–25.

_ **and** _ , "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics*, 1989, *25* (1), 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*, 2011, *100*, 514–537.

Ferson, W. E. and C. R. Harvey, "The Variation of Economic Risk Premiums," *Journal of Political Economy*, 1991, *99*, 385–415.

Fornaro, P., "Forecasting US Recessions with a Large Set of Predictors," *Journal of Forecasting*, 2016, *35*, 477–492.

Galvão, A. B., "Structural Break Threshold VARs for Predicting US Recessions using the Spread," *Journal of Applied Econometrics*, 2006, *21*, 463–487.

__ **and M. T. Owyang**, "Forecasting Low Frequency Macroeconomic Events with High Frequency Data," FRB St. Louis Working Paper No. 2020-028 2020.

Ghaderi, Mohammad, Mete Kilic, and Sang Byung Seo, "Why do rational investors like variance at the peak of a crisis? A learning-based explanation," *Journal of Monetary Economics*, 2024, *142*, 103513.

Gómez-Cram, Roberto, "Late to recessions: Stocks and the business cycle," *The Journal of Finance*, 2022, *77* (2), 923–966.

Goyal, Amit, Ivo Welch, and Athanasse Zafirov, "A comprehensive 2022 look at the empirical performance of equity premium prediction," *The Review of Financial Studies*, 2024, *37* (11), 3490–3557.

Gu, S., B. Kelly, and D. Xiu, "Empirical Asset Pricing via Machine Learning," *Review of Financial Studies*, 2020, *33*, 2223–2273.

Hamilton, J. D., "Calling recessions in real time," *International Journal of Forecasting*, 2011, *27*, 1006–1026.

Hanley, J. A. and B. J. McNeil, "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve," *Radiology*, 1982, *143*, 29–36.

__ **and** __ , "A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases," *Radiology*, 1983, *148*, 839–843.

Hansen, B. E., "Approximate asymptotic p-values for structural change tests," *Journal of Business and Economic Statistics*, 1997, *15*, 60–67.

__ , "The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity," *Journal of Economic Perspectives*, 2001, *15* (4), 117–128.

Harvey, D. I., S. J. Leybourne, and P. Newbold, "Tests for Forecast Encompassing," *Journal of Business and Economic Statistics*, 1998, *16*, 254–259.

Huang, D. and M. Kilic, "Gold, platinum, and expected stock returns," *Journal of Financial Economics*, 2019, *132*, 50–75.

__ **, F. Jiang, J. Tu, and G. Zhou**, "Investor Sentiment Aligned: A Powerful Predictor of Stock Returns," *Review of Financial Studies*, 2015, *28* (3), 791–837.

— , — , — , **and** — , "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.

Ibbotson, R. G. and R. A. Sinquefield, "Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926-1974)," *Journal of Business*, 1976, *49*, 11–47.

Jacobsen, B., B. R. Marshall, and N. Visaltanachoti, "Stock Market Predictability and Industrial Metal Returns," *Management Science*, 2019, *65*, 2947–3448.

Jiang, F., J. Lee, X. Martin, and G. Zhou, "Manager sentiment and stock returns," *Journal of Financial Economics*, 2019, *132*, 126–149.

Johansson, P., "Predicting Recession Probabilities Using the Slope of the Yield Curve," FEDS Notes. Washington: 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*, 2008, *90* (4), 777–791.

Kim, C.-J. and C. R. Nelson, "Has the U.S. Economy Become More Stable? A Bayesian Approach based on a Markov-Switching Model of the Business Cycle," *Review of Economics and Statistics*, 1999, *81*, 608–616.

Lacker, J. M., "Introduction to Special Issue on the 50th anniversary of the Treasury-Federal Reserve Accord," Federal Reserve Bank of Richmond *Economic Quarterly*, Vol. 87, 2001.

Lettau, M. and S. Ludvigson, "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 2001, *56*, 815–849.

— **and S. Van Nieuwerburgh**, "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, 2008, *21* (4), 1607–1652.

Liu, W. and E. Moench, "What predicts US recessions?," *International Journal of Forecasting*, 2016, *32*, 1138–1150.

Lustig, H. and A. Verdelhan, "Business cycle variation in the risk-return trade-off," *Journal of Monetary Economics*, 2012, *58*, 535–549.

Maio, P., "Return decomposition and the Intertemporal CAPM," *Journal of Banking and Finance*, 2013, *37*, 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*, 2000, *90*, 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*, 2014, *60* (7), 1772–1791.

Newey, W. K. and K. D. West, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 1987, *55* (3), 703–708.

— and —, "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 1994, *61*, 631–653.

Nyberg, H., "Predicting bear and bull stock markets with dynamic binary time series models," *Journal of Banking and Finance*, 2013, *37*, 3351–3363.

Pesaran, M. H. and A. Timmermann, "Selection of estimation window in the presence of breaks," *Journal of Econometrics*, 2007, (137), 134–161.

Pettenuzzo, D., A. Timmermann, and R. Valkanov, "Forecasting stock returns under economic constraints," *Journal of Financial Economics*, 2014, *114*, 517–553.

Politis, D. N. and J. P. Romano, "The Stationary Bootstrap," *Journal of the American Statistical Association*, 1994, *89* (428), 1303–1313.

Pyun, S., "Variance risk in aggregate stock returns and time-varying return predictability," *Journal of Financial Economics*, 2019, *132*, 150–174.

Rapach, D. E. and G. Zhou, "Forecasting Stock Returns," in G. Elliott and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 2A, Elsevier, Amsterdam, 2013, pp. 329–383.

—, J. K. Strauss, and G. Zhou, "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy," *Review of Financial Studies*, 2010, *23* (2), 821–862.

—, M. C. Ringgenberg, and G. Zhou, "Short interest and aggregate stock returns," *Journal of Financial Economics*, 2016, *121*, 46–65.

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*, 2009, *27*, 492–503.

—, B. P. Sack, and E. T. Swanson, "Macroeconomic Implications of Changes in the Term Premium," Federal Reserve Bank of St. Louis *Review* 2007.

Schrimpf, A. and Q. Wang, "A reappraisal of the leading indicator properties of the yield curve under structural instability," *International Journal of Forecasting*, 2010, *26*, 836–857.

Smith, S. C. and A. Timmermann, "Break Risk," *Review of Financial Studies*, 2021, *34*, 2045–2100.

Welch, I. and A. Goyal, "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 2008, *21*, 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.

Working, H., "Note on the Correlation of First Differences of Averages in a Random Chain," *Econometrica*, 1960, *28* (4), 916–918.

Wright, J. H., "The Yield Curve and Predicting Recessions," FEDS Working Paper No. 2006-7 2006.

Online Appendix for
“Equity premium predictability over the business cycle”

by Emanuel Moench & Tobias Stein

Not for Publication

A. Forecast evaluation

The quadratic probability score (QPS), the logarithm score (LS), and the diagonal elementary score (DES) are given by:

$$\text{(A.1)} \qquad \text{QPS} = \frac{2}{\tau}\sum_{j=M+1}^{T-h+1}\left[Y_{j:j+h-1} - \hat{\text{p}}_{j:j+h-1}\right]^2$$

$$\text{(A.2)} \qquad \text{LS} = \frac{1}{\tau}\sum_{j=M+1}^{T-h+1} -\ln\left[|1 - Y_{j:j+h-1} - \hat{\text{p}}_{j:j+h-1}|\right]$$

$$\text{(A.3)} \qquad \text{DES} = \frac{1}{\tau}\sum_{j=M+1}^{T-h+1} \pi\mathbb{I}\left[\hat{\text{p}}_{j:j+h-1} > \pi\right](1 - Y_{j:j+h-1}) + (1-\pi)\mathbb{I}\left[\hat{\text{p}}_{j:j+h-1} \leq \pi\right]Y_{j:j+h-1}$$

where $|\cdot|$ refers to the absolute value, $\mathbb{I}[\,\cdot\,]$ is an indicator function, π equals the unconditional probability of $Y_{j:j+h-1} = 1$ in the evaluation period, and $\tau = T - M - h + 1$. QPS assigns values between 0 and 2 and can be seen as a counterpart to the mean squared forecast error, and LS ranges between 0 to ∞.

The out-of-sample pseudo R^2 and the area under the receiver operating characteristic (AUROC) curve are:

$$\text{(A.4)} \qquad \text{pseudo } R^2 = 1 - \left[\frac{\ln L_u}{\ln L_c}\right]^{\frac{-2\ln L_c}{\tau}},$$

$$\text{(A.5)} \qquad \text{AUROC} = \frac{1}{n_0 n_1}\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} A(\hat{\text{p}}_{i:i+h-1}, \hat{\text{p}}_{j:j+h-1}|Y_{j:j+h-1} > Y_{i:i+h-1}),$$

where $\ln L_u$ ($\ln L_c$) denotes the unconstrained (constrained) log likelihood with out-of-sample forecasts from the probit model (Estrella, 1998; Chen, Chou and Yen, 2016). The number of expansions (recessions) in the out-of-sample period is denoted by n_0 (n_1) and A assigns values similar to the Mann-Whitney U statistic (Bouallègue, Magnusson, Haiden and Richardson, 2019).[1]

[1] More precisely, we have $\ln L_u = \sum_{j=M+1}^{T-h+1} Y_{j:j+h-1}\ln\left[\hat{\text{p}}_{j:j+h-1}\right] + (1 - Y_{j:j+h-1})\ln\left[1 - \hat{\text{p}}_{j:j+h-1}\right]$ and $\ln L_c = \sum_{j=M+1}^{T-h+1} Y_{j:j+h-1}\ln\left[\bar{p}_{j:j+h-1}\right] + (1 - Y_{j:j+h-1})\ln\left[1 - \bar{p}_{j:j+h-1}\right]$, where $\bar{p}_{j:j+h-1}$ denotes recession forecasts from a model with only a constant. A assigns values as follows:

$$A(\hat{\text{p}}_{i:i+h-1}, \hat{\text{p}}_{j:j+h-1}|Y_{j:j+h-1} > Y_{i:i+h-1}) = \begin{cases} 0 & \text{if } \hat{\text{p}}_{j:j+h-1} < \hat{\text{p}}_{i:i+h-1}, Y_{j:j+h-1} > Y_{i:i+h-1}, \\ 0.5 & \text{if } \hat{\text{p}}_{j:j+h-1} = \hat{\text{p}}_{i:i+h-1}, Y_{j:j+h-1} > Y_{i:i+h-1}, \\ 1 & \text{if } \hat{\text{p}}_{j:j+h-1} > \hat{\text{p}}_{i:i+h-1}, Y_{j:j+h-1} > Y_{i:i+h-1}. \end{cases}$$

B. In-sample and out-of-sample forecasts - 150 years of data

We create a dataset with data from 1871:1 to 2022:12 for the term spread and the log equity premium. The risk-free rate from 1871:1 onward is taken from the Goyal-Welch dataset. Data prior to 1926:1 for the equity premium are taken from Robert Shiller's homepage. The stock returns are based on the S&P Comp. index (monthly average data); Column J: "Real Total Return Price" including dividends. Additionally, the long interest rate from 1871:1 to 2020:12 is taken from Robert Shiller's database ("Long Interest Rate GS10)". The first NBER-dated business-cycle peak goes back to June 1857.

In-sample regressions for log excess returns are only shown for data from 1926:1 onward. Data prior to this period are monthly average data. This is known to generate serial correlation in returns - which would distort both t-statistics and R^2 statistics (Working, 1960). The out-of-sample forecasts make use of all available data from 1871:1 onward.

Table B.1 provides the in-sample (top panel) and out-of-sample (bottom panel) forecast results for the probit model with TMS_t and MA-TMS_t as regressors as well as for the corresponding OLS model. Figure B.1 provides the corresponding cumulative squared forecast error (CSFE) differences of both models relative to the historical average return benchmark. The table and chart highlight that the recession probabilities based on the probit model with TMS_t and MA-TMS_t have strong predictive power for the equity premium even considering a very long sample covering more than 150 years of data.

TABLE B.1

In-sample and Out-of-sample statistics for log equity premium forecasts

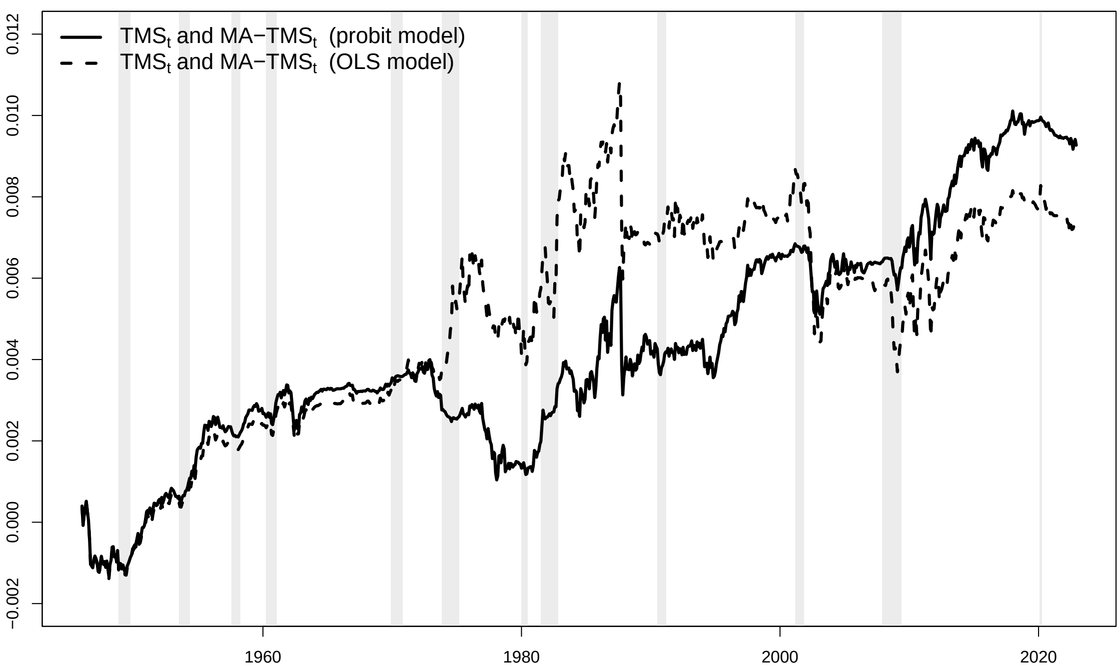
This table reports in-sample and out-of-sample results for the long sample. Panel A presents in-sample results for the slope coefficient from regressions of log excess returns ($r_{t+1:t+h}$) on recession probability forecasts; $r_{t+1:t+h} = \alpha + \hat{\mathrm{p}}_{t+1:t+h|t} + \epsilon_{t+1:t+h}$. The estimation period is 1926:1 to 2022:12. $\hat{\mathrm{p}}_{t+1:t+h|t}$ refers to recession probability forecasts from a probit model with TMS_t and MA-TMS_t. Additionally, results are shown for a linear regression model with TMS_t and MA-TMS_t (including an intercept) The t-statistics are given in parenthesis and the adjusted R^2 is shown below in %. Standard errors are Newey-West adjusted with a lag length of 12 months. Panel B shows R^2_{OS} statistics for the out-of-sample periods from 1946:1 to 2022:12. Results are reported for forecast horizons of $h = 1, 3, 6, 12$ months. $*$, $**$, $***$ denote significance at the 10%, 5%, and 1% significance levels.

(1) Variable	(2) $h = 1$	(3) $h = 3$	(4) $h = 6$	(5) $h = 12$
Panel A: In-sample results (1926:1 to 2022:12)				
Probit model				
$\hat{\mathrm{p}}_{t+1:t+h\|t}$ (TMS_t, MA-TMS_t)	-0.034**	-0.030**	-0.024***	-0.018***
t-stat.	(-2.392)	(-2.429)	(-2.429)	(-2.588)
R^2_{adj} (%)	0.57	1.59	2.80	4.80
OLS model				
TMS_t	-0.032	-0.049	-0.071	-0.004
t-stat.	(-0.157)	(-0.240)	(-0.365)	(-0.020)
MA-TMS_t	0.489*	0.510*	0.529*	0.487*
t-stat.	(1.724)	(1.788)	(1.862)	(1.758)
R^2_{adj} (%)	0.37	1.37	3.10	5.75
Panel B: Out-of-sample results (1946:1 to 2022:12)				
Probit model				
TMS_t, MA-TMS_t	0.56***	1.70***	2.83***	5.36***
OLS model				
TMS_t, MA-TMS_t	0.44**	1.11**	1.74***	2.90**

FIGURE B.1

Out-of-sample performance for the long sample

This figure presents the difference between cumulative squared forecast errors (CSFE) of the historical average and the CSFE of two different forecasting models for one-year ahead forecasts. The solid black line shows results for a model that predicts with recession probability forecasts that are based on a probit model with TMS_t and MA-TMS_t. The dashed line refers to forecasts from a linear regression model with TMS_t and MA-TMS_t. The out-of-sample period is 1946:1 to 2022:12.



C. Accounting for a structural break in the term spread

A. Evidence for a structural break in the mean of the term spread

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 e.g. 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 break is difficult to localize.[2] Other papers also document breaks in the dynamic relationship between the term spread and real growth in the U.S. 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. Here, 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 difficult to narrow down to an exact point in time - not least because of the infrequent occurrence of recessions (Wright, 2006). In contrast, as we will see below, the shift in the mean of the term spread can be identified precisely and in a more timely manner regardless of the state of economy.

Figure 2 in the main text plots the normalized term spread from 1951:3 to 2020:12. Eyeballing this time series shows that the mean of the term spread has shifted upwards in the early 1980s. This is most visible when comparing the period from 1965 to 1982 with the period from 1983 to 2020: the mean of the former period is -0.60 whereas it is 0.48 for the latter period. In what follows, we formally test the hypothesis of a structural break in the mean of the term spread.

The classical break test for coefficients in linear regression models goes back to Chow (1960). A critical limitation of the Chow-test is that the break date has to be known a priori. Here, we treat the break date as unknown and perform break tests over a grid of candidate values – namely on a fraction of the sample between $[\tau_1, \tau_2]$ with $\tau_1 = \pi_\tau T$ and $\tau_2 = (1 - \pi_\tau)T$. We refer to π_τ as the

[2] 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.

trimming value. When performing the Chow test on a sequence of dates the standard chi-square critical values are not applicable; for a discussion of this point see Hansen (2001). We estimate the following model for all z values between $[\tau_1, \tau_2]$:

$$\text{TMS}_t = \beta_1 \mathbf{I}\{t \leq z\} + \beta_2 \mathbf{I}\{t > z\} + \epsilon_t, \tag{A.6}$$

where $\mathbf{I}\{t \leq z\}$ ($\mathbf{I}\{t > z\}$) is an indicator function that equals one for $t \leq z$ ($t > z$). The coefficients β_1 and β_2 are re-estimated for a grid of z values and the SSE values are saved for each of these grid points. If there is no structural break in the coefficients the SSE values vary randomly over time. However, if there is a unique structural break, then the time series will have a well-defined global minimum near the true break date (Hansen, 2001).

The upper panel in Figure C.1 presents the SSE as a function of z with a trimming value of $\pi_\tau = 0.15$. The SSE is thus calculated for potential breaks from 1961:8 to 2010:6. The resulting SSE clearly shows a strong V-shape, indicating a well defined and unique break point. The break date corresponds to the month with the lowest sum of squared errors (Bai, 1997). This global minimum is in 1982:5. We formally test the null hypothesis of no structural break by using the Sup-F, Ave-F, and Exp-F statistics of Andrews (1993) and Andrews and Ploberger (1994). The variance-covariance matrix is estimated according to Newey and West (1994) and the p-values are computed following Hansen (1997). The null hypothesis of no structural break is rejected at the 1% significance level for Sup-F, Ave-F, and Exp-F and is robust to changes in the trimming value and pre-whitening of the residuals; details are provided in Online Appendices C and D. The middle panel in Figure C.1 presents the estimated sub-sample means in the term spread with full sample information. The upward shift in the mean is consistent with attenuated recession probabilities after the break that we have observed above.

The previous test results are based on full sample information. Would an investor have been able to identify the break in real-time? To answer this question we estimate the p-values with a recursively expanding sample from 1980:1 onwards and re-estimate the p-values each month until 2020:12. The lower panel in Figure C.1 presents the resulting series of p-values. The null hypothesis of no structural break is first rejected at the 10% critical value by the Sup-F, Ave-F, and Exp-F tests in 1986:7, 1987:3, and 1986:9. Since then, the p-values have consistently remained below 5%, providing strong evidence that the break in the mean of the term spread could have been identified in real-time as early as the mid 1980s. From 1995:1 to 2020:12 the null hypothesis is always rejected at the 1% level for each of the test statistics. The estimated break date is identical for the different test statistics as it is simply localized at the global minimum of the sum of squared errors

(Bai, 1997).[3] Overall, the results are in line with other evidence of structural breaks in the standard probit model and the estimated break date coincides with the great moderation (Kim and Nelson, 1999; McConnell and Perez-Quiros, 2000).[4]

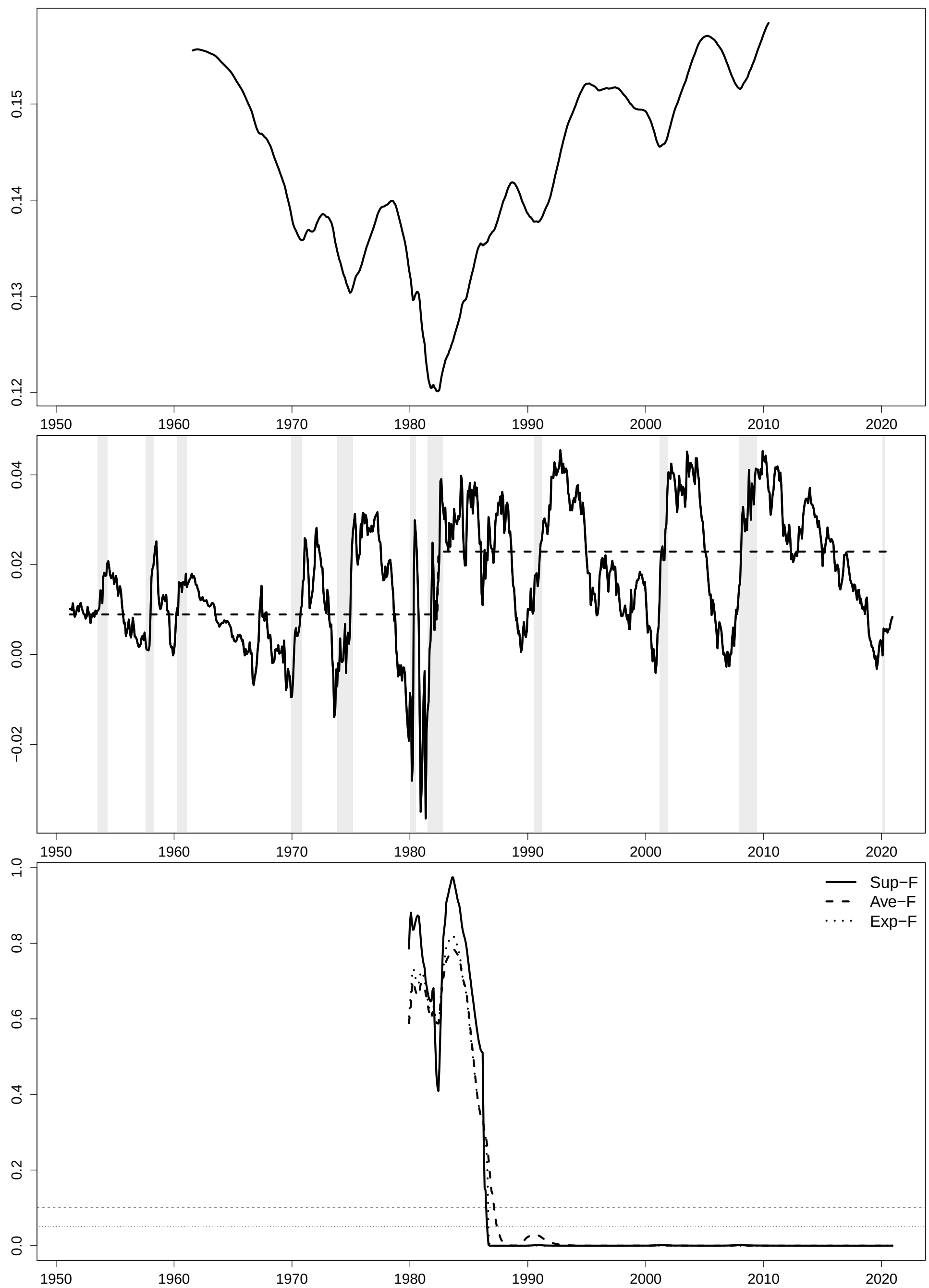
[3] Figure C.3 displays recursively estimated break dates from 1980:1 to 2020:12. This shows that the identified break dates are very stable around the full-sample break date 1982:5.

[4] We also test for multiple breaks in the mean of the term spread by applying the methods proposed in Bai and Perron (1998, 2003). The sequential tests do not provide evidence in favor of multiple breaks. Moreover, we apply the sequential tests to a longer sample that starts in 1933:4, just after the Great Depression. Interestingly, we find evidence for another break in the mean in 1947:6, in addition to the one in 1982:5. The additional break aligns well with the Treasury-Federal Reserve Accord, indicating a change in monetary policy after longer-term interest rates were pegged during wartime (Eichengreen and Garber, 1991; Carlson and Wheelock, 2014).

FIGURE C.1

SSE as a function of the break date and real-time detection of the break

This figure presents the sum of squared errors (SSE) when testing for a structural beak in the mean of the term spread (upper panel). The change point is allowed to lie between 1961:8 and 2010:6, which corresponds to a trimming value of 15%. The middle panel shows the term spread from 1951:3 to 2020:12 (solid line) and the sub-sample means from 1951:3 to 1982:5 and from 1982:6 to 2020:12 (dashed line). The estimated break date in 1982:5 corresponds to the global minimum SSE (Bai, 1997). The lower panel reports the recursively estimated p-values of the null hypothesis of no structural break (Hansen, 1997). Sup-F, Ave-F, and Exp-F refer to the Wald-type statistics of Andrews (1993) and Andrews and Ploberger (1994). The dashed (dotted) horizontal line shows the 10% (5%) level. The estimation of p-values is carried out from 1980:1 to 2020:12.



B. Predicting recessions in the presence of structural breaks

We have documented that the term spread has experienced a structural break in the mean in 1982. We now show that both the recession and equity premium forecasts improve considerably when adjusting for this break. To do so, we follow Lettau and Van Nieuwerburgh (2008). These authors find evidence in favor of multiple shifts in the mean of the dividend-price ratio. They correct for these breaks by creating a break-adjusted time series that simply subtracts the sub-sample means from the original predictor. They show that the adjusted time series has robust in-sample predictive power for the equity premium but fails to beat the historical average out-of-sample.[5]

We carry out the following steps: we first test for a break in the mean of the term spread by estimating the Sup-F statistic and by using a significance level of 10% and a trimming value of 15%. Only real-time information is used to mimic the situation of an investor. Second, if the null hypothesis of no break is rejected, we estimate the two sub-sample means and subtract them from the term spread to create a break-adjusted term spread, denoted by $\text{TMS}^{\text{break}}$. Finally, we estimate the probit model with this adjusted time series and generate out-of-sample forecasts for recession probabilities and the log equity premium. If the null hypothesis is not rejected, we predict with the unadjusted term spread.

Alternative approaches to forecasting in the presence of structural breaks have been proposed by Pesaran and Timmermann (2007). They present methods to determine the optimal estimation window when there are breaks. These are based on the insight that if the pre-break data follow a data generating process that is different from the one characterizing the post-break data, then the coefficient estimates are biased when using all data. We implement their proposed combination of forecasts from probit models estimated over different windows as follows. If the null hypothesis of no break is rejected, we estimate the probit model over an equally-spaced grid of starting values. This grid covers the beginning of our data set until the estimated break date. To reduce computing time, we only estimate models at annual increments in the starting date. This provides us with multiple recession forecasts and equity premium forecasts that only differ in the start date of the estimation window. Then, our pooled forecast is simply the average of the individual forecasts over the grid of start values, denoted as ”Pooling (average)”.

A disadvantage of the pooling approach is that one includes many forecasts with a large fraction

[5] Lettau and Van Nieuwerburgh (2008) argue that the break dates can be estimated in real-time but that the uncertainty about the shift in the mean prevents significant forecasting gains. Smith and Timmermann (2021) present a method that uses cross-sectional information and economically motivated priors to (i) better detect breaks in real-time and to (ii) estimate parameters more accurately. The latter point is especially relevant when only few post-break observations are available.

of pre-break data when the sample is long or when the break occurs relatively late in the sample. Alternatively, one can only choose the best performing grid point. We also implement this approach by performing a pseudo out-of-sample exercise over the most recent five years of data. Then we evaluate all start dates over this holdout period and select the start date that minimizes the MSFE for the equity premium. The selection of the grid point is chosen based on forecasts of the log equity premium and not on forecasts of the probit model as the former is the main purpose of this paper. Forecasts from this approach are denoted as ”Cross-validation”.

Finally, we consider one estimation strategy that only uses post-break data to estimate parameters of the probit model and to forecast recession probabilities and the equity premium. We denote this strategy as ”Post-break window”.[6] If the null hypothesis of no structural break is not rejected, then the forecasts of the break-correction methods are identical to forecasts by the unadjusted probit model. In what follows, we apply these break-adjustment methods to the probit model with TMS and MA-TMS as predictors.

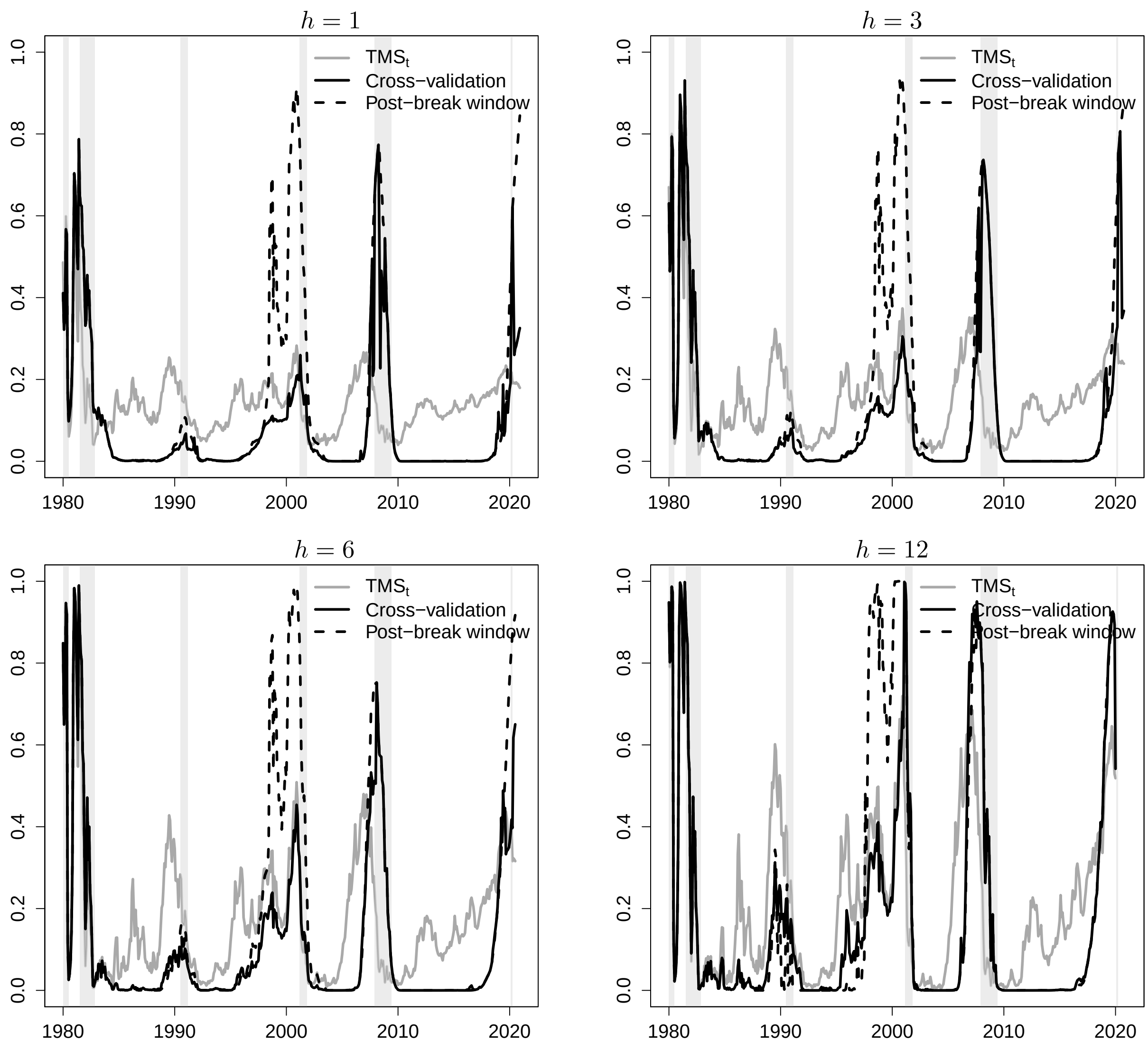
Figure C.2 shows recession probability forecasts for cross-validation and for post-break window, as well as for the standard probit model with the term spread. The estimated probabilities for cross-validation remain at fairly low levels during the 1990-91 and 2001 recessions for short-horizon forecasts. However, the probabilities for the 2008-09 recession increase substantially relative to the unadjusted full-sample models. They rise above 60% for cross-validation and are thus considerably higher than the probabilities in Figure 3 which are less than 30% for $h = 1$ and unadjusted models. Moreover, for $h = 12$ we see a substantially improved forecast performance with estimated probabilities as high as 90% for the 2001 and 2008-09 recessions. Post-break window generates out-of-sample probabilities as high as 90% for the 2001 recession, however, this approach also gives rise to a false positive in the late 1990s. Overall, we confirm previous findings that the implied recession probabilities are more pronounced when properly correcting for instabilities and breaks (Galvão, 2006; Chauvet and Potter, 2010). We provide forecast evaluation statistics for the break-correction methods in Online Appendix E.

[6] We use at least 15 years of data. Hence, while the sample of post-break data is shorter than 15 years we use the most recent 15 years of data for estimation and forecasting.

FIGURE C.2

Out-of-sample performance: (break-corrected) recession probabilities

This figure presents out-of-sample recession probability forecasts for four different forecast horizons and three different models. The forecasts denote the probability of a recession within the next $h = 1, 3, 6, 12$ months. The solid gray line shows forecasts from the standard probit model with the term spread as the only predictor (TMS_t). The solid black line depicts forecasts from a probit model with the term spread and the moving average component (MA-TMS_t), where the optimal estimation window is determined by cross-validation over a holdout period of 60 months. The dashed black line presents forecasts from a probit model with TMS and MA-TMS that only uses post-break data for coefficient estimation. Out-of-sample forecasts are recursively estimated for the sample from 1980:1 to 2020:12.



C. Recursively estimated break dates and optimal starting points for cross-validation

The Sup-F, Ave-F, and Exp-F statistics are estimated as:

$$\text{Sup-F} = \sup_{\tau_1 \leq \tau \leq \tau_2} F_T(\tau), \tag{A.7}$$

$$\text{Ave-F} = \frac{1}{\tau_2 - \tau_1 + 1} \sum_{\tau=\tau_1}^{\tau_2} F_T(\tau), \tag{A.8}$$

$$\text{Exp-F} = \ln \Big[\frac{1}{\tau_2 - \tau_1 + 1} \sum_{\tau=\tau_1}^{\tau_2} \exp F_T(\tau)\Big], \tag{A.9}$$

where $F_T(\tau)$ refers to the Wald statistic for testing $\hat{\beta}_2 = \hat{\beta}_1$ and $\tau = \tau_1, \ldots, \tau_2$.

The recursively estimated break dates for the Sup-F test are shown in the upper panel of Figure Figure C.3. The estimated location of the break is consistently between 1981 and 1983. The lower panel of Figure Figure C.3 presents the selected start values from cross-validation for one-year ahead log equity premium forecasts. At each point in time we estimate the probit model with different estimation windows, and evaluate forecasts from these models over a pseudo out-of-sample period consisting of the most recent five years of data. The optimal start value refers to the probit model with the smallest MSFE for the cumulative log equity premium with $h = 12$. It is important to mention that this analysis is feasible in real-time without any look-ahead bias in the data. Our most recently available observation at time t and $h = 12$ is $r_{t-11:t}$. The evaluation period runs from $r_{t-70:t-59}, \ldots, r_{t-11:t}$.

TABLE C.1

Estimated p-values with pre-whitening of standard errors

This table reports p-values for the null hypothesis of no structural break in the mean of the term spread. Ave-F, Exp-F, and Sup-F refer to the test statistics of Andrews (1993) and Andrews and Ploberger (1994), and p-values are estimated by Hansen (1997). The standard errors are Newey and West (1994) with an automatic bandwidth selection and different versions of pre-whitening. Results are shown for no pre-whitening of standard errors and for AR(1), AR(3), AR(6), and AR(12) pre-whitening of standard errors. The tests are carried out for trimming values of 15% (Panel A) and 5% (Panel B). The estimations are based on data from 1951:3 to 2020:12 and the break date refers to the global minimum in the sum of squared errors (Bai, 1997).

(1)	(2)	(3)	(4)	(5)
	Ave-F	Exp-F	Sup-F	break date
		Panel A: 15% trimming		
No pre-whitening	0.00	0.00	0.00	1982:5
AR(1)	0.07	0.05	0.02	
AR(3)	0.04	0.02	0.01	
AR(6)	0.01	0.01	0.02	
AR(12)	0.00	0.00	0.00	
		Panel B: 5% trimming		
No pre-whitening	0.00	0.00	0.00	1982:5
AR(1)	0.08	0.06	0.03	
AR(3)	0.04	0.03	0.02	
AR(6)	0.02	0.01	0.03	
AR(12)	0.00	0.00	0.00	

D. Break tests with pre-whitening of standard errors

Table C.1 provides p-values for the Ave-F, Exp-F, and Sup-F statistics when applying pre-whitening to the standard errors. The null of no structural break is rejected for no pre-whitening, as well as for AR(1), AR(3), AR(6), and AR(12) pre-whitening. Furthermore, the results are robust to the choice of trimming value.

E. Forecast evaluation: probit model with break-corrections

Table C.2 presents the forecast evaluation statistics for the probit model with TMS_t as well as $\frac{1}{36}\sum_{j=0}^{35}\text{TMS}_{t-j}$ as predictors, and for five break-corrected versions of this model. We can see that the break-adjusted model ($\text{TMS}_t^{\text{break}}$, $\frac{1}{36}\sum_{j=0}^{35}\text{TMS}_{t-j}^{\text{break}}$) does not improve the forecast performance. Contrarily, for $h = 1$ the other methods - cross-validation, pooling, and post-break window - have smaller (or identical) values for QPS, LS, and DES. We have seen in the main text that post-break window with at least 15 years of data in the probit model generates a false positive prior to 2000. This is canceled out when setting the minimum number of observations to 20 years. Thus, post-break window with 20 years performs even better for $h = 1, 3, 6, 12$ compared to post-break window with 15 years. Overall, cross-validation is most reliable in improving the performance relative to the unadjusted model; the statistics improve for each forecast horizon and each forecast evaluation statistic.

TABLE C.2

Out-of-sample forecasting performance - Probit model

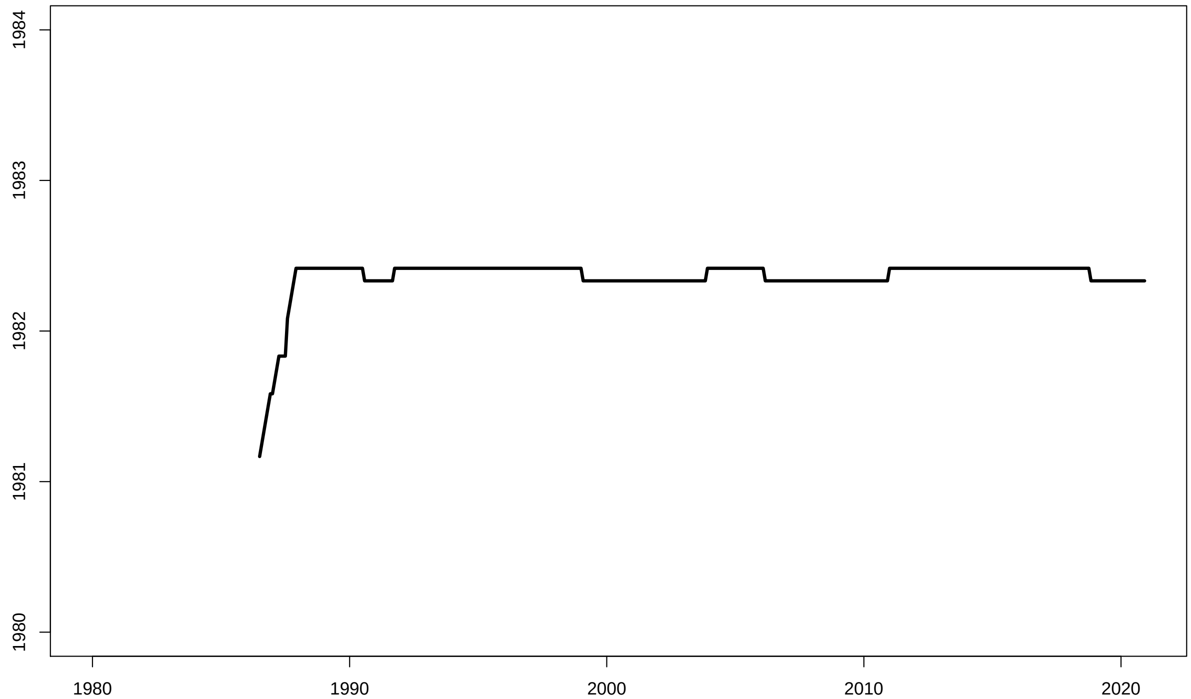
This table presents five forecast evaluation statistics for the out-of-sample recession probability forecasts, as well as the correlation between the probability forecasts and the (cumulative) log equity premium (ρ). The statistics are the quadratic probability score (QPS), logarithm score (LS), diagonal elementary score (DES), pseudo R^2, and the area under the receiver operating characteristic curve (AUROC). Results are shown for the probit model with TMS_t and $\frac{1}{36}\sum_{j=0}^{35}\text{TMS}_{t-j}$ as predictors, and for five break-corrected versions of this probit model (Pesaran and Timmermann, 2007; Lettau and Van Nieuwerburgh, 2008). 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.

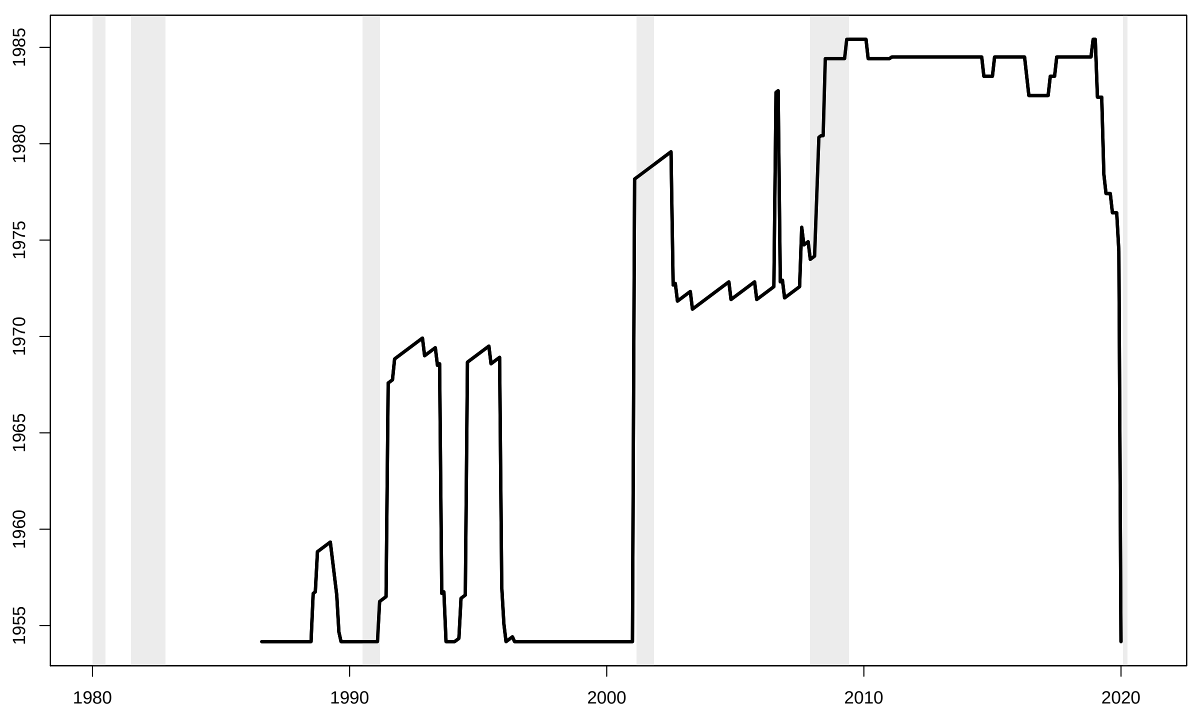
	1980:1 to 2020:12					
Variables in probit model	QPS	LS	DES	pseudo R^2	AUROC	ρ
Panel A: $\mathbf{h = 1}$						
TMS_t, MA-TMS_t	0.18	0.27	0.05	0.26	0.91	-0.12
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	0.22	0.34	0.06	0.13	0.83	-0.11
Cross-validation	0.15	0.23	0.04	0.35	0.93	-0.13
Pooling (average)	0.18	0.27	0.05	0.28	0.91	-0.13
Post-break window (15 years)	0.19	0.27	0.04	0.26	0.90	-0.14
Post-break window (20 years)	0.16	0.24	0.04	0.33	0.92	-0.14
Panel B: $\mathbf{h = 3}$						
TMS_t, MA-TMS_t	0.21	0.32	0.07	0.25	0.89	-0.20
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	0.25	0.39	0.07	0.11	0.81	-0.18
Cross-validation	0.18	0.27	0.04	0.35	0.93	-0.23
Pooling (average)	0.20	0.30	0.06	0.28	0.90	-0.22
Post-break window (15 years)	0.20	0.30	0.05	0.28	0.90	-0.25
Post-break window (20 years)	0.18	0.27	0.05	0.35	0.92	-0.25
Panel C: $\mathbf{h = 6}$						
TMS_t, MA-TMS_t	0.23	0.37	0.09	0.25	0.88	-0.23
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	0.29	0.45	0.08	0.10	0.80	-0.20
Cross-validation	0.20	0.32	0.07	0.36	0.92	-0.32
Pooling (average)	0.22	0.35	0.07	0.30	0.89	-0.26
Post-break window (15 years)	0.22	0.34	0.06	0.31	0.90	-0.30
Post-break window (20 years)	0.19	0.30	0.06	0.39	0.91	-0.32
Panel D: $\mathbf{h = 12}$						
TMS_t, MA-TMS_t	0.25	0.41	0.10	0.34	0.87	-0.36
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	0.33	0.51	0.11	0.15	0.81	-0.31
Cross-validation	0.21	0.34	0.08	0.47	0.91	-0.44
Pooling (average)	0.24	0.39	0.09	0.37	0.88	-0.40
Post-break window (15 years)	0.29	0.51	0.09	0.16	0.88	-0.41
Post-break window (20 years)	0.25	0.39	0.08	0.39	0.90	-0.43

FIGURE C.3

Recursively estimated break dates, optimal starting points for cross-validation

This figure presents the recursively estimated break dates for the Sup-F test (upper panel) and the optimal start values of the estimation window for cross-validation for one-year ahead equity premium forecasts (lower panel). The horizontal axis denotes the time of estimation and the vertical axis denotes the respective break date and start value for the probit model. As an example, from 1990 onward the estimated break date was consistently between 1982 and 1983. The optimal start value for cross-validation is estimated by performing a pseudo out-of-sample exercise over the most recent five years of data, whereby the data are recursively expanding. The selected start date refers to the probit model with the lowest MSFE for the equity premium over the holdout period.





D. Further Robustness Checks

A. Different lengths for moving average

Figure D.1 presents R^2_{OS} statistics (in %) for forecasts of the log equity premium. The predictor variable is the recession probability forecast from a probit model with TMS_t and $\frac{1}{l}\sum_{j=0}^{l-1}\text{TMS}_{t-j}$. We let the length of the moving average component, denoted by l, vary between values of 2 and 60. The panels on the left of Figure D.1 present the R^2_{OS} values for $h = 1, 6, 12$ and the panels on the right depict the Clark and West (2007) statistics. The vertical axis denotes the R^2_{OS} values and the MSFE-adjusted statistics and the horizontal axis denotes different values of l. We can see for $h = 1$ that the R^2_{OS} is above 0.50% for any backward-looking moving average between 30 to 60 months and that the MSFE-adjusted statistic is significant at the 5% level for these values. A shorter moving average of one year also has a R^2_{OS} above 0.50% and is significant at the 10%. The R^2_{OS} for one-month ahead forecasts is highest for moving averages between three to five years. For cumulative six- and twelve-month ahead forecasts the R^2_{OS} is always positive for moving averages between 2 to 60 months. A moving average between one to five years has statistically significant R^2_{OS} values above 2.50% (6%) for the six-month (twelve-month) horizon.

Table D.1 presents results for the break-correction methods when using values of $l = 12, 24, 48, 60$. The picture is similar to Figure D.1: the longer moving averages perform better for short horizon forecasts, whereas the exact choice of moving average is less important for long horizon forecasts. The R^2_{OS} values are positive for all break-correction methods and moving averages. Furthermore, the results are robust to weighted pooling (Pooling (weighted)) and to a minimum window length of 20 years for post-break window (Post-break window (20 years)).

Table D.2 presents the forecast evaluation statistics for the probit model with TMS_t and $\frac{1}{l}\sum_{j=0}^{l-1}\text{TMS}_{t-j}$ for values of $l = 12, 24, 36, 48, 60$. The statistics often improve further for $l = 12$ and $l = 24$ compared to $l = 36$ but worsen for $l = 48$ and $l = 60$. Nonetheless, the latter two values still provide superior values compared to the probit model without the moving average component. This is most salient for short-horizon forecasts and AUROC values.

FIGURE D.1

Equity premium forecasts with different values of l in the probit model

This figure presents R^2_{OS} values (in %) for the log equity premium relative to the historical average and MSFE-adjusted statistics. Forecasts are based on recession probability forecasts from a probit model with TMS_t and $\frac{1}{l}\sum_{j=0}^{l-1}\text{TMS}_{t-j}$, for l values between 2 and 60 months. The horizontal axis denotes the value of l and the vertical axis denotes the R^2_{OS} statistic (left panels) and the MSFE-adjusted statistic of Clark and West (2007) (right panels). Results are shown for out-of-sample forecasts from 1980:1 to 2020:12 and for three forecast horizons, namely $h = 1, 6, 12$. The black dashed line (dotted line) denotes the 1% (5%) critical value, and the dash-dotted line depicts the 10% critical value of the MSFE-adjusted statistic.

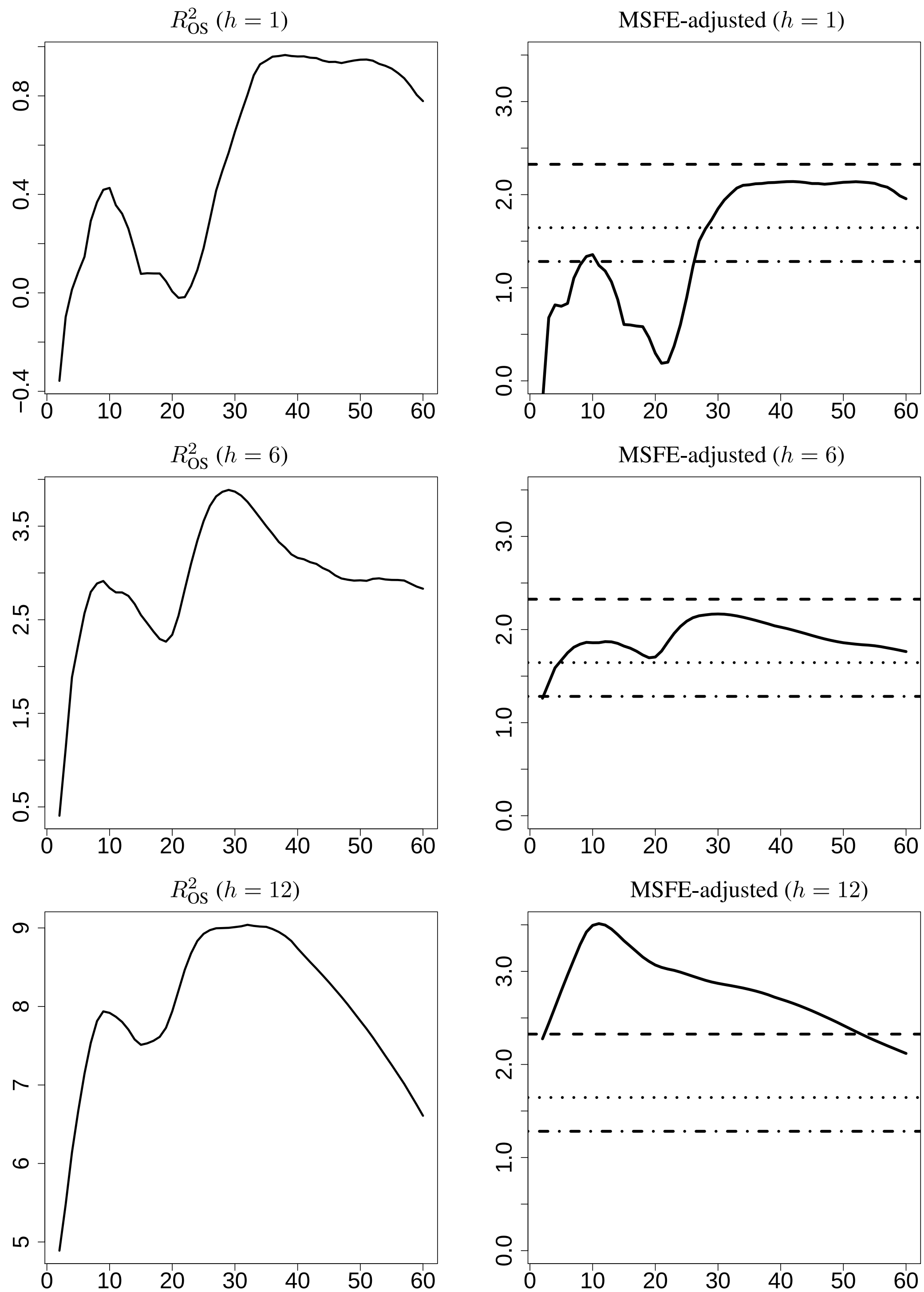


TABLE D.1

Out-of-sample R^2: Alternative lengths of moving average component

This table shows R^2_{OS} statistics (in %) for forecasts of the log equity premium relative to the historical average benchmark. The forecasts are based on a linear predictive regression model with a constant and recession probability forecasts as variables. Hereby, the probability forecasts are derived from a probit model with TMS_t and $\frac{1}{l}\sum_{j=0}^{l-1}\mathrm{TMS}_{t-j}$, and for different break-correction methods of this probit model. Results are shown for different lengths of the moving average component l, namely for $l = 12, 24, 48, 60$. The out-of-sample period is 1980:1-2020:12 and $h = 1, 3, 6, 12$ depicts the forecast horizon.

(1)	(2)	(3)	(4)	(5)
Variable	l=12	l=24	l=48	l=60
Panel A: $\mathbf{h = 1}$				
TMS_t, MA-TMS_t	0.32	0.09	0.94**	0.78**
$\mathrm{TMS}_t^{\mathrm{break}}$, MA-$\mathrm{TMS}_t^{\mathrm{break}}$	0.11	0.37	0.66*	0.23*
Cross-validation	-0.19	0.03	1.65**	0.55*
Pooling (average)	0.18	-0.04	1.06**	0.81**
Post-break window (15 years)	-0.15	-0.66	1.20**	0.65*
Panel B: $\mathbf{h = 3}$				
TMS_t, MA-TMS_t	1.36*	1.53**	2.36**	2.13**
$\mathrm{TMS}_t^{\mathrm{break}}$, MA-$\mathrm{TMS}_t^{\mathrm{break}}$	0.15*	2.03**	0.99**	0.04*
Cross-validation	1.50	0.84*	2.10**	1.30**
Pooling (average)	1.05*	1.42**	2.76***	2.33**
Post-break window (15 years)	0.62*	2.64**	3.39**	2.44**
Panel C: $\mathbf{h = 6}$				
TMS_t, MA-TMS_t	2.79**	3.34**	2.93**	2.83**
$\mathrm{TMS}_t^{\mathrm{break}}$, MA-$\mathrm{TMS}_t^{\mathrm{break}}$	0.49*	3.40**	-0.23**	-1.22*
Cross-validation	0.50*	4.23**	5.13***	2.11**
Pooling (average)	2.33**	3.27**	3.42**	3.00**
Post-break window (15 years)	1.72**	3.87**	3.32***	1.96**
Panel D: $\mathbf{h = 12}$				
TMS_t, MA-TMS_t	7.80***	8.83***	8.02***	6.61**
$\mathrm{TMS}_t^{\mathrm{break}}$, MA-$\mathrm{TMS}_t^{\mathrm{break}}$	6.23**	8.14***	5.54**	4.80**
Cross-validation	7.42***	8.60***	6.63***	4.56**
Pooling (average)	7.82***	9.89***	9.04***	7.29***
Post-break window (15 years)	8.10***	11.80***	9.77***	5.91***

TABLE D.2

OOS forecasting performance - Different moving average lengths

This table presents five forecast evaluation statistics for the out-of-sample performance of the probit model with the term spread and the moving average term spread as variables, 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). Results are shown for five different lengths of the moving average term spread, $\frac{1}{l}\sum_{j=0}^{l-1}\text{TMS}_{t-j}$, namely $l = 12, 24, 36, 48, 60$. 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.

	1980:1 to 2020:12					
Variables in probit model	QPS	LS	DES	pseudo R^2	AUROC	ρ
Panel A: $\mathbf{h = 1}$						
$l = 12$	0.15	0.31	0.05	0.19	0.87	-0.07
$l = 24$	0.16	0.24	0.02	0.33	0.92	-0.08
$l = 36$	0.18	0.27	0.05	0.26	0.91	-0.12
$l = 48$	0.20	0.32	0.07	0.16	0.87	-0.11
$l = 60$	0.21	0.37	0.08	0.05	0.78	-0.10
Panel B: $\mathbf{h = 3}$						
$l = 12$	0.17	0.33	0.05	0.21	0.86	-0.14
$l = 24$	0.18	0.27	0.03	0.34	0.92	-0.17
$l = 36$	0.21	0.32	0.07	0.25	0.89	-0.20
$l = 48$	0.22	0.37	0.09	0.15	0.85	-0.18
$l = 60$	0.24	0.41	0.09	0.05	0.77	-0.17
Panel C: $\mathbf{h = 6}$						
$l = 12$	0.18	0.36	0.06	0.28	0.87	-0.20
$l = 24$	0.20	0.31	0.05	0.37	0.91	-0.23
$l = 36$	0.23	0.37	0.09	0.25	0.88	-0.23
$l = 48$	0.25	0.41	0.09	0.16	0.84	-0.20
$l = 60$	0.26	0.46	0.09	0.08	0.76	-0.19
Panel D: $\mathbf{h = 12}$						
$l = 12$	0.18	0.35	0.06	0.45	0.90	-0.36
$l = 24$	0.22	0.35	0.07	0.45	0.91	-0.38
$l = 36$	0.25	0.41	0.10	0.34	0.87	-0.36
$l = 48$	0.24	0.44	0.10	0.30	0.85	-0.32
$l = 60$	0.28	0.47	0.11	0.23	0.80	-0.28

TABLE D.3

Out-of-sample R^2 statistics for log raw returns

This table presents R^2_{OS} statistics (in %) for forecasts of log raw returns relative to the historical average benchmark. In contrast to the main text, we do not subtract the short rate from the continuously compounded returns on the S&P 500 index. The forecasts are based on a linear predictive regression model with a constant and recession probability forecasts as variables. Hereby, the probability forecasts are derived from a probit model with TMS_t and $\frac{1}{36}\sum_{j=0}^{35}\text{TMS}_{t-j}$, and for four different break-correction methods of this probit model. Results are shown for four different out-of-sample periods, and $h = 1, 3, 6, 12$ depicts the forecast horizon. $*$, $**$, $***$ denote significance at the 10%, 5%, and 1% significance levels according to the Clark and West (2007) MSFE-adjusted statistic. "Short interest" and "Gold-to-platinum ratio" refer to the predictors of Rapach et al. (2016) and Huang and Kilic (2019).

(1)	(2)	(3)	(4)	(5)
Variable	1980:1-2020:12	1980:1-1999:12	2000:1-2020:12	1990:1-2013:12
Panel A: $\mathbf{h = 1}$				
TMS_t, MA-TMS_t	0.28	0.60	-0.03	0.28
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	0.58$*$	-0.31	1.42$*$	1.80$**$
Cross-validation	1.60$**$	0.63	2.50$*$	3.30$**$
Pooling (average)	0.40$**$	0.51	0.30	0.80$**$
Post-break window	1.08$**$	0.83$*$	1.32$*$	4.51$***$
Short interest				1.19$**$
Gold-to-platinum ratio				1.56$**$
Panel C: $\mathbf{h = 6}$				
TMS_t, MA-TMS_t	0.50$*$	2.53$*$	-0.95	0.79
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	0.49$*$	-5.75	4.92$**$	4.25$**$
Cross-validation	4.96$*$	1.28	7.52	8.34$*$
Pooling (average)	1.17$*$	2.20$*$	0.42	2.40$*$
Post-break window	3.76$**$	4.04$**$	3.57$*$	13.85$**$
Short interest				6.93$**$
Gold-to-platinum ratio				11.93$***$
Panel D: $\mathbf{h = 12}$				
TMS_t, MA-TMS_t	3.41$**$	6.77$**$	1.48	3.18$*$
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	4.03$**$	-6.29	10.24$***$	5.05$*$
Cross-validation	0.44$*$	4.99$*$	-2.06	-0.10
Pooling (average)	4.29$**$	6.46$*$	3.10$*$	4.47$*$
Post-break window	7.40$**$	12.57$***$	4.76$*$	14.50$*$
Short interest				4.26$*$
Gold-to-platinum ratio				15.82$***$

B. Forecasting log raw returns

Table D.3 presents the R^2_{OS} statistics when forecasting log raw returns instead of the log equity premium (without subtracting the short rate). The results show that recession probabilities have strong predictive power for equity market returns not only in excess of the risk-free rate.

TABLE D.4

Asset allocation exercise with proportional transaction costs

This table 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. The gains are corrected for proportional transaction costs of 50 basis points per transaction. Results are shown for one 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 out-of-sample period is 1980:1 to 2020:12. $*$, $**$, $***$ indicate significantly improved performance relative to the prevailing mean benchmark at the 10%, 5%, and 1% significance level. The p-values are obtained by using a bootstrap approach similar to DeMiguel et al. (2013) with the average block length set to three months (Politis and Romano, 1994).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: 1980:1 to 2020:12							
	ΔCER				ΔSR			
γ	3	5	3	3	3	5	3	3
ω	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$
Gains relative to prevailing mean:								
TMS_t, MA-TMS_t	2.42$**$	1.30$*$	2.15$**$	1.26$***$	0.14$***$	0.12$**$	0.12$***$	0.09$***$
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	3.54$*$	2.13$*$	3.16$*$	1.59	0.26$*$	0.24$*$	0.24$**$	0.21$**$
Cross-validation	3.72$**$	2.35$**$	3.06$**$	2.13$**$	0.21$**$	0.21$**$	0.17$**$	0.17$**$
Pooling (average)	2.88$**$	1.76$**$	2.63$**$	1.55$**$	0.17$**$	0.17$**$	0.16$**$	0.13$**$
Post-break window	5.37$***$	3.55$**$	4.55$***$	2.70$**$	0.31$***$	0.30$**$	0.27$***$	0.24$***$
Buy-and-hold	1.14$**$	0.45	1.14$**$	0.48$**$	0.08$**$	0.08$**$	0.08$**$	0.03$*$

C. Transaction costs

Table D.4 shows the gains in certainty equivalent return and Sharpe ratio relative to the prevailing mean when correcting for proportional transaction costs of 50 basis points per transaction. This is similar to the base case in Balduzzi and Lynch (1999), which assumes proportional costs rather than fixed costs per transaction. This specification has often been used in related articles to account for trading costs; see, for example, Neely et al. (2014); Jiang et al. (2019). The results show that even when proportional trading costs are taken into account, asset allocation based on equity premium forecasts using recession probabilities provides investors with an economically and statistically significant advantage relative to forecasts based on the historical average.

TABLE D.5

Predictive regressions for stock market return components - annual data

This table 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 annual data from 1980 to 2020. The t-statistics for the slope coefficients are shown in the brackets below and the standard errors are HAC-robust (Andrews, 1991). $*$, $**$, $***$ denote significance at the 10%, 5%, and 1% significance levels.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VAR variables	$\hat{\beta}^{E}$	$\hat{\beta}^{CF}$	$\hat{\beta}^{DR}$	VAR variables	$\hat{\beta}^{E}$	$\hat{\beta}^{CF}$	$\hat{\beta}^{DR}$
r, DP	-0.03	-0.07$**$	0.12	r, DP, LTY	-0.04	-0.07$*$	0.11
	[-0.71]	[-2.04]	[1.08]		[-1.01]	[-1.80]	[0.91]
r, DP, DY	-0.03	-0.08$**$	0.11	r, DP, LTR	-0.02	-0.07$**$	0.12
	[-0.90]	[-2.16]	[0.87]		[-0.65]	[-2.19]	[1.09]
r, DP, EP	-0.02	-0.17$**$	0.02	r, DP, TMS	-0.07$**$	-0.05	0.10
	[-0.61]	[-2.27]	[0.20]		[-2.16]	[-1.38]	[0.83]
r, DP, DE	-0.02	-0.17$**$	0.02	r, DP, DFY	-0.04	-0.10$**$	0.08
	[-0.61]	[-2.27]	[0.20]		[-1.24]	[-2.16]	[0.65]
r, DP, SVAR	-0.02	-0.10$*$	0.09	r, DP, DFR	-0.05	-0.06$**$	0.12
	[-0.63]	[-1.90]	[0.73]		[-1.13]	[-1.97]	[1.01]
r, DP, BM	-0.04	-0.08$*$	0.10	r, DP, INFL	-0.13$***$	-0.08$**$	0.02
	[-1.30]	[-1.76]	[0.98]		[-3.15]	[-2.30]	[0.15]
r, DP, NTIS	-0.04	-0.07$**$	0.11	r, DP, PC	-0.12$**$	-0.11$**$	-0.00
	[-0.85]	[-2.04]	[0.99]		[-2.56]	[-2.13]	[-0.04]
r, DP, TBL	-0.10$***$	-0.02	0.10				
	[-2.71]	[-0.49]	[0.80]				

D. VAR(1) decomposition for annual data

The VAR(1) decomposition is applied to non-overlapping annual data from 1980 to 2020.[7] Model-implied recession probabilities are non-overlapping one-year ahead forecasts ($h = 12$) from cross-validation.

E. Asset allocation exercise for lower frequency re-balancing

D.6 shows the gains in certainty equivalent return and Sharpe ratio when an investor only re-balances the portfolio at the same frequency as the forecast horizons of $h = 3, 6, 12$ months. Hence, non-overlapping forecasts are used in this exercise; see Rapach et al. (2016) for details.

[7]For annual data ρ equals $\frac{1}{1+exp(\overline{d_t - p_t})}$, where $\overline{d_t - p_t}$ is the mean of the annual log dividend-price ratio.

TABLE D.6

Asset allocation exercise with lower-frequency re-balancing

This table 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 forecast horizons of $h = 3, 6, 12$ months of the equity premium, where an investor re-balances at the same frequency as the forecast horizon (Rapach et al., 2016). The coefficient of relative risk aversion is denoted by γ, and ω states restrictions on the weights in the risky asset. The "Prevailing mean" shows the CER and SR values, whereas all other values denote the improvements relative to this benchmark. Results are shown for the out-of-sample period from 1980:1 to 2020:12.

Panel A: $h = 3$	ΔCER				ΔSR			
γ	3	5	3	3	3	5	3	3
ω	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$
Prevailing mean	7.12	5.31	7.12	8.13	0.43	0.40	0.43	0.49
Gains relative to prevailing mean:								
TMS_t, MA-TMS_t	2.61	1.32	2.45	1.38	0.15	0.15	0.14	0.09
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	2.79	2.07	3.15	1.85	0.17	0.18	0.19	0.19
Cross-validation	3.89	2.71	3.49	2.24	0.21	0.23	0.19	0.16
Pooling (average)	2.99	2.01	3.03	1.73	0.17	0.18	0.17	0.13
Post-break window	5.70	3.90	5.10	3.10	0.31	0.32	0.28	0.25
Buy-and-hold	1.70	0.88	1.70	0.69	0.09	0.13	0.09	0.04

Panel B: $h = 6$	ΔCER				ΔSR			
γ	3	5	3	3	3	5	3	3
ω	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$
Prevailing mean	8.05	6.14	8.05	8.50	0.48	0.45	0.48	0.52
Gains relative to prevailing mean:								
TMS_t, MA-TMS_t	2.21	1.30	2.01	1.31	0.13	0.14	0.12	0.08
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	1.14	0.77	1.46	0.85	0.08	0.09	0.10	0.09
Cross-validation	3.53	2.29	3.32	2.00	0.20	0.21	0.19	0.16
Pooling (average)	2.30	1.39	2.44	1.47	0.14	0.15	0.15	0.11
Post-break window	4.62	2.80	4.25	2.14	0.27	0.25	0.25	0.17
Buy-and-hold	1.21	0.82	1.21	0.76	0.08	0.11	0.08	0.04

Panel C: $h = 12$	ΔCER				ΔSR			
γ	3	5	3	3	3	5	3	3
ω	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$	$[-0.5, 1.5]$	$[-0.5, 1.5]$	$[0, 1.5]$	$[0, 1]$
Prevailing mean	7.92	5.85	7.92	8.30	0.48	0.44	0.48	0.50
Gains relative to prevailing mean:								
TMS_t, MA-TMS_t	2.63	1.01	2.75	1.93	0.15	0.11	0.15	0.13
$\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$	0.28	0.33	0.98	0.68	0.02	0.03	0.06	0.08
Cross-validation	2.16	0.72	2.27	1.69	0.12	0.08	0.12	0.13
Pooling (average)	1.88	0.89	2.22	1.83	0.11	0.09	0.13	0.14
Post-break window	1.80	0.02	2.21	1.64	0.10	0.04	0.12	0.14
Buy-and-hold	1.32	0.74	1.32	0.94	0.07	0.11	0.07	0.05

E. Comparison with Gómez-Cram (2022)

This section compares our results with those obtained by predicting recessions and the equity premium with the "common growth component" of Gómez-Cram (2022).[8] This component is estimated as a latent variable driving the joint comovement in various real-time business cycle indicators, see Gómez-Cram (2022) for details. With the CG component available starting in 1965, we restrict our OOS comparison to the period 1980:01-2019:12.

To enhance comparability with our baseline results using the term spread and moving average term spread, we construct recession probabilities based on both the CG component as well as its one-year backward-looking moving average (labeled CG-MA). These are provided in Figure E.1 below.

Table E.1 provides a comparison of the out-of-sample predictive ability of CG and CG-MA with the term spread and moving average term spread over the period from 1980:01-2019:12. The results in the first five columns highlight that both the GC component and its moving average predict recessions about as well as TMS and TMS-MA. That said, the last column shows that the correlation of the CG-implied recession probabilities with the equity premium is considerably less strong than that of the TMS-implied recession probabilities.

Tables E.2 and E.3 document the in-sample and out-of-sample predictability of the excess equity market return based on recession probabilities implied by CG and CG-MA, respectively. The tables show that although the common growth component and its moving average predict recessions about as well as the term spread and moving average term spread, the resulting recession probabilities do not forecast the equity premium well. We conjecture that this is due to the fact that the CG-implied probabilities spike towards the end of recessions while those based on the term spread towards the beginning of recessions which is when the equity market tends to perform the worst.

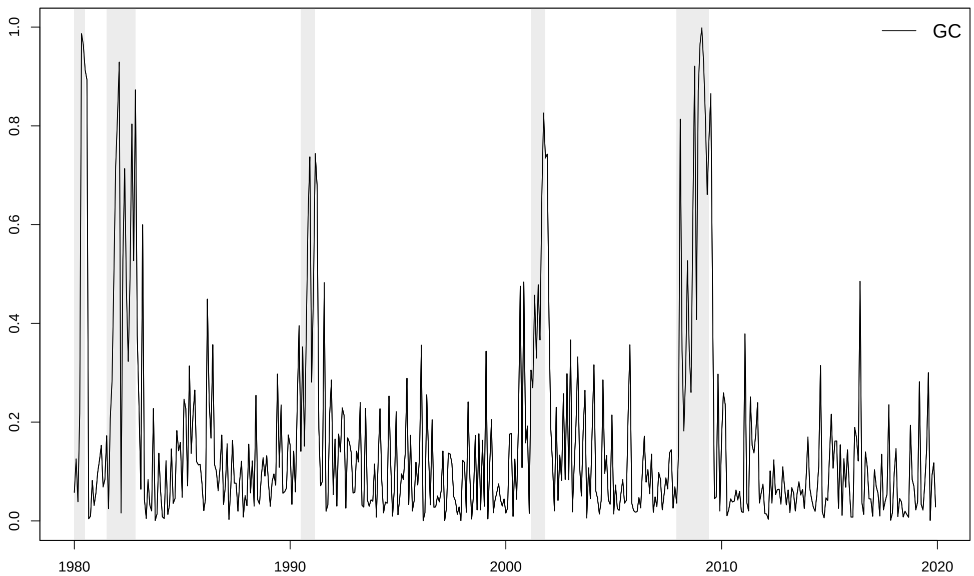
---

[8] We obtain the estimated common growth component from the replication files provided by Roberto Gomez-Cram on the website of the *Journal of Finance*.

FIGURE E.1

Recession probability forecasts based on Gómez-Cram (2022)

The out-of-sample period is 1980:1 to 2019:12 and shaded areas indicate NBER-dated recession periods. The forecast horizon is $h = 1$ month. CG-MA denotes the one-year backward-looking moving average of the common growth component estimated in Gómez-Cram (2022).



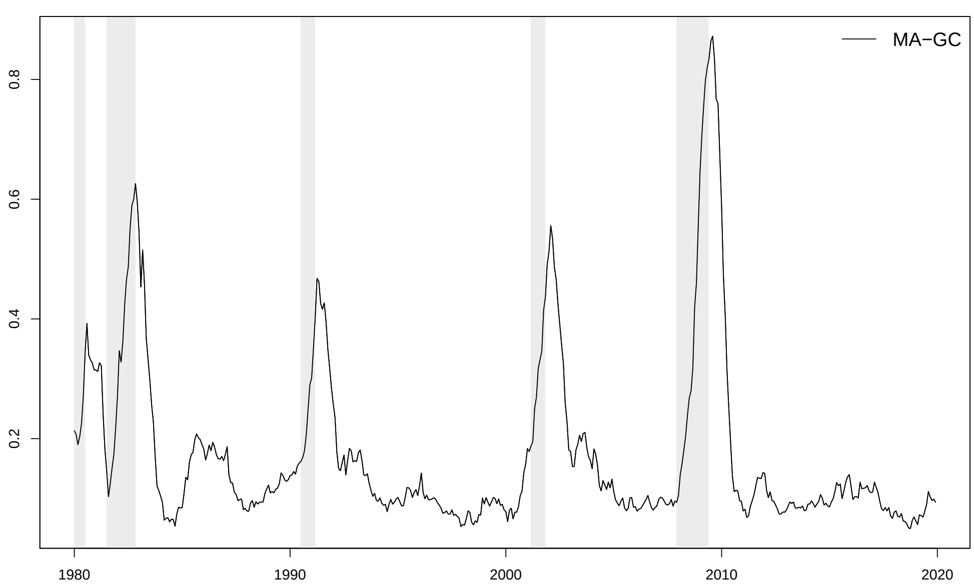


TABLE E.1

Out-of-sample recession prediction performance: Forecast evaluation

This table presents five forecast evaluation statistics for the out-of-sample performance of different probit models based on the term spread as well as the common growth component of Gómez-Cram (2022). The last column provides 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). "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 2019:12.

Variables in probit model	QPS	LS	DES	pseudo R^2	AUROC	ρ
Panel A: $\mathbf{h = 1}$						
GC_t	0.13	0.24	0.05	0.33	0.89	-0.06
MA-GC_t	0.19	0.32	0.05	0.17	0.87	0.00
TMS_t	0.23	0.40	0.12	-0.01	0.47	-0.04
TMS_t, MA-TMS_t	0.17	0.27	0.05	0.27	0.92	-0.14
Historical average	0.23	0.39	0.11	0.00	0.43	0.05
Panel B: $\mathbf{h = 3}$						
GC_t	0.16	0.29	0.07	0.29	0.85	-0.04
MA-GC_t	0.22	0.36	0.07	0.15	0.84	0.01
TMS_t	0.25	0.43	0.12	-0.01	0.54	-0.08
TMS_t, MA-TMS_t	0.20	0.31	0.07	0.25	0.90	-0.22
Historical average	0.26	0.43	0.12	0.00	0.43	0.09
Panel C: $\mathbf{h = 6}$						
GC_t	0.21	0.35	0.09	0.26	0.82	-0.03
MA-GC_t	0.26	0.42	0.11	0.13	0.80	0.00
TMS_t	0.27	0.46	0.12	0.03	0.63	-0.11
TMS_t, MA-TMS_t	0.22	0.36	0.09	0.25	0.87	-0.25
Historical average	0.30	0.48	0.14	0.00	0.43	0.11
Panel D: $\mathbf{h = 12}$						
GC_t	0.30	0.47	0.13	0.20	0.75	0.00
MA-GC_t	0.36	0.54	0.17	0.07	0.69	-0.01
TMS_t	0.29	0.48	0.12	0.18	0.73	-0.24
TMS_t, MA-TMS_t	0.25	0.41	0.10	0.32	0.86	-0.37
Historical average	0.38	0.57	0.17	0.00	0.47	0.12

TABLE E.2

In-sample equity premium forecasts

This table reports in-sample results for the slope coefficient from regressions of log excess returns ($r_{t+1:t+h}$) on recession probability forecasts; $r_{t+1:t+h} = \alpha + \beta \hat{\text{rp}}_{t+1:t+h} + \epsilon_{t+1:t+h}$. $\hat{\text{rp}}_{t+1:t+h}$ refers to recession probability forecasts from probit models with different predictive variables. The t-statistics are given in parenthesis and the adjusted R^2 is shown below in %. Standard errors are Newey-West adjusted with a lag length of 12 months. Results are reported for forecast horizons of $h = 1, 3, 6, 12$ months. $*$, $**$, $***$ denote significance at the 10%, 5%, and 1% significance levels.

(1)	(2)	(3)	(4)	(5)
Variable	$h=1$	$h=3$	$h=6$	$h=12$
GC	-0.002	0.004	0.004	0.006
t-stat.	(-0.158)	(0.432)	(0.565)	(0.997)
R^2_{adj} (%)	-0.001	-0.000	0.001	0.006
MA-GC	0.015	0.015	0.015	0.016
t-stat.	(1.125)	(1.233)	(1.275)	(1.363)
R^2_{adj} (%)	0.001	0.006	0.011	0.016

TABLE E.3

Out-of-sample equity premium forecasts

This table 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 one predictor variable. *, **, 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 smaller MSFE than the historical average benchmark. The out-of-sample period runs from 1980:1 to 2019:12.

(1)	(2)	(3)	(4)	(5)
Variable	$h=1$	$h=3$	$h=6$	$h=12$
GC_t (probit model)	-0.80	-2.82	-3.78	-4.13
GC_t (OLS model)	-0.43	-1.81	-2.93	-3.57
MA-GC_t (probit model)	-1.01	-3.02	-5.82	-9.49
MA-GC_t (OLS model)	-0.85	-2.81	-5.68	-9.55

F. Additional Results

A. Forecasting characteristics portfolios

We extend our analysis to a rich set of equity portfolios. Specifically, we follow Huang, Jiang, Tu and Zhou (2015) and analyze the predictive power of recession probability forecasts for portfolios sorted on different characteristics. We focus on 10 industry portfolios, 10 momentum portfolios, 10 size portfolios, and 10 book-to-market portfolios, all obtained from Kenneth French's homepage.[9] We predict the log excess return for these portfolios with the same recession probability forecasts as in the previous sections. Results are shown in Table F.1. We find that durable, manufacturing, energy, technology, and telecom portfolios are most predictable, whereas health, utility and nondurables are not significantly predictable. This is intuitive as the former sectors are more exposed to business cycle variation.[10] Interestingly, all of the momentum portfolios – independently of the break-correction method – have positive and significant R^2_{OS} statistics. Additionally, there is a tendency toward higher predictive power for portfolios formed on high market equity (large stocks) and low book-to-market ratios (growth stocks).

---

[9] The returns are value-weighted and include dividends.

[10] It is also consistent with Da et al. (2017), who find that electricity usage better forecasts excess returns of capital-intensive producers that are more exposed to fluctuations in the business cycle and have higher operating leverage.

TABLE F.1

Forecasting characteristics portfolios with recession probability forecasts

This table presents R^2_{OS} statistics (in %) for one month ahead equity premium forecasts of 10 industry portfolios (Panel A), 10 momentum portfolios (Panel B), 10 size portfolios (Panel C), and 10 book-to-market portfolios (Panel D). Forecasts are derived from the linear predictive regression model with recession probability forecasts as a predictor variable. These recession probabilities are estimated by a probit model with the term spread (TMS) and the backward-looking three-year moving average of the term spread (MA-TMS) - results are shown in column (2) - as well as for four break-correction methods - columns (3) to (6). These correction methods include cross-validation, pooling, post-break window, and forecasts from a break-adjusted term spread series (Pesaran and Timmermann, 2007; Lettau and Van Nieuwerburgh, 2008). The out-of-sample period is 1980:1 to 2020:12. $*$, $**$, $***$ denote significance at the 10%, 5%, and 1% significance levels according to the Clark and West (2007) MSFE-adjusted statistic.

(1) Portfolio	(2) TMS_t, MA-TMS_t	(3) Cross-validation	(4) Pooling (average)	(5) Post-break window	(6) $\text{TMS}_t^{\text{break}}$, MA-$\text{TMS}_t^{\text{break}}$
Panel A: Industry portfolios					
Nondurable	-0.27	-0.17	-0.33	-1.11	-0.20
Durable	0.12	0.89*	0.00	-0.74	-0.13
Manufacture	0.93**	1.13**	0.92**	1.02**	0.64*
Energy	0.45*	-0.17	0.56*	0.12	0.63*
Technology	0.62**	0.75**	0.60**	-0.53	0.49*
Telecom	0.12	0.31*	0.40*	1.02**	0.63*
Shop	-0.59	0.13	-0.61	-0.60	-0.54
Health	-0.06	0.11	-0.13	-0.87	-0.14
Utility	-0.01	-0.35	0.06	-0.50	-0.01
Other	0.41*	2.22**	0.62**	1.13*	0.81*
Panel B: Momentum portfolios					
Loser	0.09	1.28*	0.31	0.55	0.53
2	0.11	0.96	0.11	-0.27	-0.16
3	0.15	1.03*	0.34	0.71*	0.21
4	0.42**	1.31**	0.37*	-0.15	0.07
5	0.72**	0.91**	0.75**	0.71*	0.40*
6	0.81**	1.61**	0.99**	1.16**	0.80**
7	0.75*	0.78**	0.86*	1.04**	0.71**
8	1.03**	1.28**	1.10**	1.16**	0.77**
9	0.58*	2.11**	0.87*	1.60**	1.28**
Winner	0.65*	0.89*	0.81*	0.22	0.60*
Panel C: Size portfolios					
Small	0.23	0.82*	0.25	-0.56	0.57*
2	0.13	0.12	0.10	-0.91	0.18
3	0.19	0.63*	0.22	-0.20	0.52*
4	0.07	0.05	0.08	-0.28	0.37
5	0.14	0.44	0.14	-0.10	0.22
6	0.24	0.45	0.28	0.13	0.48
7	0.41*	1.07*	0.44*	0.46	0.49
8	0.25	1.16**	0.26	0.07	0.24
9	0.80**	1.66**	0.86**	0.82*	0.68*
Large	1.12**	1.75***	1.25***	1.43**	0.80**
Panel D: Book-to-market portfolios					
Growth	1.18**	1.52**	1.18**	0.67*	0.53*
2	0.65**	0.50*	0.57*	0.20*	0.11
3	0.88**	1.61**	0.88**	0.90**	0.45*
4	0.87**	1.38**	0.89**	0.79**	0.62**
5	1.14**	1.66**	1.33**	1.17**	1.11**
6	0.41*	1.12**	0.46*	0.15	0.59*
7	0.36*	2.68**	0.76**	2.45**	1.71**
8	-0.30	0.75	-0.04	-0.03	0.16
9	-0.22	0.44	-0.22	-0.77	0.04
Value	0.04	0.70	0.05	0.09	0.49

B. International evidence

We apply our model to four additional countries: Germany, France, Canada, and the United Kingdom. We obtain data for long-term interest rates and short-term interest rates from the OECD.[11] The long-term interest rates are for ten-year government bonds, the short-term interest rates are "based on three-month money market rates where available". Data are available from 1970:1 to 2020:12.[12] The term spread is the difference between the long-term and short-term interest rates. Recession data are from the Economic Cycle Research Institute. We use end-of-month country indices for Germany, France, Canada, and the United Kingdom from the MSCI database to calculate monthly log excess return series.

Figure F.1 shows the evolution of the cumulative one-year ahead log equity premium ($\sum_{j=0}^{11} r_{t+j}$) for the four countries around recessions. We see a clear V-shape pattern similar to the one for the U.S. The solid gray lines depict the individual recessions, which show pronounced variation both in terms of magnitude and timing. Next, we perform a pseudo out-of-sample exercise for forecasting the log equity premium from 1990:1 to 2020:12. In line with the U.S. results, we compare the performance of three models: (1) TMS_t, (2) TMS_t and TMS_{t-6}, (3) TMS_t and MA-TMS_t. Table F.2 shows the R^2_{OS} statistics for the log equity premium forecasts relative to a country-specific historical average. The values for Germany, France, and Canada are positive and statistically significant for almost all horizons when forecasting with the recession probabilities derived from the probit model with only the term spread. The R^2_{OS} statistics for the U.K. are positive only for $h = 6$ and $h = 12$. In contrast to the U.S., we do not see significant improvements by adding lagged and averaged term spread information to the probit model. The reason is that the recession probabilities for models (2) and (3) do not predict the beginning of recessions better than the simple model (1). Overall, however, the international data provide additional support for our main finding that recession

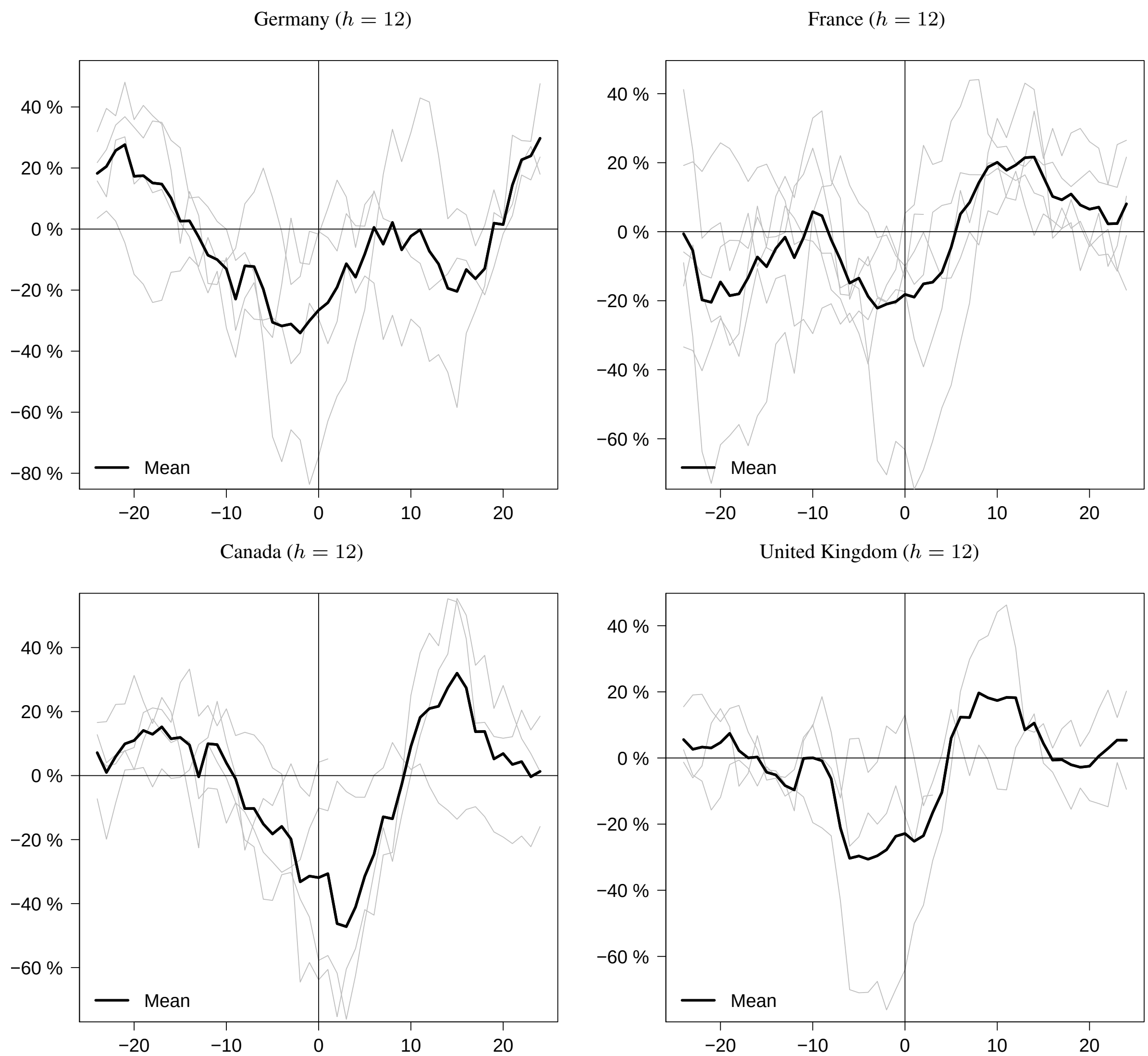
[11] `https://data.oecd.org/interest/long-term-interest-rates.htm`

[12] For the UK short-term interest rate we use the three month Treasury yield series until 2017:6 (`https://fred.stlouisfed.org/series/IR3TTS01GBM156N`).

FIGURE F.1

Log equity premium around business cycle peaks - International evidence

This figure presents the arithmetic average (solid black line) of the cumulative one-year ahead log equity premium around the recessions in the sample from 1970:2 to 2020:12. The solid gray lines depict the cumulative log equity premium of the individual recessions. The equity premium is the difference between the country-specific end-of-month index return and the country-specific short-term interest rate. The vertical axis depicts $\sum_{j=0}^{11} r_{t+j}$ for $t = -24, \ldots, -1, 0, 1, \ldots, 24$, whereby r_{t+j} is the log equity premium in 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 recession. Recession indicators are taken from the Economic Cycle Research Institute and results are shown for Germany, France, Canada, and the U.K.



probabilities derived from the term spread help to time the equity market. Interestingly, we further find that the recession probabilities estimated from U.S. data predict the equity premiums in those four countries as well. Table F.2 shows that the U.S. recession probabilities from cross-validation

TABLE F.2

Out-of-sample performance: Germany, France, Canada, United Kingdom

This table reports R^2_{OS} statistics in % for the out-of-sample predictability of (cumulative) log excess returns at the h-month ahead horizon relative to forecasts from the historical average. Results are shown for Germany, France, Canada, and the U.K. Forecasts are based on the linear predictive regression model with a constant and model-implied recession probabilities as a predictor variable. 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 model add either the term spread lagged by six-months (TMS$_{t-6}$) or the three-year moving average of the term spread (MA-TMS$_t$) as additional predictors. Cross-validation (U.S. data) refers to the probability forecasts from the U.S. data with TMS$_t$ and MA-TMS$_t$ and cross-validation as the break-correction method. *, **, 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 smaller MSFE than the historical average benchmark. The out-of-sample period is 1990:1 to 2020:12.

Variables in probit model	h=1	h=3	h=6	h=12
Germany:				
TMS$_t$	1.08**	3.17**	4.85**	7.73**
TMS$_t$, TMS$_{t-6}$	0.33	1.84*	3.67**	7.34**
TMS$_t$, MA-TMS$_t$	0.52*	1.82*	3.14*	7.53**
Cross-validation (U.S. data)	0.26	0.05	2.82	7.33**
France:				
TMS$_t$	1.00**	3.00***	4.30**	6.80***
TMS$_t$, TMS$_{t-6}$	0.84*	2.29**	3.40**	6.91***
TMS$_t$, MA-TMS$_t$	0.55*	1.77**	3.58**	6.66**
Cross-validation (U.S. data)	0.33*	0.81	5.90*	10.72**
Canada:				
TMS$_t$	0.63**	1.37**	3.08**	5.91***
TMS$_t$, TMS$_{t-6}$	0.94**	1.72**	2.24**	2.94**
TMS$_t$, MA-TMS$_t$	-0.65	-0.93	-0.10	1.88*
Cross-validation (U.S. data)	0.08	0.16	3.68	2.32**
United Kingdom:				
TMS$_t$	-0.98	-0.35	1.74**	6.04***
TMS$_t$, TMS$_{t-6}$	0.21	1.58**	2.58**	5.73***
TMS$_t$, MA-TMS$_t$	-0.14	0.76*	1.78**	2.29**
Cross-validation (U.S. data)	1.67**	4.75**	9.09**	20.02***

with TMS$_t$ and MA-TMS$_t$ often perform similarly to those of the best country-specific recession probabilities.