# Hodrick-Prescott -Filter-Defined Recession Prediction Using Yield Curve in A Logit Framework

Authors: Yihong Xu, Heli Vora

## Abstract:

In this paper, we investigate the forecasting ability of the yield curve and other financial variables by using the average difference between the interest rate on the long term and short term Treasury bills in a logit-model framework for the period from 1984: Q1 to 2015: Q4. We implemented a Hodrick-Prescott Filter to decompose GDP series data and then used this series to define when there is a recession. The forecasting power grows strongest when we forecast 11 periods (quarters) before. With 1984—2009 data as the training set and 2010—2015 as the test set, we can achieve an overall forecasting accuracy of 83.3%. The results are compared to the alternative machine learning model to prove the predicting effect of the yield curve.

Key Words: Yield Curve, Recession, Logit Model, SVM, H-P filter

## Introduction

Forecasting economic recession is a big topic in macroeconomic researches. From the macro-economic perspective, every country wants to predict the economic situation, to adjust policies to control the negative impact of a future economic recession. In the microeconomic environment, enterprises also want to see the danger of economic recession in advance and maintain their survival by adjusting their strategies.

In previous studies, scholars have found that the information contained in the yield curve can predict economic recessions to a certain extent. The spread of yield curve is the most intuitive and important curve characteristic. The interest rate spread refers to the difference of interest rate between long-term Treasury bonds and short-term Treasury bonds, which is also the difference between the far end and the near end of the yield curve in the graph. The normal spread of bonds shows an upward trend, that is, the yield curve slopes upward. When spreads turn negative, the yield curve inverts and looks downward.

One of the main factors influencing this spread is the forward rate expectations, which usually include the market's view of future short-term interest rates. If many people think there will be a recession in the future, short-term interest rates will fall during that period and long-term interest rates will fall as well. The far end of the yield curve falls, as a result, causing the entire yield curve to reverse its trend.

The reversal of the yield curve is significant for all investors because it is associated with recession risk. Banks, for example, borrow short-term money at lower interest rates in order to lend long-term at higher rates. The difference between these two interest rates, the positive interest margin, is their profit. But if a bank borrows short-term at high interest rates and lends to borrowers at low rates, the difference results in a negative spread.

In such an environment, banks lose money on loans. It doesn't necessarily apply to all loans, but it does make some loans unviable and others less profitable. This forces banks to reduce lending which in turn impacts the credit markets needed by businesses. When it becomes difficult for companies to borrow money, they have to cancel or delay projects and hiring. Weaker companies fail because they lose access to credit, which in turn leads to layoffs. When that happens, it takes on an average of about one year for the economy to slip into recession.

A further explanation is that in a recession, government bonds provide a safe haven from potential credit problems, leading to a rise in Treasury prices as investors want to avoid risks and thus accept low or even negative interest rates as a safe price. Conversely, since the interest rate on very short-term Treasury bonds is determined by the central bank's monetary intervention policy, and because some investors have portfolio preferences or hedging needs for longer maturities and maturities, the yield on Treasury bonds will tend to fall more than the bill rate, leading to a more downward yield curve.

Even when expectations are imprecise, speculators, or particularly well-informed traders may seek to capitalize on an increased probability of crisis by stocking up on the most liquid assets. They are in anticipation of trading them at favorable terms for illiquid assets that initially less informed investors may be forced to sell at the onset of a crisis (Acharya, Shin, and Yorulmazer, 2011).

As the relative prices of illiquid assets begin to weaken, speculators may bid for liquid low yielding assets as temporary stores of value until the market turns. Either for speculative leverage or for maintaining duration matches, some of these purchases of liquid assets may focus on longer maturity treasury notes and bonds, flattening or inverting the maturity yield curve (Erdogan, Bennett, and Ozyildirim, 2015).

## Literature Review

Over the past few decades, many researchers, including Laurent (1988), Harvey (1988), Stock and Watson (1989), and Chen (1991), have empirically formalized the idea that an inverted yield curve can represent a recession. Much of the researches have focused on using the term spread to predict the likelihood that the U.S. economy will fall into recession in the coming quarters.

In these studies, Estrella and Hardouvelis (1991), Estrella and Mishkin (1995, 1997) and Estrella (2005) provide the most comprehensive documentation of strong predictors of output propagation, including the ability to predict binary recession indicators in probabilistic regression. Hamilton and Kim (2002) confirmed the earlier results on the usefulness of the spread between long-term and short-term interest rates in predicting GDP growth and showed how to decompose this effect into expected effect and long-term premium effect. Estrella, Rodrigues, and Schich (2003) proved that the binary model is more stable than the continuous model used to predict the economic recession in Germany and the United States.

Moneta(2003) used the probit model, and in addition to the spread, also selected the short-term Treasury yield, OCED composite leading indicator, stock index yield, unemployment rate, and other variables to investigate the prediction effect of the euro zone's economic situation, but it is still the spread of the best prediction effect. On this basis, Bellego and Ferrara (2009) studied other financial indicator variables and found that corporate bond spread could also enhance the effect of prediction. Other scholars have conducted separate studies on countries in the Eurozone, and the predictive effect of the yield curve has been widely proved to be effective. Most of the literature suggests that Treasury maturity spreads can be predicted two to eight quarters in advance.

In most pieces of literature, the definition of economic recession (Y=1) basically adopts the time series cycle published by the National Bureau of Economic Research (NBER). The NBER is an unofficial research organization that publishes authoritative reports on the cycles of economic history. The NBER defines a recession as a complete process from a peak to a trough of a business cycle. But the problem is that the NBER's recession report is much later than the current economic situation, and the definition of the rule is not determined by data, there is no universality.

Yi Fang (2010) innovatively proposed that the Hodrick-Prescott Filter (HP Filter)(1997) could be used to decompose GDP series data into a long-term trend and a periodic fluctuation term, and then judge whether the economy is in a state of recession through the changes of fluctuation terms. In the long run, GDP will keep rising. But there have been periods when year-on-year GDP growth has slowed and not turned negative simply because of historical trends. Using the HP Filter can help see the fluctuations across history and define a recession mathematically.

## Data and Methodology

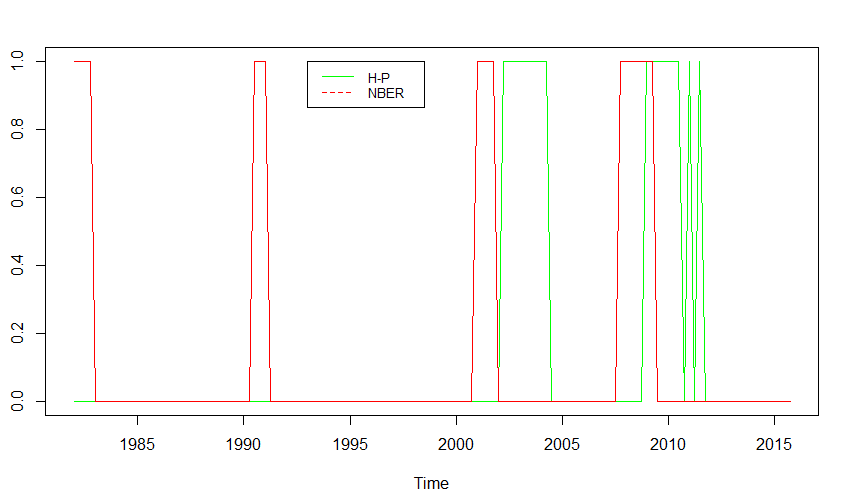
1. The definition of recession

In most works of literature, the NBER recession series is used as a standard of judging. The NBER is of high authority, but the problem is that it focuses on the great economic recession. And sometimes real GDP has been negative but nominal GDP still performed well which makes a few small recessions cannot be perceived. On the other hand, the NBER’s dates as to when U.S. recessions began and ended are based on the subjective judgment of the committee members, which raises two potential concerns. First, the announcements often come long after the event. For example, NBER waited until July 17, 2003, to announce that the 2001 recession ended in November 2001. Second, outsiders might wonder (perhaps without justification) whether the dates of announcements are entirely independent of political considerations. For example, there might be some benefit to the presidential incumbent of delaying a declaration that a recession had started or is accelerating a declaration that a recession had ended. For these reasons, it is worth exploring whether one could perform a similar function using purely objective summaries of the data.

In this paper, we use Hodrick-Prescott Filter to decompose GDP series data and then define when we can consider a recession. The purpose of using an HP filter is to separate the GDP series into a smooth long term and a fluctuation (ie.).

Then we minimize the following formula:

The first part shows the magnitude of the fluctuation and how close the actual GDP is to the long term , and the second part is how smooth the long term factor is. We want to get a balance between two of them when minimizing the whole formula and is the weight. The long term trend should be as similar to GDP as possible while the trend should also be as smooth as possible. It’s easy to imagine that when , we will get the exactly as the GDP as there’s no fluctuation. From Yi Fang（2010）suggestion, is set 1600 here. After we get , we calculate the mean and variance. When , we consider it’s a recession, .



According to the NBER’s definition, a recession is defined as a period from peak to trough. Our method, because of the calculation, will not consider the period before the trough which is in a recession in NBER’s data, and the period of gradual recovery after the trough will also be considered in a recession. So graphically our recession will be slightly later than the recession defined by the NBER

An Interesting point is that in certain periods after 2010, short-term volatility appeared in which means in the short term, the economy got through a few small recessions. The duration was short, and the frequency is high which is not consistent with the definition of the NBER significantly. We think that the reasons for the fluctuation are because our quantitative item is valid. It catches some imperceptible influence on economic development in the market. Comparing the regression of NBER and the recession on the independent variables we defined later, we can find that all the non-spread variables we selected are highly significant, while in the case of NBER sequence as the independent variable, only the Spread is significant

1. Model

Logit model has long been used and proved effective in the prediction of future economic recession. Thus we still use the model:

when the economy is in recession in period t. F(.) is :

SPREAD is the quarterly average of monthly interest rate spread data calculated as the difference between the interest rate on long term Treasury bonds and short term Treasury bills which shows the shape of the yield curve. Since we want to see the prediction effect, we need to use the spread k quarters before. In this case, we will compare a different combination of long (10, 5 years) and short (6, 3 months) term bonds to see which one has the best prediction effect. The estimate of should be negative.

RCE stands for the real consumption expenditure growth rate which reflects people’s consuming power. In the US, real consumption expenditure drives the demands. Only when the consumption expenditure increases, more industrial productions and services will be provided, which finally raise the capital. Thus, the decreasing consumption rate is not a good symbol for the national economy. The estimate of should be negative too.

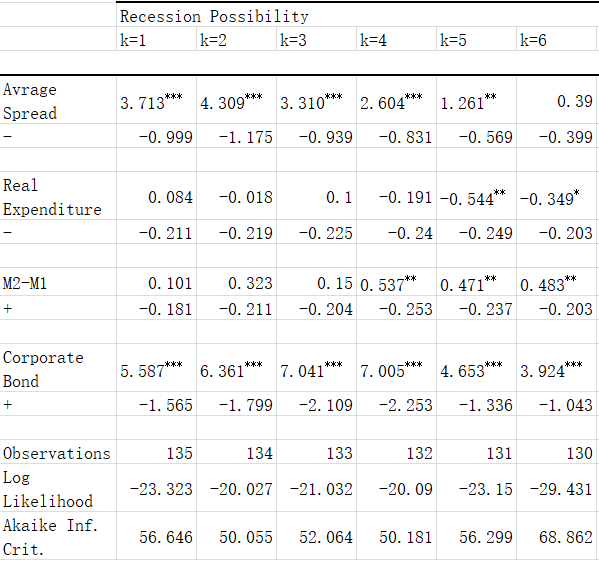
M is the difference in the growth rate between the Broad Money and the Narrow Money supply. The Narrow Money is M1 and the Broad Money is M2. M2-M1 is the majority of deposits in the banking system, which is a strong indicator of the direction of capital flow. If M is positive, people tend to save and are not optimistic about future economic prospects. The estimate of should be positive.

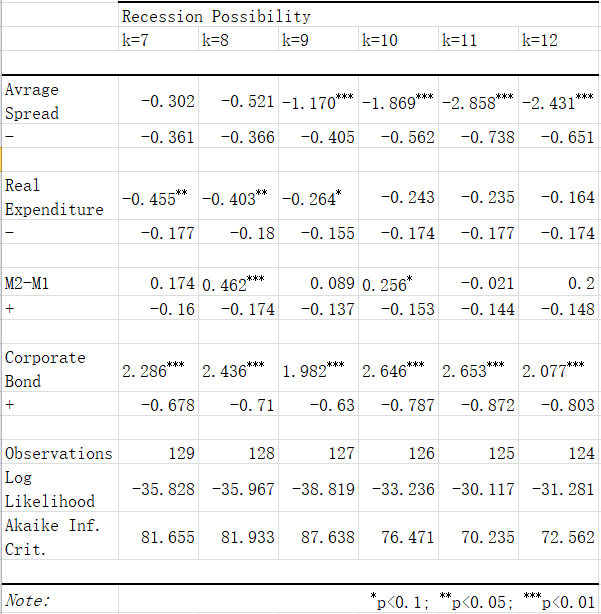
The last explanatory variable C is Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate. Compared with Treasury bonds, corporate bonds are riskier and their yield to maturity is higher. When the market expects a worse economic environment in the future, corporate bonds will be required to pay higher yields to compensate for these market risks. So credit spreads on corporate bonds should rise in proportion to future recessions.

All the data comes from FRED. We combined all the related data and made estimations from period 1984: Q1 to 2015: Q4.

## Empirical Results

In most of past papers, researchers chose different lags for their models, varying from 1 quarter to 2 years. We will compare the different combinations of long (10, 5 years) and short (6, 3 months) term bonds to see which one has the best prediction effect. There are 4 types of spread, (10-year rate - 6-month-rate), (10-year rate - 3-month-rate), (5-year rate - 6-month-rate), (5-year rate - 3-month-rate). We find that in most cases, the 4 specifications perform similarly through different time lags. To get enough information, we take the average of all 4 types.





We focus on Spread most. According to the explanation before, the coefficient should be negative, or it’s of no economic meaning. Thus we should eliminate all the cases with k <= 6.

The regression model indicates we should use different lags for different variables to improve their predicting power. After we train different models, we get the best combination as k =11 for Spread, k=7 for RCE, k=8 for M and k=7 for C.

We use data from 1984 to 2009 as our training set on which we build our model. We have considered from 2010 to 2015 as the test data set using which we have validated our model. As the historical data shows, we have far more 0(not in recession) than 1(in recession). So it’s reasonable to set the fitted value of 1(Recession=yes) =1 when fitted possibility is bigger than 0.85.

|  |  |  |
| --- | --- | --- |
|  | Forecasted No Recession | Forecasted Recession |
| True No Recession | 88 | 0 |
| True Recession | 6 | 7 |

Table 1-Forecasting results on Training Set Data of logit model

Based on our training set, the accuracy of the model is 94%. Such high accuracy means that the model is good in predicting a recession. However, the model has been trained using this dataset so it might be biased and the high accuracy maybe because of overfitting. To brush off such concerns, we used only a part of our dataset as the training set (1984-2009) and the remaining as the test set (2010-2015). The test set data is something our model has never seen before, hence its’ performance on this dataset will be a better metric to judge.

|  |  |  |
| --- | --- | --- |
|  | Forecasted No Recession | Forecasted Recession |
| True No Recession | 17 | 2 |
| True Recession | 1 | 4 |

Table 2-Forecasting Results on Test Set Data of logit model

When we tested our model on the out-of-sample data i.e. the test set, the forecasting accuracy was 87.5%. This means that the dependent variables such as the yield curve are good predictors of a recession.

## An Alternative Model – SVM

The support vector machine model is one of the advanced supervised machine learning algorithms used for binary data classification. Supervised machine learning algorithms use a sub-sample of our data, which are called the training set, to train the algorithm.

In this algorithm, we plot our data-point or the independent variable in an n-dimensional space where n is equal to the number of features or dependent variables present in the data. These data-points that help us train our algorithm are called support vectors. Our algorithm then finds the hyperplane in that dimension that classifies our data. Figure 1 shows how the hyperplane classifies our data into two separate classes.

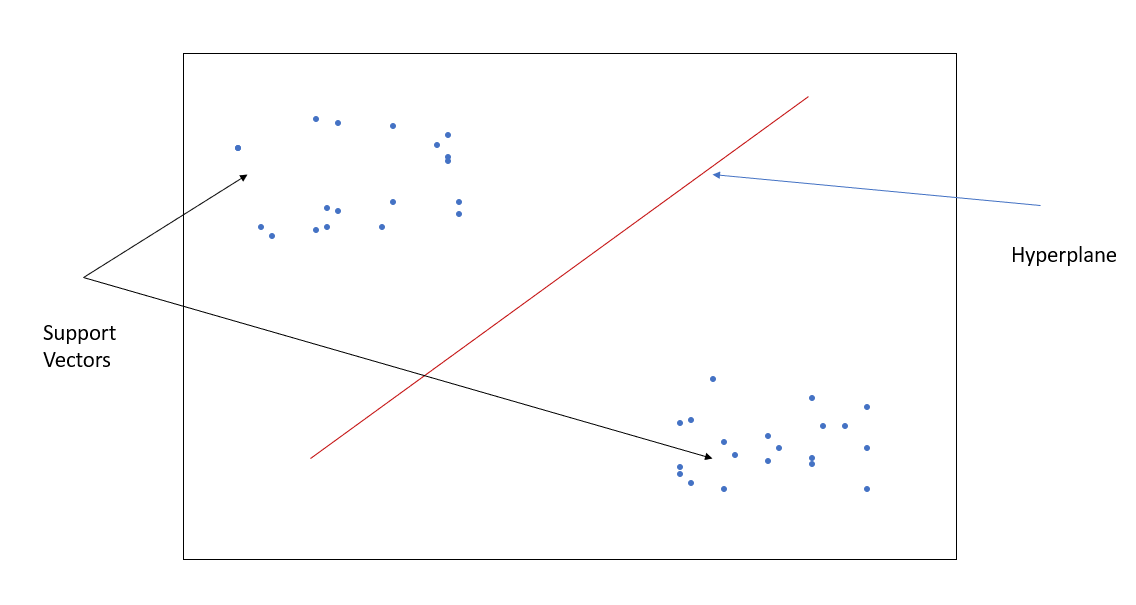


Figure -SVM Intuition

Determining the correct hyperplane:

SVM chooses the hyperplane that classifies the data correctly (Ray, 2017). Example, in the figure below SVM will choose the hyperplane A because it correctly separates our independent variable into two well-distinguished classes.

Figure -SVM: Selecting the Hyperplane, Source: Analytics Vidhya

However, there may be cases in which more than one hyperplane is able to do so. In such a case, our algorithm calculates margin (i.e. the distance of each data point from the hyperplane). The algorithm then chooses the hyperplane that has the maximum average margin. For example, in figure 3 we can see that the data is separated by three hyperplanes: A, B, and C. However, the margin for hyperplane C is higher than that of A and B (hyperplane C is more equidistant from both classes of the data), which is why the algorithm will choose hyperplane C. This is done because if we choose a hyperplane with low margin, chances of misclassification on out-of-sample data are high.

Figure - Hyperplane Selection, Source: Analytics Vidya

One major advantage of using this algorithm is that it works well with non-linear data as well. Kernel functions can transform a low dimensional input space into a high-dimensional space i.e. they can segregate data in a non-linear space. SVM uses kernel functions such as the radial kernel, the gaussian kernel, etc. to find a non-linear hyperplane that would separate the data clearly. (see below)



Figure -Using Kernels to Classify Non-Linear Data, Source: Analytics Vidya

As we described above, to implement SVM we need to determine the correct value for each parameter such as: what kernel to use, what is the coefficient of the kernel parameter to be used, etc. We implemented a grid search algorithm to find the best parameters for this function. A grid search algorithm in R takes a sample of data and runs different versions of SVM. We evaluated our data w.r.t to a linear, polynomial and radial kernel. Our grid search algorithm evaluated that for our data the best parameters to be used are: degree 2 and gamma 0.52.

We implemented this algorithm using the above parameters. Here are the results:

As we can see, the SVM model predicts with 91% accuracy on the training set and 87.5% accuracy on the test set.

|  |  |  |
| --- | --- | --- |
|  | Forecasted No Recession | Forecasted Recession |
| True No Recession | 86 | 2 |
| True Recession | 7 | 6 |

Table 3-Forecasting results on Training Set Data of SVM

|  |  |  |
| --- | --- | --- |
|  | Forecasted No Recession | Forecasted Recession |
| True No Recession | 19 | 0 |
| True Recession | 3 | 2 |

Table 4-Forecasting Results on Test Set Data of SVM

SVM is very good at identifying the possibility of a non-linear relationship between variables. Unlike logistic regression, SVM can generate complex decision boundaries. However, SVM does not generate probabilistic values w.r.t outputs rather directly gives us the decision. This reduces the interpretability of the model.

In our case, the performance of SVM is similar to that of the logit model in terms of accuracy w.r.t both in-sample and out-of-sample dataset. This further validates our claim that the yield curve, real consumer expenditure, corporate bonds, and M2-M1 are together all very strong predictors of whether a recession will occur or not in a particular timeframe.

## Conclusion

The spread of Yield curve is the most intuitive and important curve characteristic to predict an economic recession.

In this paper, we first define a recession with Hodrick-Prescott Filter which is quite different from the usual definition by NBER. HP filter separates the GDP series into a smooth long term and a fluctuation term . When , we consider it’s a recession, .

Then we investigate the forecasting ability of the yield curve and other financial variables compare logit models of different forecasting periods. We find all the 4 variables are significant but in different lags. After we check different models, we get the best combination as k =11 for Spread, k=7 for real consumption expenditure growth rate, k=8 for M2 growth rate minus M1 growth rate and k=7 for Corporate Bond minus Federal Funds Rate. With 1984—2009 data as the training set and 2010—2015 as the test set, we can achieve an overall forecasting accuracy of 87.5%. We further validated this hypothesis by using a different methodology i.e. Support Vector Machine and confirmed that the yield curve, RCE, corporate bonds, and M2-M1 are strong predictors of the recession.

## References:

Acharya, V. V., Shin, H., and Yorulmazer, T. (2011) Crisis resolution and bank liquidity, Review of Financial Studies 24, 2166–2205.

Erdogan, O., Bennett, P., and Ozyildirim, C. (2012) An early warning signal of financial crisis by using the deepness and liquidity in stock markets, The International Review of Applied Financial Issues and Economics 4-1, 58–63

Erdogan, O., Bennett, P., & Ozyildirim, C. (2015). Recession prediction using yield curve and stock market liquidity deviation measures. Review of Finance, 19(1), 407-422.

Estrella, A. (1998) A new measure of fit for equations with dichotomous dependent variables, Journal of Business and Economic Statistics 16, 198–205

Gogas, P., Papadimitriou, T., Matthaiou, M., & Chrysanthidou, E. (2015). Yield curve and recession forecasting in a machine learning framework. Computational Economics, 45(4), 635-645.

Gogas, P., Papadimitriou, T., & Chrysanthidou, E. (2015). Yield curve point triplets in recession forecasting. International Finance, 18(2), 207-226.

Harvey, C. (1988) The real term structure and consumption growth, Journal of Financial Economics 22, 305–333.

Moneta, F. (2005) Does the yield spread predict recessions in the Euro area? International Finance 8, 263–301.

Ozturk, H., & Pereira, L. F. V. N. (2013). Yield curve as a predictor of recessions: evidence from panel data. Emerging Markets Finance and Trade, 49(s5), 194-212.

Stock, J. H. and Watson, M. W. (1989) New indexes of coincident and leading economic indicators, in: O. J. Blanchard and S. Fischer (eds), NBER Macroeconomics Annual 1989, MIT Press, 352–394.

Ray, S. (2017, 9 13). Understanding Support Vector Machine algorithm from examples.

Retrieved from Analytics Vidya: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/