HELLENIC REPUBLIC

UNIVERSITY OF MACEDONIA
ECONOMIC & SOCIAL SCIENCES

Interdepartmental Program of Postgraduate Studies in Economics

Thesis

FORECASTING INFLATION RATE AND OUTPUT GROWTH

OF

Karaoulanis Dimitrios

Supervisor: Pantelidis Theologos

October 2013

Thessaloniki

CONTENTS

Abstract

This study evaluates the predictive power of various economic and financial variables for inflation and GDP growth. The research covers the 2003-2013 decade including the recent financial crisis. A simple AR model for the variables under scrutiny serves as a benchmark in our analysis. Our empirical findings suggest that the benchmark performs pretty well for short-term forecasts (up to three months). However, we observe significant improvement in the forecast accuracy of the model for long-term forecasts when the proper predictor augments the benchmark. Among the predictors considered in this study, the unemployment related variables (that is, the unemployment rate, the unemployment gap, the NAIRU and the jobless claims) and capacity utilization seem to be the optimal predictors for output growth and inflation respectively.

1. Introduction

Forecasting inflation and output growth is one of the most popular subjects in the forecasting literature. A variety of methods has been developed, especially after the 1973 crisis. Initially researchers focused on the in-sample properties of the candidate models. Later on, researchers managed to develop models that are more sophisticated. These are univariate and multivariate linear and non-linear models that include lagged values of the dependent variable and the regressors. Emphasis was given to the out-of sample forecasting performance of the models. That is instead of comparing the forecasting ability of the model in the regression sample, the sample breaks into two parts, the estimation sample and the forecasting sample. The model estimation occurs in the first part generating forecasts for the second part.

In general, there are two ways to generate the out-of-sample forecasts: The recursive and the rolling scheme. In both schemes, there is a reevaluation of the model at each forecast, since the estimation sample shifts one observation forward in time. In the recursive scheme the estimation sample increase by one observation at each step whereas in the rolling scheme the size of the estimation window is fixed since the first and the last observation shift one-step ahead.

Previous studies consider a variety of forecast horizons depending on the goal of each researcher. Many papers focus on one-step ahead forecasts of the variable of interest, while others extend the analysis to general h-step ahead forecasts (where $h > 1$). In general, there are two ways to produce multi-period forecasts. The first one is the iterated method that iterates the one-step ahead model forward to the desired h-step ahead forecast. The second one is the direct method entailing the adjustment of the model at each forecasting horizon. For a meticulous analysis of the two forecasting method one may address to Marcellino and Watson (2005), According to the authors if the model's lag specification is correct the iterated method yields lower forecasting errors compared to the direct method.

The empirical approach we follow is a linear nested framework using the rolling scheme. The extraction of the multi-step forecasts occurs using the iterated method. These tools allow us to evaluate the predictive capability of a univariate model versus a variety of more general multivariate models that include a financial or a macroeconomic variable as a predictor. To compare the relative forecasting

performance of the models considered in our own analysis, we adopt the Clark and West (2006) test, which is described analytically in subsection 3.3.

The research starts with the literature review. Section 3 describes the data, presents the results of unit root tests and briefly explains the methodology we follow to generate our forecasts. Section 4 provides the results of our forecast exercise for inflation and output growth. The last subsection of section 4 presents the main implications of our findings. Section 5 concludes.

2. Literature survey

The purpose of this section is to summarize the main aspects of some of the most significant papers that are devoted to the forecasting of inflation and output[1]. Based on the literature we choose to include the following variables in our analysis: the term spread, the interest rate, the stock returns, the housing and commodity prices, the nominal exchange rates, the capacity utilization, the jobless claims, the unemployment rate and the unemployment gap (Phillips curve models). In this section we also describe the main econometric techniques and the most significant empirical results of the literature.

2.1. Inflation

Stock and Watson (2003) confirm the prevailing view that the predictive content of many assets for inflation forecasts have diminished since the mid-80s. Estrella and Mishkin (1997), Stock and Watson (2003) and Mishkin (2007) attribute this tendency to a variety of reasons such as the country, the institutional framework and the monetary regime. They also observe that the use of the term spread as a tool of monetary policy diminishes its effectiveness as a predictor. Stock and Watson (2003) claim that in the period 1985-1999, short-term interest rates outperform the term spread for inflation forecasts in terms of the Mean Squared Forecast Error (MSFE) criterion. Mishkin (1990) investigates the predictive power of spread for the U.S. economy. He finds weak

[1] The Appendix provides a brief synopsis of each paper that we mention in this section.

short-term predictability and strong predictive content in the long-term. However, his models do not consider lagged values of the dependent variable. His sole criterion is that the maturities of the term spread have to match with the forecasting horizon under examination. Estrella and Mishkin (1997) argue that if the monetary policy is independent and the economy is unaffected by exchange rates, there is strong predictive content in the term spread for inflation forecasts (for a forecast horizon of one to two years). However, Stock and Watson (2003) in a pseudo-out of sample forecasting exercise, show that the in-sample approach followed by Estrella and Mishkin (1997) is unreliable.

Mishkin (2007) gives another perspective claiming that anchoring is an important factor that affects the level of current inflation. Anchoring is a term used to express the degree in which inflation expectations affect current inflation. He implies that during the 70's the expectations had been unanchored. For this period, New-Keynesian Phillips Curve models and the term spread contained more information about future inflation.

From the early 90's the change of the monetary policy from reactive to forward looking has affected the inflation process. The intervention to the inflation expectation via the yield curve and the monitoring of asset prices and macroeconomic variables belong to that category. This has led to a less persistent and more trend following inflation process. Low autoregressive models perform better than the term spread and the Phillips-Curve based NAIRU models that use unemployment and output gap as predictors.

These findings confirm Atkeson and Ohanian (2001) who use a random walk naïve model. Their model systematically outperforms, in terms of predictive accuracy, three different NAIRU Phillips curve model for one-year ahead forecasts. The first one of these NAIRU Phillips curve models is used by academics, the second one by central banks and the third model is developed by Stock and Watson (1999).

Stock and Watson (2007) divide the inflationary process into two components: a shock that affects the stochastic trend and a serially uncorrelated component that fluctuates around the trend. They argue that the shock of the first component is the source of persistence, since it affects the course of the trend. Mishkin (2007) correlates the two components with the inflation expectation. When the shock to the inflation process roots from the stochastic trend, the expectations are unanchored. On the other hand,

when the second component prevails, the inflation is less persistent and the expectations are anchored.

Exploiting these results Stock and Watson 2007 investigate data from 1970 to 2004 to find why Phillips Curve models have lost their gloom. They find that the only period that these models performed well was from 1970 to 1983, a period with persistent and trend-following inflation. These characteristics are the reason of their superiority. For this period Phillips Curve models consist the most accurate forecasting method for multi-period forecasts. Before 70s and after 1984 though, the shocks in the inflation process come from the serially uncorrelated component. For these periods, low autoregressive models that use Akaike Information Criterion (AIC) for the lag length selection and the Atkeson Ohanian model provide better forecasts.

The adaptability of a forecasting model from one period to another has to do with several factors.

- The inflation process and more specifically the dynamic of the volatility as explained in the previous paragraph.
- The type of the monetary policy that the central banker exercises.
- The existence of time varying parameters in the model.

These are the main reasons why multivariate forecasting models have lost their gloom and lose ground toward more accurate univariate time-varying IMA and UC-SV (unobserved component model with stochastic volatility) which are more adaptive to these time varying conditions.

Fisher, Liu, Zhou (2002) use data from 1977 to 2000. They find that NAIRU Phillips curve models are more accurate in predicting the direction of inflation especially during the 1977-1984 period (when inflation is volatile). Another advantage is the capability of these models to generate reliable long-term predictions assuming a stable monetary policy regime. In contrast, the Atkeson Ohanian (2001) model provides reliable forecasts for the period 1985-2000. In this period there is a monetary policy shift while the volatility is low; its use for forecasting purposes is valid not only for short-term but also for long-term forecasts.

Bachmeier and Leelahanon (2007), using data from 1994 to 2002, evaluate the forecasting performance of nonlinear models. Their out-of-sample inflation forecasting exercise generally shows that nonlinear models are superior to linear specifications.

Their non-parametric model becomes even more accurate when money growth and velocity enter their model.

Goodhart and Hofmann (2000) assess a linear framework in an attempt to extract the predictive power of asset prices by evaluating the statistical significance of the predictive content of GDP growth, housing prices, equity prices and yield spread. The baseline model contain an AR (1) approximation of inflation, GDP growth, exchange rates and interest rates. They perform an in-sample forecasting exercise for 4 and 8 quarters.

GDP growth seems to have predictive content for one-year forecasting horizon. Furthermore, housing prices outperform all other predictors from one to two years forecasting horizon.

Patton and Timmermann (2011) confirm these results finding that inflation is more persistent than G.D.P. growth. They argue that low AR process are appropriate not only for short but also for long horizons.

Apart from the traditional models that contain lags of the dependent variable and financial or nonfinancial independent variables, Ang, Bekaert and Wei (2006) assess the predictive content of surveys like the SPF, the Livingston and the Michigan inflation expectation. Their comparison also involves time series ARIMA models, Phillips curve based models and term spread models. They conclude that surveys are superior mainly because they aggregate the available information and because participants respond faster to changes in the data generating process.

Another alternative is the inflation sentiment indicators developed by Dohrn Schmidt and Zimmermann (2008).The construction of such indicators initiates from the fact that if a price change roots from many items of the CPI basket then this price change will be more intense in the future. On the other hand if the price change starts from the high volatility of few items then it will not last.

The authors investigate the case of US and Germany using quarterly data from 1978 to 2006. The examined indices are a skewness index, a diffusion index and a momentum index. The results show that in relation to the Phillips curve benchmark the sentiment indicators outperform the traditional methods in most cases especially for horizons up to two years. Furthermore, the supremacy of those indicators is more evident when inflation is volatile.

2.2 Output growth

Estrella and Hardouvelis (1991) find predictive content for GDP Growth using the slope of term spread up to 4 years ahead. They underlie though, that its predictive content has diminished since the mid-80s, an evidence also supported also by Dotsey (1998) and Wheelock and Wohar (2009).

Furthermore, Estrella and Hardouvelis (1991) find that linear models that include the term spread generate reliable forecasts for GDP growth. They find that the term spread outperforms the lagged GDP, the lagged inflation, the short-term interest rates, the forecast surveys and the index of leading indicators. Despite the existence of nonlinearities and structural breaks, Dotsey (1998) goes further by including the lag GDP and federal fund rate. He investigates for nonlinear structure and concludes that nonlinear models improve the long-term forecasting performance. Estrella and Mishkin (1997) find that the inclusion of real stock returns in a model with term spread as a predictor improves the forecasting accuracy at the short-term. Stock and Watson (2003) cover an extended bibliography regarding the most appropriate predictors for GDP growth. In brief, their bivariate and trivariate models with a lagged dependent variable and an individual asset price or macro variable are instable as far as their forecasting performances is concerned. They argue that a combination of predictors outperforms the linear GDP growth rate forecasts.

A general observation is that the usefulness of term spread as a predictor depends on the time horizon, the country and the persistence of GDP growth; the more persistent the progress is, the more accurate the forecasting values are. However, Patton and Timmerman (2011), who assess the forecasting performance of professional forecasters taking into account the reliability of the data collected, point out that the error is greater for forecasts concerning GDP than those of inflation confirming the already mentioned argument that the process of the GDP growth is less persistent.

Some researchers focus on forecasting recessions using the term spread. Inversions of the yield curve tend to precede a crisis. Generally, a probit model that includes the term spread is able to signal a pending recession 6 to 12 months before the event. Even so, many authors among them Wheelock and Wohar (2009), question the effectiveness of the term spread because it barely predicted the recessions of 1990 and 2001 and the recent one in 2008.

3. Data and forecasting methodology

3.1 Data manipulation

For the current analysis, we use 11 series that include 546 monthly observations starting from 1/1/1968 to 1/6/2013 and 2 series the Michigan Inflation Expectation and the U.S. Dollar Index that contain 425 observations starting from 1/1/1978[2] to 1/6/2013. The source of the data is the database of the Federal Reserve Bank of St. Louis.

We use Industrial Production (IP) as a measure of output. Output is calculated as $\Delta \log IP_t = 100 \log(IP_t/IP_{t-1})$ where t denotes the period and the factor 100 standardizes the units to percentage growth rates. As a measure of inflation, we use the Consumer Price Index (CPI). The inflation rate is calculated as follows: $\Delta \log \mathrm{CPI}_t = 100\log(\mathrm{CPI}_t/\mathrm{CPI}_{t-1})$.

The selection of monthly observations ensures a sufficient amount of observations and access to a variety of data collected in this frequency. For initial jobless claims and stock returns, we aggregate the weekly and daily data by taking the 4-weeks and the 30-days average respectively

According to Mishkin (1990), the yields that construct the spread have to match with the inflation change under examination. Since we examine the predictive content for up to two years (24 months) the appropriately specified term spread should be the difference between the 2-year treasury yield and the 1-month T-bill rate. However due to data inadequacy this research examines the closest approximation of the term spread (Spread) which is the difference between the 3 year Treasury bond and the Federal Fund Rate (FFR).

[2] Unfortunately, no data are available for these two variables before 1978.

Table 1: Description of series

Notes: Table 1 presents the series name, the original sampling frequency, the symbol used in the text and a brief description.

Series	Sampling Frequency	Symbol	Description
Capacity utilization	M	Caput	Capacity Utilization: Total Industry
Exchange rate	M	Exrate	Trade Weighted U.S. Dollar Index: Broad index
Federal fund rate	M	FFR	Effective Federal Funds Rate
Housing start	M	House	Housing Starts: Total: New Privately Owned Housing Units
Inflation rate	M	CPI	Consumer Price Index for All Urban Consumers: All Items
Initial claims	W	Claims	Unemployment Insurance Weekly Claims Report
M2	M	M2	M2 Money supply
NAIRU	M	NAIRU	Non accelerating inflation rate unemployment
Output growth	M	IP	Industrial Production Index
Stock returns	D	SnP	S&P 500 Stock Price Index
Survey	M	Survey	University of Michigan Inflation Expectation
T-Bill	M	T-bill	3-Month Treasury Bill: Secondary Market Rate
Term spread	M	Spread	The difference between 2 year Treasury and the Federal Fund rate
Treasury	M	Bond	3-Year Treasury Constant Maturity Rate
Unemployment	M	Un	Unemployment rate
	M	Un-gap	The difference between the Unemployment rate and the NAIRU

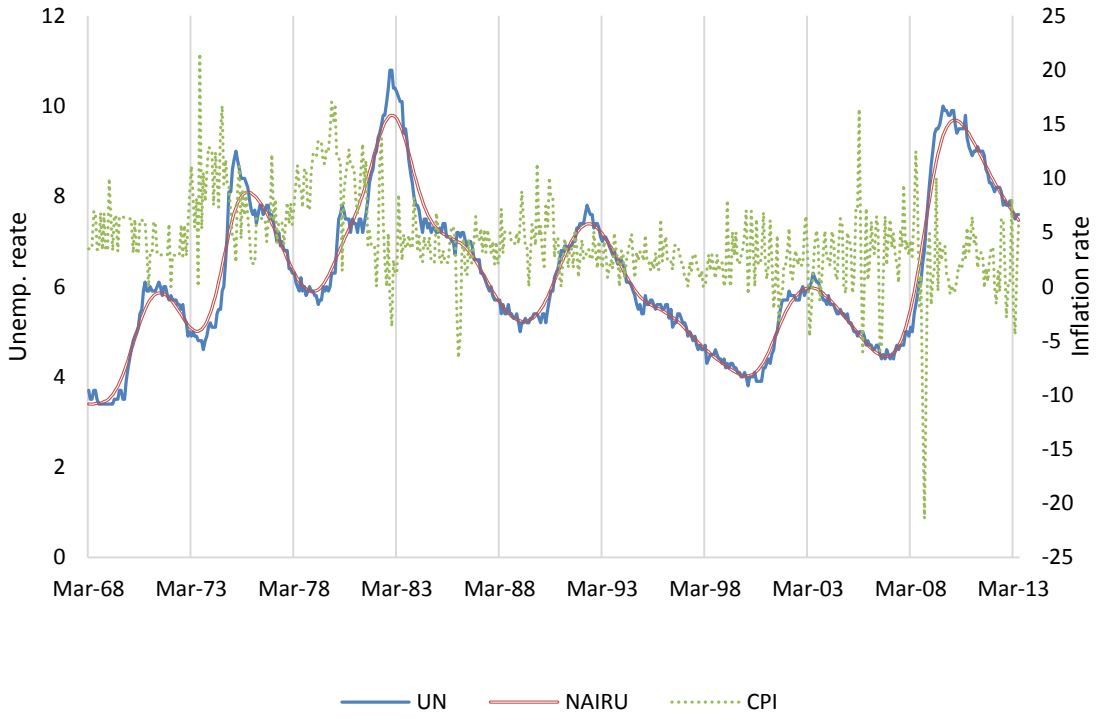
In order to assess the Phillips curve forecasting performance within the current framework, the estimation of the Non Accelerating Inflation Rate of Unemployment

(NAIRU) is required. We calculate the NAIRU on a monthly basis following the methodology described by Ball and Mankiw (2002). The construction of the NAIRU is a function of the form:

$$NAIRU_t = f_t(cpi_t, un_t, a) = UN_t + \Delta^2 \log \mathrm{CPI}_t / a$$

UN_t denotes the level of unemployment and parameter α equals the value of 0.63/12=0.0525, which is the constant annual inflation-unemployment tradeoff coefficient as calculated by Ball and Mankiw (2002) and others, divided by 12 to adjust for monthly data. Finally, the calculation of this series requires the employment of the Hodrick-Prescott filter, Hodrick and Prescott (1997), setting the Lambda smoothness parameter to its default value of 1600. Figure 1 graphs the unemployment rate, the constructed NAIRU and the inflation rate.

Figure 1
The unemployment rate (UN), the derived nairu and the inflation rate during the 1968-2013 sample



An inspection of Figure 1 that covers the period from 1/3/1968 to 1/6/2013 reveals two facts: When the unemployment rate lies above or below the NAIRU, the inflation rate moves to the opposite direction. Secondly, when the unemployment rate is in line with the NAIRU the inflation rate follows a steady rate confirming the fact that the NAIRU is the rate at which the inflation is non-accelerating. Having defined the NAIRU, the

next step is the calculation of the Unemployment gap (UN-gap) which is the difference between the unemployment rate (UN) and the NAIRU.

3.2. Unit root tests

This section presents the results of Augmented Dickey Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) unit root tests. We apply the tests to all the variables under scrutiny. This procedure guarantees the consistency of the forecasts generated from the estimation of OLS regressions.

The ADF tests for the null that the series has a unit root against the alternative that it is stationary, meaning that a rejection of the null indicates that the series is stationary. Table 2 summarizes the critical values for 1%, 5% and 10% significance levels using a constant and a trend in the specification of the test respectively.

Table 2

ADF Test: Null Hypothesis- Series has a unit root

Test Critical Values	1% level	5% level	10% level
With Constant	-3,44	-2,87	-2,57
With Constant and trend	-3,98	-3,42	-3,13

On the other hand the KPSS test examines the null hypothesis that the series is stationary meaning that a failure to reject the null indicates that the series is stationary. Table 3 summarizes the critical values of the KPSS test for 1%, 5% and 10% significance levels using both a constant and a trend specification of the test.

For the sake of robustness, the results of both tests should provide evidence that a series is stationary. The reason is that for each case a test with the null that the series is stationary and a test with the null that the series has a unit root, give stronger results than the application of one test, as De Jong and Whiteman (1992) and others observe.

Table 3

KPSS Test: Null Hypothesis-Series is Stationary

Asymptotic Critical Values	1% level	5% level	10% level
With Constant	0,74	0,46	0,35
With Constant and trend	0,22	0,15	0,12

Table 4 presents the results of the two tests for each series under examination. The first column contains the symbol of the variable and the second column depicts the appropriate transformation. The remaining four columns show the results for the ADF and the KPSS test. In both tests, we present the results for two specifications. The first one with constant and the second one with constant and linear trend.

Table 4

Unit root test

Notes: The values denoted with 1 asterisk indicate the rejection of null at a 5% confidence level, while those with 2 asterisks reject the null hypothesis at a 1% confidence level

Variable	Transformation[3]	ADF t-Statistic		KPSS LM-Stat	
		cons	trend+const.	cons	trend+const.
Caput	log	-2,95*	-4,04*	0,60*	0,13*
Claims	log	-3,21*	-3,13*	0,43	0,23**
CPI	Δlog	-5,97**	-12,92**	1,57**	0,13*
Exrate	Δlog	-14,97**	-15,00**	0,16	0,08
House	Δlog	-31,88**	-31,88**	0,07	0,04
IP	Δlog	-8,39**	-8,41**	0,07	0,04
M2	Δ^2log	-13,69**	-13,67**	0,05	0,06
NAIRU	Δ	-4,85**	-4,88**	0,09	0,08
SnP	Δlog	-18,30**	-18,27**	0,10	0,10
Spread	level	-4,50**	-4,59**	0,25	0,07
Survey	log	-3,28*	-4,98**	0,08	0,35**
T-bill	Δ	-5,99**	-6,01**	0,09	0,03
Un	gap	-6,08**	-6,08**	0,01	0,01

[3] The transformations are level=S_t log=logS_t, Δ=S_t-S_{t-1}, Δlog= logS_t-logS_{t-1}, Δ^2log = (logS_t-logS_{t-1}) - (logS_{t-1}-logS_{t-2}). The gap symbol refers to the difference between the unemployment rate and the NAIRU.

Both test results indicate that the series under examination are stationary. Table 5 reports some basic descriptive statistics for all series considered in this study.

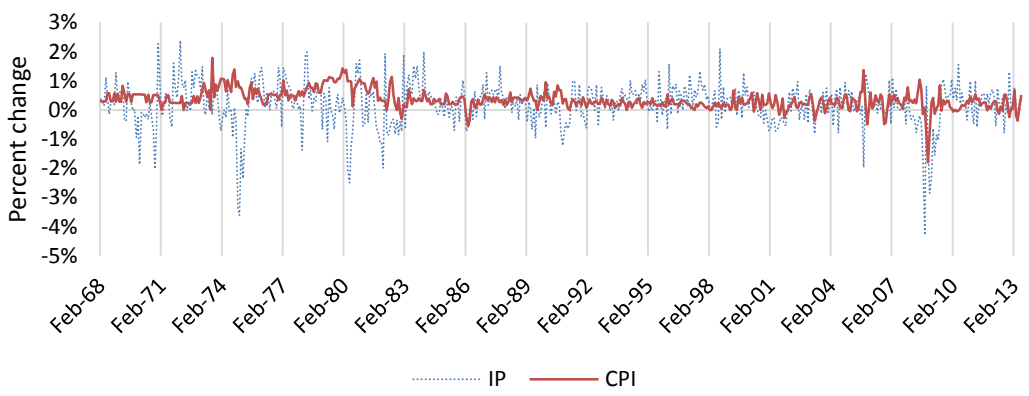
Table 5

Descriptive statistics

		Mean	Median	Max	Min	Std. Dev.	Skew.	Kurt.	J. Bera	Obs.
Caput	log	4,39	4,39	4,49	4,20	0,05	-0,63	3,59	44,14	546,00
Claims	log	12,79	12,78	13,40	12,11	0,23	-0,26	3,98	28,00	546,00
CPI	Δlog	4,23	3,53	21,53	-21,44	3,96	0,04	7,58	475,97	545,00
Exrate	Δlog	0,00	0,00	0,06	-0,04	0,01	0,11	4,27	29,50	425,00
House	Δlog	0,00	0,00	0,26	-0,31	0,08	-0,08	3,83	16,30	545,00
IP	Δlog	2,22	3,00	28,53	-51,64	8,97	-1,16	8,11	715,95	545,00
M2	Δ^2log	0,00	0,00	0,02	-0,02	0,00	-0,49	11,64	1713,25	544,00
NAIRU	Δ	0,01	-0,02	0,24	-0,15	0,08	0,80	3,16	58,46	543,00
SNP	Δlog	0,01	0,01	0,12	-0,21	0,03	-1,02	6,91	441,04	545,00
Spread	level	0,42	0,64	2,76	-5,85	1,26	-1,43	6,23	423,09	546,00
Survey	log	1,23	1,13	2,34	-0,92	0,37	0,86	7,09	349,76	426,00
T-bill	Δ	0,00	0,00	0,20	-0,39	0,06	-1,51	11,12	1707,09	545,00
UN	gap	0,00	-0,01	1,26	-0,83	0,28	0,72	5,91	238,83	544,00

As the standard deviation indicates, the two dependent variables seems to be the most volatile processes. Furthermore, IP has more than twice the variance of the CPI. Figure 2 depicts the two processes from 1968.

Figure 2
CPI-IP rate



3.3. Forecasting methodology

The evaluation of the forecasting performance occurs within a framework of nested models. That entails, the comparison between a parsimonious null model (model one) and a larger model that nests the null model (model two). The first one is a univariate autoregressive model of the variable to be predicted (CPI or Industrial Production) where the lag length is selected by means of the Schwartz Bayesian Information Criterion (SIC). The larger model enriches model one with lagged values of the candidate predictors. SIC once more, defines the selection of the lag length for model two given the lag specification of model one. It is obvious that if the extra parameters of model two are non-significant then the two models are equivalent.

The estimation of the equations described above occurs using a rolling regression window of 408 observations (R). The selection of rolling scheme allows the adjustment of the regression coefficient as the sample moves forward without distortion by the effect of an increasing sample, as it is the case in the recursive scheme. Based on our initial sample that covers the 1969:01- 2002:12 period SIC selects the optimal lag length for both models.

The basic purpose of the empirical analysis is the extraction of multi-step ahead forecasts. We follow Marcellino and Watson (2005) and apply the iterated method. To derive the forecasts we consider the following forecast horizons: 1, 3, 6, 9, 12, 18 and 24 months. That is, the h step ahead forecast stems from the iteration of the one-step ahead model. The one-step ahead model for the two specifications under investigation are

One-step ahead forecasting	
Model one	Model two
$y_{t+1} = \alpha + \sum_{i=1}^{p} \varphi_i y_{t+1-i} + \varepsilon_{t+1}$	$y_{t+1} = \beta + \sum_{i=1}^{p} \varphi_i y_{t+1-i} + \sum_{j=1}^{q} \gamma x_{t+1-j} + \varepsilon_{t+1}$

The symbols y_t and x_t denote the predicted-dependent variable and the candidate predictor respectively. The coefficients φ_i and γ_j are estimated recursively while p and q denote the optimal lag length. Iterating forward the one-step ahead models, we generate the h-step ahead forecasts.

h-step ahead forecasting	
Model one	Model two
$\hat{y}_{1,t+h\|t} = \hat{\alpha} + \sum_{i=1}^{p} \hat{\varphi}_i \hat{y}_{t+h-i\|t}$	$\hat{y}_{2,t+h\|t} = \hat{\beta} + \sum_{i=1}^{p} \hat{\varphi}_i \hat{y}_{t+h-i\|t} + \sum_{j=1}^{q} \hat{\gamma}_j \hat{x}_{t+h-j\|t}$

The rolling regression window for the one-step ahead forecasting models generates 126 out-of-sample forecasts (P) covering the period 2003:01-2013:06. As the forecasting horizon h increases, the generated forecasts are equal to 126-h+1 covering the period from 2003:01+h-1 to 2013:06.

We denote the realized change of the dependent variable within the t and t+h period as $Y_{t+h} = \sum_{k=1}^{h} y_{t+k}$.

On the other hand, the notation $\hat{Y}_{1,t+h} = \sum_{k=1}^{h} \hat{y}_{1,t+k|t}$and $\hat{Y}_{2,t+h} = \sum_{k=1}^{h} \hat{y}_{2,t+k|t}$ refers to the h-period ahead return forecasted by model one and two respectively. In this way, it is feasible to measure the forecasting accuracy of our models.

A common tool in the forecasting literature that measures the forecasting performance is the Mean Squared Forecasting Error (MSFE). This is the squared difference between the forecast and the actual value of the predicted variable at each period in the forecasting sample $MSFE_{1,t} = \left(Y_{t+h} - \hat{Y}_{1,t+h}\right)^2 / (P - h + 1)$ for model one and $MSFE_{2,t} = \left(Y_{t+h} - \hat{Y}_{2,t+h}\right)^2 / (P - h + 1)$for model two while P-h+1 is the number of forecasts. The ratio of these squared differences is the Relative Mean Squared Forecasting Error (RMSFE).

$$RMSFE_{t+h} = \frac{\left(Y_{t+h} - \hat{Y}_{2,t+h}\right)^2 / (P - h + 1)}{\left(Y_{t+h} - \hat{Y}_{1,t+h}\right)^2 / (P - h + 1)}$$

A RMSFE below unity suggest that the predictor included in model two improves the forecasting performance of the model for the variable of interest.

We can also test the null hypothesis that the two models are equivalent in terms of their forecasting accuracy for the variable of interest, against the alternative that model two outperforms model one by means of the methodology developed by Clark and West (2006)

We adopt the Clark and West (2006) one-sided t-statistic. The functional form of the test is:

$$\hat{f}_{t+h} = \left(Y_{t+h} - \hat{Y}_{1t,t+h}\right)^2 - \left[\left(Y_{t+h} - \hat{Y}_{2t,t+h}\right)^2 + \left(\hat{Y}_{1t,t+h} - \hat{Y}_{2t,t+h}\right)^2\right]$$

To test the null of equal MSFE, one regresses $\hat{f}_{t+h}$ on a constant and tests its statistical significance. One rejects the null hypothesis of equal MSFE for a 10% and 5% confidence level if the t-statistic value is greater than 1,282 and 1,645 respectively. The validity of these numbers depend on the P/R ratio as Clark and West (2006) underlie. Their empirical approach takes the value of 0.3 approximately for this ratio. In this study, the ratio equals 0.308 for one-step ahead forecasting and decreases as the horizon increases.

4. Empirical results and implications

Section 4 presents the empirical results of our forecast exercise. Subsection 4.1 focuses on forecasting inflation while subsection 4.2 focuses on forecasting output growth. In subsection 4.3, we summarize the main findings and underlie the main differences between the two macroeconomic variables. Based on the forecasting results, we draw some implications regarding the usefulness of the autoregressive models and the examined predictors for policymaking.

4.1 Forecasting inflation rate

Initially we present some figures associated with the in-sample explanatory power of the model. An attempt to draw conclusions regarding the usefulness of these models in the forecasting process based on these figures would be futile though. The reason is that in many cases in the literature the performance of the in-sample fit is not related to the out-of-sample forecasting accuracy of the models under investigation. Our goal is to examine the relation between the in-sample explanatory power of our models and the ability to generate accurate out-of-sample forecasts for the variable of interest.

We begin our forecast exercise by selecting an AR (p) model for inflation. This model (model one) serves as a benchmark in our analysis. The selected lag order by the SIC is $p = 3$. The estimated AR (3) model is given in the first row of Table 6.

Table 6

In-sample estimation results: CPI

Notes: The regression sample ends on 1/12/2002 including 408 observations. An exception is the exrate and the survey whose estimation sample includes 298 observations. Values denoted with one asterisk reject the null of a zero coefficient at a 5% level while, values with two asterisks reject the null hypothesis at a 1% level. Standard errors are reported in parentheses.

		α	φ_1	φ_2	φ_3	γ_1	R^2
Model 1		0,08**	0,42**	0,20	0,18**		0,49
		(4,22)	(4,19)	(1,79)	(3,04)		
Caput	log	-2,77*	0,40**	0,18	0,18**	0,65**	0,50
		(-2,28)	(3,90)	(1,74)	(3,01)	(2,35)	
Claims	log	0,86	0,42**	0,20	0,19**	-0,06	0,49
		(1,36)	(4,11)	(1,80)	(3,14)	(-1,23)	
Exrate	Δlog	0,06**	0,58**	-0,01	0,26**	0,55	0,57
		(3,23)	(7,00)	(-0,13)	(4,04)	(0,46)	
House	Δlog	0,08**	0,42**	0,20	0,17**	-0,20	0,49
		(4,37)	(4,19)	(1,86)	(2,91)	(-1,44)	
IP	Δlog	0,07**	0,42**	0,20	0,20**	0,02	0,49
		(3,37)	(4,14)	(1,81)	(3,18)	(1,38)	
M2	Δ^2 log	0,08**	0,42**	0,18	0,20**	-8,47**	0,50
		(4,23)	(4,19)	(1,60)	(3,30)	(-2,62)	
NAIRU	Δ	0,08**	0,42**	0,19	0,18**	0,07	0,49
		(3,94)	(4,18)	(1,78)	(2,91)	(0,44)	
SNP	Δlog	0,08**	0,43**	0,20	0,18**	0,19	0,49
		(3,89)	(4,17)	(1,81)	(3,00)	(0,57)	
Spread	level	0,15**	0,37**	0,16	0,15**	-0,04**	0,52
		(4,49)	(3,52)	(1,61)	(2,33)	(-3,50)	
Survey	log	-0,14*	0,45**	-0,10	0,14	0,25**	0,60
		(-1,97)	(5,68)	(-1,24)	(1,64)	(2,94)	
T-bill	Δ	0,09**	0,38**	0,19	0,21**	0,06**	0,52
		(4,75)	(3,66)	(1,72)	(3,78)	(2,47)	
UN	gap	0,08**	0,41**	0,19	0,18**	-0,06	0,49
		(4,31)	(4,05)	(1,76)	(3,01)	(-1,32)	

The model explains 49% of the variability of inflation. We then augment the specification of model one with lags of one of our predictors. In all cases, we choose (based on SIC) to include one lag of the predictor in our model. It turns out that the lag of the predictor is statistically significant in only five (out of 12) cases. Specifically, caput, M2, survey, spread and T-bill enter in model two with a statistically significant coefficient. On the other hand, all other predictors are not important (in-sample). In most cases, the R^2 of our regression remain close to 49%. The only exception are the models that include either survey or exrate. In these cases, R^2 increases to 60% and 57% respectively.

Table 7 reports the RMSFE for all the predictors considered in this analysis and for forecast horizon h=1, 3, 6, 9, 12, 18 and 24 months. Entries with an asterisk highlight cases where the forecasting performance of model two is statistically superior to that of model one.

As the forecast horizon increases, the $MSFE_{1,h}$ increases too. We observe a significant increase in the $MSFE_{1,h}$when the forecast horizon exceeds six months.

The main findings for each predictor considered in our analysis are as follows:

- Capacity utilization: Models that include caput outperform the benchmark in all cases. The superiority of model two over model one is statistically significant for all forecast horizons and it is more evident for long-term forecasts. Specifically for $h = 1$, RMSFE equals 0.98. As h increases, RMSFEs decrease reaching $0{,}43$ for $h = 24$. Caput seems to be the second best predictor among the ones considered in this study.
- Exchange rate: Models that embed exrate outperform the benchmark at h= 1, 12, 18 and 24 forecast horizons. For $h = 1$, RMSFE equals 0.95, while from $h = 12$ and on, RMSFEs decrease reaching 0.81 for $h = 24$. This evidence is in line with the theory that inflation of open economies is susceptible to exchange rates.

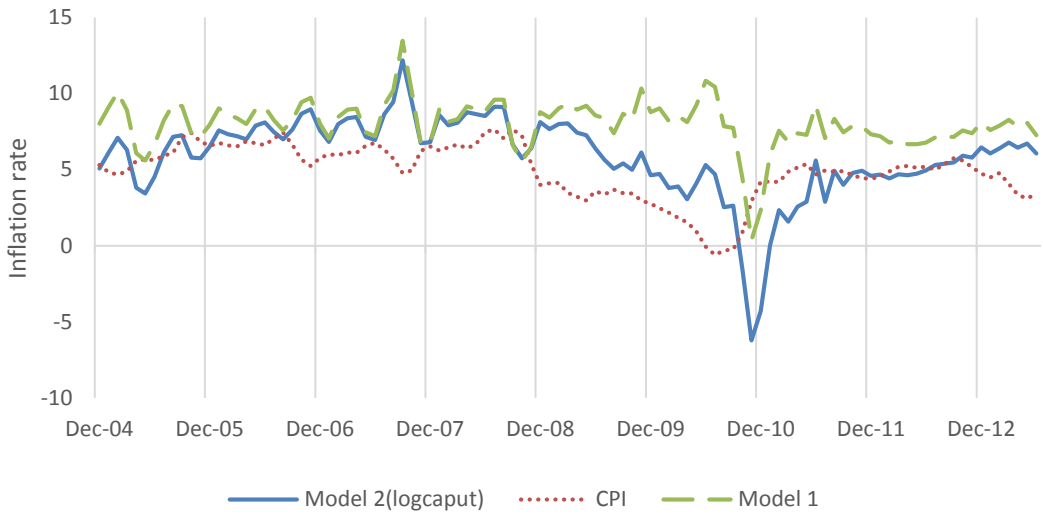
Table 7

RMSFE and Clark and West test results

Notes: The second row depicts the MSFE values of model one for the forecasting horizons under examination. The first column show the candidate predictor. The entries in bold are the RMSFEs and the entries in bracket are the Clark and West t-statistic. One asterisk corresponds to rejection of the null hypothesis of equal MSFE at 10% confidence level. Two asterisks correspond to rejection of the null hypothesis of equal MSFE at 5% confidence level.

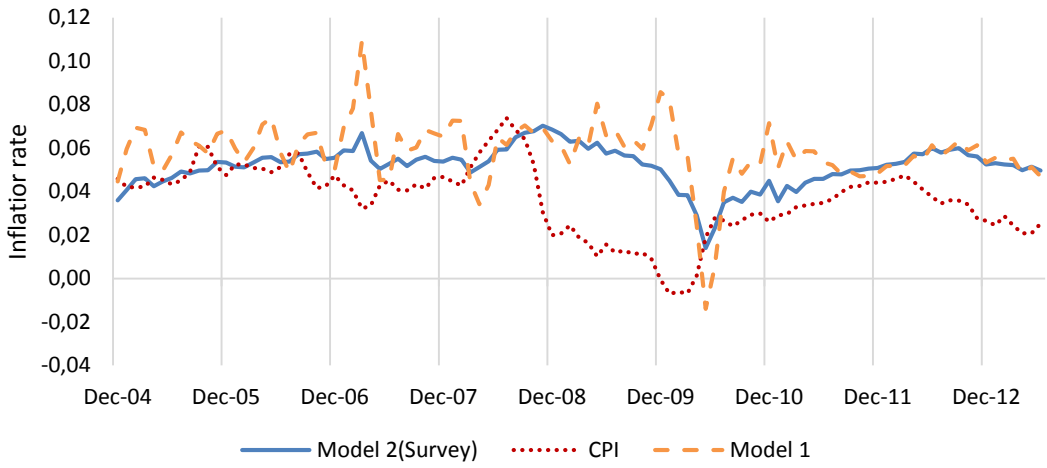
Forecast horizon		1	3	6	9	12	18	24
Model 1 $MSFE_{1,h}$		0,00	0,01	0,02	0,04	0,06	0,09	0,16
Caput	log	**0,98***	**0,95****	**0,90****	**0,84****	**0,76****	**0,56****	**0,43****
		(1,57)	(2,25)	(2,07)	(2,24)	(2,62)	(3,32)	(3,43)
Claims	log	**1,01**	**1,01**	**1,03**	**1,04**	**1,05**	**1,04**	**1,02**
		(-1,02)	(-1,10)	(-1,01)	(-0,98)	(-0,89)	(-0,39)	(0,01)
Exrate	Δlog	**0,95****	**0,99**	**1,00**	**0,98**	**0,96****	**0,88****	**0,81****
		(1,81)	(0,58)	(0,48)	(1,20)	(2,31)	(4,15)	(5,33)
House	Δlog	**0,99***	**1,00**	**1,02**	**1,04**	**1,05**	**1,09**	**1,11**
		(1,58)	(-0,26)	(-0,53)	(-0,86)	(-1,02)	(-1,69)	(-2,30)
IP	Δlog	**1,02**	**1,02**	**1,02**	**1,00**	**0,97**	**0,88**	**0,83****
		(-0,65)	(-0,27)	(-0,05)	(0,32)	(0,65)	(1,55)	(2,11)
M2	Δ^2log	**1,02**	**0,99**	**0,98**	**0,98**	**0,99**	**1,00**	**1,00**
		(-0,53)	(0,65)	(1,26)	(1,21)	(0,98)	(0,50)	(0,81)
NAIRU	Δ	**1,01**	**1,05**	**1,12**	**1,21**	**1,29**	**1,44**	**1,54**
		(-0,82)	(-1,08)	(-1,27)	(-1,45)	(-1,65)	(-2,01)	(-2,36)
SNP	Δlog	**0,97**	**0,96**	**0,95**	**0,93**	**0,92**	**0,93**	**0,92***
		(1,05)	(0,95)	(0,87)	(0,96)	(1,04)	(1,06)	(1,39)
Spread	level	**0,96****	**0,92****	**0,90****	**0,92****	**0,96****	**1,06**	**1,11**
		(2,19)	(3,00)	(2,50)	(2,28)	(1,88)	(0,89)	(0,45)
Survey	log	**0,93***	**0,84****	**0,69****	**0,63***	**0,58****	**0,51****	**0,46****
		(1,66)	(3,04)	(3,54)	(3,02)	(3,36)	(3,71)	(4,08)
T-bill	Δ	**1,00**	**0,98**	**0,98**	**0,98**	**0,98**	**0,97**	**0,96**
		(0,74)	(1,40)	(0,88)	(0,83)	(0,80)	(0,87)	(0,98)
UN	gap	**1,00**	**1,01**	**1,04**	**1,06**	**1,07**	**1,07**	**1,05**
		(-0,01)	(-0,34)	(-0,72)	(-0,94)	(-1,04)	(-0,88)	(-0,56)

Figure 3
Inflation forecasting 24 months ahead
Model two: caput



- Term Spread: Models that include Spread outperforms the benchmark up to 12 months ahead. The lowest RMSFE equals 0.90 for 6 months ahead forecasts.
- Michigan Inflation Survey: Although, caput performs slightly better for h>24 months, survey yields the lowest RMSFEs for all other forecast horizons under examination. For $h = 1$ RMSFE equals 0.93 and its superiority against benchmark becomes more evident for longer forecast horizons. Ang, Bekaert and Wei (2006) confirm the superiority of surveys in inflation forecasting. They argue that in terms of forecasting accuracy, Michigan inflation survey, survey of professional forecasters and Livingston survey outperform the macroeconomic variables, the financial variables and the nonlinear model they exam. Figure 4 graphs the inflation rate and the forecasts of model one and model two at h=18.
- Models that include either IP or SnP or T-Bill outperform the benchmark for long-term forecasts. More specifically, model two with IP performs better than model one for h=18 and h=24. On the other hand, SnP and T-Bill are superior for h=24.

Figure 4
Inflation forecasting 18 months ahead
Model two: survey



- Models that include either Claims or House or NAIRU or M2 and UN-gap are inferior to benchmark in terms of forecasting performance as the respective RMSFEs indicate.

4.2 Forecasting output growth

This subsection presents the results of the forecasting exercise for the variable of Industrial Production. The selected lag order of SIC for the benchmark model (model one) is $p = 3$.

All the parameters of the benchmark model are significant at a 5% significance level. The lag order of the predictors included in model two varies from one to four. The coefficient of the first lag is statistically significant at a 5% confidence level for five (out of 12) predictors. Claims, House, NAIRU, Spread and UN-gap belong to that category. The coefficient of the second lag is statistically significant at a 5% confidence level for all nine predictors that contain two lags or more. The third lag is statistically significant for three (out of 5) predictors. House, NAIRU and SnP are those predictors. Finally, models that includes NAIRU and survey are the only predictors that have a fourth lag. Among them, only the coefficient of NAIRU is statistically significant. Model one explains 18% of the variability of IP. Most models have a similar R^2 . However, models that include NAIRU or UN-gap have much higher than model's one R^2. The percentage rate of the former equals 34% and of the latter 25%. Table 8 includes the regression estimates, the t-statistic values and the R^2-adjusted of each model.

Table 8

In-sample estimation results: IP

Notes: The regression sample ends on 1/12/2002 including 408 observations. An exception is the exrate and the survey whose estimation sample includes 298 observations. Values denoted with one asterisk reject the null of zero coefficient at a 5% level while, values with two asterisks reject the null hypothesis at a 1% level.

IP		β	φ_1	φ_2	φ_3	γ_1	γ_2	γ_3	γ_4	R^2
Model 1		0,09*	0,31**	0,12*	0,13**					0,18
		(2,10)	(3,81)	(2,05)	(2,41)					
Caput	log	0,27**	0,28**	0,11	0,13**	0,06	-0,47**			0,21
		(4,90)	(3,81)	(1,94)	(2,42)	(0,49)	(-3,12)			
CPI	Δlog	0,06	0,42*	0,12*	0,12**	-11,71				0,18
		(1,03)	(2,07)	(2,01)	(2,34)	(-0,63)				
Claims	log	0,32	0,20**	0,07	0,16**	-3,89**	3,88**			0,22
		(0,15)	(2,65)	(1,05)	(3,18)	(-3,18)	(3,32)			
Exrate	Δlog	0,10*	0,16	0,21**	0,14	1,14	-6,33*	-1,99		0,13
		(2,24)	(1,59)	(4,51)	(1,77)	(0,32)	(-2,05)	(-0,65)		
House	Δlog	0,10**	0,26**	0,09	0,16**	1,07*	2,31**	1,10*		0,22
		(2,53)	(3,43)	(1,50)	(3,48)	(2,00)	(3,86)	(2,05)		
M2	Δ^2log	0,09*	0,31**	0,12*	0,12*	-7,12				0,18
		(2,11)	(3,78)	(2,06)	(2,32)	(-0,69)				
NAIRU	Δ	0,22**	0,08	-0,06	0,00	1383,31**	-3895,00**	3632,98**	-1123,13**	0,34
		(5,84)	(1,19)	(-1,08)	(0,10)	(5,39)	(-5,24)	(5,02)	(-4,74)	
SNP	Δlog	0,06	0,27**	0,12*	0,14**	0,70	2,43*	3,26**		0,22
		(1,46)	(3,53)	(2,00)	(2,87)	(0,63)	(2,22)	(2,96)		
Spread	level	0,06	0,26**	0,10	0,12*	0,12**				0,23
		(1,37)	(3,49)	(1,72)	(2,54)	(4,45)				
Survey	log	0,40**	0,15	0,18**	0,12	0,20	-0,57**	0,28	-0,14	0,14
		(2,85)	(1,60)	(3,81)	(1,49)	(1,28)	(-2,72)	(1,46)	(-0,83)	
T-bill	Δ	0,10*	0,29**	0,10	0,12*	0,13				0,19
		(2,21)	(3,44)	(1,77)	(2,41)	(1,58)				
UN	gap	0,09*	0,25**	0,13*	0,19**	0,08	0,63*			0,25
		(2,26)	(3,18)	(2,29)	(3,66)	(0,36)	(2,54)			

Having defined the in-sample strength of our models, we now present the results of the out-of-sample forecast exercise. Table 9 presents the RMSFE and the Clark and West t-statistic.

The first row that contains the MSFEs of model one for all forecast horizons reveals that the error term increases sharply beyond the three-month forecast horizon. This is in line with findings that the univariate specifications are more appropriate for short-term forecasting. The bias of the estimators and the inadequacy of AR models to describe periods of increased volatility make model one incapable of providing long-term forecasts.

The main findings for each predictor considered in our analysis are as follows:

- NAIRU: Models that include NAIRU as a predictor outperform the benchmark for all forecast horizons and produce the lowest MSFEs compared to all other predictors. The superiority of model two is significant for all forecast horizons and it is more distinct for long-term forecasts. For $h = 1$ RMSFE equals 0.79. As the forecast horizon increases RMSFE decreases to the value of 0.16 for $h = 24$. Despite the strong statistical significance of these RMSFEs, model two is inferior compared to model one before August 2007 and after Augoust of 2010. However, during the 2008 crisis model two tracks the course of the IP growth almost perfectly while at the same time model one continues to vary around its mean. Had we exclude it from our exercise the period of crisis, RMSFEs values would have been much closer to unity. Figure 5 graphs the performance of the two competing models and the real course of CPI for h=24 months.
- Initial jobless claims and Unemployment gap: An important finding of our analysis is that during the last decade, unemployment related variables are the best predictors of IP growth. Models that include either Claims or UN-gap generate low RMSFEs for all forecast horizons. Claims at forecast horizon $h = 1$ yield RMSFE equal to 0.93 decreasing to $0{,}67$ at forecast horizon $h = 24$. UN-gap has statistically significant RMSFEs for all forecast horizons. The smaller RMSFE equals 0.48 for h=9.

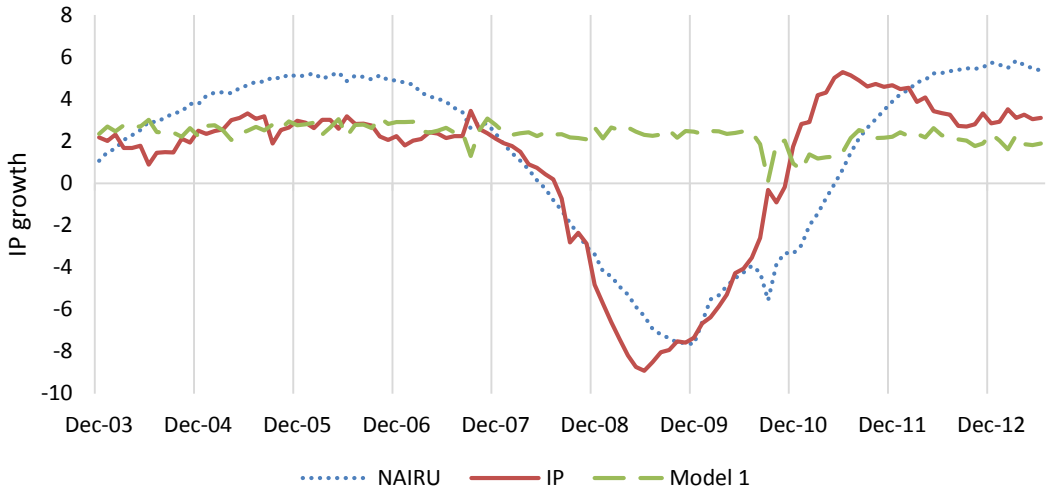
Table 9

RMSFE and Clark and West test results

Notes: The second row depicts the MSFE values of model one for the forecasting horizons under examination. The first column show the candidate predictor. The entries in bold are the RMSFEs and the entries in bracket are the Clark and West t-statistic. One asterisk corresponds to rejection of the null hypothesis of equal MSFE at 10% confidence level. Two asterisks correspond to rejection of the null hypothesis of equal MSFE at 5% confidence level.

Forecast horizon		1	3	6	9	12	18	24
Model 1 $MSFE_{1,h}$		0,01	0,02	0,07	0,16	0,28	0,52	0,74
Caput	log	**1,06**	**1,23**	**1,32**	**1,33**	**1,33**	**1,36**	**1,45**
		(-0,36)	(-1,00)	(-1,21)	(-1,34)	(-1,50)	(-1,80)	(-2,26)
CPI	Δlog	**1,01**	**1,34**	**1,60**	**1,79**	**1,91**	**2,07**	**2,16**
		(-0,49)	(-2,53)	(-2,46)	(-2,46)	(-2,49)	(-2,78)	(-3,11)
Claims	log	**0,93****	**0,79****	**0,70****	**0,68****	**0,67****	**0,67****	**0,67****
		(2,14)	(2,29)	(2,32)	(2,18)	(2,18)	(2,31)	(2,53)
Exrate	Δlog	**0,93****	**0,91****	**0,94**	**0,96**	**0,96**	**0,99**	**1,00**
		(2,18)	(1,82)	(1,24)	(1,13)	(1,10)	(0,75)	(0,36)
House	Δlog	**0,91****	**0,74****	**0,64****	**0,61****	**0,58****	**0,53****	**0,47****
		(2,63)	(2,28)	(2,22)	(2,07)	(2,14)	(2,32)	(2,65)
M2	Δ^2log	**1,02**	**0,99**	**0,99**	**0,99**	**1,00**	**1,00**	**1,00**
		(-0,95)	(1,67)	(1,19)	(1,90)	(0,98)	(0,07)	(-0,43)
NAIRU	Δ	**0,79****	**0,45****	**0,28****	**0,22****	**0,19****	**0,17****	**0,16****
		(3,09)	(2,11)	(1,96)	(1,93)	(2,00)	(2,24)	(2,54)
SNP	Δlog	**0,84****	**0,65****	**0,52****	**0,45****	**0,41****	**0,38****	**0,38****
		(2,53)	(1,92)	(1,89)	(1,91)	(1,98)	(2,13)	(2,41)
Spread	level	**1,00**	**1,04**	**1,07**	**1,05**	**1,03**	**1,00**	**0,97**
		(0,39)	(0,18)	(0,12)	(0,33)	(0,63)	(1,08)	(1,38)
Survey	log	**0,97***	**0,99**	**1,01**	**1,00**	**0,99**	**0,99**	**1,01**
		(1,44)	(1,00)	(0,36)	(0,56)	(0,94)	(1,05)	(0,44)
T-bill	Δ	**1,00**	**0,99**	**0,97***	**0,95****	**0,93****	**0,90****	**0,88****
		(0,36)	(0,86)	(1,65)	(2,19)	(2,16)	(2,19)	(2,37)
UN	gap	**0,91****	**0,68****	**0,52****	**0,48****	**0,51****	**0,59****	**0,69****
		(2,59)	(2,23)	(2,18)	(2,13)	(2,10)	(2,06)	(1,96)

Figure 5.
Output growth forecasting 24 months ahead
Model two: NAIRU



• SnP and Housing start: Both variables have lower than unity RMSFEs, which are statistically significant for all forecast horizons. Once more, the forecasting accuracy is notably higher for long-term forecasts.

• T-Bill: Models that include this variable outperform the benchmark for forecast horizon of 6 months or higher. As the horizon increases, its superiority becomes more obvious.

• Exrate: Models that embed lagged values of Exrate outperform the benchmark for one to three months forecasts. The statistically significant RMSFEs are 0.93 and 0.91 respectively.

Finally, survey outperforms slightly the benchmark for h=1 and spread improve long-term forecasts ($h = 24$). Caput, CPI and M2 fail to outperform the benchmark for all forecast horizons.

Generally, our benchmark tracks the IP growth pretty well especially for short-term forecasting. In case of a persistent shock though, it is unable generate reliable forecasts. This case necessitates the involvement of the appropriate predictor to obtain better forecasts. This is in line with the findings in the literature about the behavior of these models during the 1973-1984 period. Another implication that comes straight from the results of table 9 is that the T-bill rate is significantly better predictor compared to the term spread. This finding is in line with the finding of Stock and Watson (2003) who

claim that the importance of T-bill as a predictor has risen during the recent years outperforming the term spread.

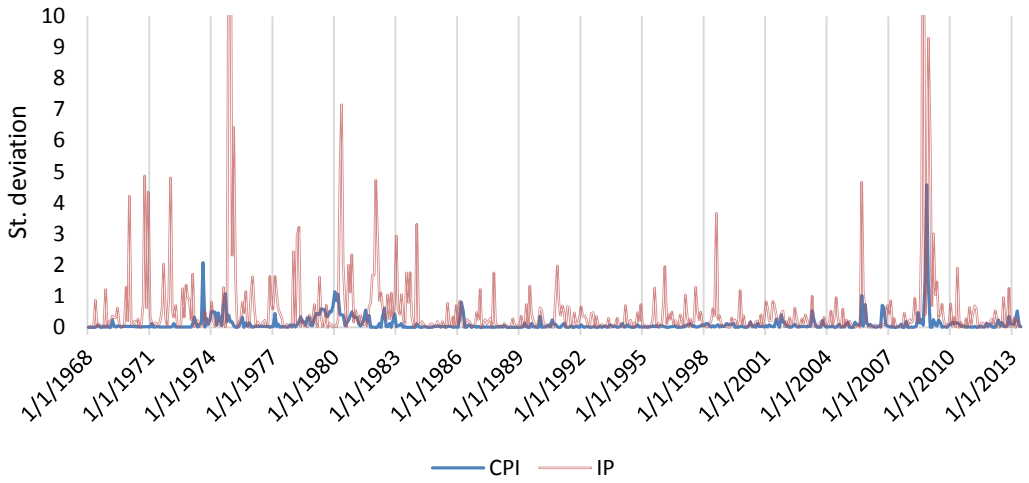
4.3 Comparison and implications regarding the forecasting of inflation rate and output growth.

The current subsection summarizes the main aspects regarding the forecasting of the two variables for the decade 2003-2013. Additionally we underlie the main facts that one can infer from the literature and the present empirical analysis.

An interesting result that comes straight from our forecast exercise is the relation between the in-sample fit and the out-of-sample forecasts. We generally observe that models that outperform the benchmark out-of-sample, have strong in-sample fit. The coefficients of the predictors are significant for all lags and the R^2 of these models is higher than that of the benchmark. The reverse though is not true. For example, despite the fact that the CPI model has strong in-sample fit for the explanation of IP, the out-of-sample forecasts are inferior to the benchmark. The related literature states that strong in-sample fit does not guarantee accurate out-of-sample forecasts and generally, there are models that show poor in-sample fit and generate accurate out-of-sample forecasts.

One of the most significant differences between IP and CPI is the MSFE of model one. The inflation rate is much less volatile than the industrial production index; one can attribute this to the top priority that the Central Bank has set to the price stability since the mid-80's and a forward looking stance toward that policy. Figure 8 depicts the standard deviation of the two processes from 1968 to 2013.

Figure 6.
St. Deviation CPI-IP



Without doubt, the volatility plays a significant role in the effectiveness of model one. As the literature highlights in times of low volatility, a low AR model is unbeatable in terms of forecasting. In the 2003-2013 forecasting sample, from 2003:12 to 2007:08 and from late 2009 to the end of the sample model one is more accurate than model two for both cases. Whereas as with the 78-84 crisis during the 2008 crisis the ability of the specific model breaks down.

For long-term forecasts, we find that model one has important limitations. For CPI, the error term increases abruptly for forecasts of 6 months or higher. The forecasts of IP show sharp a increase in their error term for $h \geq 3$. The cumulative forecast error in combination with the high variance, get the benchmark out of track three months earlier for IP relative to CPI. Nonetheless, the inclusion of a significant predictor, like the NAIRU for the IP case, corrects this deficiency as highlighted by the performance of model two. In general, model one performs well for short-term forecasts (i.e. up to 3 months for IP and up to 6 months for CPI) but its performance deteriorates for long-term forecasts.

The effect of volatility is also responsible for the number of significant predictors at each case. For the CPI case, only three out of 16 variables outperform model one while the respective analogy for IP is 10 out of 16. Marcellino (2008) states that simple linear specifications are generally robust benchmarks. However, in case of structural breaks, like the 2008 crisis, more sophisticated models seems to be more appropriate.

5. Conclusions

This study examines the ability of various predictors to improve forecasts for the inflation rate and the GDP growth rate. We use a parsimonious AR model as a benchmark and test its forecasting accuracy for the variable of interest relative to models that enrich the specification of the AR model with lags of one of the candidate predictors. The results indicate that the univariate model is superior when volatility is low. During the 2008 crisis though, the univariate model breaks down while the bivariate one that includes the appropriate predictor outperforms the former in terms of forecasting accuracy.

The most significant predictors for inflation are capacity utilization, exchange rates, the Michigan inflation survey and Industrial production for long-term forecasts.

Regarding the industrial production forecasts, the most significant predictors are the unemployment related variables, that is the unemployment rate, the unemployment gap, the NAIRU, and the jobless claims. Furthermore, predictive content exists in the exchange rates, the housing start, the stock returns and the T-Bill rate.

Appendix.

Forecasting output and inflation: the role of asset prices. J. Stock M Watson. (2003)

The paper summarizes the literature that investigates the predictive power of certain asset prices for forecasting output and inflation. The variables under examination are the following: the stock returns, the term spread, the interest rates, the dividend yields and the exchange rates. The usefulness of the predictors varies with the forecast horizon, the country, the financial framework and the type of shock that affect the economy.

As far as the output growth is concerned, the combination of predictors seems to yield more stable forecasts than the use of a single predictor. The same is not valid for the case of inflation whose predictability depends a lot on the appropriate selection of the above-mentioned factors.

The authors find that the in-sample Granger causality test is misleading when it comes to out-of-sample forecasting. They suggest that the forecasting performance should be tested using pseudo out-of-sample forecasting tests and break tests, which are capable of detecting instability.

What does the term structure tell us about future inflation? F.S. Mishkin (1990)

Mishkin investigates the predictive power of the term spread on inflation using US treasury maturities from 1 to 5 years. In accordance with earlier results, he finds that for short-term maturities there is little information on expected inflation. He finds though information about the real interest rates, which are not directly observable.

On the other hand, as the number of years under consideration increase the term spread contains information about future inflation while information about the real rates is ambiguous. The latter makes the yield curve a good predictor for expected inflation. For example, an upward sloping curve signals a rising future inflation.

Do macro variables, asset markets, or surveys forecast inflation better? A. Ang, G Bekaert, M Wei. (2006)

Using out of sample forecasting, the authors here attempt to evaluate which method consists a better predictor for expected inflation. The alternatives are time series ARIMA models, real activity measures (Philips curve motivated), linear, nonlinear and arbitrage free term structure models and finally survey based measures (the SPF, Livingston, and Michigan surveys). The results indicate that the survey-based method is more accurate.

Secondly, they find little evidence that the combination of forecasting methods is more efficient than the use of single forecasting methods, an argument that opposes the findings of Stock-Watson 2003.

Overall, regarding the forecasting of CPI, the best time series model is ARMA (1, 1) but the survey measures outperform it. The same applies for the term structure and the Philips curve model, which performs worse than the surveys. The situation is different for PCE forecasting where the time series model outperform surveys.

The explanation for the superiority of the surveys lies on the fact that it aggregates the information of many professionals and consumers about expected inflation.

Additionally the surveys respond faster to the data generation process than the other methods.

Predictability of output growth and inflation: A multi-horizon survey approach. A.J.Patton A.Timmermann. (2011)

This paper investigates the accuracy of forecasts of GDP and inflation by comparing models that incorporate different forecast horizons. The validity of professional forecasters results are assessed taking into account the fact that in their predictions they take into account data with noise-error. Measuring the estimated MFSE and persistence of these different horizon models, the investigators find that the error is greater for forecasts concerning GDP than those of inflation. Moreover, inflation shows more persistence than GDP making it thus predictable for longer horizons. Finally yet importantly, the persistence of the above variables is accurately approximated by a low AR specification.

A linear benchmark for Forecasting GDP growth and Inflation M. Marcellino. (2008)

This paper conducts a comparison between linear and nonlinear models, which can be used as benchmarks for the evaluation of output and inflation forecasting. In general, well-specified linear benchmarks are more appropriate than nonlinear specifications. A plausible explanation is that nonlinearities are limited and insignificant in inflation and output growth rate. An exception is the 1970-1980 decade where a nonlinear benchmark that contain time varying parameters perform better than the respective linear specifications.

Do asset prices help to predict consumer price inflation? C. Goodhart B. Hofmann

This cross-country time series investigation uses a linear framework to extract the predictive power of asset prices by evaluating the statistical significance of housing prices equity prices and yield spread. The baseline model contain an AR (1) approximation of inflation, GDP growth exchange rates and interest rates. The forecasting horizons is 4 and 8 quarters ahead. From the above only the past inflation seems to provide predictive content. When each of the assets enters the equation, the

information set increases especially for the two-year forecast. Additionally the use of housing prices outperforms the other two.

Inflation dynamics F.S. Mishkin (June 2007)

In this article Mishkin claims that inflation in recent years has become less persistent resulting in a more trend following pattern allowing thus the current inflation to be more reliant to long-term inflation expectation, Mishkin uses the term anchoring to refer to the aforementioned relation.

Furthermore, the flattening of the new-Keynesian Philips curve has also contributed to this effect. Oil prices, unemployment rate and capacity utilization do not affect inflation as they used to in the 70's

Overall, by monitoring inflation expectations, and intervening when necessary the monetary authority is able to control future inflation, something that any forecaster should take into account.

The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank A. Estrella F.S. Mishkin (1997)

The aim of the specific paper is to examine whether the yield curve is able to reveal evidence about the monetary policy in the U.S. and Europe. Additionally, it examines its predictive power on inflation and output. The empirical analysis refers to the 1970-1995 period.

In short, as far as the monetary policy part is concerned, the evidence shows that the term structure is not the only instrument and its usefulness relates to the degree of commitment that a central bank can achieve.

The application of term structure for forecasting purposes varies from country to country and it is stronger for those authorities that conduct independent monetary policy and are immune to the effect of the exchange rates. The latter is more important for inflation.

Overall, for the U.S. the predictive power of the yield curve is strong from mid-term to long-term forecast (one to two years), while for the European countries the only robust

forecasting concerns long-term prediction (two years ahead). The above results are valid for inflation forecasts using the Fisher equation and GDP growth.

Inflation Forecasting with Inflation Sentiment Indicators R.Dohrn C.M. Schmidt T. Zimmermann (2008)

This paper presents an alternative way to forecast core CPI using inflation sentiment indicators. The construction of such indicators roots from the assumption that if a price change is a function of many items of the CPI basket then this price change will be more intense in the future. On the other hand, if the price change is a result of the high volatility of few items then it will not last long.

The study investigates the case of US and Germany using quarterly data from 1978 to 2006. The examined indices are a skewness index, a diffusion index, and a momentum index.

The results show that in relation to the Philips curve benchmark, the sentiment indicators outperform the traditional methods in most cases especially for longer horizons (up to two years). Furthermore, the supremacy of those indicators is more distinct when inflation is more volatile.

Why has U.S. inflation become harder to forecast? J.H. Stock M.W. Watson (2007)

According to Stock and Watson's paper, which refers to data from 1970 to 2004, the structure of the inflation generating process has changed. More specifically from 1970 to 1983, the forecasting process (for 1 to 8 quarters ahead) using Phillips curve models was the most accurate method. However, their performance deteriorated from 1984 to 2004 compared to AR (AIC) and Atkeson Ohanian models.

The above changes in the suitability of the forecasting models from one period to another is caused by the inflation's changing variance, the existence of time varying parameters, monetary policy and the fact that the stochastic trend and the transitory uncorrelated component vary their influence on inflation overtime.

This is why multivariate forecasting models have lost their gloom toward more accurate univariate time-varying IMA and UC-SV (unobserved component model with

stochastic volatility), which are more adaptive to the conditions described in the previous paragraph.

Are Phillips curves useful for forecasting inflation? A. Atkeson L.E. Ohanian (2001)

Using a simplistic model as a benchmark, the authors compare the predictive accuracy of three NAIRU Phillips curve-forecasting models used by academics and central banks. The benchmark model assumes that the next year inflation rate will be the same with the current inflation rate. The forecasting horizon is one year and the results are striking. The examined models not only fail systematically to beat the benchmark but also in most cases are inferior with the exception of the model developed by Stock and Watson.

The above result refers to the period from 1970 to 1984 and from 1984 to 1999. The blame for the Phillips curve forecasting inaccuracy is the flattening of the curve during this period and the change of the parameters characterizing the equation.

When can we forecast inflation? J.D.M. Fisher, Chin Te Liu, R. Zhou (2002)

In this paper, the authors compare the NAIRU Phillips curve forecasting models to the Atkeson Ohanian (AO) naive model.

The results show that the AO model hold for the period 1985-2000 where there is a monetary policy shift and it is characterized by low volatility. In this period the forecasts are accurate at short-term and at the long-term as well. In addition, we get some information about the direction after adding a directional variable.

On the other hand, for the 1977-2000 NAIRU Phillips curve models are accurate predicting the direction of inflation especially in the 1977-1984 period, which is characterized by high volatility, and they are more accurate for long-term predictions assuming stable monetary policy regime.

Money Growth and Inflation in the United States. Bachmeier Leelahanon (2007)

This paper presents the superiority of nonlinear models relative to linear models when the purpose of the investigator is to forecast inflation. Using data from 1994 to 2002, the model attempts to forecast inflation using monetary aggregates. Incorporating in a

nonlinear model money growth and velocity yields more accurate forecasts and reduces the Mean Squared Forecast Error up to 40%

The predictive content of the interest rate term spread for future economic growth Michael Dotsey (1998)

This article supports the findings that the term spread and more specifically the difference between the 10-year treasury bonds and the 3-month T-bill rate contain predictive content for output, though its accuracy has diminished over the last years of the analysis.

The research includes in-sample and out of sample forecasting using data from 1955 to 1997.

According to the findings of this study, the forecasting capability of the model is up to 6 quarters and it does better for longer horizons. Moreover, the inclusion of lag GDP, nonlinear structure and federal fund rate do not seem to improve the forecasting accuracy.

Finally yet importantly, the term spread has been useful in predicting the pending recessions during this period with 83% success.

The predictive power of the term structure of interest rates in Europe and the United States. E. Arturo F.S. Mishkin (1997)

In this article, the term spread is considered to be the best out of sample predictor of recessions. Additionally, GDP forecasts are more accurate for short horizons if the stock indices price movement is involved.

The term structure as a predictor of real economic activity Estrella G.A. Hardouvelis (1991)

This paper explores the predictive power of the term structure in the period 1955-1988 using quarterly data of the spread between 10-year treasury bonds and 3-month T-bill rates.

The evidence show that the slope of the yield curve predicts the cumulative changes in GNE for up to 4 years and marginal changes in output for 1.5 years.

The authors conclude that the forecasting usefulness of the spread is not solely a result of the monetary policy. However, its effectiveness in forecasting will be dubious if the monetary authorities decide to use it as a tool for monetary policy.

Can the term spread predict output growth and recessions? A survey of literature D. C. Wheelock and M. E. Wohar (2009)

According to this survey, the forecasting accuracy of term spread on GDP depends on the period and the country under examination. In general, they argue that the use of linear forecasting models can approximate future GDP growth adequately from six to 12 months. Moreover, there is evidence of non-linearity and structural breaks in the relationship between the two variables. Recessions can be predicted with high probability.

Finally, there is a consensus that the monetary regime and the persistence of the inflation play significant role to the usefulness of the term spread and its slope as a predictor.

References

Ang Andrew & Bekaert Geert & Min Wei, (2007). *"Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?*" Journal of Monetary Economics Vol. 54 pp. 1163–1212

Atkeson Andrew, Ohanian E. Lee (2000) "*Are Phillips Curves Useful for Forecasting Inflation?"* Federal Reserve Bank of Minneapolis Quarterly Review Vol. 25, No.1, pp. 2-11

Ball Laurence, Mankiw Gregory N. (2002) *"The NAIRU in Theory and Practice*" Journal of Economic Perspectives Vol. 16, No. 4, pp. 115–136

Clark E. Todd, West D. Kenneth (2006) *"Approximately Normal Tests for Equal Predictive Accuracy in Nested Models*" Journal of Econometrics Vol. 138, pp. 291–311

De Jong N. David and Whiteman H. Charles (1991) *"Reconsidering Trends and Random Walks in Macroeconomic Time Series"* Journal of Monetary Economics Vol. 28, pp. 221-254

Dohrn Roland, Schmidt M. Christoph and Zimmermann Tobias (2008) "*Inflation Forecasting with Inflation Sentiment Indicators"*, Ruhr Economic Papers #80.

Dotsey Michael (1998) "*The Predictive Content of the Interest Rate Term Spread for Future Economic Growth"* Federal Reserve Bank of Richmond Economic Quarterly Vol. 84, No. 3, pp. 31-51

Estrella Arturo, Hardouvelis A. Gikas (1991) "*The Term Structure as a Predictor of Real Economic Activity*" The journal of finance Vol. 46 No. 2, pp. 555-576

Fisher D. M. Jonas, Chin Te Liu and Ruilin Zhou (2002) "*When can we Forecast Inflation?"* Federal Reserve Bank of Chicago Vol. (2002) Q1 p.p. 32-44

Goodhart Charles and Hofmann Boris (2000) *"Do Asset Prices Help to Predict Inflation?"* Manchester School Supplement pp. 122-40

Hodrick J. Robert, Prescott C. Edward (1997) *"Postwar U.S. Business Cycles: An Empirical Investigation"* Journal of Money, Credit and Banking Vol. 29, No. 1, pp. 1-16

Marcellino Massimiliano (2008) "A Linear Benchmark for Forecasting GDP Growth and Inflation?" Journal of Forecasting. Forecast. Vol. 27, pp. 305–340

Marcellino Massimiliano, Stock H. James and Watson W. Marc (2005) "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series" Working Paper n. 285

Mishkin Frederic S. (1990) "The *Information in the Longer Maturity Term Structure About Future Inflation*" The Quarterly Journal of Economics, Vol. 105, No. 3 , pp. 815-828

Mishkin Frederic S. (2007) "*Inflation Dynamics*" NBER Working Paper No. 13147JEL No. E31, E50

Mishkin Frederic S. and Estrella Arturo (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 pp. 1375-1401

Patton J Andrew and Timmermann Allan (2011) "*Predictability of Output Growth and Inflation: A Multi-Horizon Survey Approach*", Journal of Business & Economic Statistics Vol. 29, No. 3, pp. 397-410

Stock, J.H., and M.W. Watson (1999) *"Forecasting Inflation"* Journal of Monetary Economics. Vol. 44, No. 2, pp. 293-335

Stock, J.H., and M.W. Watson (2003), "*Forecasting Output and Inflation: The role of asset prices*", Journal of Economic Literature Vol. 41, pp. 788-829.

Stock, J.H., and M.W. Watson (2007) "*Why Has U.S. Inflation Become Harder to Forecast?*" Journal of Money, Credit and Banking, Supplement to Vol. 39, No. 1

Wheelock C. David and Wohar E. Mark (2009) "*Can the Term Spread Predict Output Growth and Recessions? A Survey of the Literature*" Federal Reserve Bank of St. Louis Review, September/October 2009, 91(5, Part 1), pp. 419-40