Finance, Economics and Monetary Policy
Discussion Papers

CALIBRATED RECESSION PROBABILITY MODEL

Arturo Estrella

Discussion Paper No. 2301
February 2023

Calibrated Recession Probability Model

Arturo Estrella

Finance, Economics and Monetary Policy Discussion Paper No. 2301

February 2023

JEL Classification: C25, C53, E32

Abstract

The 3-month U.S. Treasury yield has persistently exceeded the 10-year yield before every recession since 1968, but at no other times in that period. A probit model developed in 1989 established statistically that the difference between these two yields is a reliable leading indicator of recessions, but the model requires an exogenous benchmark to gauge the severity of the recession probability signals. This paper presents an alternative calibrated recession probability model that makes use of statistical analysis of the historical data to derive intuitive probability values endogenously.

Keywords: Yield curve, forecasting, coin toss level

Arturo Estrella
Professor of Economics, Emeritus
Rensselaer Polytechnic Institute
estrea@rpi.edu

1. Introduction

A large body of research has demonstrated that spreads between long-term and short-term government interest rates are reliable leading indicators of future recessions. These spreads have been used in dichotomous dependent variable (DDV) models to forecast recessions one year ahead, with strong in-sample and out-of-sample performance. The fitted variables in these models correspond to probabilities that a recession will occur at a specific time in the future.

One drawback of the original recession probability model is that the magnitudes of the spikes in estimated probabilities that precede recessions tend to be lower than intuition would suggest. For example, recessions are accurately forecast if a benchmark level of 30% is used to gauge the strength of the signal, rather than 50% or higher, as might be expected. The predictive power of the model is not impaired, but the interpretation of the probabilities requires some adjustment.

This paper proposes a statistical modeling approach designed to improve on existing models in two important respects. First, it starts with statistical analysis of the historical data for the recession and predictive variables, examining their individual and joint time-series properties. Second, it restructures both sides of the original probit equation to calibrate the model endogenously and produce more intuitive recession probabilities. A “coin toss level” form of the explanatory side facilitates incorporation of results from the statistical analysis of the series.

The resulting model can serve as a warning system to gauge the level of recession risk at any time during the business cycle. The model allows for reliable forecasts that identify recessions well in advance with readily interpretable probability estimates. For example, the recession probabilities derived from the model since 1968 reach well above 50% in the periods leading up to each of the eight recessions, but remain under 50% at all other times.

The paper first examines the original recession probability model, as well as critiques and extensions designed to produce more intuitive calibration of the model's probabilities. Generally, the drawback of these earlier proposals is that they are applicable in practice only within the estimation sample and do not fare well in terms of out-of-sample forecasting.

The next step is to analyze the time-series properties of the term spread and the recession DDV as well as their empirical relationship. The calibrated recession probability model integrates those statistical results to produce endogenous probability estimates that conform well with intuition.

2. The original recession prediction model, extensions and critiques

The association of the yield curve with the business cycle goes back almost as far as the concept of the cycle itself. Mitchell (1913), in a work that set the groundwork for the modern definition of the cycle, tabulated the relationship between long- and short-term interest rates and subsequent economic slowdowns, which he called "depressions." Later, Kessel (1965) showed renewed interest in the topic, examining how rates of different maturities varied over the cycle.

In both cases, however, the approach was descriptive, employing graphs, tables and general discussion rather than statistical methods or models. A descriptive approach was also employed later by Fama (1986) and Stambaugh (1986), who demonstrated graphically that forward interest rates tended to increase before expansions and to decrease before recessions.

In the late 1980s, research turned to more explicit models using the term spread as a predictor of future real economic growth. Published work included Laurent (1988), who used lags of the spread between 20-year Treasury yields and the federal funds rate to predict real GNP growth, and Harvey (1988), who used lags of estimated real Treasury rates at 3-month and 10-year maturities to forecast real consumption growth.

Estrella and Hardouvelis (1989, 1991) introduced the first formal statistical model that used the term spread to forecast recessions. Estimates showed that the spread was strongly significant in a probit model to predict recessions four quarters ahead. Other contemporaneous variables such as the real federal funds rate, the index of leading indicators, real GDP growth and inflation were not significant singly or jointly when added to the equation.

The analysis also showed that there was a clear spike in the estimated recession probabilities preceding each of the recessions that started in 1970, 1974, 1980 and 1981. Since that time, similar spikes have been observed prior to the four subsequent recessions (1990, 2001, 2008, 2020). Moreover, this indicator has not given any false positive signals in the period since 1968, that is, there have not been other spikes comparable in magnitude to those observed before recessions.

The probit equation introduced in the 1989 paper is

$$P\left(R_{t+k}=1 \mid S_t\right)=F\left(\beta_0+\beta_1 S_t\right), \tag{1}$$

where R_t is a recession indicator with value 1 if period t is in a recession and 0 otherwise, S_t is the spread between a long-term rate and a short-term rate at time t and F is the Gaussian cumulative distribution function.

Estrella and Hardouvelis (1989, 1991) used quarterly data and a one-year predictive horizon, $k=4$, to estimate the model. The recession indicator was derived from the business cycle turning point dates of the National Bureau of Economic Research (NBER).[1] The long-term

[1] The NBER announces dates of peaks and troughs of the business cycle. The standard assumption is that recessions start on the month or quarter following the peak and end on the month or quarter of the trough.

rate was the yield on 10-year U.S. Treasury securities and the short-term rate was the bond-equivalent yield on 3-month U.S. Treasury bills. Both yields were quarterly averages of monthly average rates.

The Estrella-Hardouvelis modeling approach has been employed in much of the subsequent literature, including applications to other countries as in Bernard and Gerlach (1998) and Duarte et al. (2005), variations in the form of the model as in Chauvet and Potter (2005), Kauppi and Saikkonen (2008) and Österholm (2012), and testing against other predictive variables as in Estrella and Mishkin (1998), Filardo (1999) and Rudebusch and Williams (2009).

It seems remarkable that a single economic variable that is easily and broadly accessible contains so much information about the future prospects for a recession, which by definition adversely affects the entire economy. Given its discernible successes, the term spread has attracted interest beyond academic researchers to reach financial practitioners and the press, though not always with the requisite level of rigor.

Notwithstanding the success and popularity of the original probit model, it has two characteristics that have raised questions about its calibration and interpretability. Chauvet and Potter (2005) identified these properties, demonstrated their effects, and suggested various fixes.

The first characteristic is that the parameter estimation method treats monthly or quarterly recession observations as independent events, making no provision for serial correlation in either the recession DDV or the predictive term spread variable. Parameter estimates are statistically consistent, but for the purpose of statistical inference, standard errors need to be adjusted for serial correlation. Empirically, both variables in the model exhibit strong positive serial correlation.

The second characteristic is that, although the probabilities produced by the model contain clear peaks that correspond to forecasts of actual recessions, the magnitudes of those peaks vary substantially across recessions. A probability level of 30% has been suggested as a time-consistent threshold for a reliable recession signal, but this level is not ideal for an intuitive interpretation of the recession probabilities.

The 30% benchmark is meant to be applied to the estimated probability of the event that a single future monthly observation will be in a recession. Every instance since 1968 in which the 30% threshold has been crossed has been followed by an actual recession a few months later. So, taking each recession as a single event, application of the 30% threshold has been perfectly accurate in predicting whether a recession will soon follow.

The calibrated model proposed in the present paper deals expressly with the issues of serial correlation and interpretability. First, the time series of the dependent and predictive variables are analyzed both separately and jointly in order to develop the modeling strategy. Second, the structure of the probit model is reformulated to allow for calibration that reflects the findings of the time series analysis and produces recession probabilities that accord more closely with intuition. In addition, the model is subjected to temporal stability tests to ensure that the results are consistent with the current properties of the data.

The calibrated model retains the excellent real-time predictive properties of the original model at a 12-month forecast horizon. However, the level of the estimated probabilities is more in line with intuition and does not require the additional step of identifying a benchmark for the interpretation of the results.

Before proceeding to the construction of the calibrated model, consider first the solution proposed by Chauvet and Potter (2005) (CP) to deal with the two issues. They employ a latent variable interpretation of the probit model in which the latent variable Y_t^* has the dynamic form

$$Y_{t+k}^* = \beta_0 + \beta_1 S_t + \beta_2 Y_{t+k-1}^* + \varepsilon_{t+k} .$$

If $\beta_2 = 0$ and ε_t is serially independent and normally distributed with mean 0 and variance 1, the model reduces to equation (1) with $P\left(R_{t+k} = 1 | S_t\right) = P\left(Y_{t+k}^* > 0 | S_t\right)$.

CP allow for nonzero β_2 and heteroskedastic ε_t, the latter implemented by estimating different β parameters for each individual business cycle. The complicated likelihood function is derived from the joint distribution of the $\left\{\varepsilon_s\right\}$ series and is evaluated using the Gibbs sampler to estimate the parameters of the model.

The in-sample empirical results produce a much closer fit with the data than the original model. Recession probabilities line up very well with actual recessions and the peaks are consistent across recessions. The flexibility provided by the inclusion of the adaptable latent variable and the allowance for heteroskedasticity clearly enhances the fit of the model.

For out-of-sample forecasting, however, the model is limited by the unobservability and unpredictability of both the latent variable and the distribution of the disturbance. For example, CP report that, in contrast with the original model, the CP model missed the 2001 recession. After a thorough case study of the period 2001-2002, they argue that "The fact that the simple models predict a recession in 2001 is not necessarily a virtue. The forecasts from these simple models did not allow for any updating of the probability of a recession as 2001 proceeded."

The CP model has the edge in within-sample analysis and historical classification of recession periods after the fact, but for real-time forecasting purposes, accuracy has to be considered a virtue and the original model performs better.

3. Serial correlation

The likelihood function of the original probit model treats each observation as independent. It does not make use of the fact that recession months or quarters occur consecutively in the data or that the term spread is strongly serially correlated. Econometric theory indicates that parameter estimates are nonetheless consistent, that is, they should converge to the correct values as the number of observations increases, but they do not directly incorporate all the known properties of the data. How prejudicial is this omission to the accuracy of the model?

Consider the serial correlation of the variables in the model by themselves and in connection with the original model. It is clear that the dichotomous recession variable is serially correlated, essentially by construction. The NBER (2022) states that "The NBER's traditional definition of a recession is that it is a significant decline in economic activity that is spread across the economy and that lasts more than a few months."

The definition ensures that recession months occur consecutively and thus gives rise to serial correlation. The solid curves in Figure 1 show that this autocorrelation persists for about 12 months and that it is driven mainly by a first-order partial autocorrelation of .891.[2]

[2] These and all subsequent estimates in the paper are based on monthly NBER recession dates and monthly averages of the 10-year and bond-equivalent 3-month Treasury yields.

The variable used as a predictor, the term spread, is the difference between two interest rates that are both serially correlated and correlated with one another, again resulting in serial correlation. However, it is important to note that this autocorrelation is not necessarily detrimental to the predictive power of the model since it could help model the autocorrelation of the dependent variable.

Figure 1, where the spread is represented by the dotted lines, shows a very persistent autocorrelation that remains positive for about two years. Again, the driver is a first-order partial autocorrelation, in this case .956, which is somewhat higher than the corresponding statistic for the dependent variable but is comparable in magnitude.

The original probit model produces recession probabilities by applying a nonlinear function to the term spread as shown in equation (1), with the parameters evaluated so as to fit the dependent variable as closely as possible using the likelihood function as metric.[3] Thus, by construction, these probabilities inherit the serial correlation of the model's variables, as shown by the dashed lines in Figure 1. The autocorrelation of the fitted recession probabilities is slightly less persistent than that of the spread and is driven by first-order partial autocorrelation of .934. As may be expected, these results lie between those of the dependent and independent variables.

The CP paper maintains that "in the standard probit model conditional on the observed yield, the probabilities of recession states are independent of each other, which follows directly from the assumption of independent errors." This claim is clearly disproven by the empirical results in Figure 1. In the framework of the CP model, the original model is estimated as if the disturbance ε_t is serially independent, but the probabilities on either side of equation (1) clearly need not be.

[3] Model estimate is reported in Table 5, line 1.

More generally, the fact that the recession variable is serially correlated is not necessarily detrimental to the accuracy of the model. To the extent that the time-series properties of the dependent variable are shared by the independent variable at a stable lag, the existence of these dynamic properties can serve to enhance the predictive power of the model.

There are three main ways of adjusting the original model to incorporate serial correlation directly: adding lags of the dependent variable as in Kauppi and Saikkonen (2008), adding a latent variable and its lags as in Chauvet and Potter (2005) and adding lags of the explanatory variable. The latter option is rarely seen because only one lag of the term spread tends to be significant in recession prediction models.

The problem with the other two alternatives is that data for the added components are not available for use in out-of-sample forecasts. NBER turning point dates are announced well after the fact and other forecasters do not fare any better. Forecasting the lag of the DDV is just as hard as forecasting the DDV itself. The latent variable, of course, is never observed. Even if the model is successful in fitting the latent variable within the estimation sample, forecasting it out of sample is much more difficult, as the CP results suggest.

To answer the question posed at the beginning of this section, the absence of explicit dynamics in the original model does not seem to be materially prejudicial to its accuracy. The fact that the independent and dependent variables are comparably serially correlated at a stable lag contributes to the accuracy of the model. Moreover, in view of the lack of viable alternatives for out-of-sample forecasting, the use of the standard probit estimator seems reasonable.

4. Negative term spreads and recessions

A closer look at the relationship between the term spread and the recession variable suggests a strategy for calibrating the recession probability model without compromising its ability to produce accurate out-of-sample forecasts. In fact, it is at this level of analysis that the term spread has a perfect predictive record since 1968, and the corresponding quantitative statistical results may be brought to bear on the issue of model calibration.

The term spread is specified as the monthly average of the difference between 10-year and 3-month bond-equivalent U.S. Treasury yields. A look at the history of this spread in Figure 2 shows why interest has focused on yield curve inversions, that is, on negative values of the spread, as predictors of recession. Estrella and Hardouvelis (1989, 1991) presented evidence that the spread had been negative in anticipation of the four recessions prior to 1989. Figure 2 shows that, in addition, the spread was negative before each of the four recessions that have occurred since then.

Table 1 lists every month in which the spread was negative between 1968 and 2020. The table confirms that inversions occurred in anticipation of the eight recessions in this period, with some variation in lead times and in the persistence of the inversions. In the cases of the first four recessions, the spread continued to be negative for varying amounts of time after the start of the recession. In the last four cases, the spread was no longer negative by the time the recession started. One possible explanation for this change is that the Federal Reserve followed a different monetary policy strategy in the two subperiods.

Yield curve inversions tend to occur when the Federal Reserve tightens the stance of monetary policy beyond a certain level, which may vary over time. For example, Furmann and Otrok (2013) present evidence that a major influence on the term spread is the Federal Reserve reaction to news shocks, in particular those connected with total factor productivity. Viewed in this light, Table 1 suggests that Federal Reserve reactions were more forceful and persistent in the earlier part of the period.

Table 2 provides a count of the number of negative monthly spreads connected with each recession in Table 1, followed by the length of the recession. The table suggests that there is a relationship between the persistence of negative spreads before and during a recession and the length of the subsequent recession. Specifically, a recession in this sample period lasted more than 9 months if and only if there were more than 9 months of negative spreads.

The only exception to this rule is the recession that started in February 1980, which is the first of the “twin recessions” of 1980-1982. During this episode, monetary policy targets were set with reference to the money supply and the direction of the federal funds rate changed abruptly more than once, whipsawing from 19.96% in early April 1980 to 7.65% in early August and back to 22.36% during July 1981. The economy responded with a brief recovery before plunging into a second deeper recession. Taken together, the twin recessions lasted a total of 22 months and were associated with 30 months of negative spreads.

A key question in the construction of a method to predict recessions using the term spread is whether there is a benchmark level of the spread that helps distinguish recessions from soft landings or other non-recessionary periods. The data in Table 1 shows that recessions are always preceded by negative spreads, but is zero the best benchmark level?

Two questions can help establish a functional benchmark. First, what range of values of the term spread precedes recessions within a reasonable horizon? Second, what range of values does the spread exhibit when it has reached a local minimum, but a recession has not ensued within a reasonable horizon? Of course, the concept of reasonable horizon must be defined more precisely.

Table 3 addresses the first question by examining the magnitude and lead time of two types of low values of the spread, the first negative value prior to each recession and the lowest value observed in the 12 months prior to the start of the recession.

Tabulation of the first negative value helps address the question of forecast horizon. Among the eight recessions, the first negative occurred between 6 and 17 months prior to the recession start, with a median lead time of 12 months. These first negative values range in magnitude from -.52% to -.09%.

The tabulation of the lowest value imposes a time horizon of 12 months, but these minima occur closer to the start of the recession than the first negative values and the horizon constraint is for the most part not binding. The only case in which a minimum value lies slightly outside the 12-month horizon is the August 1990 recession. The lead time of the minimum values ranges from 2 to 10 months, with a median value of 6.

Based on the information in this table, which spans 53 years and contains 8 recessions, we may conclude that whenever the monthly value of the spread fell below -.08%, a recession ensued within 12 months. While there is no guarantee that this result will hold uniformly in the future, the sample is large enough to suggest that -.08% is a reasonable practical benchmark going forward.

The second question requires the identification of local minima of the monthly spread within in the 53-year sample, whether or not those minima consist of negative values or are related to a recession. Table 4 lists all local minima, defined as the lowest values of the spread in a period ranging between 12 months before and 12 months after the observation.

Of the 13 cases in Table 4, seven of the spreads are negative, and they all correspond to pre-recession minima listed in Table 3. As noted earlier, the minimum value of the spread in anticipation of the August 1990 recession occurred outside the 12-month horizon, and we see in Table 4 that it was observed 14 months before the recession started.

For all other cases, the local minima of the spread are positive and the time to the next recession is at least 31 months, suggesting that relatively low positive values of the spread are not indicative of an impending recession. This pattern is also confirmed by the fact that the negative spreads are followed by increases in the unemployment rate during the following 24 months, whereas the positive values are followed by decreases in unemployment, as shown also in Table 4.

Thus, when the local minimum of the spread was .08% or higher, a recession did not start within the following 30 months. Putting together this result with the parallel finding from Table 3, we see that recessions aways followed within 12 months when the spread was -.08% or lower, whereas recessions did not follow within 30 months when the spread was .08% or higher. Any benchmark value that falls between those two levels would be able to discriminate perfectly between recessions and non-recessions with a lead time of one year to the start of the recession.

This level of discrimination is so strong that standard methods do not allow for the calculation of a single benchmark within the range $-.07 \le S_t \le .07$. For example, application of models like probit or logit fails when discrimination is perfect, and discriminant analysis is designed for continuous random variables rather than for dichotomous variables such as the recession indicator.

In the following section, the statistical results of this section are combined with a modified specification of the probit model to construct a calibrated recession probability model that combines accuracy and interpretability.

5. Calibrated recession prediction model

The calibrated model modifies both sides of the probit specification of the original model, equation (1), in order to make the resulting probabilities of recession more intuitive. The alternative dependent variable is more adaptable than the original and the mathematical form of the predictive function makes it easier to apply the statistical analysis of the previous section.

The dependent variable of the calibrated model is defined as an index of the event that a recession starts within the following 12 months.[4] This specification provides more flexibility than the very focused dependent variable of the original model, which requires forecasting whether or not a single observation one year ahead is in a recession. Empirical estimates confirm that a closer fit may be obtained with the more inclusive dependent variable of the calibrated model.

However, use of the alternative dependent variable requires some care. First, the flexibility of the variable makes it difficult to select an appropriate forecast horizon. In the original model, the forecast focuses on a single future period regardless of the horizon, which may be extended one month at a time until the best fit is obtained.

With the alternative dependent variable, extending the forecast horizon generally increases the chances of obtaining a correct classification and tends to improve the fit of the model. The reason is that the probability of the composite event corresponding to the DDV tends to increase with each additional month included in the horizon.[5] Thus, a simple comparison of results for different forecast horizons may be misleading.

---

[4] An analogous dependent variable is used in Engstrom and Sharpe (2018) with quarterly data. Leamer (2022) defines a DDV based on an event that is formally equivalent to the one used here, described differently, with other additional conditions.

[5] The probability that a recession starts in month t or in month $t+1$ is essentially the sum of the two probabilities because the joint probability of the two events is virtually zero.

The analysis of the previous section is helpful in dealing with this issue, since it demonstrates that a 12-month horizon is reasonable for predicting the start of a recession using the term spread. Results showed that a benchmark level of the spread in the range $-.07 \le S_t \le .07$ identifies recessions perfectly in the sample period with a 12-month horizon.

This choice of horizon is also supported by sensitivity analysis of the original model, in which the timing of the DDV is more precise. Model (1) was estimated with data from January 1968 to December 2019 and with horizons ranging to *k* from 1 to 24 months. The end date for the estimates is set to accommodate the longest forecast horizon and the delay in NBER turning point announcements. The value $k = 12$ maximizes the likelihood of the model.

A second issue with the new dependent variable is that the resulting empirical estimates are not as stable over time as those obtained with the original model. Estrella, Rodrigues and Schich (2003) showed that tests for unknown breakpoints are not able to detect a break in the original model using data from either the United States or Germany, where the model fit is very strong. The solution to this issue is to apply breakpoint testing and, where the tests find statistically significant evidence of a break, the estimation period is adjusted so as to employ estimates consistent with the most recent data.[6]

To summarize, the left hand side of the calibrated model equation takes the form $P\left(B_t^{12} = 1 | S_t\right)$, where B_t^{12} is an index variable that takes on a value of 1 if an NBER-dated recession begins in any month from $t+1$ to $t+12$ and is otherwise 0.

The form of the right-hand side of the probit equation is also adjusted so as to make effective use of the statistical analysis of the previous section. Instead of the standard probit form

[6] This strategy is somewhat analogous to the treatment of heteroskedasticity in Chauvet and Potter (2005), but only one break is required, as the empirical estimates will show.

$F\left(\beta_0+\beta_1 x\right)$, the argument of the normal cumulative distribution function is expressed as $F\left(m\left(x-c\right)\right)$. If the model is estimated without constraints, the two forms are mathematically equivalent with $m=\beta_1$ and $c=-\beta_0/\beta_1$.

The parameter *c* is interpreted as the coin toss level (CTL) of the predictor variable *x*. At this level, the probability that the dependent variable equals 1 is exactly one half, as in a fair coin toss. The multiplicative factor *m* scales the probabilities for other values of *x*. The form of the calibrated model is thus

$$P\left(B_t^{12}=1 \mid S_t\right)=F\left(m\left(S_t-c\right)\right). \tag{2}$$

The analysis of the previous section implies that if $-.07 \le c \le .07$, the beginning of each recession will be preceded by observations with probabilities above 50%, whereas for observations in which the probability is less than 50%, a recession will not begin within the forecast horizon.

The strategy then is to estimate model (2) subject to the restriction that $-.07 \le c \le .07$. To verify that the resulting estimate is consistent with the data, we need to test that the restriction on *c* is not statistically rejected and also to confirm the stability of the model over the estimation period. The base sample covers the 53-year period from January 1968 to December 2020, with the end of the period determined by the length of the forecast horizon plus an allowance of 12 months in recognition of lags in turning point announcements.

For reference, the first line of Table 5 provides estimates of the original Model (1) over the base sample period. Parameter estimates are shown in the table in unconstrained CTL form. The fit of the model is in line with past estimates and the results are stable over time in a sup-

Wald breakpoint test analogous to those of Estrella, Rodrigues and Schich (2003).[7] However, the estimated CTL c is -.76 and the restriction that $-.07 \le c \le .07$ is rejected with p value .047, which is consistent with the need for a benchmark lower than 50% to interpret the recession probabilities derived from the unconstrained original model.[8]

The second line of Table 5 shows results for the recession probability Model (2), estimated without constraints. Again, the point estimate of the CTL falls outside the benchmark region from the earlier analysis, but in this case a test of the restriction that $-.07 \le c \le .07$ has p value .220 and the constraint is not rejected. However, there is evidence of instability over time as the sup-Wald test has a p value very close to zero.

The third line of the table imposes on Model (2) the restriction that $-.07 \le c \le .07$ and produces similar results. R^2 is .362, just slightly lower than the unconstrained model, and the CTL is estimated at the lower bound of the benchmark interval. However, the problem of temporal instability remains, with a sup-Wald test strongly indicating that there is a breakpoint at October 1981.[9]

These results suggest that the constrained Model (2) should be estimated over the sample period that starts from the breakpoint identified by the test. For reference, the next line of the table also presents estimates of the unconstrained model over this sample period, with two

---

[7] Andrews (1993) shows that a portion of the observations must be excluded as candidate breakpoints at each end of the sample so that the estimates converge, and suggests .15 of the sample. With a DDV with many consecutive equal values, such as the recession indices, the portion excluded frequently has to be larger to guarantee convergence. The present tests use .25 as in Estrella, Rodrigues and Schich (2003) or .35, depending on the sample period.

[8] This test and others reported in this section are performed with standard errors adjusted for heteroskedasticity and autocorrelation with a flat window and lags to 12 months.

[9] When the model is constrained, the CTL constraint is imposed for each of the subsample estimates in the breakpoint test.

notable results: the estimated value of c is very close to the benchmark restriction range and there is no significant evidence of instability within this sample period.

The final line of the table uses the sample period starting from the breakpoint and imposes the CTL benchmark restriction $-.07 \le c \le .07$. The CTL is estimated at the upper end of the range and the constraint is not rejected, with a p value of .894. The model fits the data well, with $R^2 = .386$. Moreover, there is no further evidence of structural instability within the sample period. This specification is selected as the calibrated recession probability model.

Figure 3 compares the probability functions derived from the original model and from the calibrated model, corresponding to lines 1 and 5 of Table 5. These probabilities pertain to two different types of events, so they should be different. The main point of the figure is to show that the calibrated model approach is better suited to an intuitive interpretation of the probabilities in connection with an upcoming recession as a whole, rather than with each recession month individually.

For example, when the spread is zero, the original model gives a probability of .30, whereas the calibrated model probability is .55. As noted earlier, the .30 level has been proposed as a benchmark for the original model in connection with a recession as a whole, but the .55 level is more reflective of the actual likelihood of an upcoming recession given historical data. At the estimated CTL level of the spread, .07, the calibrated probability is .50 by construction, whereas the original model probability is only .28.

The autocorrelation analysis of Figure 1 provides additional support for the calibrated model, represented by the curve with dashes and dots. In particular, the figure shows that the autocorrelation properties of the calibrated model resemble those of the dependent variable more closely than those of either the spread data or the original model.

Finally, Figure 4 presents the historical performance of the calibrated probability model since 1968, as well as out-of-sample forecasts starting in January 2021. Whenever the probability that a recession will start within the following 12 months exceeds .50, in other words, the start of a recession is more likely than not, a recession did start within the predictive horizon. In fact, the peak probability levels in anticipation of each of the 8 recessions in the sample were well above .50 in all cases.

6. Conclusions

The original recession probability model from Estrella and Hardouvelis (1989, 1991) has performed well over the years, but the probability levels arising from the model are best interpreted with reference to an exogenous benchmark level. The calibrated recession probability model of the present paper has an equivalent level of accuracy and a closer fit with the alternative dependent variable. In addition, the predictive recession probabilities produced by the model accord more directly with intuition without the need for supplementary benchmarks.

The calibrated model produces very strong signals before each of the 8 recessions that have been identified by the NBER since 1968. In anticipation of each recession, the model's estimates of the probability that a recession will start within 12 months have spiked to levels ranging from .67 to 1.00. At other times in the sample period, the probabilities given by the model have been below .50. As of this writing, the model strongly suggests that a recession will start in 2023, as Figure 4 shows.[10]

[10] Results of the calibrated probability model and the original model are updated monthly at financeecon.com/yc.html.

References

Andrews, Donald W.K. (1993) “Tests for parameter instability and structural change with unknown change point.” Econometrica 61(4): 821-856.

Bernard, Henri and Stefan Gerlach (1998) “Does the term structure predict recessions? The international evidence.” International Journal of Finance and Economics 3: 195-215.

Chauvet, Marcelle and Simon Potter (2005) “Forecasting recessions using the yield curve.” Journal of Forecasting 24: 77-103.

Duarte, Agustin, Ioannis A. Venetis and Iván Paya (2005) “Predicting real growth and the probability of recession in the euro area using the yield spread.” International Journal of Forecasting 21 (2): 261–77.

Engstrom, Eric and Steven Sharpe (2018) “The near-term forward yield spread as a leading indicator: A less distorted mirror.” Finance and Economics Discussion Series 2018-055.

Estrella, Arturo (1998) “A new measure of fit for equations with dichotomous dependent variables.” Journal of Business and Economic Statistics 16(2): 198-205.

Estrella, Arturo (2003) “Critical values and p values of bessel process distributions: Computation and application to structural break tests.” Econometric Theory 19: 1128-1143.

Estrella, Arturo and Gikas Hardouvelis (1989) “The term structure as a predictor of real economic activity.” Federal Reserve Bank of New York Working Paper no. 89-07, May.

Estrella, Arturo and Gikas Hardouvelis (1991) “The term structure as a predictor of real economic activity.” Journal of Finance 46(2): 555.576.

Estrella, Arturo and Frederic S. Mishkin (1998) “Predicting U.S. recessions: Financial variables as leading indicators.” Review of Economics and Statistics 80(1): 45-61.

Estrella, Arturo, Anthony P. Rodrigues and Sebastian Schich (2003) “How Stable Is the Predictive Power of the Yield Curve? Evidence from Germany and the United States.” Review of Economics and Statistics 85(3): 629–644.

Fama, Eugene F. (1986) "Term Premiums and Default Premiums in Money Markets," Journal of Financial Economics 17: 175-96.

Filardo, Andrew J. (1999) “How reliable are recession prediction models?” Federal Reserve Bank of Kansas City Economic Review 84: 35-55.

Harvey, Campbell R. (1988) “The real term structure and consumption growth.” Journal of Financial Economics 22: 305-333.

Kauppi, Heikki and Pentti Saikkonen (2008) “Predicting U.S. recessions with dynamic binary response models.” Review of Economics and Statistics 90(4): 777-791.

Kessel, Reuben A. (1965) “The cyclical behavior of the term structure of interest rates.” National Bureau of Economic Research Occasional Paper No. 91.

Kurmann, Andre and Christopher Otrok (2013) “News shocks and the slope of the term structure of interest rates.” American Economic Review 103(6): 2612-2632.

Laurent, Robert (1988) “An interest rate-based indicator of monetary policy.” Federal Reserve Bank of Chicago Economic Perspectives 12: 3-14.

Leamer, Edward E. (2022) “A new way of forecasting recessions.” National Bureau of Economic Research Working Paper 30247.

Mitchell, Wesley C. (1913) Business Cycles. Berkeley: University of California Press.

National Bureau of Economic Research (2022) “Business Cycle Dating Procedure: Frequently Asked Questions” Revised August 15, 2022. Available online at https://www.nber.org/research/business-cycle-dating/business-cycle-dating-procedure-frequently-asked-questions.

Österholm, Pär (2012) “The limited usefulness of macroeconomic Bayesian VARs when estimating the probability of a U.S. recession.” Journal of Macroeconomics 34 (1): 76–86.

Rudebusch, Glenn D. and John C. Williams (2009) “Forecasting recessions: The puzzle of the enduring power of the yield curve.” Journal of Business and Economic Statistics 27(4): 492-503.

Stambaugh, Robert F. “The information in forward rates: Implications for models of the term structure.” Journal of Financial Economics 21: 41-70.

Table 1. Negative values of monthly average term spread, 1968 to 2020
Time to start of recession (months) and monthly recession indicator (0/1)

Dec 1968	-0.11	13	0
Jan 1969	-0.28	12	0
Feb 1969	-0.11	11	0
Apr 1969	-0.12	9	0
Jun 1969	-0.07	7	0
Jul 1969	-0.51	6	0
Aug 1969	-0.51	5	0
Sep 1969	-0.16	4	0
Oct 1969	-0.13	3	0
Nov 1969	-0.34	2	0
Dec 1969	-0.44	1	0
Recession started Jan 1970			
Jan 1970	-0.35	0	1
Feb 1970	-0.12	-1	1
Jun 1973	-0.52	6	0
Jul 1973	-1.16	5	0
Aug 1973	-1.59	4	0
Sep 1973	-1.50	3	0
Oct 1973	-0.67	2	0
Nov 1973	-1.37	1	0
Recession started Dec 1973			
Dec 1973	-0.96	0	1
Jan 1974	-1.05	-1	1
Feb 1974	-0.39	-2	1
Mar 1974	-1.03	-3	1
Apr 1974	-1.12	-4	1
May 1974	-0.94	-5	1
Jun 1974	-0.63	-6	1
Aug 1974	-1.25	-8	1
Sep 1974	-0.30	-9	1
Nov 1974	-0.04	-11	1
Nov 1978	-0.15	15	0
Dec 1978	-0.41	14	0
Jan 1979	-0.61	13	0
Feb 1979	-0.58	12	0
Mar 1979	-0.73	11	0
Apr 1979	-0.65	10	0
May 1979	-0.74	9	0
Jun 1979	-0.49	8	0
Jul 1979	-0.64	7	0
Aug 1979	-0.86	6	0
Sep 1979	-1.35	5	0
Oct 1979	-1.92	4	0
Nov 1979	-1.67	3	0
Dec 1979	-2.20	2	0
Jan 1980	-1.75	1	0
Recession started Feb 1980			
Feb 1980	-1.07	0	1
Mar 1980	-3.28	-1	1
Apr 1980	-2.38	-2	1
Oct 1980	-0.39	10	0
Nov 1980	-1.74	9	0
Dec 1980	-3.51	8	0
Jan 1981	-3.26	7	0
Feb 1981	-2.39	6	0
Mar 1981	-0.90	5	0
Apr 1981	-0.70	4	0
May 1981	-3.14	3	0
Jun 1981	-2.04	2	0
Jul 1981	-1.47	1	0
Recession started Aug 1981			
Aug 1981	-1.43	0	1
Sep 1981	-0.16	-1	1
Jun 1989	-0.16	14	0
Jul 1989	-0.13	13	0
Aug 1989	-0.06	12	0
Nov 1989	-0.08	9	0
Dec 1989	-0.05	8	0
Recession started Aug 1990			
Jul 2000	-0.09	9	0
Aug 2000	-0.44	8	0
Sep 2000	-0.38	7	0
Oct 2000	-0.55	6	0
Nov 2000	-0.63	5	0
Dec 2000	-0.70	4	0
Jan 2001	-0.13	3	0
Recession started Apr 2001			
Aug 2006	-0.21	17	0
Sep 2006	-0.22	16	0
Oct 2006	-0.32	15	0
Nov 2006	-0.47	14	0
Dec 2006	-0.42	13	0
Jan 2007	-0.35	12	0
Feb 2007	-0.45	11	0
Mar 2007	-0.51	10	0
Apr 2007	-0.31	9	0
May 2007	-0.10	8	0
Recession started Jan 2008			
Jun 2019	-0.14	9	0
Jul 2019	-0.08	8	0
Aug 2019	-0.36	7	0
Sep 2019	-0.23	6	0
Feb 2020	-0.05	1	0
Recession started Mar 2020			

Table 2. Duration of Recession and Persistence of Negative spreads

Recession start	Duration	Negative spreads
Jan 1970	11	13
Dec 1973	16	16
Feb 1980	6	18
Aug 1981	16	12
Aug 1990	8	5
Apr 2001	8	7
Jan 2008	18	10
Mar 2020	2	5

Notes: Recession duration is the number of months from NBER peak to trough. Negative spreads is the number of months before and during each recession that the term spread was negative.

Table 3. Negative spreads before recessions

	First negative			Minimum in prior 12 months		
Recession start	Month	Spread	Lead	Month	Spread	Lead
Jan 1970	Dec 1968	-0.11	13	Aug 1969	-0.51	5
Dec 1973	Jun 1973	-0.52	6	Aug 1973	-1.59	4
Feb 1980	Nov 1978	-0.15	15	Dec 1979	-2.20	2
Aug 1981	Oct 1980	-0.39	10	Dec 1980	-3.51	8
Aug 1990	Jun 1989	-0.16	14	Nov 1989	-0.08	9
Apr 2001	Jul 2000	-0.09	9	Dec 2000	-0.70	4
Jan 2008	Aug 2006	-0.21	17	Mar 2007	-0.51	10
Mar 2020	Jun 2019	-0.14	9	Aug 2019	-0.36	7

Notes: First negative is first instance of a negative monthly spread in the period leading to each recession. Lead time is time in months from first or minimum negative spread to start of the following recession.

Table 4. Spread Local Minima: Month, Minimum Spread, Lead Time

Month of minimum	Spread	Next recession (months)	Change in unemployment
Aug-69	-0.51	5	2.6
Aug-73	-1.59	4	3.6
Dec-80	-3.51	8	3.6
Sep-84	1.72	71	-0.3
Mar-86	1.02	53	-1.5
Jun-89	-0.16	14	1.6
Oct-93	2.24	90	-1.3
Nov-95	0.42	65	-1.0
Sep-98	0.08	31	-0.7
Dec-00	-0.70	4	2.1
Mar-07	-0.51	10	4.3
Jul-12	1.43	92	-2.0
Aug-19	-0.36	7	1.5

Notes: Window for local minimum spreads is 25 months. Lead time is time in months from minimum spread to start of the following recession and the change in unemployment is over the 24 months following the month of the minimum.

Table 5. Estimates of Recession Probability Models

Model (1) $P\left(R_{t+12}=1 \mid S_t\right)=F\left(\beta_0+\beta_1 S_t\right)$, $c=-\beta_0/\beta_1$, $m=\beta_1$

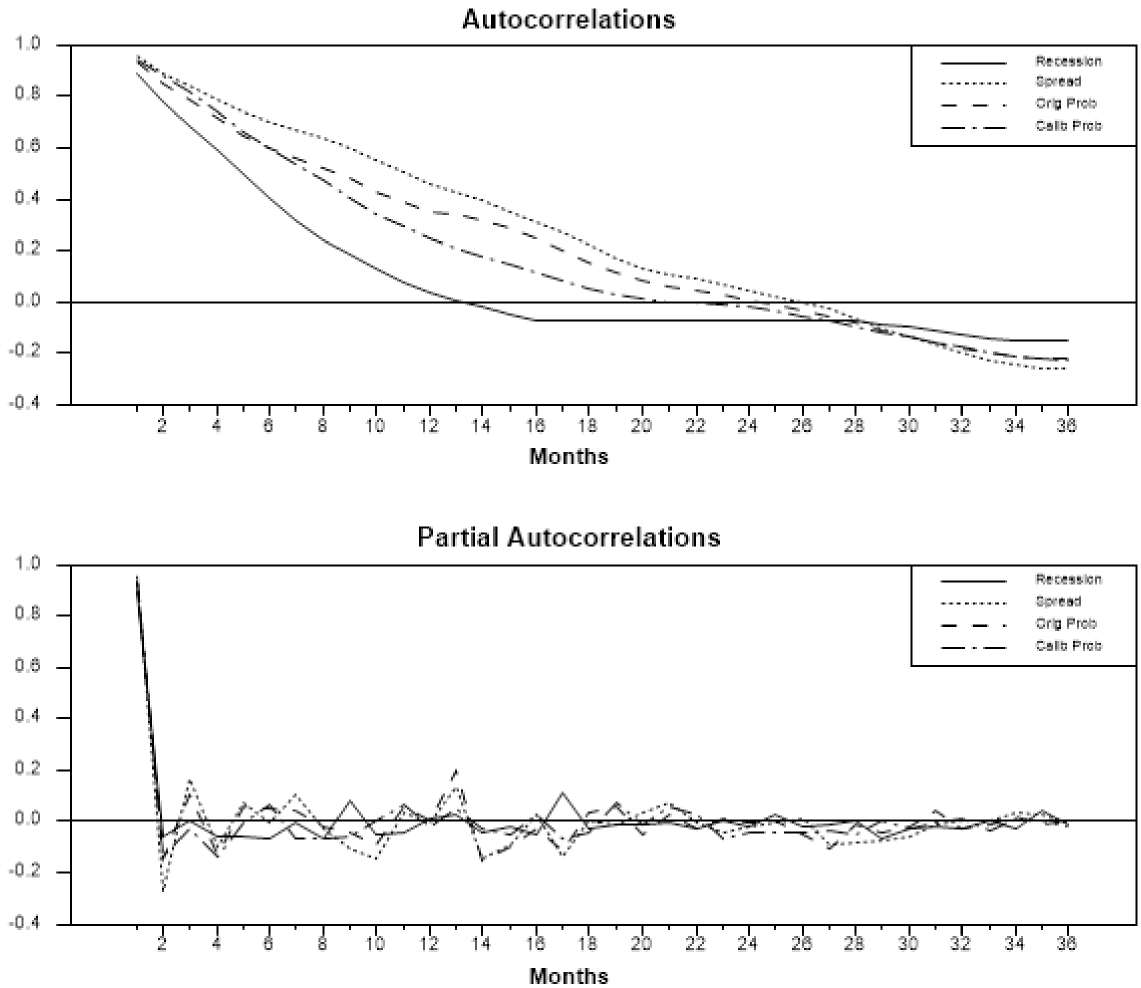
Model (2) $P\left(B_t^{12}=1 \mid S_t\right)=F\left(m\left(S_t-c\right)\right)$, Constraint: $-.07 \le c \le .07$

Model	Sample starts	Constr.	c	m	R^2	sup-Wald	Breakpoint
(1)	Jan 1968	No	-.764(.350)	-.685(.144)	.247	3.08(.786)	None
(2)	Jan 1968	No	-.337(.218)	-.958(.155)	.376	52.7(.000)	Mar 2007
(2)	Jan 1968	Yes	-.07(na)	-1.07(.161)	.362	23.8(.000)	Oct 1981
(2)	Oct 1981	No	.091(.182)	-1.90(.497)	.386	5.62(.275)	None
(2)	Oct 1981	Yes	.07(na)	-1.86(.455)	.386	1.34(.953)	None

Notes: All samples contain monthly data ending December 2020. HAC standard errors (flat window, 12 lags) in parentheses; not assigned if constraint is binding. Estrella (1998) R^2. Breakpoint test is based on sup-Wald with exact p values (Estrella 2003) in parentheses; .25 excluded for full sample and .35 for Oct 1981 sample.

Figure 1. Autocorrelations and partial autocorrelations

Recession, spread and fitted probability variables



Note: Calculations use monthly data from January 1968 to December 2020. Estimates based on monthly NBER recession dates and monthly averages of the 10-year and bond-equivalent 3-month Treasury yields. Spread is the difference between the two rates.

Figure 2. Term Spread

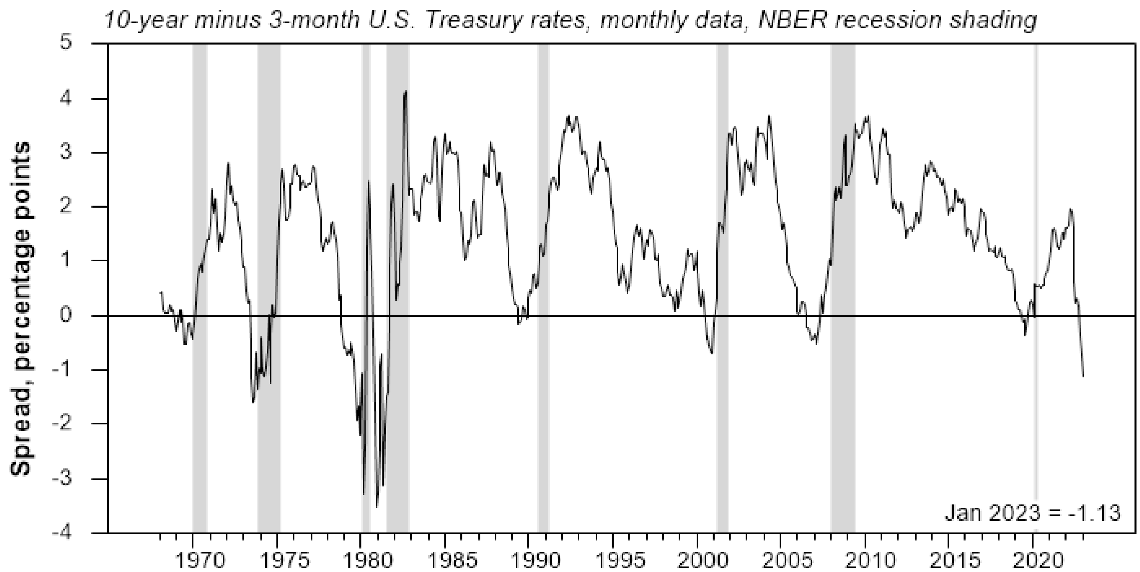
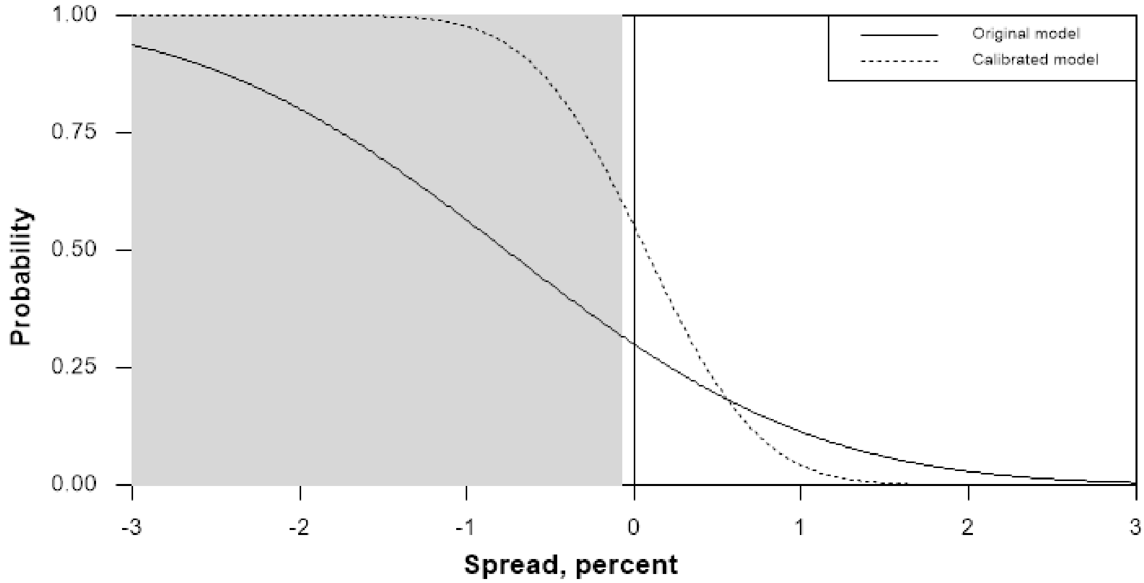


Figure 3. Recession Probability as Function of Term Spread



Notes: The probabilities correspond to the events $R_{t+k} = 1 | S_t$ for the original model and $B_t^{12} = 1 | S_t$ for the calibrated model. The shaded region encompasses all values of the spread that anticipate recessions in Table 3.

Figure 4. Calibrated Model: Probability That a Recession Will Start Within 12 Months