

A re-evaluation of the term spread as a leading indicator

Vasilios Plakandaras ^{a,*}, Periklis Gogas^a, Theophilos Papadimitriou^a and Rangan Gupta^b

^a Department of Economics, Democritus University of Thrace, Greece

^b Department of Economics, University of Pretoria, South Africa

*Corresponding author. University Campus, Komotini, Greece. email: vplakand@econ.duth.gr

Abstract

Forecasting the evolution path of macroeconomic variables has always been of keen interest to policy makers and market participants. A common tool used in the relevant forecasting literature is the term spread of Treasury bond yields. In this paper, we decompose the term spread into an expectation and a term premium component and evaluate the informational content of each component in forecasting the GDP growth rate and inflation in various forecasting horizons. In doing so, we employ alternative decomposition procedures and introduce the Support Vector Regression (SVR) methodology from the field of Machine Learning, coupled with linear and non-linear kernels as a novel forecasting method in the field. Using rolling windows in producing point and conditional probability distribution forecasts we find that neither the term spread, nor its decomposition components possess the ability to accurately forecast output growth or inflation. Our findings extend the existing literature, since they are focused on an explicit out-of-sample evaluation in contrast to most existing empirical studies that produce only in-sample forecasts. To strengthen our findings, we also consider several control variables suggested in the relevant literature without significant qualitative differences from the initial results. The main innovation of our approach stems from the use of the non-linear Support Vectors Machine methodology, that is introduced for the first time in this line of research for forecasting out-of-sample.

Keywords: Inflation; GDP; Forecasting; Support Vector Machines; Term Premium

JEL classification: C22; C53; E47

1. Introduction

There is a voluminous literature that empirically studies the relationship between macroeconomic and financial evolution. One such case, more specifically, is the potential relation between the term spread (the difference between long-run and short-run yields of treasury bonds) and the observed fluctuations of output growth and inflation. [Kessel \(1965\)](#), first explores the term structure of Treasury bonds along with the business cycle suggesting a direct linkage between the two. Since then, the ability of the term spread to forecast various macroeconomic variables such as the economic activity, inflation and consumption has become a “stylized fact” among economists. Nonetheless, in a survey paper, [Wheelock and Wohar \(2009\)](#) argue that the ability of the term spread to forecast economic activity has declined over the years. In a similar context, [Rossi and Sekhposyan \(2010\)](#) detect a significant drop in the forecasting ability in the mid-1970s. Using rolling windows to forecast output growth and considering many potential regressors, the authors find that the forecasting ability of the term spread is significant in the pre-1970s period and declines ever since.

In this paper, we re-access the ability of the term spread to forecast a) economic activity in terms of GDP growth and b) inflation, based -for the first time in the relevant literature- on a machine learning methodology, namely, the Support Vector Regression (SVR). Our work is directly motivated by the continuously emerging successful applications of machine learning in the economics and finance

literature.

Blumenstock, Cadamuro, and On (2015) use data from mobile phones in Rwanda to provide estimates of personal income at the individual level, while on a similar context Glaeser, Kominers, Luca, and Naik (2018) estimate block-level income in New York City and Boston based on images from Google Street View. Antweiler and Murray (2004) use machine learning in identifying text messages on online financial boards as bullish, bearish, or neither. The application on a sample of 1.5 million messages shows that online boards dedicated to the financial markets help explain volatility on the stock market. Another application of machine learning is credit scoring; Bjorkegren and Grissen (2018) forecast the likelihood of loan repayment based on mobile phone data.

In her recent paper, Athey (2018) highlights the rapid growth of the machine learning literature in economics, suggesting that the use of machine learning in forecasting and decision making in economics will become pervasive. Mullainathan and Spiess (2017) review much of the recent interaction between machine learning and economics arguing that unlike typical econometrics, the focus of machine learning is on forecasting rather than causal interactions and structure identification. Thus, they conclude that machine learning could become a useful analysis tool that would provide not only new data that cannot be constructed and observed by earlier methods, but could shift research to new questions that could not be formed up to this point. Ng (2017) proposes the usefulness of machine learning in dealing with modern, real-life, big data analysis in a fashion that can be of use to central banking and monetary policy determination. Athey (2018) goes as far as stating that with the growing number of applied economics papers in machine learning, in a decade machine learning will be included in every economics textbook.

From the field of Machine Learning, the Support Vector Regression (SVR) methodology, seems to produce superior forecasting performance compared to traditional statistical and econometric methodologies, according to the relevant literature. Rubio, Pomares, Rojas, and Herrera (2011) compare the forecasting ability of SVR to traditional Ordinary Least Squares (OLS) models on a variety of artificial and real datasets. The data sets used are selected to cover a range of possible practical situations: linear or non-linear, deterministic or stochastic and discrete or continuous time series. The empirical findings of Rubio et al. (2011) suggest that the SVR models produce a lower forecasting error than the OLS ones for all data series. Härdle, Lee, Schäfer, and Yeh (2009) find that the discrete dependent variable Support Vector Machines (SVM) models outperform the traditional logit/probit models in forecasting the default risk of companies. Gogas, Papadimitriou, and Agrapetidou (2018) in a recent application of SVM models in forecasting bank failures, reach similar conclusions and report an overall out-of-sample forecasting accuracy of 99.22% in an extended sample of 1443 U.S. banks.

Plakandaras, Gupta, Gogas, and Papadimitriou (2015a) compare the forecasting ability of SVR and OLS regressions in forecasting house prices, concluding in favor of the SVR models. Finally, Plakandaras, Papadimitriou, and Gogas (2015b) compare the forecasting ability of several machine learning and econometric methodologies in forecasting selected exchange rates. The empirical findings once again suggest that the forecasting superiority of machine learning methodologies over their econometric counterparts is pervasive across exchange rates. Driven by the aforementioned studies, the selection of the SVR methodology is justified by several factors: it does not impose initial assumptions on the data, it is robust in non-linear and stochastic data structures, the forecasting performance is higher and the computational ease in training and selecting the best among a variety of models is superior to alternative methods.

Nevertheless, despite all above advantages of ML, its applications in the macroeconomics literature are sparse. Among the very few, Gogas, Papadimitriou, Matthaïou, and Chrysanthidou (2015) investigate the ability of the yield curve to directionally forecast output and inflation gaps around their long-run trends by comparing several SVM and logit models. They use a variety of short-term and long-term interest rate combinations under both machine learning and traditional econometrics. Their empirical findings suggest that the machine learning models outperform the econometric logit and probit in terms of out-of-sample forecasting accuracy.

Our paper introduces several innovations. First, we employ SVR models to assess the impact of the yield spread on economic activity that -to the best of our knowledge- has not been used before in the literature. Second, we decompose the term spread into an expectation and a term premium component. Third, while previous approaches were merely focused on directional forecasting, in this paper we attempt to forecast the actual levels of the GDP. Finally, we evaluate the ability of the term spread and the term premium to act as leading indicators of inflation and output growth fluctuations, in a strictly out-of-sample forecasting exercise. In doing so, we build on the macro-finance category of models and more specifically on the family of affine¹ models of the term spread, in order to extract an estimate of the otherwise unobserved term premium. In contrast to previous approaches that complete their analysis to the beginning of the 2008 global financial crisis, we use an extended data sample covering the period 1971Q4-2014Q4. Additionally, we evaluate the forecasting ability of the models constructing both point and conditional density forecasts. Density forecasts are currently reported by most Central Banks (see for instance the reports of the Bank of England) and present information not for the mean value of the (point) forecast but also for the uncertainty entailed to the forecast.

Our empirical findings regarding GDP growth are in line with the existing literature. As in Rossi and Sehkoposyan (2010) and Dewachter et al. (2012), we detect a significant decline in the ability to forecast GDP growth both based on the term spread or its components. This result is highlighted especially in the post-2008 period. In contrast to Wright (2011), inflation forecasts based on the term spread are characterized by excess uncertainty, a feature revealed by our density forecasts. The difference between our findings and the ones reported in the relevant literature could be attributed to our explicit focus on out-of-sample forecasting and the evaluation of the entire forecasted distribution, in contrast to the point forecasts of all previous efforts. Thus, we are able to uncover valuable information on the underlying mechanism of the reversal in the forecasting ability of the term spread and its components by evaluating the entire forecasting distribution. Finally, we argue that given the prevalent evidence against the ability of the term spread and its components to forecast GDP growth and inflation, the term spread and its components could be considered as irresolute indicators regardless of the modeling approach.

¹ By “affine” we refer to models where the term premium and short-term expectations to maturity are all affine functions of the term spread.

2. Literature review

The literature on forecasting economic activity and inflation based on the yield curve dates back to [Kessel \(1965\)](#). The expectations hypothesis states that the interest rate of a given maturity should equal the average of expected short-term rates over the period until maturity. In other words, bonds of different maturities are assumed perfect substitutes which, in turn, implies that the term spread is zero. Empirically, however, the yield curve, on average, tends to slope upwards ([Mishkin, 2007](#)). Given that the short-term interest rates reflect the implemented monetary policy while the market's expectations over future economic activity are embodied in the interest rates on long-term bonds, the slope of the yield curve should have predictive power over economic activity and inflation. Nonetheless, given the usual positive slope of the yield curve, the expectations theory would imply that almost always agents expect an economic downturn in the future.

The failure of the expectations hypothesis to explain the positive slope of the yield curve is addressed by the liquidity premium theory. Investors prefer to hold bonds of shorter maturity to minimize the interest rate risk. In order to hold longer maturity bonds, investors require a liquidity (risk) premium to compensate for the additional uncertainty of locking away their money for longer periods. The term premium, which increases with the maturity of the bond, explains the stylized fact of an upward sloping yield curve. Thus, the term premium is the difference between the long-term interest rate and the average (expected) short-term rates to maturity.

The existing literature in forecasting macroeconomic variables typically segregates the term spread into an expected short-term interest rate and a term premium component, according to the maturity of the long-term bond used in constructing the term spread. The informational content of the two components is exploited separately or jointly in forecasting output, inflation and output growth. [Hamilton and Kim \(2002\)](#) propose a model with time-varying expectations for the decomposition of the term spread and argue that both components are important in forecasting the GDP growth rate, with the expected short-term component being slightly more important than the latter. [Estrella and Mishkin \(1997\)](#) train a probit model to evaluate the forecasting power of interest rates, interest rate spreads, stock prices and indices, monetary aggregates and other individual macroeconomic indicators for the period 1959Q-1995Q1 to forecasting U.S. recessions. Based on a recursive out-of-sample exercise for the period 1971Q1-1995Q1 they find that the term spread of the 10-year with the 3-month Treasury bond has the highest predictive ability for recessions up to 3 quarters ahead. In a similar vein, [Stock and Watson \(1989\)](#) observe an increase in the forecasting ability of an experimental leading indicator index, but only when the term spread is included in the construction of the index.

The decomposition of the term spread into an expected short-term interest rate and a term premium component is not straightforward and in the relevant finance literature have been proposed various approaches. [Duffee \(2002\)](#) decomposes the term spread using a dynamic three-factor model of the level, the slope and the curvature of the yield curve. In this approach the factors are selected from the yield curve according to the maturity of the term-spread under examination. He finds that the actual short-term interest rate and the expected short-term component can forecast the future evolution of bond prices during financial tranquil periods, while the term premium captures “fly-to-quality” phenomena during recessions and economic turmoil. Extending the work of [Duffee, \(2002\)](#), [Kim and Wright \(2005\)](#) estimate an arbitrage-free three-factor model that adopts a different normalization of the factors based on survey data. The extracted term premium and the expected short-term rate indicate that the significant drop in the interest rates in the post 2004 period should be attributed to falling term premiums rather than decreasing expected short-term rates.

Using a different approach, [Cochrane and Piazzesi \(2005\)](#) estimate bond term premiums by regressing excess bond returns on the term structure of forward rates. The excess bond returns are constructed as the difference of a bond with an n -year maturity over a bond of annual maturity. This gives a “tent-shaped” function that explains up to 44 percent of the variation in excess bond returns of different maturities. In their seminal paper [Joslin, Singleton, and Zhu \(2011\)](#) propose an arbitrage-free, three factor dynamic affine structure model that imposes certain restrictions to the distributions of bond yields. The so-called “canonical model” allows for the zero-coupon yields to act as salient observable states rather than latent (not observable) factors used in all previous models. The empirical findings from the application to forecasting future treasury bills indicate that the “canonical model” adheres more closely to the evolution of interest rates than the term spread. [Adrian, Crump, and Moench \(2013\)](#) extend the approach of [Joslin et al. \(2011\)](#) to an affine term structure model based on observable state variables extracted from regressions on market portfolios, alleviating the need for the existence of zero-coupon bonds in all maturities.

The decomposition of the yield curve to the expected short-term rate and the term premium spurred several empirical efforts in evaluating the forecasting relevance of each individual component on inflation and economic activity. In an in-sample forecasting exercise, [Favero, Kaminska, and Söderström \(2005\)](#) find that the best forecasting model of GDP growth rate is the one including contemporaneously the expected short-term rate, the term spread and the term premium as regressors. Nevertheless, the value of the expected short-term interest rate coefficient is hard to explain in a macroeconomic context given that macroeconomic variables do not have the necessary time to adjust to short-term fluctuations of the term spread. [Ang et al. \(2006\)](#) propose the use of unobserved risk factor models in forecasting the U.S. GDP, based on Principal Component Analysis (PCA). More specifically, the GDP growth rate is modelled as a function of the first two principal components of interest rates of different maturities and lagged values of GDP growth rate. The two principal components are treated as unobserved risk factors and are fed into a Vector Autoregressive (VAR) model. The use of only two risk factors reduces the complexity of the model, while the empirical findings suggest that in comparison to the term spread and the interest rates, only the risk factors forecast consistently the GDP growth rate consistently in-sample and out-of-sample.

[Dewachter, Iania, and Lyrio \(2012\)](#) examine an affine term structure model in forecasting in-sample the output gap, the GDP growth rate and the GDP deflator of the U.S. economy for the period 1971Q2 to 2008Q4. In this approach, the term premium is estimated as a function of macroeconomic and financial risk factors. Although the decomposition of the term spread does not seem important in describing the evolution path of the GDP growth rate and the output gap, it adheres very closely to inflation in various forecasting horizons. The inclusion of the short-term interest rate and the lagged values of the dependent as control variables does not affect the

forecasting ability of the models, leading to the conclusion that the forecasting performance stems from the risk factors. [Lange \(2018\)](#) builds on the macro-finance literature decomposing term spreads into investors' expectations about the path of future short-term interest rates plus a time-varying term premium, based on a bivariate VAR model. His findings suggest a strong relationship between financial latent factors and macroeconomic variables, since almost 50 percent of the term spread's variance is attributed to shocks to the short-term treasury bond interest rate and 20 percent to an inflation shock. Exploiting the dynamic structure of the VAR model, [Evgenidis, Philippas, and Siriopoulos \(2019\)](#) extend the aforementioned approach to an international level, arguing that the linkages between financial and macroeconomic variables are also strong and robust not only at the national but at the international level as well.

[Wright \(2011\)](#) decomposes forward five-to-ten² year interest rates into an expected short-term interest rate and a term premium component for 10 OECD countries. In doing so, he builds on the empirical findings of [Ang et al. \(2006\)](#) and [Joslin et al. \(2011\)](#) and develops risk factor models based on the first three principal components of interest rates of zero-coupon bonds augmented with macroeconomic variables. The macroeconomic variables are considered as "unspanned" variables in the sense that they are important in forecasting future interest rates but do not possess explanatory power over contemporaneous yield curves. Then, using the extracted term premiums and survey data, he performs panel regressions in order to uncover the data generating process of quarterly inflation for the period 1981Q1 – 2009Q4. He concludes that neither the risk factors nor the survey data can efficiently explain the evolution path of inflation for the entire time period and to all countries. In a later comment on the work of [Wright \(2011\)](#), [Bauer, Rudebusch, and Wu \(2014\)](#) argue that the maximum likelihood estimation approach followed by Wright is mis-specified and propose a bias correction procedure. Although the bias corrected term premiums appear to possess countercyclical behavior consistent with the empirical and theoretical suggestions on the behavior of term premiums, the proposed correction does not improve the overall forecasting accuracy. Recently, [Morell \(2018\)](#) examines the predicting ability of the term spread and the term premium in forecasting economic output in the U.S. using a Dynamic Stochastic General Equilibrium (DSGE) model, with data spanning the period 1966Q1-2006Q4. His results indicate that the predictive power of the term spread has weakened from the mid-1980s to present, since the focus of the monetary policy by the Federal Reserve shifted from promoting economic growth to controlling prices.

Overall, the literature does not reach to a definite conclusion concerning the ability of the term spread and its components to forecast output growth and inflation. The empirical findings suggest that the forecasting capability changes over time, while it is dependent on the selected decomposition scheme and the forecasting model.

3. Affine models of term premiums

In order to obtain the expected short-term and the term premium we use the affine Gaussian Dynamic Term Structure Model (DTSM) of [Joslin et al. \(2011\)](#). The interest rate of an n -period zero-coupon bond can be split into the expected short-term interest rate and a term premium as follows:

$$y_t^{(n)} = \frac{1}{n} \sum_{\tau=0}^{n-1} E_t [y_{t+\tau}^{(1)}] + x_t^{(n)} \quad (1)$$

where the first part of equation (1) represents the expected average of the one period short-term interest rates and the second is the term premium. The expected short-term interest rate is the average rate if the investor chooses to invest iteratively his money on bond with one period maturity for the entire period n . The term premium is the excess return he demands for losing this flexibility and locking away his money for n -periods. Following [Ludvigson and Ng \(2009\)](#) the term premium component can be written as the average risk premium from holding one period zero-coupon bonds to maturity:

$$rx_t^{(n)} = \frac{1}{n} \sum_{\tau=0}^{n-1} E_t [rx_{t+\tau,t+\tau+1}^{(n-\tau)}] \quad (2)$$

$$\text{Subject to } \begin{cases} rx_{t,\tau+1}^{(n)} = \ln \left(\frac{P_{t+1}^{(n-1)}}{P_t^{(n)}} \right) - y_t^{(1)} \\ y_t^{(n)} = -\ln(P_t^{(n)})/n \end{cases}$$

where $rx_{t+\tau,t+\tau+1}^{(n-\tau)}$ denotes the one period excess return of an n -period bond, $P_t^{(n)}$ is the price at time t and $y_t^{(n)}$ its yield. Assuming a lognormal nominal pricing kernel m_t , it must be $P_t^{(n)} = E_t [\prod_{j=1}^n m_{t+j}]$. Following the class of affine structure models of [Duffee \(2002\)](#) and [Cochrane and Piazzesi \(2005\)](#), the price $P_t^{(n)}$ is given by an affine structure model of the state variables X_t , according to

$$P_t^{(n)} = \exp(a_n + B_n' X_t) \quad (3)$$

² The term five-to-ten year interest rates refers to the expected interest rate of a five year interest rate bond that will be bought in five years from now, in the sense of a forward interest rate. The term is used commonly in the relevant literature in order to describe the expectations over a future interest rate.

$$\text{Subject to } \begin{cases} m_{t+1} = \exp\left(-\delta_0 - \delta_1' X_t - \frac{1}{2} \lambda_t' \Sigma \lambda_t - \lambda_t' \varepsilon_{t+1}\right) \\ \lambda_t = \lambda_0 + \lambda_1 X_t \\ \varepsilon \sim N(0, \Sigma) \end{cases}$$

with $\lambda_t = \lambda_0 + \lambda_1 X_t$ expressing an affine structure model of state variables. If the matrix of the state variables is observable and the sign restrictions imposed in the model are not more than the necessary for its identification, then it follows a VAR model:

$$X_{t+1} = \mu + \Phi X_t + \varepsilon_{t+1} \quad (4)$$

and thus a_n would be a vector $m \times 1$ and B_n an $m \times m$ matrix that satisfy the recursions:

$$a_{n+1} = -\delta_0 + a_n + B_n'(\mu - \Sigma \lambda_0) + \frac{1}{2} B_n' \Sigma B_n \quad (5)$$

$$B_{n+1} = (\Phi - \Sigma \lambda_1)' B_n - \delta_1 \quad (6)$$

starting from $a_1 = -\delta_0$ and $B_1 = -\delta_1$. The state variables could be macroeconomic or financial variables or simply risk factors of the yield curve, assumed to be observable. Given that we can estimate μ , Σ and Φ from the VAR model of equation (4) the remaining variables can be estimated by minimizing the sum of square errors between the actual and the fitted yields according to:

$$\{\tilde{\lambda}_0, \tilde{\lambda}_1, \tilde{\delta}_0, \tilde{\delta}_1\} = \underset{\lambda_0, \lambda_1, \delta_0, \delta_1}{\operatorname{argmin}} \sum_t \Sigma_t (y_t(n) - \tilde{y}_t(n))^2 \quad (7)$$

given that $\tilde{y}_t(n) = -(a_n + B_n' X_t)/n$ are the model implied yields. The difference between the model implied yields and the average (expected) short-term interest rate for the entire period n to maturity is the term premium. In this paper, we examine both macroeconomic and financial state variables. As macroeconomic variables we consider the GDP deflator (π_t) and the output gap (\tilde{y}_t), while as financial the effective federal funds rate as the central bank policy rate (i_t^{cb}) and the first three principal components of zero-coupon Treasury bonds.³ The first three principal component represent the level, the slope and the curvature of the yield curve (Joslin et al., 2011). Thus, the state variable vector is:

$$X_t = [\pi_t, \tilde{y}_t, i_t^{cb}, l_1, l_2, l_3] \quad (8)$$

4. Data and methodology

4.1. The data

We compile a dataset of quarterly observations for the U.S. spanning the period 1971Q4-2014Q4. The GDP, GDP deflator and the Effective Federal Funds rate are obtained from the Federal Reserve Bank of St. Louis database. The selection of the quarterly GDP deflator instead of the monthly CPI is dictated by the frequency of our macroeconomic variables that we considered as “unspanned” variables. The output gap is the difference between the nominal potential GDP projections of the Congressional Budget Office (CBO) and the actual nominal GDP. The three latent yield curve risk factors are the first three principal components of the zero-coupon treasury bonds of maturity 1, 4, 8, 12, 20 and 40 quarters, where the first five maturities are from the Fama-Bliss database and the 40 quarters treasury bonds from Gurkaynak, Sack, and Wright (2007).⁴ According to Joslin et al. (2011) the first three principal components include almost all the variability of the treasury bonds and there is no excess gain in including more principal components. The GDP, the GDP deflator and the output gap are transformed into their (natural) logarithmic forms.⁵ In our empirical analysis we consider the 3-month Treasury Bill (TB3) and the Chicago Fed National Activity Index (CFNAI) compiled by the Federal Reserve Bank of St. Louis as control variables, to isolate monetary policy and economic activity spillovers to the estimation.

Applying the Bai and Perron (2003) multiple structural breaks and the Enders and Lee unit root (2012) tests we find that the GDP, the GDP deflator and the TB3 series follow a unit root process, while the CFNAI is stationary in levels.⁶ As a result, in the empirical section of our analysis we use the first differences of the natural logs for the GDP, the GDP deflator (i.e., quarter-on-quarter inflation rate) and the TB3, while we stick to levels of the CFNAI.

In order to test for the existence of nonlinearities in the data we apply the Regression Error Specification Test (RESET) proposed by

³ Apart from the OLS estimation of the VAR we also examined the bias corrected methodology of Bauer et al. (2014). In all situations the results were quantitatively similar and are presented in the online Appendix.

⁴ These data are based on fitted Nelson-Siegel-Svensson curves and are available at <http://www.federalreserve.gov/Pubs/feds/2006/200628/feds200628.xls>.

⁵ Due to the space restrictions of a scientific journal, the descriptive statistics of all timeseries are not included here and are available upon request.

⁶ The results of the structural break and unit root tests are reported in the online Appendix.

Ramsey (1969, pp. 350–371) on the output and inflation growth series. The test examines the null hypothesis that the underlying data generating mechanism is linear, against the alternative that the data generating mechanism is not. The residuals of a linear model that describes a linear process should not be correlated with the regressors or polynomial functions of the dependent variable. Therefore, a regression of the residuals on polynomials of the dependent variable and the initial regressors (of the linear model) should have little explanatory power if the true process is linear. In Table 1 we report the F-test statistics of the regression of the residuals of a linear model on different polynomial expressions of the dependent variable.

The results reveal that we can reject the null hypothesis of linearity for polynomial of higher order to the squared real GDP growth, while we cannot reject for polynomial order higher than the squared GDP deflator growth rate. This empirical finding suggests that we should proceed with nonlinear forecasting models that could adhere more closely to the nonlinear data generating mechanism.

4.2. Support Vector Regression

The Support Vector Regression is a direct extension of the classic Support Vector Machines algorithm. The algorithm proposed by Vapnik, Boser, and Guyon (1992) and later extended by Cortes and Vapnik (1995) originates from the field of statistical learning. When it comes to regression, the basic idea is to find a function that has at most a predetermined deviation from the actual values of the dataset. In other words, errors are not of interest if they don't violate a predefined threshold ε ; only errors higher than ε are penalized. The vectors that define the "error tolerance band" are identified through a minimization procedure and are called "Support Vectors" (SV).

One of the main advantages of SVR in comparison to other machine learning techniques is that it yields a convex minimization problem with a unique global minimum, avoiding local minima. The model is built in two steps: the training and the testing step. In the training step, the largest part of the dataset is used for the estimation of the Support Vectors that define the band. In the testing step, the generalization ability of the model is evaluated by checking the model's performance in the small subset that was left aside during training. Using cross-validation techniques a universal and not sample-specific solution is achieved, addressing the issue of possible overfitting.

For a training dataset $D = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$, $x_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}$, $i = 1, 2, \dots, n$, where x_i is a vector of independent variables and y_i is the dependent variable the linear regression function takes the form of $y = f(x) = w^T x + b$. This is achieved by solving:

$$\min \left(\frac{1}{2} w^2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*) \right) \quad (9)$$

$$\text{subject to} \begin{cases} y_i - (w x_i + b) \leq \varepsilon + \zeta_i \\ (w x_i + b) - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases}$$

where ε defines the width of the tolerance band, and ζ_i, ζ_i^* are slack variables controlled through a penalty parameter C (see Fig. 1). All the points inside the tolerance band have $\zeta_i, \zeta_i^* = 0$. System (9) describes a convex quadratic optimization problem with linear constraints and it has a unique solution. The first part of the objective function controls the generalization ability of the regression, by imposing the "flatness" of our model controlled through the Euclidean norm w . The second part of the objective function controls the regression fit to the training data (by increasing C we penalize with a bigger weight on any point outside the tolerance band i.e. with $\zeta_i \geq 0$ or $\zeta_i^* \geq 0$). The key element in the SVR concept is to find the balance between the two parts in the objective function that are controlled by the ε and C parameters.

Using the Lagrange multipliers in System (9) the solution is given by:

$$w = \sum_{i=1}^n (c_i - c_i^*) x_i \quad (10)$$

and

$$y = \sum_{i=1}^n (c_i - c_i^*) x_i^T x$$

with the coefficient $c_i, c_i^* = 0$ for all non SVs. Thus, the SVR model is defined solely by its SVs.

Table 1
RESET F-test values.

Variable/Powers	2	3	4
$\Delta(\ln(rGDP))$	0.044	5.742***	3.383**
$\Delta(\ln(GDP \text{ deflator}))$	3.245*	1.645	1.520

Note: *, ** and *** denote rejection of the null hypothesis regarding linearity at 10%, 5% and 1% level of significance.

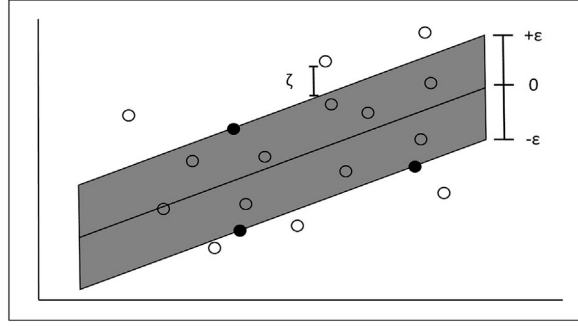


Fig. 1. Upper and lower threshold on error tolerance indicated with letter ε . The boundaries of the error tolerance band are defined by the Support Vectors (SVs) denoted with the black filled points. Forecasted values greater than ε get a penalty ζ according to their distance from the tolerance accepted band.

The underlying data generating processes of real life phenomena are rarely linear. Thus, formulating linear models to describe them often fail to describe correctly the data generating process. In order to tackle with this drawback, SVR models are coupled with kernel functions. The so-called “kernel trick” follows the projection idea while ensuring minimum computational cost: the dataset is mapped in an inner product space, where the projection is performed using only dot products within the original space through “special” kernel functions, instead of explicitly computing the mapping of each data point. The produced SVR model coupled with non-linear kernels is non-linear as well. In our simulations we test three kernels: the linear, the radial basis function (RBF) and the polynomial kernel. The mathematical representation of each kernel is:

$$\text{Linear } K_1(x_1, x_2) = x_1^T x_2 \quad (12)$$

$$\text{RBF } K_2(x_1, x_2) = e^{-\gamma \|x_1 - x_2\|^2} \quad (13)$$

$$\text{Polynomial } K_3(x_1, x_2) = (\gamma x_1^T x_2 + r)^d \quad (14)$$

with factors d, r, γ representing kernel parameters.

5. Empirical results

In this section we study the ability of the term spread and its components to forecast GDP growth rate and inflation. More specifically, we study the decomposition of the term spread S into a spread expectations component E and a term premium component π . The term spread refers to the spread between the interest rates of a long-term bond with maturity n quarters and the bond with maturity of one quarter as:

$$S_t^{(n)} = y_t^{(n)} - y_t^{(1)} = E_t^{(n)} + \pi_t^{(n)} \quad (15)$$

In Fig. 4 we depict the calculated term premium between a bond of 40-quarter maturity and a bond with 1-quarter maturity.⁷ The term premium exhibits a significant drop during the 1990s as compared to the pre-1984 period, which could be attributed to the monetary policy of the Federal reserve in targeting inflation during the so-called “Great Moderation Period” and the greater integration of the bond markets in later years that reduced market segmentation. Nevertheless, during the 2008 financial crisis the term premium returns to higher levels as a result of increased uncertainty and gradually declines as the economy stabilizes. For comparison reasons, in Fig. 2 we also depict the term premiums of Adrian et al. (2013),⁸ and Kim and Wright (2005).⁹ The term premium we extracted in this study is similar to the one of Adrian et al. (2013) despite the differences in the two decomposition schemes. In contrast, the term premium of Kim and Wright (2005) is quite different, due to the different normalization schemes between their decomposition methodology and the one we follow in this paper. These findings are comforting, given that despite that the three decomposition schemes follow different paths in extracting the components of the term spread, the two converge to similar components. Thus, our decomposition is not based on a “data-mining” mechanical process that always converges to the same results with any other previous study. Moreover, there is replicability between studies, meaning that there is some proximity with the “true” data generating process.

⁷ The decomposition of the term spread for bonds of smaller than 40-quarters maturity provide quantitative equivalent results and are available from the authors upon request.

⁸ The data are available at http://www.newyorkfed.org/research/staff_reports/sr340.html.

⁹ The data are available at <http://www.federalreserve.gov/pubs/feds/2005/200533/200533abs.html>.

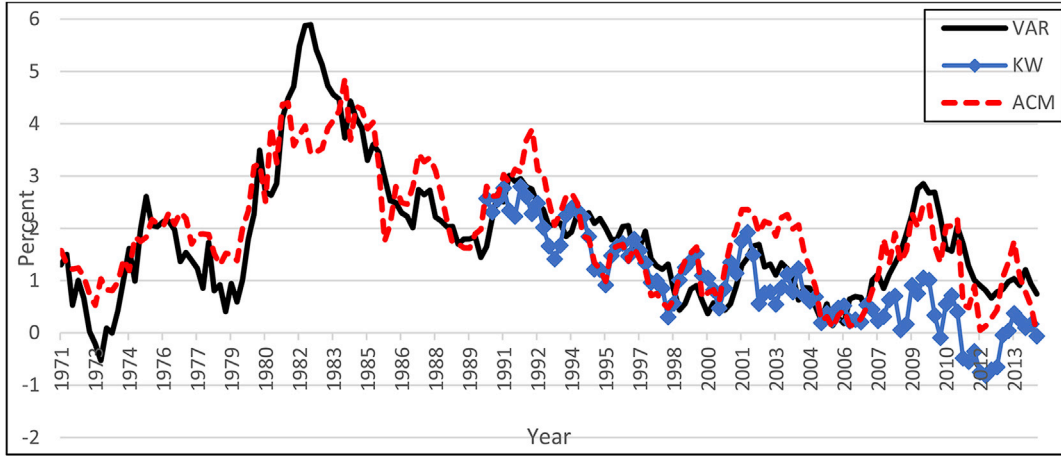


Fig. 2. The term premium for the three decomposition schemes. The VAR scheme refers to the term premium according to the decomposition methodology proposed by [Wright \(2011\)](#) that is used in this paper. ACM is according to the [Adrian et al. \(2013\)](#) and KW to the [Kim and Wright \(2005\)](#) term premiums schemes respectively.

5.1. GDP growth rate point forecasts

In order to forecast the GDP growth rate h periods ahead, we estimate the following general model:

$$g_{t+h} = c + \beta_1 S_t^{(h)} + \beta_2 g_{t-1} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h} \quad (16)$$

where c is a constant term, S_t expresses the term spread, $y_t^{(1)}$ expresses the one quarter interest rate (TB3) and $CFNAI_t$ is the CFNAI index. As reported in the literature review section, several authors claim that the decomposition components (the expectation and the term premium component) are able to capture the data generating process of GDP growth ([Ang et al., 2006](#); [Favero et al., 2005](#); [Hamilton & Kim, 2002](#)). In order to test for the informational content of each component in forecasting GDP growth, we alternatively examine each component separately in the model, replacing the term spread.

The lagged value of the GDP growth of the previous quarter (g_{t-1}), the short-term interest rate and the economic activity index are typically used in the literature as control variables. The selection of the lag order of the GDP growth rate and of the short-term interest rate is based on previous studies (see for instance [Dewachter et al., 2012](#) and [Rossi & Sekhposyan, 2010](#)). We forecast $h = 4, 8, 12, 16, 20$ and 40 quarters ahead in rolling windows of 40 observations (10 years) and consider the Random Walk (RW) model as the benchmark model to compare with. The use of rolling windows is selected to account for the existence of structural breaks in the data (see the Data section for more details) and to observe potential time variations in the forecasting accuracy of the models. As in [Dewachter et al. \(2012\)](#), the decomposition of the term spread is estimated iteratively in each rolling window. The long-term interest rate selected in each occasion is the one of the forecasting horizon (e.g. in $h = 4$ when we examine the decomposition of the four quarter with the one quarter interest rate term spread).

For comparison reasons we also estimate an autoregressive model of the GDP with the lag order selected in each rolling window according to the minimum Schwarz Information Criterion ([Schwarz, 1978](#)). Each forecasting exercise is repeated twice; once with an OLS and once with an SVR model. The latter is coupled with the linear, the RBF and the polynomial kernel. We measure the forecasting performance of all models based on the Mean Absolute Percentage Error (MAPE) given by $MAPE = 100 \times \frac{1}{n} \sum_{t=1}^{i=n} \left| \frac{\hat{g}_t - g_t}{g_t} \right|$ on the entire out-of-sample forecasted sample. The out-of-sample accuracy for each model is reported in [Table 2](#)¹⁰.

As we observe from [Table 2](#), all models outperform the RW model regardless of the selected forecasting horizon. We test for the statistical significance of all results using the [Diebold and Mariano \(1995\)](#) test. According to this test, we examine the null hypothesis of equal forecasting ability of each model with the RW model. We reject the null hypothesis for all models and in all horizons at the 5% level of significance. Since the AR model is a nested model of models 5 and 6 we also consider the [Clark and West \(2007\)](#) and the [McCracken \(2007\)](#) tests. The specific tests examine the null hypothesis that the out-of-sample forecasting accuracy of the (small) AR model is higher than the alternative (5 or 6) model with additional variables. Across all forecasting horizons but the 40-quarters ahead, the most accurate model employs the SVR methodology coupled with the linear kernel, while in the 40-quarters ahead horizon the smallest MAPE is the one employing OLS regression. Nevertheless, the forecasting performance of all structural models is qualitatively

¹⁰ In the SVR models we report only the results of the most accurate kernel. The results for all models are reported in the online Appendix available at <https://pithos.oceanos.grnet.gr/public/6Ef0YSP5YbgLIJSJrf14Z6>. In the same Appendix the interested reader can find details regarding the in-sample fit of our models. The forecasting accuracy between the in-sample and out-of-sample forecasts is similar, suggesting that the reported results are not a statistical artifact due to possible overfitting issues.

Table 2
MAPE of Point GDP forecasts.

Forecasting Horizon	h = 4	h = 8	h = 12	h = 16	h = 20	h = 40
RW model	$g_{t+h} = g_t$					
	0.829	0.824	0.788	0.783	0.779	0.774
AR model	$g_{t+h} = c + \beta g_{t-k} + \varepsilon_{t+h}$					
OLS	0.500	0.479	0.468	0.473	0.473	0.466
SVR-linear	0.480	0.486	0.454	0.467	0.453	0.435
SVR-RBF	0.525	0.507	0.491	0.464	0.461	0.445
SVR-Poly	0.503	0.547	0.609	0.515	0.472	0.459
Model 1	$g_{t+h} = c + \beta S_t^{(h)} + \varepsilon_{t+h}$					
OLS	0.474	0.449	0.442	0.478	0.470	0.453
SVR-linear	0.464	0.454	0.434	0.465	0.452*	0.436
SVR-RBF	0.472	0.465	0.490	0.486	0.493	0.457
SVR-Poly	0.488	0.600	0.509	0.505	0.540	0.519
Model 2	$g_{t+h} = c + \beta (E_t^{(h)} + x_t^{(h)}) + \varepsilon_{t+h}$					
OLS	0.471	0.448	0.459	0.502	0.500	0.519
SVR-linear	0.460	0.438*	0.436	0.474	0.476	0.489
SVR-RBF	0.465	0.460	0.460	0.490	0.468	0.439
SVR-Poly	0.498	0.593	0.620	0.738	0.894	0.676
Model 3	$g_{t+h} = c + \beta E_t^{(h)} + \varepsilon_{t+h}$					
OLS	0.472	0.454	0.436	0.463	0.479	0.419*
SVR-linear	0.458	0.459	0.423	0.444*	0.475	0.425
SVR-RBF	0.492	0.484	0.441	0.495	0.534	0.451
SVR-Poly	0.464	0.527	0.469	0.489	0.527	0.526
Model 4	$g_{t+h} = c + \beta x_t^{(h)} + \varepsilon_{t+h}$					
OLS	0.467	0.447	0.440	0.463	0.494	0.508
SVR-linear	0.457*	0.454	0.420*	0.461	0.462	0.463
SVR-RBF	0.493	0.505	0.475	0.553	0.517	0.490
SVR-Poly	0.601	0.578	0.714	1.690	2.539	0.627
Model 5	$g_{t+h} = c + \beta_1 S_t^{(h)} + \beta_2 g_{t-1} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h}$					
OLS	0.535	0.496 ⁺	0.543 ⁺	0.557	0.583	0.623 ⁺
SVR-linear	0.499	0.517 ⁺	0.483 ⁺	0.515	0.531	0.558
SVR-RBF	0.476 ⁺	0.446 ⁺	0.442 ⁺	0.454 ⁺	0.461 ⁺	0.432
SVR-Poly	0.550	0.661	0.649 ⁺	0.577	0.583	0.575 ⁺
Model 6	$g_{t+h} = c + \beta_1 (E_t^{(h)} + x_t^{(h)}) + \beta_2 g_{t-1} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h}$					
OLS	0.492 ⁺	0.488 ⁺	0.499	0.530	0.503	0.523
SVR-linear	0.483	0.482 ⁺	0.465 ⁺	0.495	0.492	0.488
SVR-RBF	0.481	0.450 ⁺	0.431 ⁺	0.482 ⁺	0.470 ⁺	0.453 ⁺
SVR-Poly	0.511 ⁺	0.566	0.564	0.549 ⁺	0.556	0.484

Note: ⁺ denotes rejection of the null hypothesis of the Clark and West (2007) and the McCracken (2007) tests of equal forecasting accuracy with the autoregressive model only for models 5 and 6. The forecasting accuracy of all models against the RW model is measured based on the Diebold-Mariano statistic and the null hypothesis of equal forecasting accuracy is rejected for all tests. All tests are performed at the 5% level of significance. The smallest MAPE for each forecasting horizon is denoted with an * and depicted in grey. All MAPE values are percentages.

similar (with differences observed after the 4th decimal¹¹). Thus, the use of the actual or the fitted term spread and the decomposition into components does not improve the out-of-sample forecasting accuracy. Moreover, the inclusion of the control variables provides qualitatively similar results. This finding corroborates with the empirical results of Dewachter et al. (2012).

5.2. Inflation point forecasts

As with GDP forecasts, we examine the ability of the term spread and its components in forecasting inflation. The latter is measured by the growth rate of the GDP deflator. Faust and Wright (2011) evaluate 17 different models in forecasting various inflation measures. They report that especially for the GDP deflator, the model of Atkeson and Ohanian (2001) based on a RW (RW-AO) model outperforms most of the alternatives considered in their study. The RW-AO model suggests that the future inflation can be successfully approximated

¹¹ all MAPE values are percentages.

by the average rate of inflation during the last 4 quarters $\pi_{t-3,t} = \frac{1}{4} \sum_{j=0}^3 \pi_{t-j}$. Thus we include the $\pi_{t-3,t}$ as a control variable along with the one quarter interest rate and the CFNAI index. As with the GDP growth estimation, we forecast the future inflation based on alternative forms of the general model:

$$\pi_{t+h} = c + \beta_1 S_t^{(h)} + \beta_2 \pi_{t-3,t} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h} \quad (17)$$

The results per model and methodology are reported in Table 3.¹²

According to the Diebold-Mariano test, all models outperform the RW model at the 5% level of significance. The SVR methodology is more accurate than the OLS in all cases, but overall the difference in the forecasting performance between all models is again

Table 3
MAPE of Point inflation forecasts.

Forecasting Horizon	h=4	h=8	h=12	h=16	h=20	h=40
RW model	$\pi_{t+h} = \pi_t$					
	2.328	2.275	2.229	2.195	2.200	1.963
AR model	$\pi_{t+h} = c + \beta \pi_{t-k} + \varepsilon_{t+h}$					
OLS	0.910	1.129	1.186	1.152	1.083	0.736
SVR-linear	0.915	1.082	1.158	1.111	1.006	0.704
SVR-RBF	1.033	1.141	1.046	1.067	1.002	0.695*
SVR-Poly	0.982	1.111	1.583	1.535	1.124	0.772
Model 1	$\pi_{t+h} = c + \beta S_t^{(h)} + \varepsilon_{t+h}$					
OLS	1.184	1.187	1.193	1.085	0.747	1.184
SVR-linear	1.145	1.176	1.170	0.984	0.722	1.145
SVR-RBF	1.297	1.226	1.373	1.012	0.802	1.297
SVR-Poly	1.224	1.191	1.254	1.320	0.734	1.224
Model 2	$\pi_{t+h} = c + \beta (E_t^{(h)} + x_t^{(h)}) + \varepsilon_{t+h}$					
OLS	1.141	1.093	0.998	0.834	0.782	1.141
SVR-linear	1.058	1.066	0.944	0.837	0.773	1.058
SVR-RBF	0.985	0.998	1.018	0.896	0.765	0.985
SVR-Poly	1.340	1.088	0.979	1.593	1.149	1.340
Model 3	$\pi_{t+h} = c + \beta E_t^{(h)} + \varepsilon_{t+h}$					
OLS	1.238	1.159	1.018	0.887	0.805	1.238
SVR-linear	1.140	1.109	0.959	0.829	0.745	1.140
SVR-RBF	1.180	1.141	0.926	0.920	0.864	1.180
SVR-Poly	1.343	1.238	1.055	0.908	0.978	1.343
Model 4	$\pi_{t+h} = c + \beta x_t^{(h)} + \varepsilon_{t+h}$					
OLS	1.182	1.015	0.885	0.856	0.794	1.182
SVR-linear	1.102	0.956	0.828*	0.821	0.771	1.102
SVR-RBF	1.091	1.003	0.861	0.803	0.817	1.091
SVR-Poly	1.565	1.413	2.408	97.175	1.440	1.565
Model 5	$\pi_{t+h} = c + \beta_1 S_t^{(h)} + \beta_2 \pi_{t-3,t} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h}$					
OLS	0.992 ⁺	0.986 ⁺	0.889 ⁺	0.948 ⁺	1.068 ⁺	0.992 ⁺
SVR-linear	0.981 ⁺	1.018 ⁺	0.870 ⁺	0.794 ⁺	0.916 ⁺	0.981 ⁺
SVR-RBF	0.846	0.937 ⁺	0.923 ⁺	0.880 ⁺	0.700* ⁺	0.846 ⁺
SVR-Poly	1.071	1.116 ⁺	0.838 ⁺	0.992	0.928	1.071 ⁺
Model 6	$\pi_{t+h} = c + \beta_1 (E_t^{(h)} + x_t^{(h)}) + \beta_2 \pi_{t-3,t} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h}$					
OLS	0.958 ⁺	0.940 ⁺	0.887 ⁺	0.795 ⁺	0.718 ⁺	0.958 ⁺
SVR-linear	0.948 ⁺	0.963 ⁺	0.846 ⁺	0.767* ⁺	0.724 ⁺	0.948 ⁺
SVR-RBF	0.841* ⁺	0.859* ⁺	0.925 ⁺	0.917	0.708 ⁺	0.841
SVR-Poly	1.000	0.996 ⁺	0.878 ⁺	0.862 ⁺	0.832	1.000

Note: ⁺ denotes rejection of the null hypothesis of the Clark and West (2007) and the McCracken (2007) tests of equal forecasting accuracy with the autoregressive model only for models 5 and 6. The forecasting accuracy of all models against the RW model is measured based on the Diebold-Mariano statistic and the null hypothesis of equal forecasting accuracy is rejected for all tests. All tests are performed at 5% level of significance. The smallest MAPE for each forecasting horizon is denoted with an * and depicted in grey. All MAPE values are percentages.

¹² In the SVR models we report only the results of the most accurate kernel. The results for all models are reported in the online Appendix.

insignificant as with the GDP forecasts, since we observe differences after the 3rd decimal,¹³ while the addition of the control variables improves the forecasting accuracy of the models only marginally. The fact that the AR models do not outperform the structural ones is rather interesting given the large body of the literature arguing that univariate AR models consistently outperform structural ones in inflation forecasting (see among others [Stock & Watson, 2008](#) and [Rossi & Sekhposyan, 2010](#)). Moreover, in contrast to [Dewachter et al. \(2012\)](#), we do not find evidence in favor of the decomposition of the term spread over the use of the actual term spread in terms of forecasting performance. The discrepancy between our findings and earlier studies stems from the fact that in our forecasting approach we focus explicitly on out-of-sample forecasts while [Dewachter et al. \(2012\)](#) and [Wright \(2011\)](#) measure the in-sample fitting ability of the forecasting models. In order to elaborate on this, we depict in [Fig. 3](#) the time varying adjusted R^2 values of the most accurate OLS models for the term spread, the term premium and the expectations component. The selection of the models and the OLS methodology is motivated by the reported results in [Dewachter et al. \(2012\)](#) in forecasting inflation.

As with previous studies, we observe the existence of time periods where the selected models are able to capture the evolution of inflation. Thus, the differences between our findings and those of the previous studies are based on the selection of different evaluation samples (in-sample vs out-of-sample forecasting). Given that out-of-sample evaluation measures the ability of the model to forecast truly unknown (up to that time) data, we consider our inferences more robust than those in previous studies.

5.3. Conditional probability forecasts of the gross domestic product growth rate

While point forecasts are useful in providing a certain estimate of the future evolution of a variable, monetary authorities are becoming more and more interested in coupling point forecasts with an explicit description of the uncertainty of the forecast. Typically point estimates are accompanied with confidence intervals to evaluate the uncertainty of the forecast. In our approach, we are interested in the entire conditional probability distribution (or predictive density as reported in the literature) in order to obtain information on the likelihood of the appearance of economic phenomena. For instance, a monetary authority is not interested only in the likelihood of a low output growth rate in the future and its statistical significance, but also on the ex-ante probability (according to the forecasting model) for the output growth rate to be low. Thus, many monetary authorities publish fan charts that include the uncertainty around the forecast, depicting the quantiles around the mean (point) value forecast (see for example the relevant publications of Bank of England, Bank of Italy and those of the Fed).

In order to examine the uncertainty around our point forecasts we evaluate the respectful conditional probability density forecasts. Based on the residuals for each rolling window, we construct conditional density forecasts where the point forecast is the mean and the error deviation is the standard deviation of the distribution. Formally, under the normality assumption of the error distribution with a given information set at time t expressed as \mathfrak{S}_t and R the length of the rolling window, we estimate the conditional probability function (PDF) $\{\hat{\phi}_{t+h}(\hat{g}_{t+h}|S_t)\}_{t=R}^n$ of a normal distribution in higher conditional moments in order to obtain the conditional forecast PDF. We then test whether the forecasted distribution $\hat{\phi}_{t+h}$ matches the actual unobserved distribution that generates the data \hat{g}_{t+h} with the specification test for conditional predictive densities proposed by [Rossi and Sekhposyan \(2019\)](#), RS hereafter). The RS specification test examines whether the empirical conditional predictive density is specified correctly in matching the actual data distribution. In other words, the test examines whether the forecasted and the actual data originate from the same distribution. A rejection of the null hypothesis implies that the model cannot capture the true data generating mechanism. Given that the test uses the full distribution and not only a point forecast provides higher power to the test in comparison to the [Diebold and Mariano \(1995\)](#), the [Clark and West \(2007\)](#) or the [McCracken \(2007\)](#) test.

For each PDF $\hat{\phi}_{t+h}$ the Probability Integral Transformation (PIT) is the Cumulative Density Function (CDF) evaluated at:

$$z_{t+h} = \int_{-\infty}^{\hat{g}_{t+h}} \hat{\phi}_{t+h}(u|\mathfrak{S}_t) du = \hat{\phi}_{t+h}(\hat{g}_{t+h}|S_t) \quad (18)$$

Following the CDF of the PDF given in equation (18), we are interested in estimating the probability of the out-of-sample forecasts of length P :

$$\Psi_P(r) = P^{-1/2} \sum_{t=R}^n \xi_{t+h}(r) \quad (19)$$

$$\text{given. } \begin{cases} \xi_{t+h}(r) = (1\{\phi_{t+h}(\hat{g}_{t+h}|\mathfrak{S}_{t-R+1}^t) \leq r\} - r) \\ r \in [0, 1] \end{cases}$$

The null hypothesis could be formally stated as $H_0 : \hat{\phi}_{t+h}(\hat{g}_{t+h}|\mathfrak{S}_{t-R+1}) = \hat{\phi}_0(\hat{g}_{t+h}|F_t)$ where F_t is the information included in the true data generating process, \mathfrak{S}_{t-R+1} the information included in the estimation model and $\hat{\phi}_0(\hat{g}_{t+h}|F_t) = \Pr(\hat{g}_{t+h} \leq g|F_t)$ the probability distribution of the null hypothesis. We are interested in the test statistics:

$$K_p^{CS} = \sup(\Psi_P(r)^2), \quad r \in [0, 1] \quad (20)$$

¹³ all MAPE values in [Table 2](#) are percentages. Thus the reported 0.846% is actually 0.00846 and the MAPE differences are in the 3rd decimal.

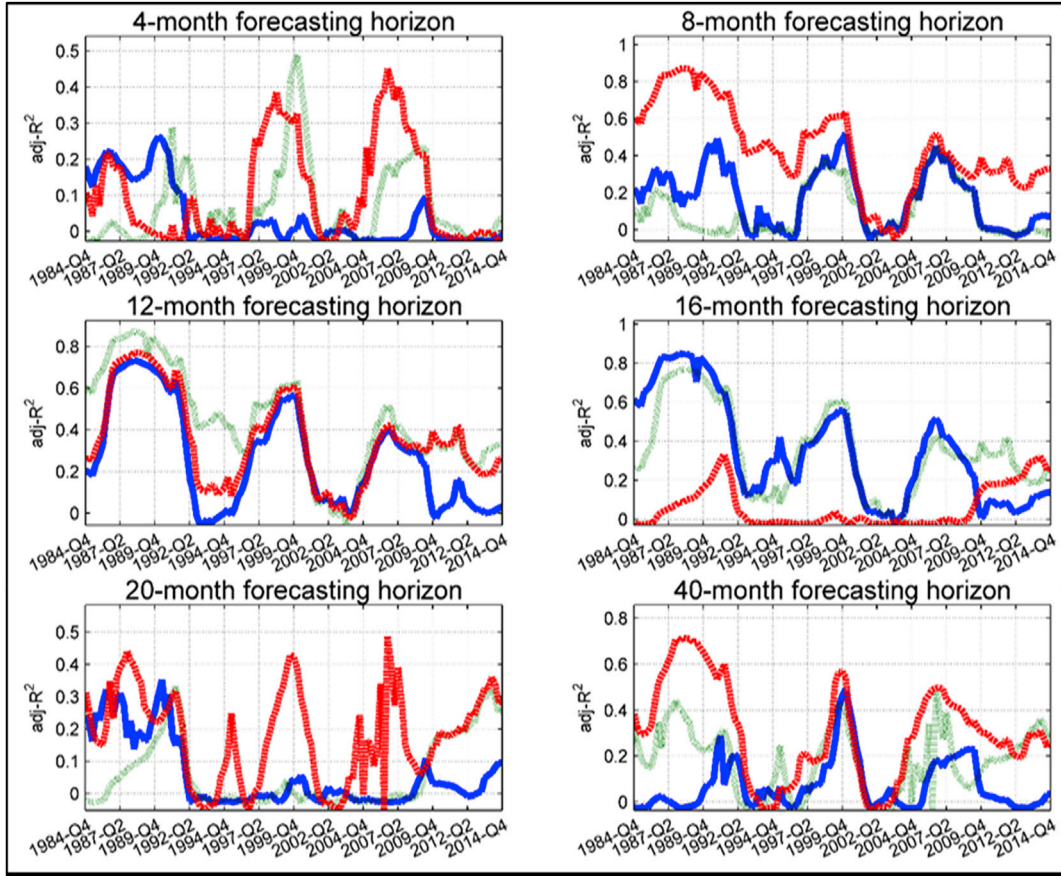


Fig. 3. The solid (blue) line represents the adjusted R^2 of the term spread model. The dotted (green) line represents the adjusted R^2 of the expectations component model and the dashed (red) line expresses the adjusted R^2 of the term premium model. Colors refer only to the online version of the paper. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

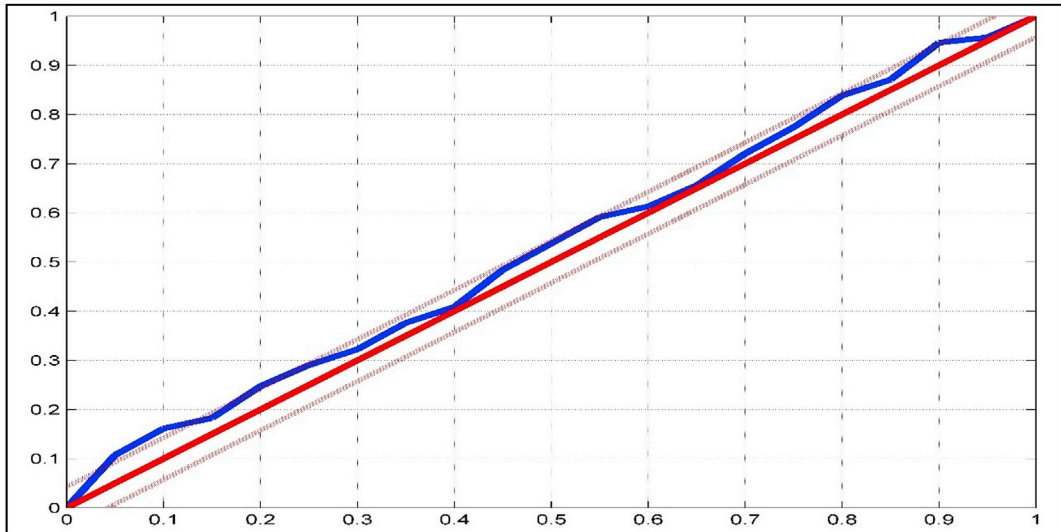


Fig. 4. K_p^{CS} test statistic values with their 95% confidence intervals (dashed lines) of the 40 quarter spread in forecasting the GDP growth with the SVR model. The (red) straight 45° degrees line represents the theoretical CDF, while the (blue) curved line the empirical one. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 4Conditional Probability of GDP forecasts- $K_{0.05,P}^{CS}$ statistics.

Forecasting Horizon	h = 4	h = 8	h = 12	h = 16	h = 20	h = 40
AR model	$g_{t+h} = c + \beta g_{t-k} + \varepsilon_{t+h}$					
OLS	1.015	1.624	1.652	1.659	1.594	0.568
SVR-linear	2.410	4.587	4.366	4.808	6.049	0.993
SVR-RBF	7.016	11.364	8.754	9.352	11.597	2.813
SVR-Poly	1.053	3.961	4.032	3.534	3.817	1.471
Model 1	$g_{t+h} = c + \beta S_t^{(h)} + \varepsilon_{t+h}$					
OLS	9.804	9.196	6.789	9.017	14.218	3.625
SVR-linear	7.250	7.162	5.922	6.050	9.880	2.882
SVR-RBF	7.488	8.981	5.575	10.752	5.776	2.445
SVR-Poly	6.117	6.374	5.967	7.485	6.376	1.705
Model 2	$g_{t+h} = c + \beta(E_t^{(h)} + x_t^{(h)}) + \varepsilon_{t+h}$					
OLS	14.025	12.660	9.520	7.335	3.022	0.624
SVR-linear	11.521	10.655	6.932	3.638	2.206	0.854
SVR-RBF	12.431	11.606	11.514	14.329	6.760	5.319
SVR-Poly	4.885	6.972	5.197	3.534	1.709	1.673
Model 3	$g_{t+h} = c + \beta E_t^{(h)} + \varepsilon_{t+h}$					
OLS	11.820	11.066	11.331	8.962	3.022	3.096
SVR-linear	8.997	7.500	8.331	6.005	1.451	2.224
SVR-RBF	9.531	11.424	8.915	4.260	1.885	2.642
SVR-Poly	7.976	8.611	7.816	3.925	2.705	1.842
Model 4	$g_{t+h} = c + \beta x_t^{(h)} + \varepsilon_{t+h}$					
OLS	13.699	14.634	7.367	5.095	1.384	2.445
SVR-linear	10.933	9.523	6.837	3.034	0.802	1.815
SVR-RBF	5.691	8.822	4.753	3.329	1.152	2.255
SVR-Poly	5.484	7.451	4.329	1.836	0.488	0.952
Model 5	$g_{t+h} = c + \beta_1 S_t^{(h)} + \beta_2 g_{t-1} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h}$					
OLS	7.488	13.302	4.442	4.569	1.473	1.094
SVR-linear	6.117	12.034	5.967	2.844	1.451	0.912
SVR-RBF	20.561	25.641	15.639	17.870	13.178	7.336
SVR-Poly	3.296	9.634	6.280	3.431	3.891	3.017
Model 6	$g_{t+h} = c + \beta_1(E_t^{(h)} + x_t^{(h)}) + \beta_2 g_{t-1} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h}$					
OLS	4.885	9.196	5.661	4.971	0.720	0.515
SVR-linear	3.622	7.795	3.888	1.762	0.643	0.515
SVR-RBF	13.699	17.116	8.226	10.334	13.867	5.463
SVR-Poly	2.690	9.088	2.253	3.397	4.153	1.417

Note: All values denote rejection of the null hypothesis that the forecasted and the actual data come from the same distribution at the 5% level of significance. The critical value for a rolling window of $R = 40$ observations and $P = 168$ observations is 0.048 for the K_P^{CS} test and 0.053 for the C_P^{CS} test at the 5% level of significance (Rossi & Sekhposyan, 2019, p. 646, Table 2).

$$C_P^{CS} = \int_0^1 \Psi_P(r)^2 dr \quad (21)$$

The critical values of the test statistic K_P^{CS} are estimated asymptotically. In Table 4 we report the K_P^{CS} test statistics at 5% level of significance.¹⁴

We reject the null hypothesis of correct model specification for all models at the 5% level of significance. Thus, regardless of the superiority of all models over the RW model, no model can actually capture the actual distribution of GDP growth rates.

Apart from the use of critical values, the test has also a graphical implementation. After plotting the CDF of the PITs together with the CDF of the uniform r (the 45° line) and the critical value lines $r \pm \sqrt{K_{0.05,P}^{CS}/P}$ one can test whether the test statistic exceeds the critical values of the test. In Fig. 4 we plot the K_P^{CS} test statistic along with its 95% confidence intervals of the mis-specified Model 6, estimated with the SVR methodology at the 40 quarters ahead forecasting horizon. As we observe, the continuous (blue) line of the empirical CDF crosses the dashed confidence intervals lines, rejecting the null hypothesis of proper model specification.

In Fig. 5 we depict the fan charts of the forecasted density of the same model, in order to describe the uncertainty around the (point) forecasts.

As we observe, certain actual observations lie outside the density forecasts, meaning that according to the forecasting model there is no probability for the appearance of this value. For instance, the actual GDP growth rate in 2008 has a very low probability of appearance, while forecasts during the period 1996–1999 are more probable. Thus, the model would never expect for the actual growth rates in 2008, resulting in misleading policy implications. This finding can be associated with Dewachter et al. (2012) who also report a

¹⁴ The results for the C_P^{CS} statistic are quantitative similar and are available upon request. The reported results for the SVR models refer only to the kernel with the higher forecasting accuracy in terms of MAPE. All other results are available in the online Appendix.

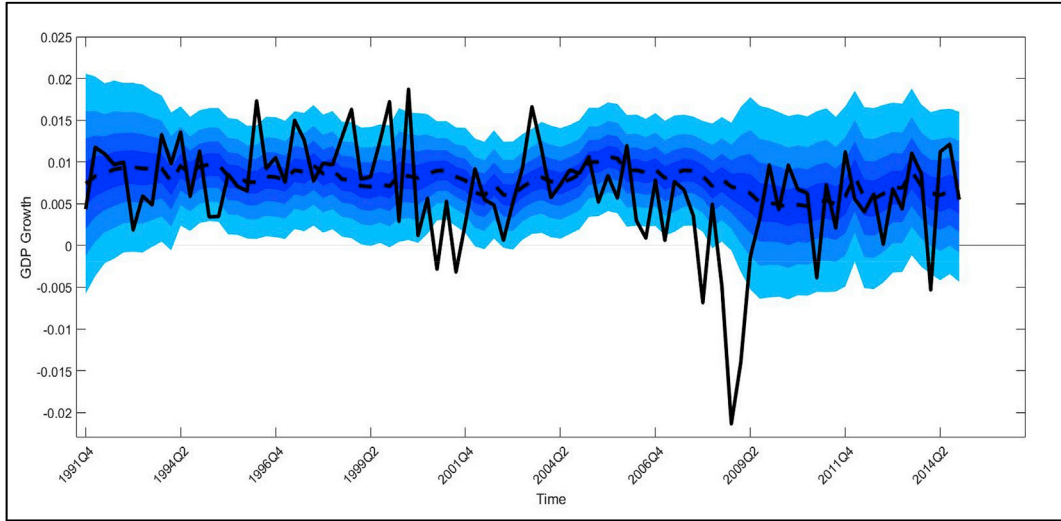


Fig. 5. Fan chart of the 40-quarter spread in forecasting the GDP growth with the SVR model. The colored areas represent the 10th, 20th, 30th, 40th, 50th, 60th, 70, 80th and the 90th deciles. The dashed line in the center of the deciles depicts the forecasted median (point forecasts) while the volatile line are the actual values of the series.

Table 5

Conditional Probability of inflation forecasts- $K_{0.05,P}^{CS}$ statistics.

Forecasting Horizon	h = 4	h = 8	h = 12	h = 16	h = 20	h = 40
AR model	$g_{t+h} = c + \beta g_{t-k} + \varepsilon_{t+h}$					
OLS	2.410	5.803	6.417	6.893	6.858	1.164
SVR-linear	2.410	4.587	4.366	4.808	6.049	0.993
SVR-RBF	7.016	11.364	8.754	9.352	11.597	2.813
SVR-Poly	1.053	3.961	4.032	3.534	3.817	1.471
Model 1	$g_{t+h} = c + \beta S_t^{(h)} + \varepsilon_{t+h}$					
OLS	0.686	1.200	1.894	0.824	0.829	0.865
SVR-linear	0.840	1.201	3.043	0.894	1.154	1.464
SVR-RBF	0.561	1.513	0.541	0.380	0.290	0.954
SVR-Poly	0.467	1.624	1.641	0.993	0.449	0.683
Model 2	$g_{t+h} = c + \beta (E_t^{(h)} + x_t^{(h)}) + \varepsilon_{t+h}$					
OLS	1.282	1.352	0.755	2.099	0.370	0.276
SVR-linear	1.415	1.300	0.599	1.194	0.736	0.340
SVR-RBF	2.165	2.450	3.924	4.129	3.151	4.934
SVR-Poly	1.712	1.512	0.381	1.597	0.862	1.580
Model 3	$g_{t+h} = c + \beta E_t^{(h)} + \varepsilon_{t+h}$					
OLS	1.415	1.404	1.104	0.728	0.906	1.366
SVR-linear	2.350	1.800	1.208	1.694	1.094	1.619
SVR-RBF	1.072	0.882	1.676	0.435	1.205	0.735
SVR-Poly	2.062	0.648	0.858	0.682	1.094	1.366
Model 4	$g_{t+h} = c + \beta x_t^{(h)} + \varepsilon_{t+h}$					
OLS	1.610	1.352	1.198	1.505	0.829	0.404
SVR-linear	1.841	1.860	1.188	0.711	1.226	0.471
SVR-RBF	0.931	0.924	1.007	0.742	0.696	0.553
SVR-Poly	1.282	1.104	1.869	1.471	1.046	0.130
Model 5	$g_{t+h} = c + \beta_1 S_t^{(h)} + \beta_2 g_{t-1} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h}$					
OLS	0.816	0.545	0.260	3.358	0.469	1.090
SVR-linear	0.437	0.480	0.361	2.099	0.342	1.026
SVR-RBF	7.795	8.064	4.677	3.412	6.084	2.557
SVR-Poly	1.700	2.738	1.657	2.085	1.620	2.144
Model 6	$g_{t+h} = c + \beta_1 (E_t^{(h)} + x_t^{(h)}) + \beta_2 g_{t-1} + \beta_3 y_t^{(1)} + \beta_4 CFNAI_t + \varepsilon_{t+h}$					
OLS	0.738	0.612	0.613	2.333	0.342	0.634
SVR-linear	1.262	1.152	0.675	1.372	0.583	0.450
SVR-RBF	5.952	9.112	5.079	3.762	5.923	3.133
SVR-Poly	1.363	1.985	2.035	0.937	0.864	1.865

Note: All values denote rejection of the null hypothesis that the forecasted and the actual data come from the same distribution at the 5% level of significance. The critical value for a rolling window of $R = 40$ observations and $P = 168$ observations is 0.048 for the K_p^{CS} test and 0.053 for the C_p^{CS} test at the 5% level of significance (Rossi & Sehkpasyan, 2019, p. 646, Table 2).

low performance of the term spread and its components in forecasting output growth, although their approach is limited to point forecasts and lacks the “uncertainty” element of our approach.

5.4. Conditional probability forecasts of inflation

We access the conditional predictive densities of the forecasting models for the GDP deflator on a similar approach to the one used with GDP growth forecasts. The test statistics for the RS test are reported in Table 5.

At the 5% level of significance all models reject the null hypothesis that the forecasted conditional predictive densities adhere to the actual data distribution. The results are quantitative similar for the C_p^{CS} statistic. In Fig. 6 we plot the K_p^{CS} statistic values with the 95% confidence intervals per rolling window for the mis-specified Model 6 estimated with the SVR methodology.

As we observe, the test statistic crosses the confidence intervals rejecting the null hypothesis. For comparison reasons with the GDP forecasts, in Fig. 7 we plot the fan chart for the model. The deciles of the forecasted probability density are wider to growth rates, denoting larger forecasting errors. Again several observations are forecasted with low probability of appearance, while other values evolve closely to the median (point) forecast of the model.

Overall, the evaluation of the conditional predictive densities of the forecasting models do not provide empirical evidence in favor of the use of the term spread or its components in forecasting inflation. The RS test results reveal that neither the decomposition, nor the term spread models capture the actual distribution of the data. Moreover, we do not find empirical evidence in favor of either the AR or the structural models.

6. Discussion

While theory suggests that the term spread should act as a leading indicator in forecasting economic activity and inflation, the empirical validation suggests otherwise. In line with our findings, Dewachter et al. (2012), state that the presence of a time-varying term premium invalidates the expectations hypothesis and in the context of macroeconomic forecasting, lowers the ability of the term spread to forecast economic growth. Morell (2018) finds that the accuracy of the spread’s predictive power will also be affected by changes in the conduct of monetary policy and more specifically by fluctuations in investment, that is inconsistent with the notion that the term spread is a leading procyclical variable.

One additional explanation for the failure of the term spread and its components to forecast economic activity and inflation, that has not been vigorously studied, could be that the macro-finance models completely ignore the conditions of the financial market. In a recent paper, Pragidis and Panagiotis Tsintzos and Vasilios Plakandaras (2018) evaluate the effect of the financial cycle on fiscal policy and its relevance to the monetary policy. They suggest that the link between financial conditions and macroeconomic variables could be tighter than usually considered in the relevant literature. Thus, a natural extension of the class of affine structure models would be to include financial uncertainty as a determinant of term premiums. We leave this strand for future research.

Nevertheless, despite the overwhelming evidence against the forecasting ability of the term spread or the term premium in macroeconomic forecasting, these variables are still considered as leading indicators for upcoming recessions or inflation and are closely watched by monetary authorities. Moreover, macro-finance models are broadly used in standalone applications or as parts of larger Dynamic Stochastic General Equilibrium models. As stated in the introduction section, the motivation of this study is to depart from the

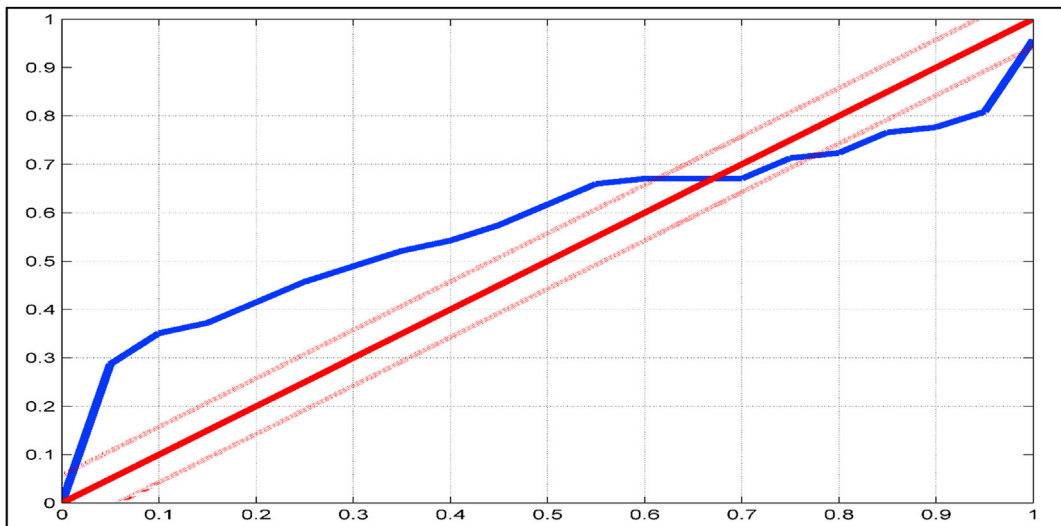


Fig. 6. K_p^{CS} test statistic values with their 95% confidence intervals (dashed lines) of the 40 quarter spread in forecasting the inflation with the SVR model. The (red) straight 45° degrees line represents the theoretical CDF, while the (blue) curved line the empirical one. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

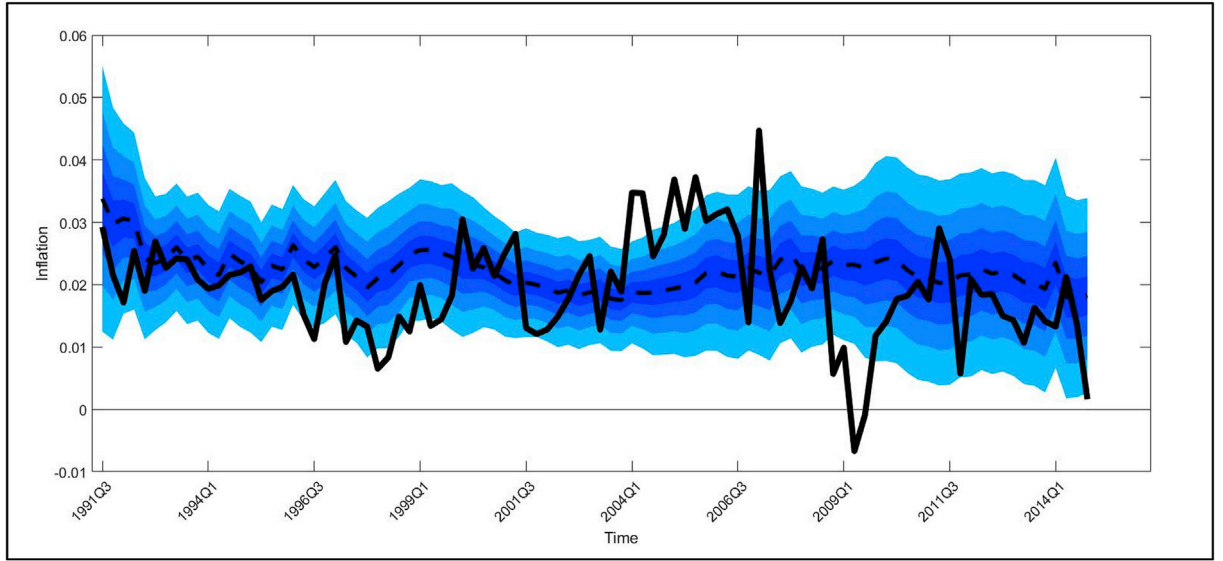


Fig. 7. Fan chart of the 40-quarter spread in forecasting the inflation with the SVR model. The colored areas represent the 10th, 20th, 30th, 40th, 50th, 60th, 70, 80th and the 90th deciles. The dashed line in the center of the deciles depicts the forecasted median (point forecasts) while the volatile line are the actual values of the series.

linear OLS regressions used in training macro-finance affine structure models and introduce a new category of SVR models coupled with linear and nonlinear kernels. Thus, we examine whether the reported failure in forecasting output growth and inflation should be attributed to the methodological approaches or to the predictors.

Our empirical findings suggest that neither the term spread, nor its components (expected short-term interest rates or the term premium) can reliably forecast economic growth or inflation regardless of the forecasting methodology used. As expected, the SVR outperformed the traditional econometric approaches, but the gains in terms of forecasting accuracy were only marginal. These findings hold even if we examine the entire forecast densities. The inclusion of control variables in order to isolate the effect of the term spread and its components on the dependent variable do not improve the forecasting ability.

Evaluating our empirical findings, we conclude that while the macro-finance models can be used successfully in forecasting the future values of the term spread, they have little to contribute to macroeconomic forecasting. Although the term-spread and the term premium are included in many DSGE models, they fail to capture the data generating process of economic growth and inflation. Thus, as suggested in [Ireland \(2015\)](#) and supported by our results, macro-finance models should have little value in macroeconomic forecasting and our focus should be on the DSGE models that provide a structural description of the economy including various aspects that influence macroeconomic evolution.

7. Conclusions

In this paper we examine the forecasting ability of the term spread in forecasting output growth rate and inflation. We decompose the term spread of Treasury bonds for various maturities into an expected short-term (to maturity) and a term premium component and evaluate the informational content of the term spread and its components in forecasting the GDP growth rate and the GDP deflator over the period 1971Q4-2014Q4. In doing so, we consider different functional forms and the application of the OLS and the SVR methodology in rolling regressions. The latter is coupled with the linear and two non-linear kernels. Departing from the commonly examined in-sample forecasting framework, we evaluate point and probability distributions on an explicit out-of-sample evaluation. Our empirical findings suggest that neither the term-spread nor the decomposition components can be used effectively as leading macroeconomic indicators of output growth rate and inflation regardless of the methodology applied and their use by the monetary authorities in shaping policies should be treated with caution, leaving room for further research.

References

- Adrian, T., Crump, R., & Moench, E. (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics*, 110(1), 110–138.
- Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, 131(1–2), 359–403.
- Antweiler, W., & Murray, Z. F. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59(3), 1259–1294.
- Athey, S. (2018). The impact of machine learning on economics. An Agenda by Ajay K. Agrawal, Joshua Gans, and Avi Goldfarb. National Bureau of Economic Research.
- Research. Atkeson A. and Ohanian L.E. (2001) Are Phillips curves useful for forecasting inflation? Federal Reserve Bank of Minneapolis Quarterly Review. In, Vol. 25. *The economics of artificial intelligence* (pp. 2–11), 1.
- Bai, J., & Perron, P. (2003). Critical values for multiple structural change tests. *The Econometrics Journal*, 6, 7 2 –78.
- Bauer, M. D., Rudebusch, G., & Wu, C. (2014). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset: Comment. *The American Economic Review*, 140(1), 323–337.
- Bjorkgren, D., & Grissen, D. (2018). Behavior revealed in mobile phone usage predicts loan repayment. *American Economic Association Papers and Proceedings*, 4.
- Blumenstock, J. E., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073–1076.
- Clark, T., & West, K. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138, 291–311.
- Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *The American Economic Review*, 95(1), 138–160.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
- Dewachter, H., Iania, L., & Lyrio, M. (2012). Information in the yield curve: A macro – finance approach. *Journal of Applied Econometrics*, 29, 4 2 –64.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13, 253–263.
- Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57(1), 405–443.
- Enders, W., & Lee, J. (2012). A unit root test using a fourier series to approximate smooth breaks. *Oxford Bulletin of Economics & Statistics*, 74(4), 574–599.
- 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.
- Evgenidis, A., Philippas, D., & Siriopoulos, C. (2019). Heterogeneous effects in the international transmission of the US monetary policy: A factor-augmented VAR perspective. *Empirical Economics*, 56, 1549–1579.
- Favero, C., Kaminska, I., & Söderström, U. (2005). *The predictive power of the yield spread: Further evidence and a structural interpretation*, CEPR discussion paper (Vol.4910). London: Centre for Economic Policy Research.
- Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry*, 56(1), 114–137.
- Gogas, P., Papadimitriou, T., & Agrapetidou, A. (2018). forecasting bank failures and stress testing: A machine learning approach. *International Journal of Forecasting*, 34(3), 440–455.
- 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.
- Gurkaynak, R. S., Sack, B., & Wright, J. H. (2007). The US treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54, 2291–2304.
- Hamilton, J. D., & Kim, D. H. (2002). A reexamination of the predictability of economic activity using the yield spread. *Journal of Money, Credit, and Banking*, 34(2), 340–360.
- Härdle, W., Lee, Y.-J., Schäfer, D., & Yeh, Y.-R. (2009). Variable selection and oversampling in the use of smooth support vector machines for predicting the default risk of companies. *Journal of Forecasting*, 28(6), 512–534.
- Ireland, P. (2015). Monetary policy, bond risk premia, and the economy. *Journal of Monetary Economics*, 76, 124–140.
- Joslin, S., Singleton, K., & Zhu, H. (2011). A new perspective on Gaussian dynamic term structure models. *Review of Financial Studies*, 24(3), 926–970.
- Kessel, R. (1965). The cyclical behavior of the term structure of interest rates. In *NBER occasional paper no. 91*. National Bureau of Economic Research.
- Kim, D. H., & Wright, J. H. (2005). *An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates*, board of governors of the federal reserve System. Finance and Economics Discussion Series. no. 33.
- Lange, R. H. (2018). The term structure of liquidity premia and the macroeconomy in Canada: A dynamic latent-factor approach. *International Review of Economics & Finance*, 57, 164–182.
- Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies*, 22(12), 5027–5067.
- McCracken, W. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics*, 140, 719–752.
- Morell, J. (2018). The decline in the predictive power of the US term spread: A structural interpretation. *Journal of Macroeconomics*, 55, 314–331.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *The Journal of Economic Perspectives*, 31(2), 87–106.
- Ng, S. (2017). *Opportunities and challenges: Lessons from analyzing terabytes of scanner data*. National Bureau of Economic Research. Working Paper No. 23673.
- Plakandaras, V., Gupta, R., Gogas, P., & Papadimitriou, T. (2015a). Forecasting the U.S. real house price index. *Economic Modelling*, 45, 259–267.
- Plakandaras, V., Papadimitriou, T., & Gogas, P. (2015b). Forecasting daily and monthly exchange rates with machine learning techniques. *Journal of Forecasting*, 34(7), 560–573.
- Pragidis, I., & Panagiotis Tsiotzos and Vasilios Plakandaras. (2018). Asymmetric effects of government spending shocks during the financial cycle. *Economic Modelling*, 68, 372–387.
- Ramsey, J. B. (1969). *Tests for specification errors in classical linear Least squares regression analysis* (Vol. 31), 2.
- Rossi, B., & Sekhposyan, T. (2010). Have Economic Models’ forecasting performance for US output and inflation changed over time and when? *International Journal of Forecasting*, 26(4), 808–835.
- Rossi, B., & Sekhposyan, T. (2019). Alternative tests for correct specification of conditional predictive densities. *Journal of Econometrics*, 208(2), 638–657.
- Rubio, G., Pomares, H., Rojas, I., & Herrera, L. J. (2011). A heuristic method for parameter selection in LS-SVM: Application to time series prediction. *International Journal of Forecasting*, 27(3), 725–739.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.
- Stock, J. H., & Watson, M. W. (1989). *New indexes of coincident and leading indicators*, NBER macroeconomic annual (Vol. 4). Cambridge, MA: MIT.
- Stock, J. H., & Watson, M. W. (2008). *Phillips curve inflation forecasts*. NBER Working Paper. No. 14322.
- Vapnik, V., Boser, B., & Guyon, I. (1992). *A training algorithm for optimal margin classifiers*. Pittsburgh, ACM: Fifth Annual Workshop on Computational Learning Theory.
- Wheelock, D., & Wohar, M. (2009). Can the term spread predict output growth and recessions? A survey of the literature. September/October *Federal Reserve Bank of Saint Louis Review*, 91(5), 419–440. Part 1).
- Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *The American Economic Review*, 101(4), 1514–34.