

The Informational Content of the Term-Spread in Forecasting the U.S. Inflation Rate: A Nonlinear Approach

Periklis Gogas[†], Theophilos Papadimitriou⁺, Vasilios Plakandaras⁺ and Rangan Gupta*

⁺ Department of Economics, Democritus University of Thrace, Greece.

*Department of Economics, Pretoria University, South Africa.

Abstract

The difficulty in modelling inflation and the significance in discovering the underlying data generating process of inflation is expressed in an ample literature regarding inflation forecasting. In this paper we evaluate nonlinear machine learning and econometric methodologies in forecasting the U.S. inflation based on autoregressive and structural models of the term structure. We employ two nonlinear methodologies: the econometric Least Absolute Shrinkage and Selection Operator (LASSO) and the machine learning Support Vector Regression (SVR) method. The SVR has never been used before in inflation forecasting considering the term--spread as a regressor. In doing so, we use a long monthly dataset spanning the period 1871:1 – 2015:3 that covers the entire history of inflation in the U.S. economy. For comparison reasons we also use OLS regression models as benchmark. In order to evaluate the contribution of the term-spread in inflation forecasting in different time periods, we measure the out-of-sample forecasting performance of all models using rolling window regressions. Considering various forecasting horizons, the empirical evidence suggests that the structural models do not outperform the autoregressive ones, regardless of the model's method. Thus we conclude that the term-spread models are not more accurate than autoregressive ones in inflation forecasting.

Keywords: U.S. Inflation, forecasting, Support Vector Regression, LASSO.

JEL Codes: C22, C45, C53, E31, E37

[†] [Corresponding author](#): Periklis Gogas, Department of Economics, Democritus University of Thrace, Komotini, University Campus, Greece. Email: pgkogkas@ierd.duth.gr Tel: +302531039555.

1. Introduction

“Inflation is hard to forecast”. With widespread empirical evidence, this statement of Stock and Watson (2003) is considered as a stylized fact in macroeconomics. Beyond academic interest, with the given commitment of the Federal Reserve (FED) to price stability, the creation of models that can describe correctly the underlying data generating mechanism of inflation is of outmost importance to policy authorities. In the voluminous relevant literature numerous approaches have been proposed: among others, household and professional surveys on inflation expectations, latent factor models, autoregressive models and structural models based on the New Keynesian Phillips curve and the term-spread¹.

There is an extensive literature in forecasting inflation rates based on the term-spread. By rearranging the terms of the Fisher equation on real interest rates, Mishkin (1990a, b) and Jorion and Mishkin (1991) attempts to forecast the U.S. inflation, using the informational content of the interest rates. In this line of research the term-spread is decomposed into three components: the expected real rate change, the expected inflation change, and the term premium. The goal is to link the variations in the term-spread to the variations in the expected inflation and ultimately to use the term-spread to forecast inflation. The empirical findings suggest that only the long term rates (interest rates longer than a year) contain valuable information on inflation expectations. Another view as to why the term-spread can forecast inflation is exposted in Estrella and Mishkin (1997), and is related to the fact that the term-spread is an indicator of the stance of the monetary policy. Hence, a low spread reflects relatively restrictive monetary policy (because the spread is low when short-term interest rates are high relative to long-term interest rates), which in turn, implies that in response to the contractionary monetary policy, real activity will slow down and inflation will decrease. Stock and Watson (2003, 2008) test the ability of the term-spread in forecasting the U.S. inflation for a variety of forecasting horizons in an out-of-sample exercise. They conclude that the univariate autoregressive model outperforms a structural model based on the term structure in the post 1984 period. Ang

¹ For a detailed exposition of the existing literature see Stock and Watson (2008, 2010).

et al., (2007) forecast CPI-all, CPI-core, CPI-ex housing and PCE inflation on a yearly forecasting horizon, using a no-arbitrage yield spread model. Evaluating constant and regime switching coefficients they examine two alternative forecasting periods, the 1985-2002 and the 1995-2002. They conclude that no structural model outperforms the univariate ARMA(1,1) model in terms of out-of-sample forecasting accuracy.

More recently, Berardi (2009) developed a structural model using yield spreads of real interest rates for various forecasting horizons up to 2-years ahead. The author concludes that no structural model outperforms an autoregressive one, regardless of the forecasting horizon. Rossi and Sehkosyan (2010) examine the 3-month Treasury bill as the short and the 1-year, the 5-year and the 10-year Constant Maturity Treasury bill as the long term end of the yield spread, covering the period 1960:1-2005:12. Evaluating regressions on rolling windows, the authors discover that the informational content of the term-spread in forecasting the U.S. inflation varies over time. In the early windows of the sample the structural model outperforms the autoregressive one, while all predictive ability of the structural models disappears in the post-1984 period, the start of the Great Moderation period. Overall, considering the average performance on out-of-sample forecasting of all windows, Rossi and Sehkosyan (2010) conclude that models based on the term-spread do not outperform the autoregressive models. Summarizing the existing literature regarding U.S. inflation forecasting based on the term-spread, models based on the term structure outperform the autoregressive ones only sporadically, and in general all researchers conclude that the difference between long and short term interest rates are an inferior predictor to the historical values of inflation.

An interesting issue emerges when reviewing the relevant literature is that the vast majority of U.S. inflation forecasting studies employ only linear models. Among the exceptions, Ascari and Marrocu (2003) develop a Self-Exciting Threshold Autoregressive (SETAR) model to forecast the U.S. inflation based on the non-accelerating-inflation-rate-of-unemployment (NAIRU) Phillips curve assumption. They conclude that the SETAR model cannot outperform the RW model in out-of-sample forecasting. Marcellino (2008) forecasts the U.S. CPI-all index with Smooth Transition Autoregressions and Neural Networks on a rolling window and a recursive scheme. He

concludes that linear models outperformed nonlinear ones. Inoue and Killian (2008) consider CPI-all forecasts using various nonlinear factor models. The factors come from the extracted components of principal component analysis on a dataset of 30 economic indicators. Examining the period 1983:3-2008:9 they find that the bagging ridge regression methodology is the most accurate in terms of out-of-sample forecasting accuracy. More recently Koop and Korobilis (2012) uses dynamic model averaging (DMA) techniques applied to a time-varying predictive-regression framework with thirteen exogenous predictors, which also included term-spread. Using the GDP deflator and personal consumption expenditures based on the inflation rate, the study shows that the DMA method outperforms all the other competing models in forecasting inflation, with inflation expectations, three-month Treasury bill rate and housing starts being the best predictors. Manzan and Zerom (2015) provide some evidence of superior forecasting of core inflation for the US using quantile autoregressions, which in turn, are also corroborated by Korobilis (2015) by applying Bayesian model averaging on multivariate quantile regressions (which involves thirty-two exogenous predictors) to forecast US inflation. Small and large-scale time-varying vector autoregressions are also observed to forecast inflation accurately, relative to alternative forecasting models in Bekiros and Paccagnini (2013, forthcoming) and Koop and Korobilis (2013), respectively. However, using genetic programming methods Álvarez-Díaz and Gupta (2015) find it difficult to beat linear models in forecasting the U.S. CPI. To the best of our knowledge, there are no additional studies that employ nonlinear models and the term-spread in forecasting the U.S. inflation.

In this paper we develop nonlinear econometric and machine learning models that exploit the informational content of the term-spread in forecasting. Within a rolling regressions framework for out-of-sample forecasting, we employ two alternative methodologies: a) the Least Absolute Shrinkage Selection Operator (LASSO); and b) a Support Vector Regression (SVR) methodology from the area of machine learning. The SVR has never been used before in inflation forecasting based on the term-spread. For comparison reasons we also provide a similar analysis using the linear OLS regression that has been the workhorse in the inflation forecasting literature. These techniques are used to forecast inflation both within an autoregressive and a structural model. We also use a long dataset

spanning the period 1871:1 – 2015:3. Finally, changes in the forecasting performance of all models over time are assessed with the Fluctuation test of Giacomini and Rossi (2010).

2. Methodology and Data

2.1 Support Vector Regression

The Support Vector Regression is a direct extension of the classic Support Vector Machine algorithm. The specific machine learning methodology has attracted significant interest in forecasting economic and financial time series (Rubio *et al.*, 2011; Härdle *et al.*, 2009; Ögüt *et al.*, 2012; Khandani *et al.*, 2010; Plakandaras *et al.*, 2015). The algorithm proposed by Vapnik *et al.* (1992) and latter 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 as long as 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 the 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, avoiding overfitting of the model.

For a training dataset $D = [(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)]$, $\mathbf{x}_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}$, $i = 1, 2, \dots, n$, where \mathbf{x}_i is a vector of independent variables and y_i is the dependent variable the linear regression function takes the form of $y = f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$. This is achieved by solving:

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

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

where ε defines the width of the tolerance band, and ζ_i, ζ_i^* are slack variables controlled through a penalty parameter C (see Figure 1). All the points inside the tolerance band have $\zeta_i, \zeta_i^* = 0$. System (1) 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 $\|\mathbf{w}\|$. The second part of the objective function controls the regression fit to the training data (by increasing C we penalize with a bigger weight 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.

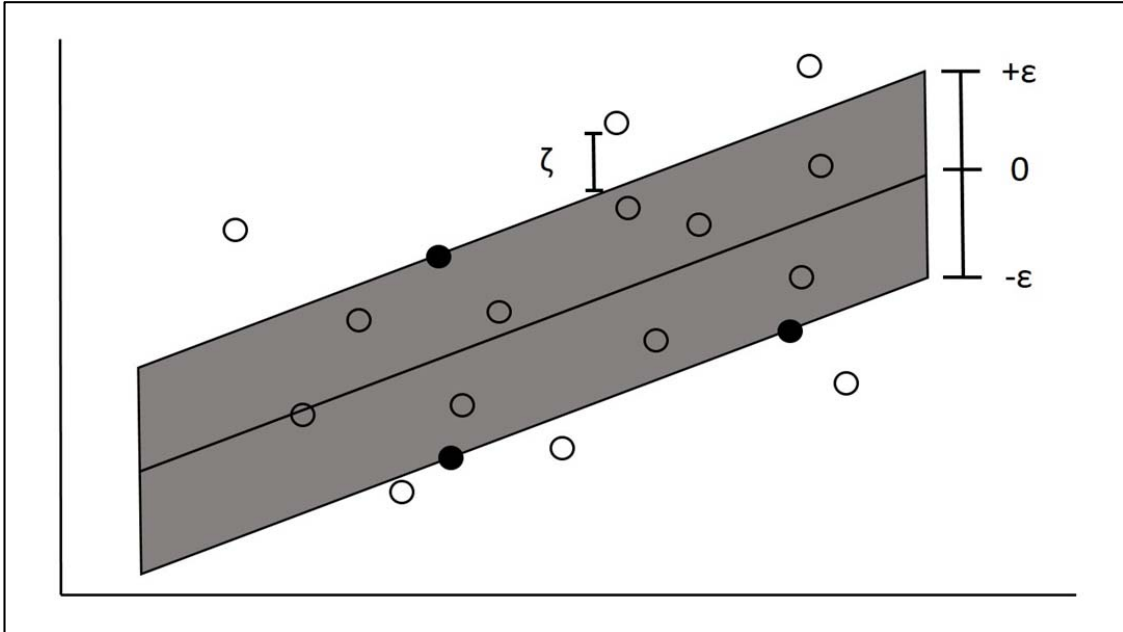


Figure 1: Upper and lower threshold on error tolerance indicated with letter ε . The boundaries of the error tolerance band are defined by the Support Vectors (SVs) denoted with the black filled points. Forecasted values greater than ε get a penalty ζ according to their distance from the tolerance accepted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

Using the Lagrange multipliers in System (1) the solution is given by:

$$\mathbf{w} = \sum_{i=1}^n (a_i - a_i^*) \mathbf{x}_i \quad (2)$$

and

$$y = \sum_{i=1}^n (a_i - a_i^*) \mathbf{x}_i^T \mathbf{x} \quad (3)$$

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

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, SVM/R 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. When the kernel function is non-linear, the produced SVR model is non-linear as well. In our empirical estimations we employed two alternative kernels: the linear and the radial basis function (RBF), with the latter being a purely nonlinear kernel. The mathematical representation of each kernel is:

Linear
$$K_1(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T \mathbf{x}_2 \quad (4)$$

RBF
$$K_2(\mathbf{x}_1, \mathbf{x}_2) = e^{-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|^2} \quad (5)$$

with factor γ representing the kernel parameter.

2.2 Least Absolute Shrinkage and Selection Operator

The Least Absolute Shrinkage and Selection Operator (LASSO) is a regularization and variable selection method proposed by Tibshirani (1996) that has been extensively studied in the forecasting literature (for a survey on LASSO applications the interested reader is referred to Bai and Ng, 2008). When using the typical OLS linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} \quad (6)$$

$$\text{with } \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{1}^T \\ \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_p^T \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ip} \end{pmatrix} \text{ and } \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^T$$

one tries to estimate the values of vector of the coefficients $\boldsymbol{\beta}$ based on the minimum residual squared error and are the independent variables. LASSO applies one more restriction to the model, attempting not only to minimize the squared error of the residuals, but also to eliminate uninformative regressors through weighting. In order to find the optimal solution in this trade-off between parsimony and high forecasting ability, LASSO minimizes the following

$$\min_{\beta_0, \boldsymbol{\beta}} \left[\frac{1}{2n} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^p |\beta_j| \right] \quad (7)$$

where \mathbf{x}_i is a vector of p -values at observation i and λ is a regularization parameter. The imposition of the regularization parameter λ defines the parsimony of the model, with the number of the nonzero elements in the coefficients' vector decreasing as λ increases. The optimization problem can be solved with any quadratic programming optimization method.

2.3 Description of the Fluctuation Test

The Fluctuation test relies on the null hypothesis that the forecasting error of the structural model is equal to the autoregressive model and does not vary over time against the alternative that the two models exhibit different forecasting performance. With a rolling window estimation for the two models, we compare their out-of-sample forecasting error over time and extract asymptotic p -values for evaluating the null hypothesis. Starting from an autoregressive model for h periods ahead forecasting $\mathbf{y}_{t+h} = \beta_0 + \beta_2(L)\mathbf{y}_t + \boldsymbol{\eta}_{t+h}$, with L representing the lag polynomial of the model, the relative Mean Square Forecasting Error (rMSFE) of the structural model $\mathbf{y}_{t+h} = \beta_0 + \beta_1(L)\mathbf{x}_t + \beta_2(L)\mathbf{y}_t + \boldsymbol{\epsilon}_{t+h}$ can be computed as

$$rMSFE = \frac{1}{m} \left(\sum_{j=t-\frac{m}{2}}^{t+\frac{m}{2}} (\hat{\epsilon}_{j+h}^2 - \hat{\eta}_{j+h}^2) \right) \quad (8)$$

where m represents the length of the rolling window, η_{t+h} is the forecasting error of the autoregressive and ϵ_{t+h} the forecasting error of the structural model. The Fluctuation test is a rescaled version of the rMSFE, as follows:

$$F_{t,m}^{OOS} = \hat{\sigma}^{-1} m^{-1/2} \left(\sum_{j=t-m/2}^{t+m/2} \hat{\epsilon}_{j+h}^2 - \sum_{j=t-m/2}^{t+m/2} \hat{\eta}_{j+h}^2 \right) \quad (9)$$

for $t = m + h + m/2, \dots, T - m/2 + 1$ where $\hat{\sigma}^2$ is a Heteroskedasticity and Autocorrelation Consistent (HAC) estimator of the asymptotic variance $\sigma^2 = \text{var}(P^{-1/2} \sum_{j=m+h}^T (\hat{\epsilon}_j^2 - \hat{\eta}_j^2))$, with $P = n - m$.

The null hypothesis of the test is that the forecasting performance of both models is the same at each point in time, that is:

$$H_0: E(\hat{\epsilon}_t^2 - \hat{\eta}_t^2) = 0, \quad t = m + h, \dots, T \quad (10)$$

Critical p -values for various significance levels and various window and sample sizes are provided by Giacomini and Rossi (2010).

2.4 The Data

We compile monthly observations for the time period spanning January, 1871 to March, 2015 for the CPI, the short and long-term interest rates. The CPI and the ten-year constant maturity Treasury bills rate are obtained from the data segment of Professor Robert J. Shiller's personal webpage². While, the short-term interest rate till 2013:12 comes from the website of Professor Amit Goyal³, and then updated till 2015:3 from the FRED database of the Federal Reserve Bank of St. Louis. Note that, the short-term rate is measured by the three-month Treasury bill rate from 1920 onwards, and prior to that is based on an estimation, as in Goyal and Welch (2008), using the Commercial paper rates for New York City, which in turn, are obtained from the National Bureau of Economic Research

² <http://www.econ.yale.edu/~shiller/data.htm>.

³ <http://www.hec.unil.ch/agoyal/>.

(NBER) Macroeconomy data base. In Table 1 we report the descriptive statistics of the series for the natural-logarithms of the CPI and the Term-spread of the two interest rates.

Table 1 Discriptive Statistics of the series		
	<u>CPI</u>	<u>Term-spread</u>
Mean	53.154	0.909
Median	17.800	0.940
St. Deviation	65.412	1.395
Skewness	1.475	-0.625
Kurtosis	3.778	5.236
J-B test (p-value)	0.000***	0.000***

Note: *** denote rejection of the null hypothesis about normality according to the J-B test.

According to the Jarque-Bera test (Jarque and Bera, 1987) the null hypothesis about normality is rejected at 1% level of significance. In what follows all values of the CPI refer to its natural logarithm. In Figures 1 to 4 we depict the evolution of these series.

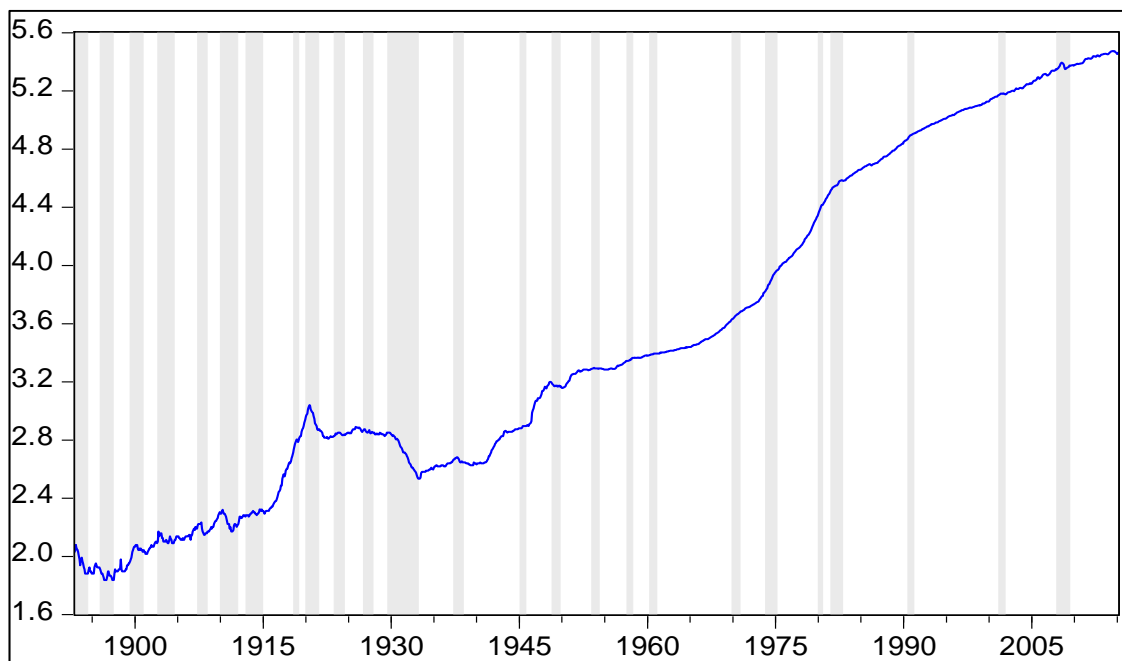


Figure 1: Monthly logarithmic CPI-all series. Grey areas denote NBER recessions.

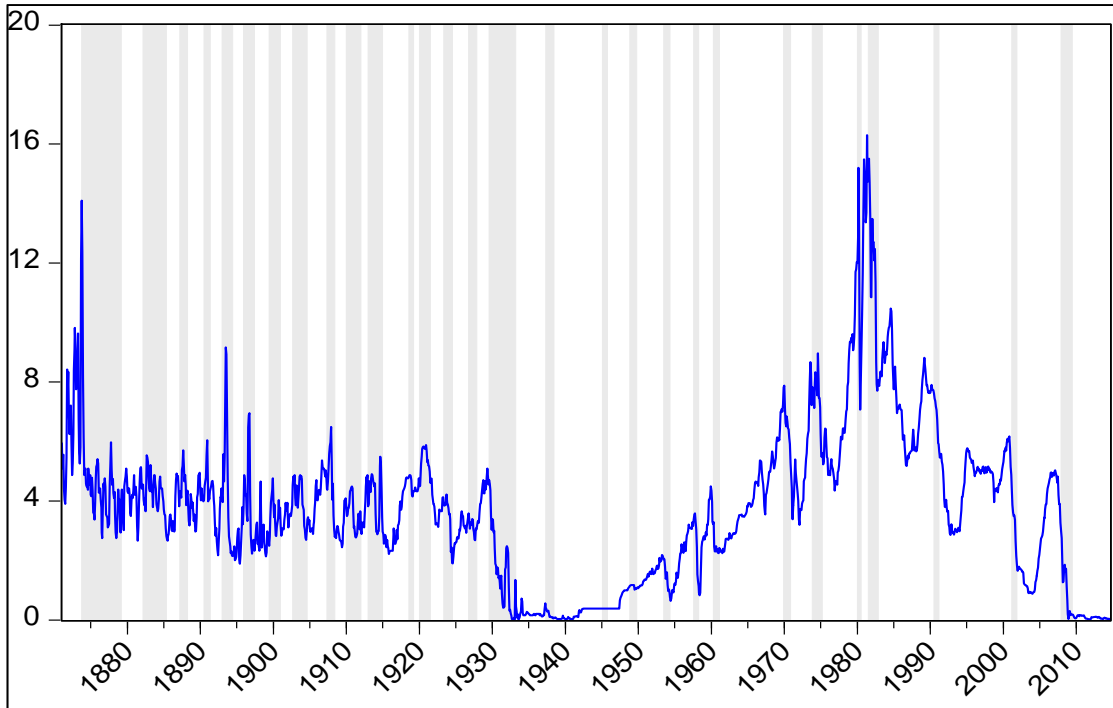


Figure 2: Monthly short-term interest rate. Grey areas denote NBER recessions.

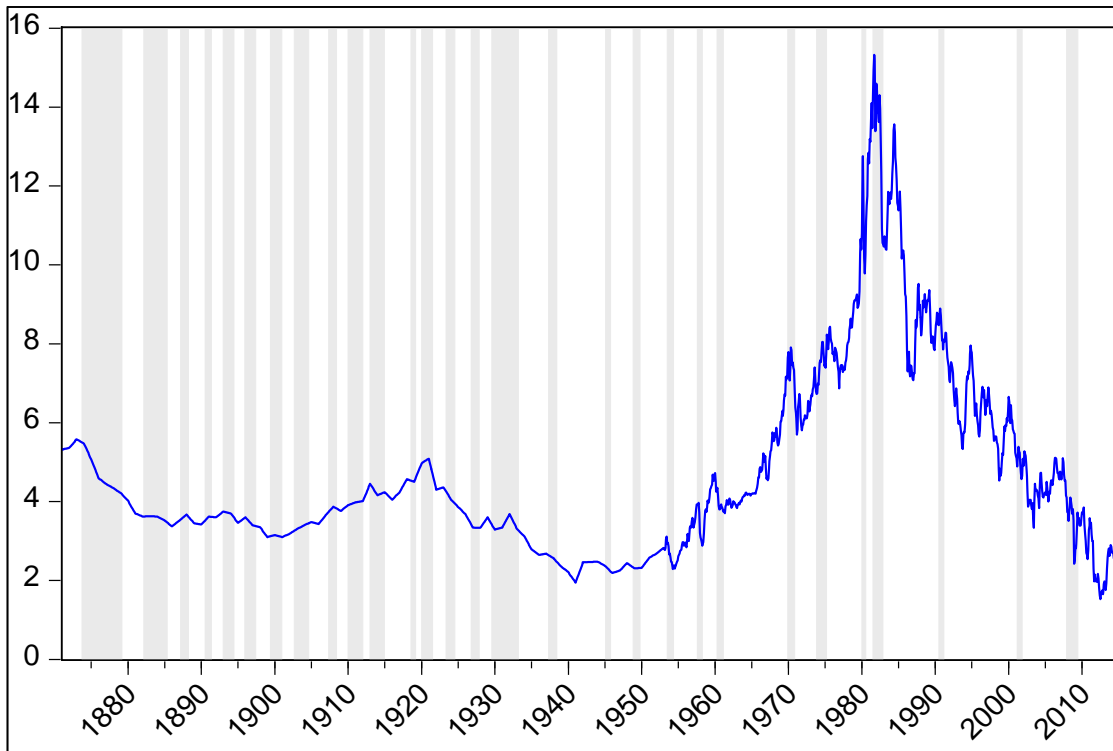


Figure 3: Monthly long-term interest rate. Grey areas denote NBER recessions.

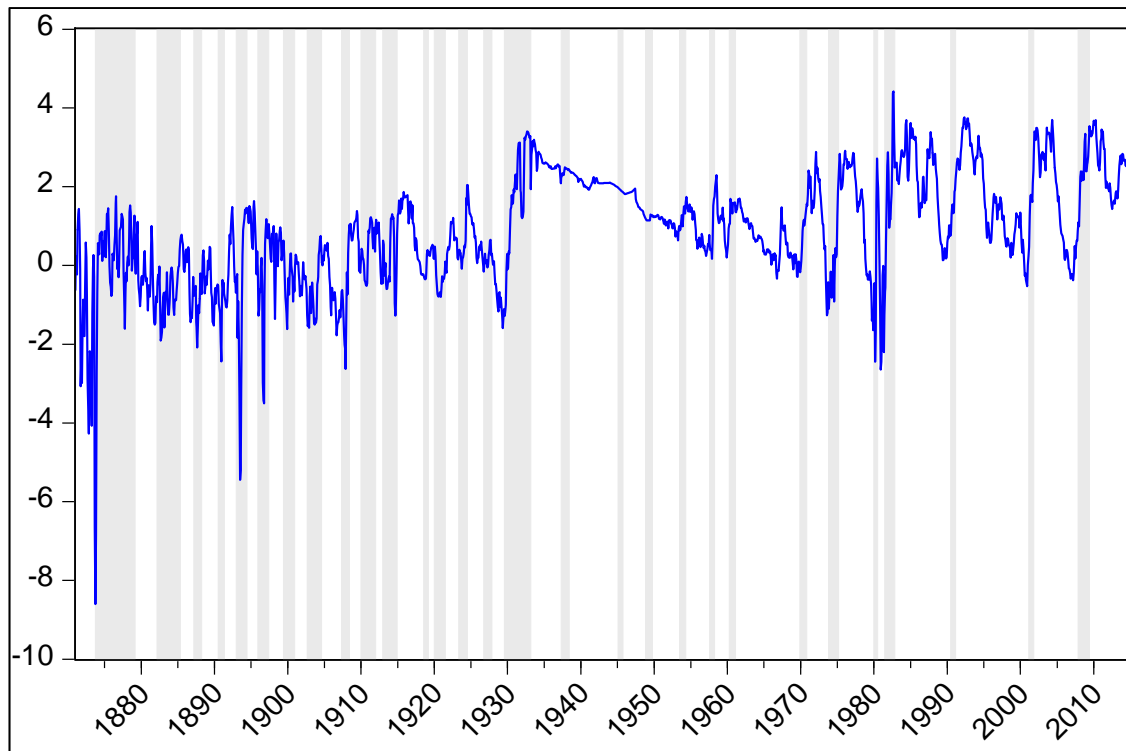


Figure 4: Monthly term-spread series. Grey areas denote NBER recessions.

After the collapse of the Bretton Woods fixed exchange rate system the CPI exhibits a significant trend in the post -1971 period. From Figures 3 and 4 the two selected interest rates appear to follow a unit root process and are cointegrated⁴, while the term-spread (Figure 5) appears mean reverting. Since all variables appear to exhibit structural breaks we apply the Bai-Perron (2003) multiple structural breaks test in order to test the null hypothesis of no structural breaks against the alternative of the existence of a fixed number of structural breaks according to different information criteria (Table 2).

⁴ The results of the maximum likelihood cointegration test (Johansen, 1991) are available from the authors upon request.

Table 2: Bai-Perron Multiple Break test results			
Breaks	Log-Likelihood	Schwarz Criterion	LWZ Criterion
<u>Panel A: CPI</u>			
0	-9658.891	8.365	8.373
1	-7977.967	6.425	6.449
2	-7430.351	5.799	5.839
3	-7313.307*	5.672	5.727*
4	-7293.486	5.657*	5.729
5	-7293.147	5.665	5.754
<u>Panel B: Term-Spread</u>			
0	-3021.358	0.669	0.678
1	-2623.713	0.217	0.241
2	-2582.788	0.178	0.218
3	-2490.788	0.080	0.136
4	-2442.871*	0.033*	0.105*
5	-2553.554	0.170	0.258

Note: Selected number of breaks for each criterion are denoted with an *.

As we observe from Table 2, both series present multiple structure breaks. According to the Schwarz Information Criterion (Schwarz, 1978) the CPI series has 4 structural breaks, while the Liu-Wu-Zidek (LWZ) criterion (Liu *et al.*, 1997) and the log-likelihood detect 3 structural breaks. In Panel B we report the results on the term-spread series. According to all three criteria the term-structure series has 4 structural breaks. The existence of a structural break renders common unit root tests such as the Augmented Dickey-Fuller (Dickey-Fuller, 1981), the Phillips-Perron (Phillips and Perron, 1988) and the Kwiatkowski-Phillips-Schmidt-Shin test (Kwiatkowski *et al.*, 1992) irrelevant. Moreover, the existence of more than one structural break does not allow for the application of the commonly cited in literature Zivot-Andrews (Zivot and Andrews, 1992) and Perron (Perron, 1997) unit root tests. Both these tests examine the null hypothesis of the existence of a unit root in the presence of only one unknown structural break, which is not suitable to our cause. Thus, we apply the recently proposed unit root test of Enders and Lee (2012).

In contrast to Zivot-Andrews and Perron tests, the Enders and Lee (2012) test approximates the series with low frequency components of the Fourier expansion without the need to define the points of the breaks. Moreover, the test is able to capture unit root

processes in the presence of multiple linear and nonlinear forms of structural breaks. Following Enders and Lee (2012) we use a small number of frequency components to avoid over-fitting the series and allow the evolution of the nonlinear trend to be gradual. In Table 3 we report the Lagrange Multiplier (LM) test statistic of the unit root test for up to three frequency components.

Table 3: LM statistics of the Enders and Lee Unit Root Test					
		<u>Frequency Components</u>			<u>Decision</u>
		1	2	3	
ln(CPI)	Levels	-0.593	-1.536	-1.714	I(1)
	First Differences	-22.512***	-22.920***	-23.455***	I(0)
Term-spread	Levels	-8.366***	-8.706***	-8.975***	I(0)

Note: *** denotes rejection of the null hypothesis about non-stationarity at 1% level of significance.

As we observe from Table 3, the natural logs of CPI exhibits a unit root in levels but is stationary in first differences. In contrast the term-spread is stationary in levels. Thus in the empirical part of our analysis we use first differences of the natural logs of CPI (i.e., month-on-month inflation rate) and levels of the term-spread.

3. Empirical Results

Following Rossi and Sehkosyan (2010) we select the width of the rolling window for model estimation $m=120$ (which is equivalent to a 10 year period) sliding at one observation each time. The rolling window procedure is appropriate as: a) it ensures that intertemporal fluctuations in the forecasting performance of the models employed are revealed, b) we use only recent information that is most relevant to inflation forecasts and, c) using overlapping windows results in a smooth course over time. As with Stock and Watson (2003) we develop both an autoregressive and a structural model of each methodology. The lag order for both the autoregressive and the structural model is selected according to the Schwarz Information Criterion (Schwarz, 1978) in every window. We consider alternative forecasting horizons of 1, 3, 6, 12, 18 and 24 months ahead and evaluate the forecasting performance of each model according to the the Mean Absolute Percentage Error (MAPE) criterion, which is

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (11)$$

where \hat{y}_i is the forecast of the actual inflation rate and n is the total number of the out-of-sample observations. We use the MAPE instead of the MSE as our metric for forecast comparison across the models, since the latter is affected by the level of measurements. Our dataset spans a very long period with a non-constant mean with respect to time (see Figure 1). Since we employ a rolling window approach to our forecasts, any MSE-based inference is rendered misleading. In Table 4 we report the MAPE calculated on the entire out-of-sample period for the RW, the OLS regression, the LASSO and the SVR model.

Table 4 Forecasting results based on the MAPE criterion							
		Forecasting Horizon					
		1	3	6	12	18	24
RW		0.619	1.514	2.630	4.503	6.227	7.714
OLS	AR	0.551*	0.575	0.580	0.582	0.600	0.593
	Term-Spread	0.560	0.581	0.586	0.595	0.611	0.599
LASSO	AR	0.552	0.577	0.580	0.581*	0.594*	0.589
	Term-Spread	0.559	0.574*	0.583	0.588	0.595	0.588
SVR-	AR	0.554	0.581	0.574*	0.587	0.595	0.586
LINEAR	Term-Spread	0.560	0.579	0.579	0.593	0.599	0.587
SVR-	AR	0.574	0.604	0.591	0.603	0.599	0.585*
RBF	Term-Spread	0.597	0.610	0.600	0.605	0.589	0.593

Note: * denotes the smallest forecasting error on each forecasting horizon, while in bold cases where the structural model outperforms the autoregressive one. All numbers express percentages.

As we observe from Table 4, all models outperform the RW in their respective forecast horizon and the structural models outperform the autoregressive ones only sporadically. More specifically, the structural OLS model does not outperform the autoregressive one in any forecasting horizon. When it comes to the LASSO, the structural model is more accurate only on the 3 and 24-months ahead forecasting horizon. The structural SVR model coupled with the linear kernel outperforms its respective autoregressive model only on the 3-months ahead forecasting horizon, while the structural SVR-RBF model outperforms the autoregressive one only on the 18-months ahead forecasting horizon. Considering the different forecasting horizons, the autoregressive OLS model produces

the lowest forecasting error in 1-month ahead forecasts, while in 3-months ahead the most accurate model is the structural LASSO. The autoregressive SVR-LINEAR model adheres more closely to inflation in 6-months ahead forecasting, while in 12 and 18-months ahead the autoregressive LASSO outperforms all the alternative models. In 24-months ahead forecasting the most accurate model is the autoregressive SVR-RBF. Nevertheless, the differences in the forecasting errors between the linear and the nonlinear methodologies in all forecasting horizons are very small. Thus we do not find evidence in favour of applying the complex nonlinear methodologies examined in this paper as alternatives to the simpler linear OLS regression. Finally, the use of the term-spread as a potential regressor in forecasting the U.S. inflation does not improve the forecasting performance in comparison to the autoregressive models.

Departing from comparisons on the entire out-of-sample forecasted dataset, we perform a point-to-point accuracy evaluation with the use of the Fluctuation test. With the specific test we may be able not to pinpoint periods where the term structure model is more accurate than the autoregressive one. In Figures 5-8 we present the results of the Fluctuation test for the OLS, the LASSO, the SVR-LINEAR and the SVR-RBF models respectively.

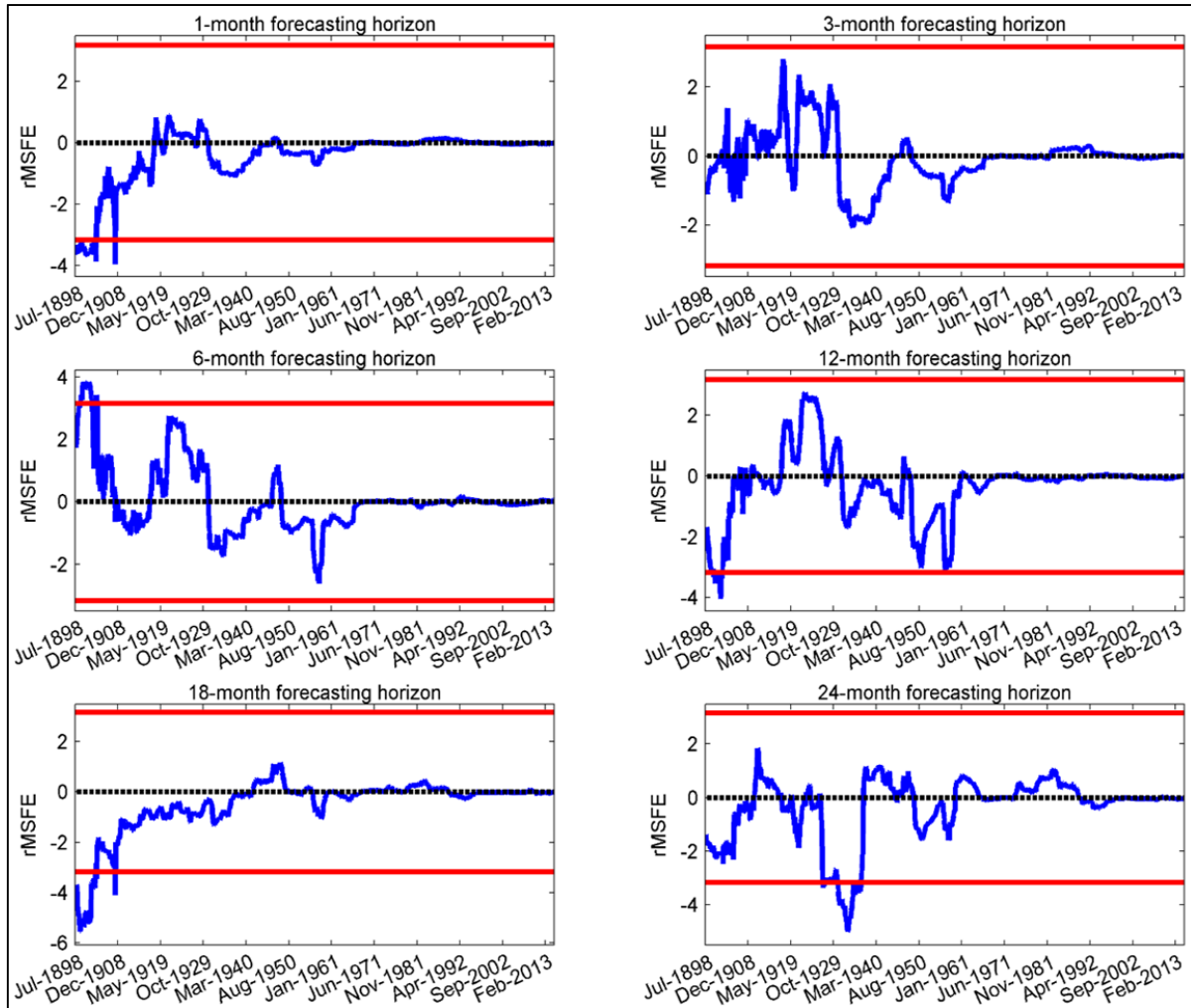


Figure 5: Fluctuation test results for the OLS models. The fluctuating line represents the values of the test statistic while the upper and lower straight lines represent the 90% confidence band of the null hypothesis that the forecasting accuracy of both models is equal.

As we observe from Figure 5 the structural model outperforms the autoregressive with statistical significance scarcely. The structural OLS is more accurate than the autoregressive OLS model before 1900 in 1, 12 and 18-months ahead forecasting and the autoregressive model in 6-months ahead forecasting. Interestingly in the 24-months forecasting horizon the structural OLS outperforms the autoregressive model with statistical significance from the entire 1929 Great recession period and up to 1940, the beginning of WWII. In Figure 6 we report the Fluctuation test results for the LASSO models.

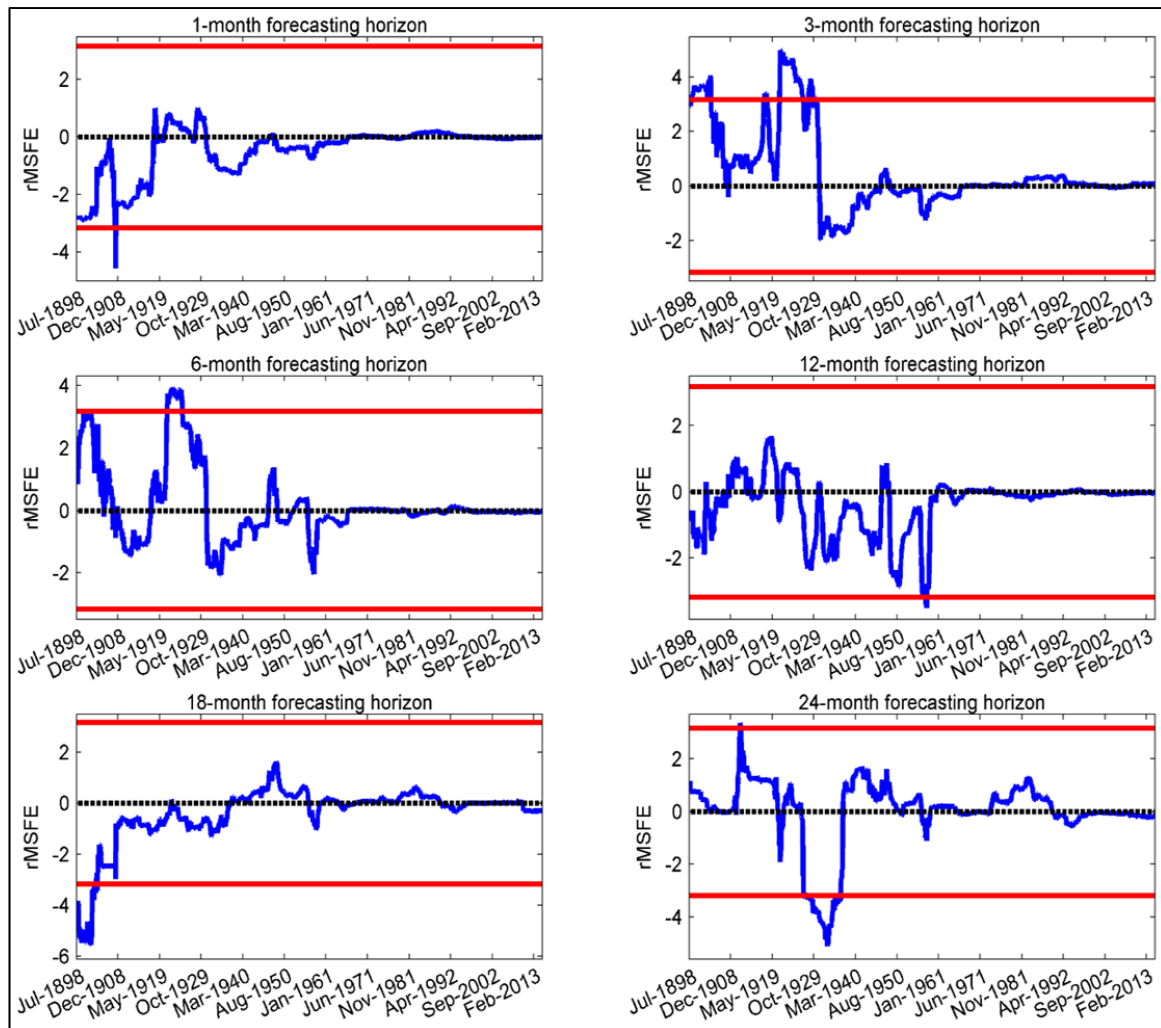


Figure 6: Fluctuation test results for the LASSO models. The fluctuating line represents the values of the test statistic while the upper and lower straight lines represents the 90% confidence band of the null hypothesis that the forecasting accuracy of both models is equal.

The results of the Fluctuation test for the LASSO models (Figure 6) reveal that as with the OLS models we can only scarcely reject the null hypothesis of equal forecasting ability of the two models and for only a small number of rolling windows. We observe some instances where the autoregressive model is more accurate than the structural one with statistical significance and the 1929-1940 period where the structural model outperforms the autoregressive one. On all other instances the null hypothesis cannot be rejected.

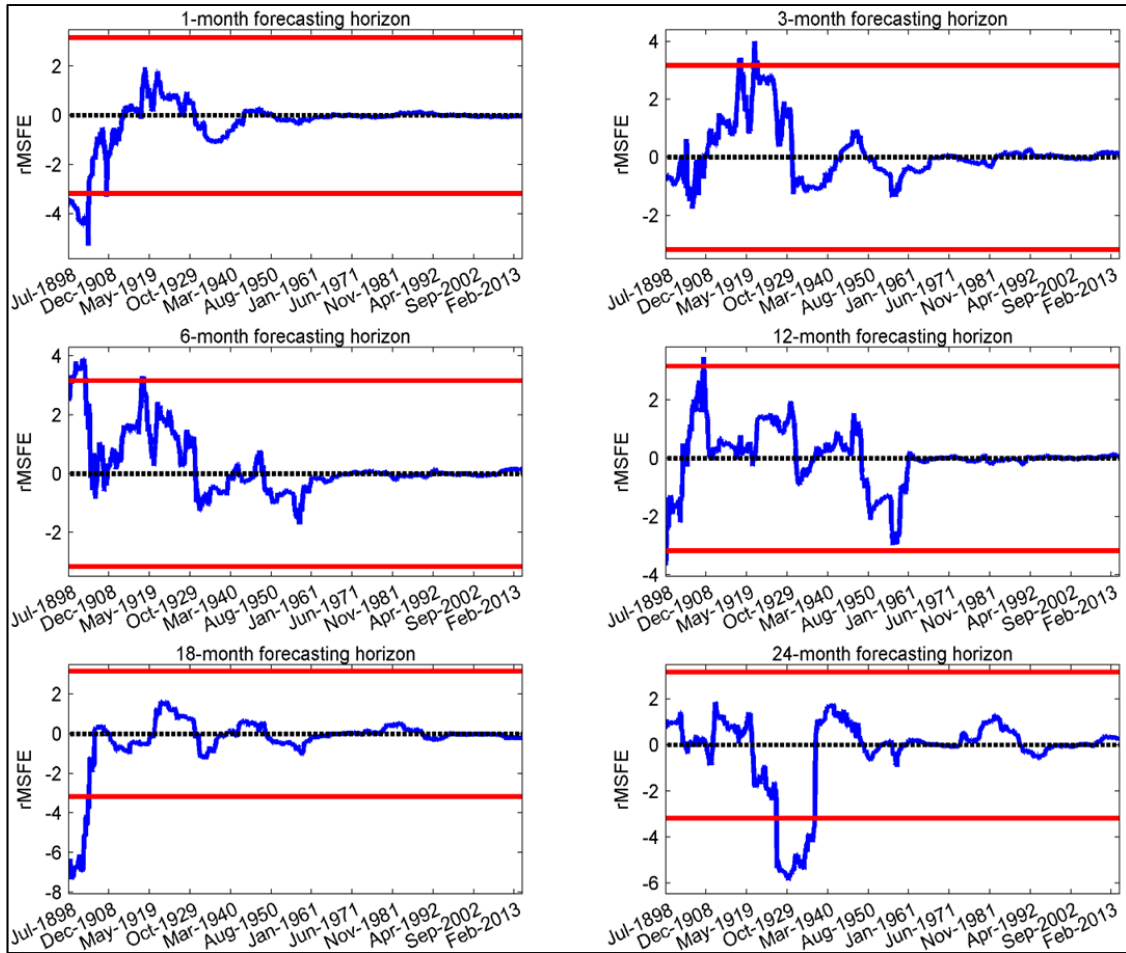


Figure 7: Fluctuation test results for the SVR-LINEAR models. The fluctuating line represents the values of the test statistic while the upper and lower straight lines represents the 90% confidence band of the null hypothesis that the forecasting accuracy of both models is equal.

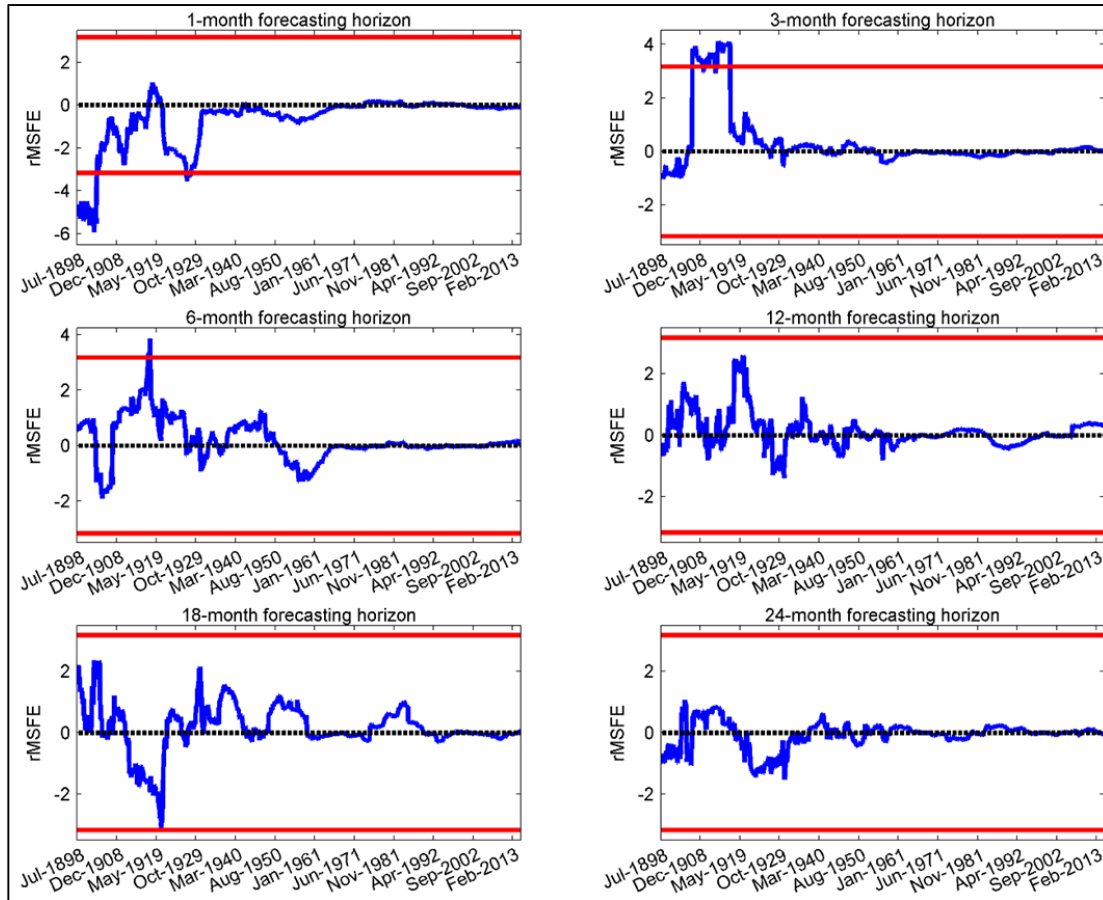


Figure 8: Fluctuation test results for the SVR-RBF models. The fluctuating line represents the values of the test statistic while the upper and lower straight lines represents the 90% confidence band of the null hypothesis that the forecasting accuracy of both models is equal.

Apparently, the results of the Fluctuation test for the SVR-LINEAR (Figure 7) and the SVR-RBF (Figure 8) are very similar with the ones from the OLS and the LASSO models. We reject the null hypothesis of equal forecasting accuracy between the structural and the autoregressive models on a very small number of rolling windows. The only common observation between the OLS, LASSO and the SVR-LINEAR models is the period 1929-1940 in the 24-months ahead forecasting horizon, where the structural outperforms the autoregressive model in all methods. As we observe from Figure 9, during that period the FED lowers the short term interest rate close to zero, while the long-term interest rate is lowered in a smaller degree. Thus, the term-spread reflects almost explicitly the long term interest rate and thus the expectations of the market participants in the future rate of the U.S. inflation. Interestingly, we observe analogous

term-spread values in the post-2008 period (Figure 10), but the term-spread cannot capture the evolution of inflation more accurately than the autoregressive model.

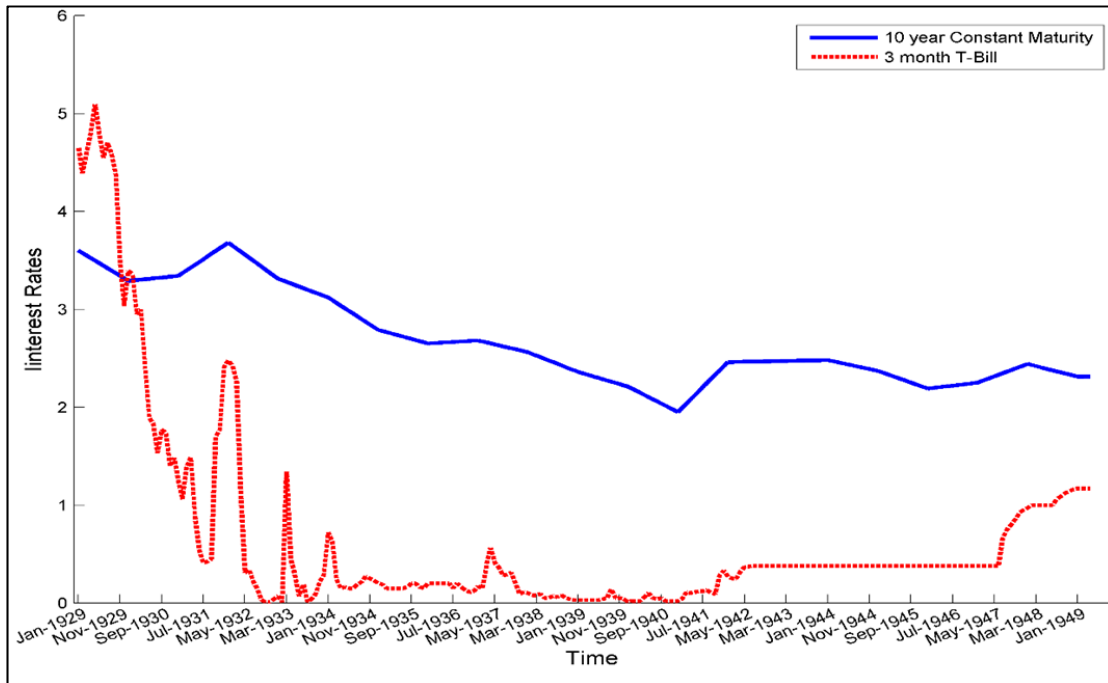


Figure 9: Interest rates evolution during the period 1929-1949.

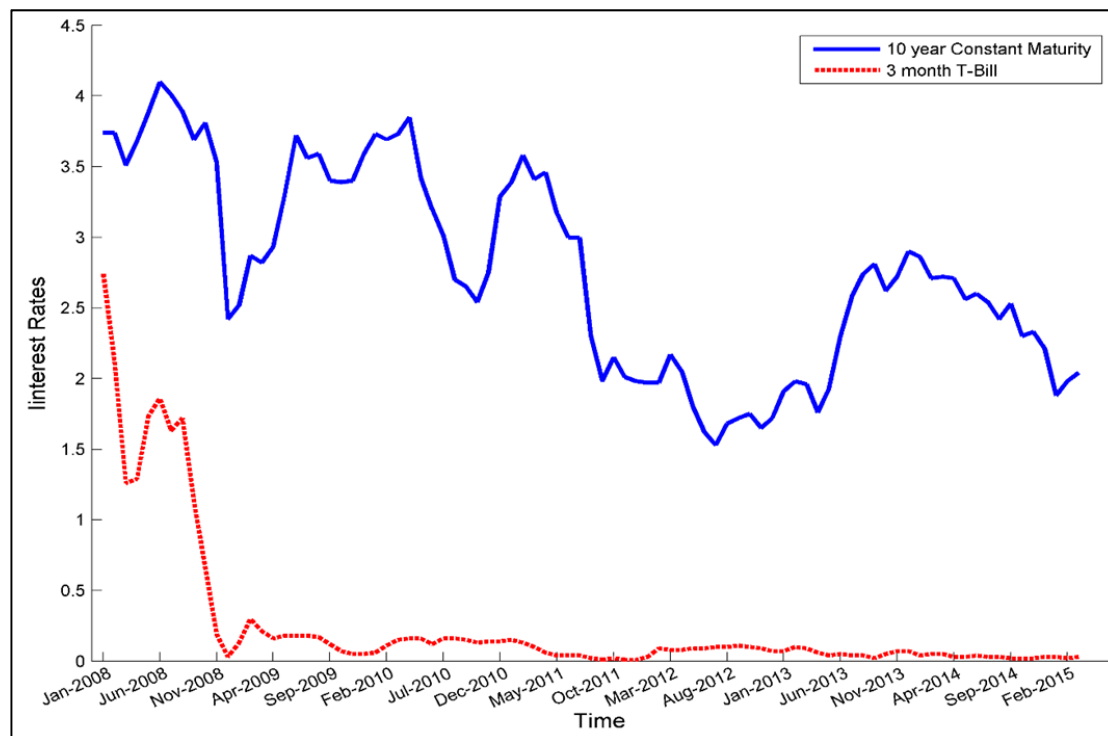


Figure 10: Interest rates evolution during the period 2008-2015.

Overall, the analysis of the Fluctuation test suggests that only in a few rolling windows we are able to forecast more accurately the evolution path of the U.S. inflation based on a term-structure model than an autoregressive one. Our analysis on different linear and nonlinear methodologies and forecasting horizons provides robustness to our results.

4. Conclusions

In this paper we attempt to forecast the U.S. inflation with a very long (covering nearly a century and half) monthly dataset by employing both classical econometric and machine learning methodologies. In doing so, we use the linear OLS, the nonlinear LASSO and the nonlinear SVR models in both an autoregressive and a structural forecasting setting. Our dataset spans the period from January 1871 to March 2015 in monthly frequency. The structural models exploit the informational content of the term-spread in forecasting inflation. In order to evaluate the contribution of the term-spread in forecasting over time we use a 10 year rolling window. Our empirical findings suggest that when we consider the entire out-of-sample period, all models outperform the RW model in terms of the MAPE criterion. The autoregressive and structural models exhibit similar forecasting performance, regardless of the forecasting methodology. This result is also validated and from point-to-point examination based on the Fluctuation test of Giacomini and Rossi (2010), with the exception of the period 1929-1940. Overall, we conclude that models based on the term-spread are not efficient in point-forecasting the U.S. inflation and that the linear models should be preferred over the more complex nonlinear ones. A future line of research would be to evaluate the forecasting ability of the nonlinear models in terms of density forecasts for inflation in comparison to the linear OLS models.

Acknowledgements

Dr. Theophilos Papadimitriou, Dr. Periklis Gogas and Vasilios Plakandaras have 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**, under grant number MIS 380292. Investing in knowledge society through the

European Social Fund. We would like to thank Professor Rossi for providing us with the code of her own paper and Professor Shiller and Professor Goyal for the available datasets.

References

- Álvarez-Díaz, M., and Gupta R. (2015). Forecasting the US CPI: Does Nonlinearity Matter? *Department of Economics, University of Pretoria, Working Paper No. 2015-12*.
- Ang A., Bekaert G. and M. Wei (2007), Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better? , *Journal of Monetary Economics*, vol. 54, pp. 1163-1212.
- Ascari G. and Marrocu E. (2003), Forecasting Inflation: A Comparison of Linear Phillips Curve Models and Nonlinear Time Series Models, *Working Paper CRENoS 00307*, Centre for North South Economic Research, University of Cagliari and Sassari, Sardinia.
- Bai J. and Ng. S (2008) Forecasting economic time series with targeted predictors, *Journal of Econometrics*, vol. 146, pp.304-317.
- Bai J. and Perron P. (2003) Critical values for multiple structural change tests, *Econometrics Journal*, vol. 6, pp. 72-78.
- Bekiros, S and Paccagnini, A. (2013). On the Predictability of Time-Varying VAR AND DSGE Models, *Empirical Economics*, 45 (1), pp. 635-664.
- Bekiros, S and Paccagnini, A. (Forthcoming). Microprudential Policy and Forecasting Using Hybrid DSGE Models with Financial Frictions and State-Space Markov-Switching TV-VARs, *Macroeconomic Dynamics*.
- Berardi A. (2009), Term Structure, Inflation, and Real Activity, *Journal of Financial and Quantitative Analysis*, vol. 44(4), pp. 987-1011.
- Cortes C. and Vapnik V. (1995) Support-Vector Networks, *Machine Learning*, vol 20, pp. 273-297.
- Dickey D. and Fuller W. (1981), Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, vol. 49, pp. 1057-1072.
- Enders W. and Lee J. (2012) A Unit Root Test Using a Fourier Series to Approximate Smooth Breaks, *Oxford Bulletin of Economics and Statistics*, vol. 74 (4), pp. 574-599.
- Estrella, A. and 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*, vol. 41(7), pp. 1375-1401.
- Giacomini R. and Rossi B. (2010), Forecast Comparisons in Unstable Environments, *Journal of Applied Econometrics*, vol. 25 (4), pp. 595-620.

- Goyal, A., and Welch I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, *Review of Financial Studies*, vol. 21(4), pp. 1455-1508.
- Härdle, W., Lee Y-J, Schäfer D. and 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*, vol.28(6), pp. 512-534.
- Inoue A. and Killian L. (2008), How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation, *Journal of the American Statistical Association*, vol. 103 (482), pp. 511-522.
- Jarque C. M., and A. K. Bera. (1987), A Test for Normality of Observations and Regression Residuals, *International Statistical Review*, vol. 5 (2), pp. 163–172.
- Johansen, S.(1991).Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*,vol.59, pp. 1551–1580.
- Jorion P. and F. S. Mishkin (1991), A Multi-Country Comparison of Term Structure Forecasts at Long Horizons, *Journal of Financial Economics*, vol. 29, pp. 59 – 80.
- Khandani, A E., A J. Kim, and A. W. Lo (2010), Consumer credit-risk models via machine-learning algorithms, *Journal of Banking & Finance*, vol. 34(11), pp. 2767-2787.
- Koop, G. and Korobilis, D. (2012). Forecasting Inflation Using Dynamic Model Averaging, *International Economic Review*, vol. 53(3), pp. 867-886.
- Koop, G. and Korobilis, D. (2013). Large Time-Varying Parameter VARs, *Journal of Econometrics*, vol. 177 (2), pp. 185-198.
- Korobilis, D. (2015). Quantile forecasts of inflation under model uncertainty, *MPRA Paper No. 64341*.
- Kwiatkowski D., Phillips, P. C.B., Schmidt, P. and Shin, Y. (1992), Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, *Journal of Econometrics*, vol. 54, pp. 159-178.
- Liu, J., Wu, S. and Zidek, J. V. (1997). On Segmented Multivariate Regression, *Statistica Sinica*, Vol. 7, pp. 497–525.
- Manzan, S. and Zerom, D. (2015). Asymmetric quantile persistence and predictability: the case of U.S. inflation, *Oxford Bulletin of Economics and Statistics*, vol. 77(2), pp. 297-318.
- Marcellino, M. (2008), A Benchmark Model for Growth and Inflation, *Journal of Forecasting*, vol. 27(4), pp. 305-340.
- Mishkin F. S. (1990a), What Does the Term Structure Tell Us About Future Inflation?, *Journal of Monetary Economics*, vol. 25, pp. 77 – 95.
- Mishkin F. S. (1990b), The Information in the Longer-Maturity Term Structure About Future Inflation, *Quarterly Journal of Economics*, vol. 55, pp. 815 – 828.

- Öğüt, H, M. Mete Doğanay, N.B. Ceylan, and R. Aktaş (2012), Prediction of bank financial strength ratings: The case of Turkey. *Economic Modelling*, vol. 29(3), pp. 632-640.
- Perron, P. (1997), Further Evidence on Breaking Trend Functions in Macroeconomic Variables, *Journal of Econometrics*, vol. 80 (2), pp.355-385.
- Phillips P.C.B and P. Perron (1988),"Testing for a Unit Root in Time Series Regression, *Biometrika*, vol. 75, pp. 335–346.
- Plakandaras V., Gupta R., Gogas P. and Papadimitriou T. (2015), Forecasting the U.S, Real House Price Index, *Economic Modelling*, vol. 45, pp. 259-267.
- Rossi B. and Sekhposyan T. (2010) Have Economic Models' forecasting performance for US output and inflation changed over time and when?, *International Journal of Forecasting*, vol. 26(4), pp. 808-835.
- Rubio G., Pomares H., Rojas I. and Herrera L. J.(2011) A heuristic method for parameter selection in LS-SVM: Application to time series prediction, *International Journal of Forecasting*, vol. 27(3), pp. 725-739.
- Schwarz G. E. (1978), Estimating the dimension of a model, *Annals of Statistics*, vol. 6 (2), pp. 461–464.
- Stock J. H. and M. W. Watson (2003), Forecasting Output and Inflation: The Role of Asset Prices, *Journal of Economic Literature*, vol. 41(3), pp. 788-829.
- Stock J. H. and M. W. Watson (2008), Phillips Curve Inflation Forecasts, *NBER Working Paper*, No. 14322.
- Stock, J. H. and Watson, M. W. (2010). Modeling inflation after the crisis, *Proceedings - Economic Policy Symposium - Jackson Hole*, pp. 173–220.
- Tibshirani R. (1996), Regression shrinkage and selection via the lasso, *Journal of the Royal Statistics Society B*, vol. 58 (1), pp. 267-288.
- Vapnik, V., Boser, B. and Guyon, I. (1992) A training algorithm for optimal margin classifiers, *Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, ACM, pp.144–152.
- Zivot E. and Andrews K. (1992), Further Evidence On The Great Crash, The Oil Price Shock, and The Unit Root Hypothesis, *Journal of Business and Economic Statistics*, vol. 10 (10), pp. 251–70.