Comput Econ (2015) 45:635–645
DOI 10.1007/s10614-014-9432-0

Yield Curve and Recession Forecasting in a Machine Learning Framework

Periklis Gogas · Theophilos Papadimitriou · Maria Matthaiou · Efthymia Chrysanthidou

Accepted: 10 March 2014 / Published online: 27 March 2014

Abstract In this paper, we investigate the forecasting ability of the yield curve in terms of the U.S. real GDP cycle. More specifically, within a Machine Learning framework, we use data from a variety of short (treasury bills) and long term interest rates (bonds) for the period from 1976:Q3 to 2011:Q4 in conjunction with the real GDP for the same period, to create a model that can successfully forecast output fluctuations (inflation and output gaps) around its long-run trend. We focus our attention in correctly forecasting the instances of output gaps referred for the purposes of our analysis here as recessions. In this effort, we applied a Support Vector Machines technique for classification. The results show that we can achieve an overall forecasting accuracy of 66.7 and 100 % accuracy in forecasting recessions. These results are compared to the alternative standard logit and probit model, to provide further evidence about the significance of our original model.

Keywords Yield curve · Machine learning · SVM · Forecasting · GDP · Recession

1 Introduction

The yield curve depicts the relationship of the yields of bonds with different maturities issued by the same entity, usually a sovereign government. The stylized facts show that it is usually upward slopping reflecting a positive liquidity premium. An inverted yield curve (negative slope) is considered a sign of an upcoming recession.

After the recent crisis in the economy worldwide, the issue of forecasting output gaps is very topical and has significant implications in terms of economic policy. The prompt warning of an upcoming output downturn is of course extremely important

P. Gogas (✉) · T. Papadimitriou · M. Matthaiou · E. Chrysanthidou
Department of Economics, Democritus University of Thrace, 69100 Komotini, Greece
e-mail: perrygogas@gmail.com

to policymakers as they can swiftly adjust fiscal and monetary policy to either steer the economy away from a recession or dampen its impact to the real economy. The yield curve provide the information of bonds' returns of various maturities; short term interest rates have a direct correlation with monetary policy implementation while long term bonds' yield-to-maturity reflect the expectations of investors on the future economic activity (Estrella and Mishkin 1996, 1997). For this reason many studies have tried to exploit the information included in the yield curve in forecasting the economic activity of a country or a region.

The methodology used by these studies can be distinguished in two groups in terms of methodology employed to forecast the GDP. In one hand, we have linear models as we can see in studies such as Ang et al. (2006); Estrella and Mishkin (1997), Bernanke et al. (2005), Diebold and Li (2006), Calvao (2006) and Moench (2008) that use methods based on Ordinary Least Square (OLS) and Vector Auto Regression (VAR). Other studies have used non-linear models; see for example Ang et al. (2006), Chauvet and Potter (2002),Chionis et al. (2009), Estrella and Hardouvelis (1991), Estrella and Mishkin (1996),Christiansen (2012), Moneta (2005), Nyberg (2010) and Wright (2006) use probit models in order to forecast economic output by defining the value of a dummy variable where value "1" indicates an output gap and value "0" an inflationary gap.

Most of the studies use the spread of interest rates and more specifically the spread between a long and a short term rate, usually a 10 year government bond along with the 3 month Treasury bill, to forecast the economic activity. These empirical results provide evidence supporting the use of the yield curve in forecasting future economic activity. Christiansen (2012) studying a group of six countries uses a probit model in his effort to examine the forecasting ability of the yield curve in simultaneous recessions (i.e. simultaneous recessions occur if at least half of the investigated countries are in recession). The countries are Australia, Canada, the U.K., the U.S. and Japan. He concludes that the yield spreads of both Germany and the U.S. are leading indicators of simultaneous recessions. In his influential paper, Moneta (2005) employed a probit model to test the forecasting power of the yield curve. The study's interest lies on the prediction of recessions within the euro area. The results reveal that yield spreads are better recession forecasting tools, compared to other competitors. Nyberg (2010) introduces a new dynamic probit model which verifies the forecasting accuracy of the term spread. Moreover, he indicates that stock market returns appear to also have a strong predictive power. The study of Estrella (2005) confirms that the yield curve is a leading indicator for accurate forecasts of output and inflation, although he points out that the performance of the yield spread depends on the monetary authority's behavior.

There is also a number of studies claiming that the inclusion of macro variables and financial variables can improve the forecasting accuracy of such a model. Bordo and Haubrich (2008) use both slope and level (a short term interest rate such as the Federal Funds Rate) of the curve in their forecasting model. Ang and Piazzesi (2003) estimate models that include macroeconomic factors of inflation and economic growth and also measures the level, curvature and slope of the yield curve called latent factors. The study finds that macro factors explain up to 85 % of the forecast variance for long forecast horizons, while 60 % is explained for a one-month forecast horizon. In a later work Ang et al. (2006) use a dynamic model in order to characterize the expectations of GDP. The model that includes a short one-month term rate, a five years spread and lagged GDP is more efficient in forecasting real output. While the majority of the studies provides evidence that the slope of the yield curve has the ability to correctly forecast recessions, these recessions are basically localized based on past data. Estrella and Trubin (2006) provide a guideline for the formation of a yield curve indicator, to act as a real time forecasting tool. Rudebusch and Williams (2009) also point out that a real-time forecasting model based on the yield spread, outperforms professional forecasters in forecasting recessions a few quarters ahead. The work of Calvao (2006) takes as granted that the slope of the yield curve is a leading indicator of future output and examines four different models (a VAR, a threshold VAR, a structural break VAR and a structural break threshold VAR), to locate the best in terms of forecasting accuracy. The results show that the structural break threshold VAR model is superior.

Diebold et al. (2006) construct a model with the latent factors (level, slope and curvature) of the yield curve and macroeconomic variables of real activity, inflation and the stance of monetary policy, in order to find the dynamic interactions between the macroeconomy and the yield curve. The results show that the macroeconomic variables have a significant effect in future movements of the yield curve. Another study that includes a variety of macro variables is that of Moench (2008) which uses the common components of a large number of macroeconomic variables and the short rate term in order to estimate the model. In this article no latent factors are used and it concludes in a 50 % reduction of the root mean squared forecast errors for short and 20% reduction for long maturities. Finally, Chionis et al. (2009) include in their model the European Central Bank's euro area government benchmark bonds of various maturities, as well as non-monetary policy variables (the unemployment and a composite European stock price index). The results show that the yield curve augmented with the composite stock index has significant forecasting power in terms of the EU15 real output.

Nevertheless, there are some recent studies pointing out that the predictive power of the yield curve is weakened over time. Chinn and Kucko (2010) use an AR(1) model to show that although the yield curve is indeed a leading indicator of economic activity compared to other indicators, its predictive power the last decade was decreased. In a related study Schrimpf and Wang (2010) use a model of window selection/multiple break tests, to explore the predicting ability of the yield curve. The results suggest that while it can still forecast the economy's real activity better than other indicators, its forecasting ability is diminished.

The purpose of this study is to create a forecasting model using information from the yield curve that will be able to forecast future economic activity (real GDP above or below the potential GDP) by means of a Support Vector Machine (SVM) classifier. The SVM are binary classifiers that to the best of our knowledge are employed for the first time to forecast the GDP cycle using the information provided by the yield curve. For this reason the dependent variable series data are classified in two classes: (a) Class "−1" denoting an output gap (real output below long-run trend) and (b) Class "1" that denotes an inflationary gap (real output above the long run trend). The results show that both the monetary policy signals and the expectations of investors about future economic activity play an important role in forecasting future deviations of real

GDP from the long-run trend and our model is especially successful in forecasting future recessions.

The rest of the article is structured as follows: Sect. 2 outlines the data set used and describes the methodology of the SVM. The empirical results from out model selection process are presented in Sect. 3. Section 4 includes the comparison of our model with alternative models and finally Sect. 5 concludes.

2 Data and Methodology

Both short-term and long-term U.S. federal government interest rates were used in our study. The short-term interest rates are from Treasury-Bills with maturities of 3 and 6 months. The long-term interest rates are from the U.S government bonds with maturities of 2, 3, 5, 7 and 10 years. Real seasonally adjusted U.S. GDP is used as the basis for our binary dependent variables. The data span the period from 1967:Q3 to 2011:Q4 and they were available from the database of the Federal Reserve Bank of Saint Louis, Federal Reserve Economic Data (FRED). The real GDP figures were transformed into natural logarithms. We define recession in our study as deviations of GDP under the long-term trend (output gap). To decompose the GDP series and obtain the cyclical component and the trend series we use the Hodrick–Prescott filter by setting the λ parameter equal to 1,600 (Hodrick and Prescott 1997). The cyclical component series is then transformed into a binary (dummy) variable that has the value of 1 whenever the cyclical component is greater than zero and -1 elsewhere. We focused on forecasting the cyclical component of real GDP in 1, 2 and 3 quarters ahead forecasting windows using the SVM methodology.

The SVM represent a supervised machine learning (ML) method used for two-class data classification. Essentially, SVM try to locate a small number of data points from our dataset, called Support Vectors (SV) that can define a hyperplane separating the

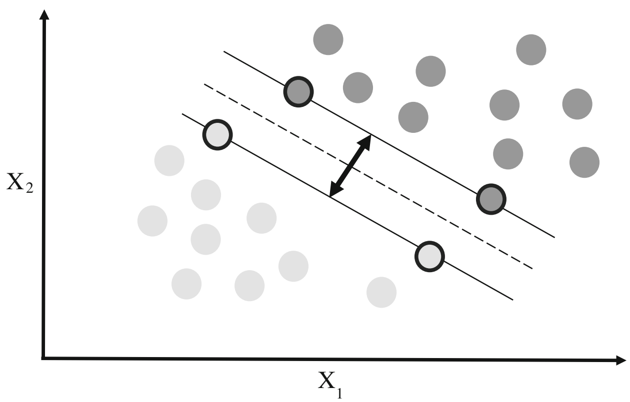

Fig. 1 Hyper plane selection and support vectors. The SV's are represented with the pronounced *black contour*, the *margin lines* are represented with the *continuous lines* and the hyper plane is represented with the *dotted line*

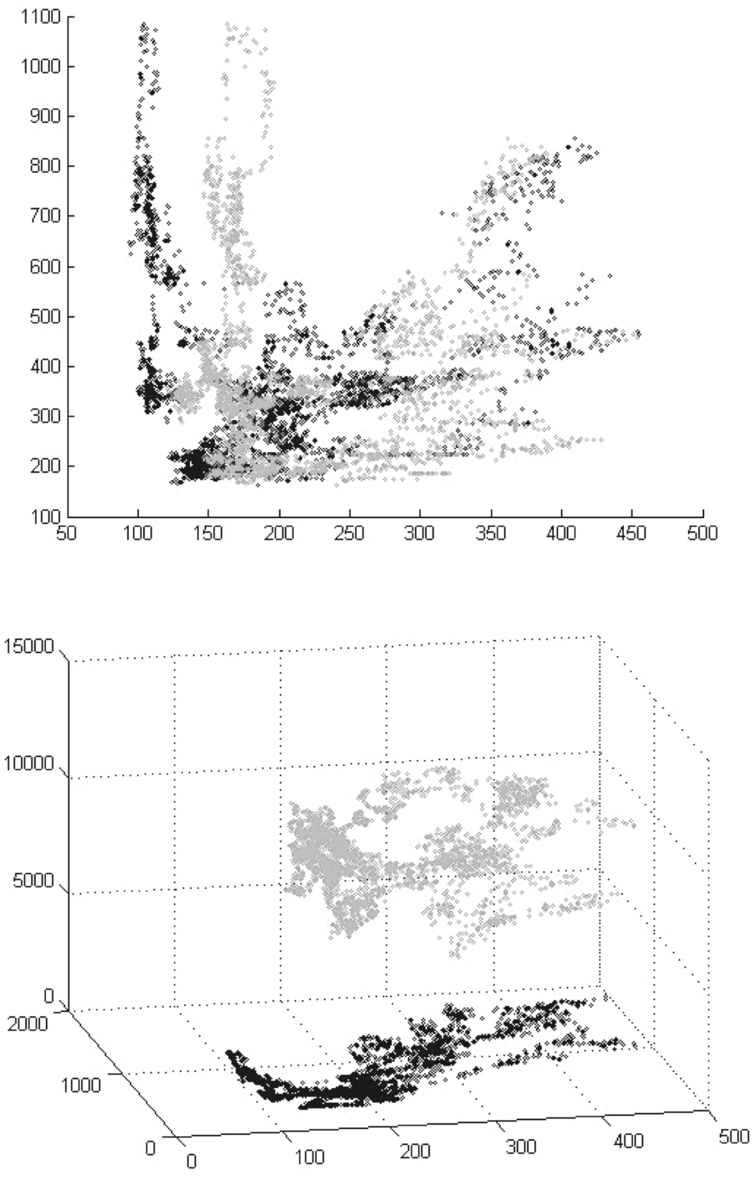

Fig. 2 Top the data space. The grey and black classes cannot be separated linearly. Bottom the feature space. The data points are projected into a richer space and the two classes are linearly separable

two classes' data points perfectly or as perfect as possible[1] (see Fig. 1). The method has two basic steps: the training step and the testing step. During the training step, the largest part of the dataset is used for the estimation of the separating hyperplane

[1] Imposing a weight parameter C for each erroneously classified data point in the minimization procedure.

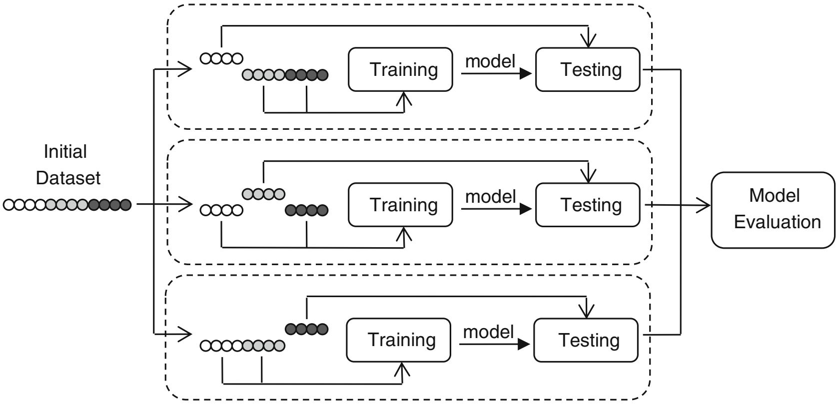

Fig. 3 Overview of a threefold cross validation evaluation system

Table 1 Descriptive statistics

	GDP	3 months interest	6 months interest	2 years interest	3 years interest	5 years interest	7 years interest	10 years interest
Mean	9,083	5,320	5,437	6,224	6,411	6,722	6,961	7,123
Median	9,080	5,035	5,105	5,935	6,190	6,425	6,700	6,855
Minimum	8,546	0,020	0,050	0,280	0,470	1,060	1,620	2,150
Maximum	9,505	15,020	14,740	15,540	15,500	15,410	15,330	15,150
SD	0,301	3,368	3,342	3,485	3,378	3,179	3,042	2,9110
C.V.	0,033	0,633	0,614	0,559	0,526	0,473	0,4369	0,408
Asymmetry	−0,126	0,592	0,529	0,469	0,487	0,570	0,620	0,686
Cumulant	−1, 337	0, 302	0, 128	−0, 079	−0, 076	−0, 055	−0, 083	−0, 099

though a minimization process; in the testing step, the generalization ability of the model is evaluated by investigating the model's performance in the small subset that was left aside in the first step.

In cases that the dataset is not linearly-separable, then SVM is coupled with a non-linear Kernel mapping procedure, projecting the data points to a higher dimensional space, called feature space, where the classes are linearly separable (see Fig. 2).

In our tests, we used (a) the Linear kernel and (b) Radial basis function (RBF). Their mapping function is:

- *Linear*: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^T \boldsymbol{x}_j$
- *Radial basis function- RBF:* $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma ||\boldsymbol{x}_i - \boldsymbol{x}_j||^2\right), \quad \gamma > 0$

The error tolerance parameter C and the RBF parameter γ were investigated through a grid search.

In a ML scheme, training results are overfitted when the model produced is significantly affected by possible noise in the sample in hand instead of the true underlying relationship that describes the phenomenon. Usually, overfitting yields a very high performance on the training step and significantly lower accuracy on the testing step.

k-fold cross validation, is adopted to avoid overfitting. The dataset is cut into k chunks and the training-testing steps are repeated k times. In each turn a different chunk is used as the test dataset, while the rest k-1 chunks are grouped together to form the training dataset. The model is evaluated by averaging the performance of the model on every fold (Fig. 3).

As it is indicated by its name—*Machine Learning*—the training procedure is responsible for the creation of the forecasting model (hyperplane), so we use the largest sample of our dataset for that cause. We enhanced this algorithm by adding a further step of out-of-sample evaluation. In this final step the trained model performance is tested against a totally unknown data chunk that was left out of the cross validation training step. As train data we used data from 1976:Q3 to 2007:Q2 while the out-of-sample data where from 2007:Q3 to 2011:Q4 (Table 1).

It is important to note that the period in which we try to forecast the GDP cycle is the period of the global financial crisis. This later crisis was a really unusual event in economic history. We decided to include it in our out-of-sample forecasting as if our model accurately forecasts the performance of the economy during this turbulent period with rare economic events, its value will be higher.

3 Empirical Results

We conducted 3,501 tests using the linear kernel and 293,371 tests using the RBF. The best accuracy was achieved by the RBF model with error tolerance parameter $C = 34.5$ and $\gamma = 0.6$. The Cross Validation test accuracy (in-sample) was 73.3 %, and the out-of-sample overall accuracy was 66.7 %. The model triggered 6 false alarms (cases that the model erroneously provides a "recession" flag). However this is a burden we can tolerate, since in the same time it forecasted correctly all the true "recession" cases (Table 2).

The same out of sample forecasting score was achieved from a linear model though with lower Cross Validation Accuracy (63.3 %).

Our results provide important new insights into the forecasting ability of the yield curve. Policy makers, the government and the central bank, can use such a model to forecast the deviations of real output from its long-run trend, the full employment output. By doing so they can efficiently design and implement the necessary monetary and fiscal policy mix that can steer the economy to the full employment GDP. The results show that this scheme was able to identify and forecast one quarter ahead all instances of actual output gaps below the long run trend. Our out-of-sample data included eight quarters of actual output below full employment and our model forecasted correctly all of them. Moreover, we correctly forecasted four quarters of inflationary gaps. The only downside of this model is that it produced some false alarms in the case of below-trend

Table 2 The estimated cases and the ground truth for SVM tests

	True growth	True recession
Forecasted growth	4 (40 %)	0 (0 %)
Forecasted recession	6 (60 %)	8 (100 %)

output: even though the model forecasted all eight such instances of unemployment gap it also provided six false alarms. These are cases where the model forecasted an unemployment gap and the actual output was above trend. Evaluating the overall performance of this model we can state that it has significant power to forecast upcoming recessions so that the government and the central bank can react swiftly and effectively to dampen the effects of the downturn in real output or even avoid it all together. As the model in the effort to forecast all recessionary events produces some false alarms, we can expect that the expansionary monetary and fiscal policy associated with those false warnings will result in some inflationary pressure to the economy.

4 Comparison with Alternative Models

In this section we compare the forecasting accuracy of the SVM model developed in the previous section to alternative models. Two well established econometric methods that are widely used in the empirical forecasting literature on the issue for binary forecasting/classification. These are the logit and probit models.

The probit model is a regression that estimates the probability:

$$P_r(y = 1|\boldsymbol{x}, \boldsymbol{\beta})\Phi(\boldsymbol{\beta}^T\boldsymbol{x}) \tag{1}$$

Where $P_r(y = 1)$ is the probability that an event y will occur or not and is assumed to be determined by a set of independent variables $\boldsymbol{x}$. y is a dummy variable, taking the value of 1 or 0. In our case, a value of zero (0) indicates an economy below the trend (recession) and the value of one (1) indicates an economy above trend. Φ denotes the cumulative distribution function (CDF) of the standard normal distribution, y is the dependent binary variable and $\boldsymbol{\beta}$ is the vector to be estimated.

Logit models also examine the probability that an event occurs and it is based upon the CDF of the logistic distribution:

$$P_r(y = 1) = e^{\boldsymbol{\beta}^T\boldsymbol{x}}/(1 + e^{\boldsymbol{\beta}^T\boldsymbol{x}}) \tag{2}$$

The explanatory variables used are the short-term interest rates (T-Bills) with 3 and 6 months maturities and the long-term interest rates of U.S. government bonds with maturities of 2, 3, 5, 7 and 10 years for the period from 1976:Q3 to 2011:Q4.

We first applied the logit and probit model for the data from 1976:Q3 to 2007:Q2 (in-sample-data) and then we test the forecasting ability of both models by comparing the forecasting outcomes with the actual observations out-of-sample from 2007:Q3 to 2011:Q4. Both models achieved a 50 % overall forecasting accuracy. Compared to the SVM models, the overall forecasting performance of both probit and logit models is inferior. However the recession accuracy (i.e. the accuracy of the correctly forecasted recessions) is 100 % for both probit and logit models (Tables 3, 4).

We also used all three methodologies (SVM, logit and probit) discussed in Sects. 3 and 4 above for two alternative datasets: (a) we use as the dependent variable the monthly recession dates with the same seven interest rates as explanatory variables but in monthly frequency and (b) with the original binary series for the dependent variable

Table 3 Forecasting performance of the logit model

	True growth	True recession
Forecasted growth	1 (10 %)	0 (0 %)
Forecasted recession	9 (90 %)	8 (100 %)

Table 4 Forecasting performance of the probit model

	True growth	True recession
Forecasted growth	1 (10 %)	0 (0 %)
Forecasted Recession	9 (90%)	8 (100%)

Table 5 Summary of alternative forecasting models' forecasting efficiency

Model	Cycle definition	Interest rates	Freq.	Methodology	Accuracy total	Accuracy growth	Accuracy recession
1	H-P	Yield curve	Quarterly	SVM Linear	66.7	40	100
2	H-P	Yield curve	Quarterly	SVM RBF	66.7	40	100
3	H-P	Yield curve	Quarterly	Probit	50	10	100
4	H-P	Yield curve	Quarterly	Logit	50	10	100
5	NBER	Yield curve	Monthly	SVM Linear	73.52	100	0
6	NBER	Yield curve	Monthly	SVM RBF	73.52	100	0
7	NBER	Yield curve	Monthly	Probit	73.52	100	0
8	NBER	Yield curve	Monthly	Logit	73.52	100	0
9	H-P	Gurkaynak et al.	Quarterly	SVM Linear	40	71.42	12.5
10	H-P	Gurkaynak et al.	Quarterly	SVM RBF	46.67	28.57	62.5
11	H-P	Gurkaynak et al.	Quarterly	Probit	43.75	12.5	75
12	H-P	Gurkaynak et al.	Quarterly	Logit	43.75	12.5	75

we use as explanatory variables the twenty-seven implied yields from Gürkaynak et al. (2007). The eight additional models estimated in this way provided an out-of-sample forecasting accuracy that was significantly lower than our original SVM models trained.These results are summarized in Table 5.

5 Conclusion

This paper is the first empirical investigation on the relation between the yield curve and an economy's real output, using an SVM classifier. Our goal is to create a forecasting model that can accurately inform us about future output gaps. Such a model can be a very useful tool for policy makers: governments and central banks. It can be employed for the efficient and prompt implementation of the fiscal and monetary policy mix in aiming to minimize deviations of real output from the long-term trend. We apply the SVM methodology, using information from the yield curve; i.e. the interest rates on U.S. federal government treasury bills for 3 and 6 months maturity and long-term government bonds with maturities 2, 3, 5, 7 and 10 years. We used both the linear and the RBF kernels in an out-of-sample one-quarter-ahead forecasting scheme. Our out-of-sample data included eight quarters of below-trend output and ten quarters of inflationary gaps. The proposed methodology reached a 73.3 % overall accuracy using the RBF kernel. Nonetheless, focusing on the incidents of an unemployment gap (an actual GDP value below the long-run trend) our model had a perfect forecasting (eight out of the eight cases), though it produced six false alarms: cases where the model forecasted an unemployment gap but the actual GDP was above trend. It is noteworthy that the perfect recession forecasting was reached using an out-of-sample part including data from the current global financial crisis. The proposed methodology outperformed classic econometric methods on overall forecasting accuracy. Thus, this can be effectively exploited by policy makers within the government and the central bank and by implementing the necessary policy mix they can dampen the effects of the recession, minimize its duration, or steer the economy away from it all together. As the model provides some false alarms, we expect that implementing fiscal and monetary policy in this manner may put some inflationary pressure to the economy.

Acknowledgments "This research has been co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: THALES. Investing in knowledge society through the European Social Fund (MIS 380292)."

References

Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics, 50*(4), 745–787.

Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about GDP? *Journal of Economics*, *131*, 1–2.

Bordo, M. D., & Haubrich, J. G. (2008). Forecasting with the yield curve; level, slope, and output 1875–1997. *Economic Letters, 99*(1), 48–50.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quart J Econ*, *120*, 387–422.

Calvao, A. B. (2006). Structural break threshold VARs for predicting US recessions using the spread. *Journal of Applied Econometrics*, *21*, 463–487.

Chauvet, M., & Potter, S. (2002). Predicting a recession: Evidence from the yield curve in the presence of structural breaks. *Economic Letters*, *77*(2), 245–253.

Christiansen, C. (2012). Predicting severe simultaneous recessions using yield spreads as leading indicators. *Journal of International Money and Finance, 32*, 1032–1043.

Chinn, M. D., & Kucko, K. J., (2010). *The predictive power of the yield curve across countries and time*. Working Paper 16398, NBER.

Chionis, D., Gogas, P., & Pragidis, I. (2009). Predicting European union recessions in the euro era: The yield curve as a forecasting tool of economic activity. *International Advances in Economic Research*, *16*(1), 1–10.

Diebold, F., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Economics*, *130*, 337–364.

Diebold, F. X., Rudebusch, G. D., & Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics, 131*, 309–338.

Estrella, A. (2005). Why does the yield curve predict output and inflation. *Economic Journal*, *115*, 722–744.

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *Journal of Finance*, *46*, 555–576.

Estrella, A., & Mishkin, F. S. (1997). The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank. *European Economic Review*, *41*(7), 1375–1401.

Estrella, A., & Mishkin, F. S. (1996, June). The yield curve as a predictor of U.S recessions. *Federal Reserve Bank of New York: Current Issues in Economic and Finance, 2*(7), 1–6.

Estrella, A., & Trubin, M. R. (2006). The Yield Curve as a Leading Indicator: Some Practical Issues. *Current Issues in Economics and Finance, Federal Reserve Bank of New York*, *12*(1), 1–7.

Gürkaynak, R. S., Sack, B., & Wright, J. H. (2007). The US treasury yield curve: 1961 to the present. *Journal of Monetary Economics 54*(8), 2291–2304.

Hodrick, R., & Prescott, E. P. (1997). Postwar business cycles: An empirical investigation. *J Money Credit Bank*, *29*(1), 1–16.

Moench, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. *Journal of Econometrics*, *146*(1), 26–43.

Moneta, F. (2005). Does the yield spread predict recessions in the Euro area? *International Finance*, *8*(2), 263–301.

Nyberg H (2010) Dynamic probit models and financial variables in recessions forecasting. Journal of Forecasting 29: 2115–230

Rudebusch, G. D., & Williams, J. C. (2009). Forecasting recessions: The puzzle of the enduring power of the yield curve. *Journal of Business and Economics, Statistics*, *27*(4), 492–503.

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

Wright, J., (2006, February). The yield curve and predicting recessions. Finance and Economics Discussion Series, Federal Reserve Board 2006–7.