SINGLE EQUATION MODELS FOR INFLATION FORECASTING IN RWANDA

Pascal Kayisire

Supervisor: Professor G. Farrell

Co-supervisor: Mr. Herbert Hove

A Research Report submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science.

School of Statistics and Actuarial Science

University of Witwatersrand

May, 2014

DECLARATION

I declare that this Research report is my own, unaided work. It is being submitted for the Degree of Masters of science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Pascal Kayisire

28th day of May 2014

i

ABSTRACT

This study evaluates Phillips curve forecasts of inflation for Rwanda. The study relies on the use of various single equation prototype Phillips curve models, as described by Stock and Watson (2008). Pseudo out-of-sample comparison tests are used to evaluate the forecast performance of these Phillips curve forecasts relative to the AR (autoregression) benchmark forecasts. In this regard, tests of equal forecast accuracy based on mean square forecast error and those based on forecast encompassing as used by several scholars (for example, Clark and McCracken (2001, 2005), Rapach and Weber (2004)) are reported. Furthermore, the results from forecasts using inflation in levels and in differences as the dependent variable are reported, to check the sensitivity to this specification issue. The study finds that the Phillips curve and augmented Phillips curve forecasts outperform the AR benchmark forecasts at one- and two-quarter horizons. The output gap, exchange rate and money supply (M3) are found to be good predictors of inflation in Rwanda in the generalised Phillips curve context. It is therefore strongly recommended that Rwandan economic policymakers take into consideration these variables when forecasting inflation.

ACKNOWLEDGEMENTS

First and foremost, I thank the Almighty God, who guided me and gave me strength during my Masters studies at the University of the Witwatersrand.

My deep and sincere thanks go to my supervisor, Professor Greg Farrell who has been a constant and reliable source of inspiration and encouragement throughout the process of writing this report. I also would like to express profound gratitude to the co-supervisor of this research, Mr. Herbert Hove, for his invaluable support, follow-up and useful suggestions on this work at different stages.

I am especially indebted to my parents, Gabriel Gahorihori and Léonie Nyiragwiza, who brought me into this world and did all they could for me to be who I am today. Their love and support throughout my life means a lot to me.

A special word of gratitude goes to my elder brother Paul Muzaliwa Gahorihori for his love and care to me and his invaluable contribution to my growth in general, and to my studies in particular. I also say thank you to my old friends, Emmanuel Sebuhoro and Innocent Mupenzi, for their continuous encouragement through weekly calls. Equally, I am grateful to all the Rwandan colleagues studying at the University of the Witwatersrand, particularly Amini, Liliane, Emmanuel and Jean Claude for their cooperation and encouragement.

Last but not least, special thanks go to the Government of Rwanda which, through the Rwanda Education Board (REB), funded my studies.

TABLE OF CONTENTS

DECLARATION .... **Error! Bookmark not defined.**

ABSTRACT ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF FIGURES v

LIST OF TABLES v

ABBREVIATIONS AND ACRONYMS vi

1. Introduction 1
   - 1.1. Background of the research 1
   - 1.2. Brief review of the Rwandan economy 2
   - 1.3. Problem statement 2
   - 1.4. Objective of the study 3
   - 1.5. Research layout 3
2. Literature review 5
   - 2.1. Phillips curves 5
   - 2.2. Forecasting methods 6
   - 2.3. Methods for evaluating forecasts and predictive content 7
     - 2.3.1. In-sample methods for measuring predictive content 7
     - 2.2.1. Pseudo out-of-sample methods for measuring predictive content 8
   - 2.2. Overview on applications and performance of Phillips curves 10
3. Methodology 12
   - 3.1. Model framework 12
   - 3.2. Generating the forecasts 14
     - 3.2.1. Direct forecasts 14
     - 3.2.2. Pseudo out-of-sample forecasting 15
   - 3.3. Forecast evaluation 16
     - 3.3.1. Background on forecast evaluation 16
     - 3.3.2. Performance evaluation 16
     - 3.3.3. Bootstrap algorithm 21
4. Results and discussion 23
   - 4.1. Data description and their transformation 23
   - 4.2. Empirical results 26

5. Conclusion and recommendations 38
References 39
Appendix A 43
Appendix B 44
Appendix C 46

LIST OF FIGURES

Figure 1: HP filtering for the Output gap 26

Figure 2: The out-of-sample fit of competing forecasts relative to observed data using inflation in levels 31

Figure 3: The out-of-sample fit of competing forecasts relative to observed data using change in inflation. 33

Figure 4: The out-of-sample fit of competing forecasts relative to observed data using inflation in levels for 8 Quarters. 35

Figure 5: The out-of-sample fit of competing forecasts relative to observed data using change in inflation for 8 Quarters. 37

LIST OF TABLES

Table 1: Summary of variables and their corresponding transformation 23

Table 2: Unit root test 25

Table 3: Pseudo out-of-sample test results, using inflation in levels 30

Table 4: Pseudo out-of-sample test results, using change in inflation 32

Table 5: Pseudo out-of-sample test results, using inflation in levels for 8 Quarters 34

Table 6: Pseudo out-of-sample test results, using change in inflation for 8 Quarters 36

ABBREVIATIONS AND ACRONYMS

ADF : *Augmented Dickey-Fuller*

ACF : *Autocorrelation function*

ADL : *Autoregressive distributed lag*

CPI : *Consumer price index*

DF : *Dickey-Fuller*

GDP : *Growth domestic product*

HAC : *Heteroskedasticity and autocorrelation consistent*

MSFE : *Mean square forecast error*

MINECOFIN : *Ministry of Finance and Economic Planning*

BNR : *National bank of Rwanda*

NAIRU : *Non-accelerating inflation rate of unemployment*

RMSFE : *Root mean square forecast error*

SIC : *Schwartz information criteria*

SER : *Standard error of regression*

UK : *United Kingdom*

HP : *Hodrick Prescott*

MAFE : *Mean absolute forecast error*

NISR : *National institute of statistics of Rwanda*

NGDP : *Nominal growth domestic product*

OLS : *Ordinary least square*

US : *United States of America*

1. Introduction

1.1. Background of the research

Monetary policy in Rwanda is now focused on price stability, while taking into account the implications of policy decisions for economic activity as a whole. The National Bank of Rwanda (BNR) cites price stability as a crucial precondition for sustained economic growth (National Bank of Rwanda, 2012). The view that stable prices foster growth in the Rwandan context is supported by researchers such as Sayinzoga and Simson (2006), and Berg, Charry, Portillo and Vlcek (2013), of the International Monetary Fund (IMF), who have encouraged the central banks of East African community (EAC) countries to consider moving towards forward-looking monetary policy frameworks such as inflation targeting.

Over the past two decades, several central banks have adopted inflation targeting as a framework for guiding monetary policy[1]. An important implication of this framework is that monetary policy becomes more forward looking. Since monetary policy instruments act with a lag via the transmission mechanism, policymakers need to allow for this when making decisions. This implies that greater emphasis is given to the role of forecasts of inflation. Under inflation targeting, the inflation forecast is often viewed as an intermediate target for central banks (Svensson, 1997). In this context, it is clear that good forecasts aid good decisions.

In Rwanda, the NBR operates a flexible monetary framework with a broad money aggregate as an intermediate target and inflation as the ultimate goal.[2] In order to set policy consistent with future price stability targets, it is therefore important to have good inflation forecasts. This study aims to contribute to the literature on forecasting inflation in Rwanda, and hence to help support a more effective monetary policy regime in the country. The choice between approaches to modelling inflation depends in part on the characteristics of the country being studied. This study uses semi-structural augmented Phillips curve models to forecast inflation since the models can be relatively easily applied to the type of time series data that are available for Rwanda. In particular, the available variables for inflation forecasting in Rwanda are the consumer price index (CPI), the output gap (generated from real GDP), the money supply and the exchange rate.

[1] Inflation targeting was introduced in New Zealand in 1990, followed by developed countries that include Canada in 1991, the United Kingdom in 1992 and Sweden in 1993. Developing countries such as Chile and Israel both in 1991, Mexico in 1994, Brazil in 1999, Thailand and South Africa both in 2000 followed (South African Reserve Bank, 2001).

[2] Monetary policy framework, http://www.bnr.rw/index.php?id=180

1.2. Brief review of the Rwandan economy

Rwanda is a landlocked East African country that is bordered by Uganda, Burundi, Tanzania and the Democratic Republic of Congo. Rwanda was embroiled in civil war from 1990 which led to the genocide in 1994, and this affected its economy. After the genocide Rwanda attempted to rehabilitate its economy and it has since been ranked among the fastest growing African countries and indeed the world. Since 1996, indicators show that Rwanda has experienced a stable economic recovery. The significant improvement was reported by the Ministry of Finance and Economic Planning (MINECOFIN) in 2011/12 where the Rwandan economy was characterized by strong growth in all sectors.

The Rwandan economy is driven largely by service sectors supported by industry and agriculture (MINECOFIN, 2009, September). Rwanda has experienced fairly stable economic growth over recent years with its inflation rate kept below 10%. Unlike many countries in the region, Rwanda does not face soaring inflation and currency depreciation. The moderate level of inflation was a result of decreasing imported inflation and enhancing the role of monetary policy (MINECOFIN, 2009, September). However, Rwanda is still vulnerable to external shocks due to its small economic size and the fact that it is largely a price taker in the global market.

The inflation rate in Rwanda is reported by the National Institute of Statistics using CPI (NISR, 2013, February). Historically, from 1997 until 2013, Rwanda's inflation rate averaged 6.6 per cent reaching an all-time high of 28.1 per cent in February of 1998 and a record low of -15.8 per cent in February of 1999 (Fedec, 2012).

1.3. Problem statement

Forecasting inflation is an important task all around the world, and as such has received much attention in the literature (examples include Matheson (2008) in New Zealand; Gruen, Pagan and Thompson (1999) in Australia; Stock and Watson (1999, 2008) in the US; Moser, Rumler and Scharler (2007) in Austria). However, despite the extensive international literature on forecasting inflation, little work has been done for Rwanda. This research aims to help fill this gap.

1.4. Objective of the study

The main objective of the study is to forecast inflation in Rwanda using Phillips curves as described by Stock and Watson (2008), and to evaluate the forecast performance of these models using the pseudo out-of-sample comparison tests developed by Diebold and Mariano (1995), West (1996) and McCracken (2007) for equal forecast accuracy and those developed by Harvey, Leybourne and Newbold (1998) and Clark and McCracken (2001) for forecast encompassing. The main question under consideration in this study is:

Do Phillips curves forecasts or augmented Phillips curve forecasts of inflation outperform benchmark AR forecasts for Rwanda*?*

The specific steps to address the main question are to:

1. Generate recursive out-of-sample forecasts using inflation forecasting models.

2. Compare inflation forecasts at different horizons in order to assess the performance of forecasts from the Phillips curve models used in the study. The general idea often used in the literature is to take the ratio of the h-step ahead mean squared forecast error (MSFE) for the competing Phillips curve forecasts to that of the benchmark forecasts (taken here to be from an autoregressive forecasting model). If the ratio is less than one, then the competing forecasts perform better than the benchmark, otherwise they are worse. However, Stock and Watson (2003) point out that decisions made based on these ratios may be due to sampling variability. Therefore, statistical tests are required to aid evaluation.

3. Undertake formal tests of the predictive accuracy of the models. Four statistical tests are considered, two of them are for testing the null of equal forecast accuracy as proposed by Diebold and Mariano (1995), West (1996) and McCracken (2007); and two test for forecast encompassing (proposed by Harvey *et al.* (1998) and Clark and McCracken (2001)).

1.5. Research layout

The report has five chapters. Chapter 1 contains the background to the research, a brief review of the Rwandan economy, the problem statement and the objective of the study.

Chapter 2 reviews the relevant literature related to this study. Chapter 3 sets out the models and the methodology used in the study. Chapter 4 discusses the empirical results and chapter 5 provides the conclusions and recommendations.

2. Literature review

2.1. Phillips curves

A Phillips curve is an equation that relates the unemployment rate, or some other measure of aggregate economic activity (called an *activity index* by Stock and Watson (1999)), to a measure of the inflation rate (Atkeson and Ohanian, 2001).[3] The Phillips curve has been useful in macroeconomic modelling for the past 50 years. Its usefulness in forecasting inflation attracted researchers' attention, particularly in the US. Different empirical studies have found the Phillips curve to be amazingly stable, reliable and accurate, compared to the alternatives theories (Blinder, 1997). Blinder (1997, p.241) called the Phillips curve the "clean little secret" of macroeconometrics and argued that "it merits a prominent place in the core model".

Stock and Watson (2003, p. 801) argue that "generalized Phillips curves and output gaps appear to be one of the few ways to forecast inflation that have been reliable". However, they add that this might depend on the time and country. Choosing the predictor variables for future inflation is a challenging issue for inflation forecasting. In the context of Rwanda, for example, it is impossible to establish a Phillips curve relation based on the unemployment rate because there are no reliable data on unemployment. However, Stock and Watson (1999) show that Phillips curves based on other indicators can perform as well or better than inflation forecasts based on unemployment gaps.

The output gap (the difference between actual and potential output) is often used as a measure of economic activity in Phillips curves. To estimate the output gap, let x_t denote the actual output or real GDP (real GDP is nominal GDP deflated by the CPI here) during quarter t, and μ_t its trend. Then the output gap denoted by y_t is obtained from the cycle component resulting from the decomposition of output into a trend and cycle component; that is; $x_t = \mu_t + y_t$ (Orphanides and van Norden, 2005). According to Clausen and Clausen (2010, p.3) the intuition behind employing the output gap in the Phillips curve to forecast inflation is

[3] Although the Phillips curve is named after A.W. Phillips' 1958 work, the initiator of the idea is Fisher (1926) who first established a statistical relationship between unemployment and price changes. Existence of statistical evidence in favour of the hypothesis that unemployment level and its rate of change may be predictive of the rate of change of money wage rate was investigated by Phillips (1958) for the United Kingdom. The resulting negative relationship between unemployment and the rate of change of money wage rate or the rate of inflation became the basic theory to explore for further researchers and the equations relating the unemployment rate to the inflation rate were the first to be called Phillips curves.

that, "when the output gap is positive (actual output is above potential output), inflationary pressures should increase. When the output gap is negative (output is below its trend), inflationary pressures should recede". Orphanides and van Norden (2005) also examined the reliability and practical usefulness of inflation forecasts based on output gaps.

2.2. Forecasting methods

There are at least two methods for generating multistep (h > 1) forecasts.[4] One is to iterate a single period forecast and the other is a direct method in which multiple periods are forecast simultaneously. The iterative method is used to forecast only one period at a time and then the value of the next period is forecasted recursively using the predicted value as an input toward the h-step forecast horizon. That is, forecasting necessitates only one estimation procedure but the estimates are modified for each horizon (Chevillon, 2007). By contrast, the direct method uses only past data to forecast h-steps ahead.

There are differing views on the relative merits of these approaches. For, instance Bhansali (1996) supports the direct method because it provides a lower bound on its h-step MSFE than the iterated method. Kang (2003) shows that the direct method may or may not improve forecast accuracy compared to the iterated forecast of an AR model procedure and that enhancing the forecast performance of the direct procedure relative to the iterated procedure seems to attain optimal performance in terms of order selection criteria, periods and horizons of forecasts as well as on the time series to be forecasted (Kang, 2003). Chevillon (2007) shows that the direct method can be asymptotically more efficient than the iterated one, even if the forecasting model is well specified. Hamzaçebi, Akay and Kutay (2009) argue that it is not possible to conclude a priori that the direct method gives better results for all time series forecasting problems, although their findings show the superiority of the direct method over the iterated method.

By contrast, Marcellino, Stock and Watson (2006) found that iterated forecasts typically outperform direct forecasts. This is particularly so if the models can select long-lag specifications. This is also supported by Proietti (2011) as long as the autoregressive representation is not strongly misspecified. However, Marcellino *et al*. (2006, p.505) also point out that "the sample MSFE might be less for a direct than an iterated forecast either

[4] Ben Taieb and Hyndman (2014, forthcoming) propose a new method for multi-step forecasting that combines the features of the recursive and direct forecasting strategies. They propose 'boosting' recursive linear forecasts with a direct strategy using a boosting autoregression procedure at each horizon.

because the direct forecast is more efficient in population or because of sampling variability". In this study the direct method will be used as in Stock and Watson (2008) to generate inflation forecast for Rwanda.

2.3. Methods for evaluating forecasts and predictive content

Evaluating predictive content can be based on the in-sample estimate of the model or based on the out-of-sample estimates from recursive or rolling regressions (Inoue and Kilian (2005); Stock and Watson (2003)). This study is primarily interested in forecasting inflation using the pseudo out-of sample forecasting method, although the in-sample Wald test obtained using the full sample is also reported to see whether there is a similarity in forecasting inflation from both techniques. It should be noted at the outset that good in-sample fit of a forecasting model does necessary imply good out-of-sample forecast performance (Stock and Watson, 2008).

2.3.1. In-sample methods for measuring predictive content

For in-sample methods the full sample is used to estimate the parameters of the model of interest. The null hypothesis that the predictor(s) has no predictive content is tested by the t-statistic (for simple regression) or F-statistic (for more than one regression coefficients) on their estimated parameter(s). The idea of the in-sample method can be easily illustrated by considering the example of simple linear regression used in Stock and Watson (2003). Suppose we are interested in assessing the predictive content (or the usefulness) of the current value of the output gap, x_t relating to the future value of inflation denoted by y_{t+1}, that is, $y_{t+1} = \alpha_0 + \beta_1 x_t + \varepsilon_{t+1}$, where α_0 and β_1 are unknown parameters, and ε_{t+1} is an error term. In this regression model, the interest is in testing whether the coefficient β_1 on the output gap is significantly different from zero ($\beta_1 \neq 0$) using the t-statistic under the null hypothesis that the output gap has no predictive content. The economic significance of the output gap as predictor in regression can be assessed using the regression R^2 and the standard error of the regression (SER) (Stock and Watson, 2003).

However, as Stock and Watson (2003) argue, the variance of the error term ε_{t+1} may depend on regressor x_t and/or be autocorrelated. To account for this problem the *t*-statistic should be computed using heteroskedasticity and autocorrelation consistent (HAC) standard errors. The most common HAC used in literature to handle the problem of serial correlation in the disturbance term was proposed by Newey and West (1987). Indeed lagged values of predictor x_t, also might contain useful predictive information in time series regression. The addition of x_t and its lagged values in the regression requires a joint test statistic to assess predictive content, and the appropriate test is the F-statistic for Granger causality (Stock and Watson, 2011). Moreover, in time series, the dependent variable is likely to be serially correlated, that is, the past values of the dependent variable are useful predictors. A time series regression model that includes lags of the dependent variable and lagged values of predictor x_t refers to an autoregressive distributed lag (ARDL) model.

So far, the discussion of predictive content has focussed on the full sample. A problem which is likely to appear with the in-sample approach is that the estimate coefficients might not be stable over time. If the parameters of the model change over time, the estimated coefficients obtained using the full-sample are likely to be misleading for out-of-sample forecasting (Stock and Watson, 2003). To emphasise this weakness of the in-sample approach, Stock and Watson (2003) highlight that, although the most common econometric methods rely on in-sample significance tests such as Granger causality tests to identify a potentially useful predictor, there is little assurance that the identified predictive relation is stable. Similarly Tsay (2008) also criticises the in-sample method by stressing that there is no guarantee that the best model selected using in-sample fitting will necessarily provide more accurate forecasts when out-of-sample forecasting. Therefore an alternative approach to the evaluation of predictive content that seeks to simulate more closely actual real-time forecasting has been proposed: pseudo-out-of sample forecast evaluation (Stock and Watson, 2003, 2008).

2.2.1. Pseudo out-of-sample methods for measuring predictive content

Pseudo out-of-sample forecasting is a method for simulating the real-time performance of a forecasting model (Stock and Watson, 2011). Stock and Watson (2011) explain that the reason for the prefix “pseudo” is that it is not true out-of-sample forecasting. True out-of-sample forecasting occurs in real time, using real-time data that may be revised. In the process of pseudo out-of-sample forecasting, one simulates real-time forecasting using a model, but with subsample future data against which to assess the pseudo forecasts (more detail on this method in the methodology section). Stock and Watson (2011) point out some useful characteristics of pseudo out-of-sample forecasting: (i) it gives a sense of how well the model has been forecasting at the end of the sample; (ii) it allows the estimation of the root mean square forecasting error (RMSE) that is used to quantify forecast uncertainty and to construct forecast intervals; and (iii) it allows the researcher to compare two or more candidate forecasts. The last is the focus of this report.

Stock and Watson (1999, 2003, 2007, and 2008) used the pseudo out-of-sample forecasting methodology to evaluate forecasting performance. They point out the benefit of this methodology in providing a degree of protection against over fitting and detecting model instability. Stock and Watson (2008, p. 4) argue that the benefit of using out-of-sample forecasts is that evaluation "captures model specification uncertainty, model instability, and estimation uncertainty, in addition to the usual uncertainty of future events".

The pseudo out-of-sample method is used to determine the appropriateness of a particular variable in forecasting. It is widely used to compare the relative predictive errors from either nested or non-nested models. In this method the idea is to compute the MSFE from each set of forecasts, and their relative MSFE. If the MSFE of the forecasts of model 1 is less than those of model 2, model 1 is deemed to have forecast better than model 2.

However, Stock and Watson (2003) show that this ratio could be less than one due to sampling variability. Thus statistical tests are required for testing significance. The most common and widely used test in the literature was proposed by Diebold and Mariano (1995) and West (1996), and tests the null hypothesis of equal predictive accuracy. Diebold and Mariano (1995) and West (1996) tests treat the case that the models have estimated parameters and are not nested (that is, the benchmark model is not a special case of the forecasting model) and the test is asymptotically standard normal. However, McCracken (2007) and Clark and McCracken (2001) show that the limiting distribution of the West (1996) test is non-standard, when comparing forecasts from nested models. McCracken (2007) therefore proposes a variant of the West (1996) test based on the F-statistic test for out-of-sample forecasts. Clark and McCracken (2001, 2005) and McCracken (2007) have shown that this F-type test is more powerful than the t-test of Diebold and Mariano (1995) and West (1996) for nested models.

Further tests for comparing out-of-sample forecasts are the encompassing test proposed by Harvey *et al.* (1998) and its variant introduced by Clark and McCracken (2001). The alternative encompassing test by Clark and McCracken (2001) was shown to be the most powerful out-of-sample forecast test for nested models, particularly in small samples.

Since the limiting distributions of these tests are generally not standard for nested models (Clark and McCracken, 2001, 2005; McCracken, 2007), bootstrap techniques are used for inference. Kilian (1999) proposed a parametric bootstrapping method which is suited for small-sample analysis, and has been used in a number of studies (including Rapach and Weber (2004), Clark and McCracken (2005, 2006)).

2.2. Overview on applications and performance of Phillips curves

The early empirical studies debating the usefulness of Phillips curve inflation forecasts were triggered by an interesting question regarding whether the statistical relationship between unemployment and inflation were expected to remain stable over time (Atkeson and Ohanian, 2001). The subsequent literature provides different empirical results about the performance of Phillips curves in forecasting inflation.

Stock and Watson (1999) used Phillips curves to forecast US inflation at the 12-month horizon using 168 economic indicators to assess the out-of-sample forecast of consumer price index for all items, denoted CPI-all and Personal Consumption Expenditure deflator for all items, denoted PCE-all. In their study, they used monthly data from 1959:1-1997:9 to investigate the inflation forecasts. The simulated out-of-sample forecasts using Phillips curves based on the unemployment rate was found to be better than univariate forecasting models (for both autoregressive and random walk models used as benchmarks). They also assessed the performance of Phillips curves based on unemployment as a benchmark relative to the Phillips curves forecasts based on measures of economic activity included in an *activity index*. The results showed that, generally, Phillips curve forecasts made using unemployment outperform forecasts based on macroeconomic variables, but that relying solely on them and excluding other forecasts is not advisable.

Atkeson and Ohanian (2001) surveyed the comparative accuracy of Phillips curve forecasts over a period of 15 years (1984 through 1999) in the US using quarterly data from 1959:1-1999:1. They compare the performance of three sets of competing inflation forecasts from non-accelerating inflation rate of unemployment (NAIRU) models (two of which are from Stock and Watson, 1999) to the forecast from a naïve benchmark model, that assumes the inflation over the coming year is expected to be the same as inflation over the past year; that is, $E_t(\pi_{t+4} - \pi_t) = 0$. The results differ from those of the Phillips curve models of Stock and Watson (1999) since none make more accurate inflation forecasts than those from a naïve (benchmark) model. However, the subsequent literature has shown that Atkeson and Ohanian's (2001) results were largely dependent on both the sample period and forecast horizon (Stock and Watson, 2008).

Orphanides and van Norden (2005) assessed the usefulness of alternative univariate and multivariate methods in estimating the output gap for inflation prediction. The forecast performance of the output gap was examined using three benchmark models, namely AR (autoregressive), real growth- and nominal growth-based forecasts. Their in-sample analysis results based on ex-post estimates of the output gap show that some estimates appear to be useful for predicting inflation, while the forecasts generated from out-of-sample analysis based on real-time output gap measures shows that the usefulness of output gap measures in predicting inflation may be rather deceptive. A lot of the output gap estimates made in real time fail to improve forecasts relative to the forecasts from a simple AR. Furthermore, Orphanides and van Norden (2005) stress that forecasts based on ex-post estimates of the output gap tend to exaggerate the ability of output gap to predict inflation, while the real-time forecasts based on output gap are generally not as accurate compared to those of benchmarks.

Stock and Watson (2008) undertook an empirical study aimed at unifying and assessing findings in the literature. Their study uses quarterly US data covering the period 1953:Q1 – 2008:Q1 where Q1 denotes the first quarter. In the study, 157 different models and 35 combination forecasts totalling 192 forecasting techniques were examined for their pseudo out-of-sample performance. The total of 192 forecasting techniques used six prototype models which were applied to forecast CPI-all; CPI, less food and energy, denoted CPI-core; PCE-all; PCE, less food and energy (PCE-core); and the GDP deflator inflation measures (Stock and Watson, 2008). Their main finding is the seemingly incidental performance of Phillips curve forecasts. For instance, they highlight that Phillips curve forecasts outperform the univariate forecasts in the late 1990s, while the later would have served a forecaster better in the mid-1990s.

Clausen and Clausen (2010) simulate out-of-sample inflation forecasting for Germany, the United Kingdom (UK), and the US using output gaps estimated with unrevised real-time GDP data. They found that the simple Phillips curve forecasts made using ex-post output gaps outperform the AR(1) benchmark forecasts for all three countries, while simple Phillips curve forecasts based on real-time output gaps often show the opposite result.

3. Methodology

3.1. Model framework

There are numerous theories and models to rely on when modelling inflation. The choice of the modelling approach depends to some extent on the characteristics of the country being studied. For instance de Brouwer and Ericsson (1998); Bowdler and Jansen (2004) modelled inflation using markup models; Stock and Watson (1999) used Philips curve models, and several others theoretical models, including those by Ang, Bekaert and Wei (2007). This work will rely on the approach of Stock and Watson (2008) that used various single equation prototype Phillips curve models to forecast US inflation. This is because such an approach has the advantages of being consistent with the data available for Rwanda and allowing certain key aspects of the Rwandan economy such as output, money supply and exchange rate, to be accounted for. The three single-equation inflation forecasting models used in the analysis are defined as follows.

(1) Forecasts based on past inflation (AR models) which will serve as benchmarks for evaluation. The AR models are given by

$$\pi^h_{t+h} - \pi_t = \alpha^h + \gamma^h(L)\Delta\pi_t + v^h_{t+h} \qquad (3.1)$$

(2) Philips curve models which include activity measures such as output gaps (denoted O_t here) represented as

$$\pi^h_{t+h} - \pi_t = \alpha^h + \gamma^h(L)\Delta\pi_t + \beta^h(L)O_t + v^h_{t+h}, \qquad (3.2)$$

(3) Augmented Philips curve models, which are based on an activity measure (such as output gaps) as well as other predictors such as money supply (M3) and the exchange rate

$$\pi^h_{t+h} - \pi_t = \alpha^h + \gamma^h(L)\Delta\pi_t + \beta^h(L)O_t + \lambda^h(L)x_t + v^h_{t+h}, \qquad (3.3)$$

where α^h is a constant term, $\gamma^h(L)$, $\beta^h(L)$ and $\lambda^h(L)$ are polynomials in the lag operator L that specify the number of lagged values included in the regression. These will be chosen separately by information criteria (Akaike information criteria (AIC) and Schwartz Bayes information criteria (SIC). Respective formulae are in Appendix C) with a maximum lag of 6; x_t, includes other predictors (M3, exchange rate); v^h_{t+h} is the *h*-step ahead error term; $\pi^h_t = h^{-1}\sum_{i=0}^{h-1}\pi_{t-i}$, where π_t is the quarterly rate of inflation at an annual rate. In particularly,

$\pi_t = 400\ \ln(p_t / p_{t-1})$ when using the log approximation where p_t is the price index in quarter t (taken here to be the CPI). This implies that four-quarter inflation at date t is $\pi_t^4 = 100\ln(p_t / p_{t-4})$. In Stock and Watson (1999) these models rely on a specification that imposes the restriction that the inflation has a unit root i.e is $I(1)$. However, for Rwanda there is little evidence that the inflation rate has a unit root. Thus, to ensure robustness, results for both inflation in levels i.e. $I(0)$ and in differences i.e inflation is $I(1)$ are reported. It is also assumed that x_t has already been transformed so that it is $I(0)$. To test the stationarity of the variables, the augmented Dick-Fuller (ADF) test (discussed in appendix B) was used.

In general, the importance of the output gap and exchange rate as determinants of inflation is well established. For instance, the output gap indicates excess demand which stimulates inflation while exchange rate depreciation increases the price of imports which results in domestic inflation (the exchange rate pass-through mechanism). However, a challenge in constructing the output gap is to find the potential output which is generally an unobserved variable. Therefore, potential output must be estimated. The method used here is the Hodrick-Prescott filter (Hodrick and Prescott, 1997), which is widely used for detrending time series in macroeconomics.

However HP filter is broadly applied to economic series in various studies of business cycles. It however has some limitations. Harvey and Jaeger (1993) for example criticise the HP filter that it may produce spurious cycles. They point out that the cyclical components generated from HP filter are distorted and this may lead the investigator to draw a wrong conclusion regarding the relationship between short-run movements in macroeconomics series. Cogley and Nason (1995) also criticise HP filter by arguing that when the HP filter is applied to integrated process, it produce business cycle periodicity and co-movement even if they are not in the original data. Further criticisms by Mise, Kim and Newbold (2005) showed that HP filter is suboptimal at the endpoints of the series. (More details on the HP method are given in Appendix A).

The idea of incorporating the money growth variable in the standard Philips curve model is supported by an argument put forward by Chhibber (1991) that most African countries have high budget deficits. This is mainly due to high nominal money growth relative to output growth, and hence inflation. Thus, it makes sense to include money growth as one of determinants of inflation in the Africa context. In Rwanda, the monetary aggregate is used in the implementation of policy through the "open market operation" instrument. This instrument serves to indicate whether the central bank needs to mop up or to inject liquidity in the banking system and keep the reserve money on the desired path. This implies that money is the nominal anchor of the system in Rwanda.

3.2. Generating the forecasts

3.2.1. Direct forecasts

Macroeconomic time series are often found to be non-stationary. That is, there are one or more unit roots meaning the series needs to be transformed appropriately to be made stationary. Construction of the dependent variable in multiperiod ahead forecasting models depends on the order of integration. Denote by X_t the level or logarithm of the series of interest and y_t its stationary series after differencing an appropriate number of times. If X_t is integrated of order d (i.e. $I(d)$) then $y_t = \Delta^d X_t$ where $d = 0,1\ or\ 2$. The target is to compute the forecasts of X_{t+h} using information at time t. For a direct forecasting regression model, the estimates of the parameters are obtained by OLS regression and the dependent variable y_{t+h}^h is represented by

$$y_{t+h}^h = \begin{cases} X_{t+h} & if\ \ X_t\ is\ I(0) \\ X_{t+h} - X_t & if\ \ X_t\ is\ I(1) \\ \sum_{i=1}^{h}\sum_{j=1}^{i}\Delta^2 X_{t+j} = X_{t+h} - X_t - h\Delta X_t & if\ \ X_t\ is\ I(2) \end{cases}$$

as in Marcellino *et al.* (2006).

If the direct forecasting regression model is, $y_{t+h}^h = \alpha + \sum_{i=1}^{p}\beta_i y_{t+1-i} + \varepsilon_{t+h}$, for the AR model, then the direct forecasts of y_{t+h}^h are $\hat{y}_{t+h}^h = \hat{\alpha} + \sum_{i=1}^{p}\hat{\beta}_i \hat{y}_{t+1-i}$. The direct estimator of the coefficients is obtained by the recursive estimation of the model using ordinary least squares (OLS) (Marcellino *et al.*, 2006).

The forecasts of X_{t+h} are computed from the $\hat{y}_{t+h}^h$ as

$$\hat{X}_{t+h|t} = \begin{cases} \hat{y}^h_{t+h|t} & for\ I(0) \\ \hat{y}^h_{t+h} + X_t & for\ I(1) \\ \hat{y}^h_{t+h} + X_t + h\Delta X_t & for\ I(2) \end{cases}$$

3.2.2. Pseudo out-of-sample forecasting

The process of pseudo out-of-sample forecasting consists of dividing the data into two subsamples. The first is used for estimating the forecasting relationships and the second for evaluating forecast performance. The out-of-sample forecasting procedure used here works as follows:

Let $x_1, x_2, \cdots, x_T$ be the set of data points and let the two subsamples be $\{x_1, \cdots, x_R\}$ for the estimation subsample (or in-sample) and $\{x_{R+1}, \cdots, x_T\}$ the forecasting subsample (out-of-sample), where R, is the initial forecast origin and $P = T - R$ is the number of observations in out-of-sample forecasting. In the case where $P - h + 1$ (i.e. $T - R - h + 1$) is the number of h-horizon forecasts, then the observations in the out-of-sample portion run from $R + h$ through $T = P + R$. An important decision in out-of-sample forecasting is to decide on the in-sample estimation period and the out-of-sample forecasting period. West (2001), Clark and McCracken (2001; 2005) among others discuss this. The general and advisable rule is to consider a small subsample of out-of-sample forecasts relative to the subsample used to estimate the parameters in model. The literature suggests that a reasonable forecast subsample proportion should be 0.10 – 0.2 of the full sample (e.g. Stock and Watson (2011) suggest a forecast subsample of 0.1 or 0.15 of the whole sample; Clark and McCracken (2005) used 0.2; Stock and Watson (2003) used 0.6).

The pseudo out-of-sample forecast comparisons work as follows: For two competing models, say M_1 and M_2, fit each model using the estimation subsample. Then compute the h-step ahead forecast at the forecast origin R for each model (the forecast horizon is the number of steps ahead that one is most interested in forecasting the target variable). The out-of-sample forecast error is given by $\hat{e}_{i,R+h} = x_{i,R+h} - \hat{x}_{i,R+h}$ where $i = 1,2$ and $\hat{x}_{i,R+h}$ denotes the forecasted value from model i. The next step is to advance the forecast origin by 1 (i.e. $R = R + 1$) and re-estimate each model by replicating the previous procedure (obviously the estimated parameters are different for each iteration). The iteration ends when the origin is $R = T - h$. Model estimation can either be rolling where a moving data window of a fixed size is used or recursive where the starting observation is always the same but the data window is increasing (Stock and Watson, 2008). In this study however, recursive model estimation is used since a fairly small sample size is available.

The root mean squared forecast error (RMSE) from the pseudo out-of-sample *h*-step ahead forecasts made over the period R to T using model i is

$$RMSE_i(h) = \sqrt{\frac{1}{T-R-h+1}\sum_{t=R}^{T-h}\hat{e}^2_{i,t+h}} \ ,$$

where $i = 1,2$ and $P = T - R$ is the number of observations in forecasting subsample.

3.3. Forecast evaluation

3.3.1. Background on forecast evaluation

Once the forecasts have been generated the question is how accurate forecasting is. In evaluating the accuracy of forecasts (where there are competing plausible models) the assessor wishes to discriminate and to evaluate the expected loss associated with each forecast model (Diebold and Mariano, 1995). As noted earlier, however, the best fitting model does not always produce the best forecasts. Diebold (2012) makes a similar point when arguing that comparing forecasts using the Diebold-Mariano tests is not the same as comparing model fit.

3.3.2. Performance evaluation

As stated in Stock and Watson (2003) a common way of evaluating pseudo out-of-sample forecast performance is to compute the MSFE of h-step ahead forecasts of a competing model, relative to the h-step ahead forecasts of a benchmark model. The ratio of MSFEs is given by $\frac{(T-R-h+1)^{-1}\sum_{t=R}^{T-h}\hat{e}^2_{j,t+h}}{(T-R-h+1)^{-1}\sum_{t=R}^{T-h}\hat{e}^2_{1,t+h}} = \frac{MSFE_j}{MSFE_1}$ where $MSFE_1$ and $MSFE_j$ are the mean square errors of the forecasts of the AR benchmark and the j^{th} competing model respectively. Alternatively, forecasts can be compared using the square root of the ratio of MSFEs. This commonly used metric for comparing forecasts is known as Theil's U statistic.

If the ratio is less than one, we say that the competing forecasts performed better than the benchmark forecasts, otherwise the benchmark is better. However, Stock and Watson (2003) pointed out that the results from this relative MSFE may be influenced by sampling variability. Hence statistical testing is required to determine whether the estimated relative MSFE is statistically different from one. That is,

H_0 : The relative mean squared error equals one

H_1 : The relative mean squared error is less than one

There are various statistical tests which are used to test this null hypothesis. In the present study, forecast evaluation is based on the simple relative MSFE criterion and four formal tests used by for example Clark and McCracken (2001; 2005) and Rapach and Weber (2004). Of these four tests, two are for equal forecast accuracy based on the relative MSFE criterion when assessing the forecasting power of the competing forecast model being considered and were designed by Diebold and Mariano (1995), West (1996), and McCracken (2007). The other two due to Harvey *et al*. (1998) and Clark and McCracken (2001) respectively use the idea of forecast encompassing to determine whether the competing forecast model adds value to the optimal composite forecast obtained from the AR benchmark and competing models.

When testing if the competing model performs better than the AR benchmark, that is the relative MSFE is significantly less than one, the *t*-statistic of equal MSFE developed by Diebold and Mariano (1995) and West (1996), and the F-statistic developed by McCracken (2007) are used.

Let $(\hat{e}_{1,t+h}, \hat{e}_{2,t+h}), t = R, \cdots, T-h$ be the forecast errors of the two competing models and let $g(\hat{e}_{i,t+h})$ be the associated loss function of forecast errors. Then the null hypothesis of equal accuracy for the two forecast is $E[g(\hat{e}_{1,t+h}) - g(\hat{e}_{2,t+h})] = 0$ or $E[\hat{d}_{t+h}] = 0$, where $\hat{d}_{t+h} = g(\hat{e}_{1,t+h}) - g(\hat{e}_{2,t+h})$ is the loss differential. For a specific loss function (e.g. based on mean square error or mean absolute error), the test is based on the observed sample mean of loss differential. For instance, in this study the loss differential is based on mean square of forecast errors, which can also be modified for forecast encompassing.

$\bar{d} = (T - R - h + 1)^{-1} \sum_{t=R}^{T-h} \hat{d}_{t+h} = \widehat{MSFE}_1 - \widehat{MSFE}_2$ and $\hat{S}_{dd} = \sum_{j=-J}^{J} K(j/J) \hat{\Gamma}_{dd}(j)$,where

$$\widehat{MSFE}_i = (T - R - h + 1)^{-1} \sum_{t=R}^{T-h} \hat{e}_{i,t+h}, \quad for\ i = 1,2 ,$$

$\hat{\Gamma}_{dd}(j) = (T - R - h + 1)^{-1} \sum_{t=R+j}^{T-h} (\hat{d}_{t+h} - \bar{d})(\hat{d}_{t+h-j} - \bar{d})$ and $\hat{\Gamma}_{dd}(-j) = \hat{\Gamma}_{dd}(j)$. As defined in Newey and West (1987) the sample autocovariance $\hat{\Gamma}_{dd}(j)$ is weighted by $K(j/J) = 1 - [j/(J+1)]$ (known as Bartlett weights, which decreases as j increases). Following Clark and McCracken (2005), Rapach and Weber (2004) among others, set the bandwidth at $J = [1.5 * h]$, for $h > 1$, where $[j/(J+1)]$ is the nearest integer function; for the case of $h = 1$, J is zero which implies $\hat{S}_{dd} = \hat{\Gamma}_{dd}(0)$. The Diebold and Mariano (1995) and West (1996) statistic is denoted by *MSE-T*, and is obtained by computing:

$$MSE - T = (T - R - h + 1)^{-0.5} \bar{d} \, \hat{S}_{dd}^{-0.5} \tag{3.4}$$

The null hypothesis of equal accuracy for the two forecasts is tested against the one-sided (upper-tail) alternative.

$$H_0 : MSFE_1 = MSFE_2 \quad i.e. \ \bar{d} = MSE - T = 0$$
$$H_1 : MSE - T > 0$$

In other words the null hypothesis of equal MSFEs for the AR benchmark and competing forecasts (either Phillips curve or augmented Phillips curve) is tested against the alternative hypothesis that the MSFE for the competing Phillips curve forecasts is less than the MSFE for the AR benchmark forecasts $(MSFE_1 > MSFE_2)$ so that $MSE - T > 0$.

This alternative is one-sided rather than two sided, which is different to those discussed in both West (1996) and Diebold and Mariano (1995) due to the fact that the AR benchmark and competing model are nested (i.e. the AR benchmark forecast model is a special case of the competing forecast models). West (1996) has shown that the MSE-T test is asymptotically standard normal for non-nested models, but not for nested models. Also, Kilian (1999) states that for a small sample, the asymptotic critical values for this test statistic are severely biased. Furthermore, for $h = 1$, McCracken (2007) has shown that the asymptotic null distribution of the MSE-T statistic is non-standard when comparing forecasts from nested models. Indeed, he showed that the limiting distribution of the MSE-T statistic can be expressed as functions of stochastic integrals of Brownian motion. Clark and McCracken (2005) also show that when comparing forecasts from nested models, the limiting distribution of the MSE-T statistics have non-standard distributions that depend on the parameters of the data-generating process for $h > 1$. Hence the asymptotic distributions are not of central importance and the implantation of the Kilian (1999)-type bootstrap procedure is suggested by Clark and McCracken (2005) for statistical inference.

McCracken (2007) develops an out-of-sample variant of the MSE-T statistic of equal MSFE, given by

$$\begin{aligned} MSE-F &= (T-R-h+1)(M\hat{S}FE_1 - M\hat{S}FE_2)/M\hat{S}FE_2 \\ &= (T-R-h+1)\,\bar{d}\,/M\hat{S}FE_2 \end{aligned} \tag{3.5}$$

Like the MSE-T test, Clark and McCracken (2005) show that the limiting distribution of the MSE-F test is non-normal when the models are nested under the null hypothesis. Similar to the case of the MSE-T statistics, McCracken (2007) shows that for $h=1$ the distribution of the MSE-F statistic is not standard, and Clark and McCracken (2005) also show that, for $h>1$ the distribution of MSE-F is non-standard and not asymptotically pivotal. Therefore, this test also requires a bootstrap procedure for inference purposes.

So far, the described tests are useful for the formal testing of the relative MSFE criterion. Harvey *et al.* (1998) propose an alternative test to evaluate the forecasts based on the concept of forecast encompassing. Harvey *et al*. (1998) state that it is possible to combine various combinations of forecasts as composite forecasts. Rapach and Weber (2004) consider an optimal composite of out-of-sample forecasts as a convex combination of out-of-sample forecasts for nested models defined as:

$$f_{c,t+h} = (1-\lambda)\, f_{1,t+h} + \lambda f_{2,t+h} \tag{3.6}$$

where $0 \le \lambda \le 1$. Without loss of generality in notation, following Stock and Watson (1999) the composite combination of the out-of-sample forecast of $\pi_{t+h}^{h} - \pi_t$ is constructed as:

$$\pi_{t+h}^{h} - \pi_t = \lambda f_{2,t+h} + (1-\lambda)\, f_{1,t+h} + error \tag{3.7}$$

where $f_{2,t+h}$ is the forecast of $\pi_{t+h}^{h} - \pi_t$ based on the competing model and $f_{1,t+h}$ is the forecast of $\pi_{t+h}^{h} - \pi_t$ based on the AR benchmark model. The equation (3.7) is equivalent to $e_{1,t+h} = \lambda(e_{1,t+h} - e_{2,t+h}) + error$.

The λ is obtained from estimating this regression of forecast errors using OLS. In this, it is intended that the expected mean error from the composite forecast will be smaller than that of $f_{1,t+h}$ unless the covariance between $e_{1,t+h}$ and $e_{1,t+h} - e_{2,t+h}$ is zero. If $\lambda = 0$, the AR benchmark model forecasts are said to encompass the competing model forecasts, due to the fact that the competing model does not contribute any valuable information apart from that contained in the benchmark. If $\lambda > 0$, then the competing model does contribute useful information to the formation of the optimal composite forecast. Hence, in this case, the benchmark does not encompass the competing model forecast.

The test by Harvey *et al.* (1998) can be used to test the null hypothesis of $\lambda = 0$ against the one-sided alternative hypothesis that $\lambda > 0$. Their proposed test of encompassing uses a t-statistic and is based on the covariance between $e_{1,t+h}$ and $e_{1,t+h} - e_{2,t+h}$. The test is defined as

$$ENC-T = (T-R-h+1)^{0.5}\ \bar{c}\ \hat{S}_{cc}^{-0.5} \tag{3.8}$$

Where $\hat{c}_{t+h} = \hat{e}_{1,t+h}(\hat{e}_{1,t+h} - \hat{e}_{2,t+h})$, $\bar{c} = (T-R-h+1)^{-1}\sum_{t=R}^{T-h}\hat{c}_{t+h}$, $\hat{S}_{cc} = \sum_{j=-J}^{J} K(j/J)\hat{\Gamma}_{cc}(j)$,

$\hat{\Gamma}_{cc}(j) = (T-R-h+1)^{-1}\sum_{t=R+j}^{T-h}(\hat{c}_{t+h} - \bar{c})(\hat{c}_{t+h-j} - \bar{c})$, and $\hat{\Gamma}_{cc}(-j) = \hat{\Gamma}_{cc}(j)$. Again as in MSE-T, $K(j/J) = 1 - [1/(J+1)]$, $J = [1.5 * h]$ for $h > 1$, and $\hat{S}_{dd} = \hat{\Gamma}_{dd}(0)$ for $h = 1$.

Under the null hypothesis that the benchmark forecast encompasses the competing forecast, the covariance between $e_{1,t+h}$ and $e_{1,t+h} - e_{2,t+h}$ will be less than or equal to zero. Under the alternative that the competing model adds information, the covariance should be positive. Hence this test is a one-sided (upper tail) test. While West (1996) shows that this test has an asymptotic distribution for non-nested forecast models, Clark and McCracken (2001) show that the ENC-T statistic has a nonstandard limiting distribution for nested models.

Clark and McCracken (2001) developed a variant of the ENC-T statistic in which the covariance between $e_{1,t+h}$ and $e_{1,t+h} - e_{2,t+h}$ is scaled by the variance of one of the forecast errors rather than an estimate of the variance of $\bar{c}$. This was proposed after arguing that this feature of the ENC-T may adversely affect the small-sample properties of the test. This is due to the fact that the population forecast errors from forecasts 1 and forecasts 2 are the same under the null and hence the sample variance in the denominator of ENC-T is heuristically equal to zero (Clark and McCracken, 2001). The test is denoted ENC-NEW and is defined as

$$ENC-NEW = (T-R-h+1)\,\bar{c}\,/\,M\hat{S}FE_2 \tag{3.9}$$

As pointed out by Clark and McCracken (2001; 2005) the asymptotic distribution of this test is nonstandard for nested model forecasts under the null. Therefore, Clark and McCracken (2005) recommend a bootstrap procedure for inference on both ENC-T and ENC-NEW.

3.3.3. Bootstrap algorithm

To maintain the assumption of the null hypothesis of equal accuracy of the competing Phillips curve compared to the AR benchmark forecasts, the bootstrap data-generating process is obtained by replicating the technique used by Rapach and Weber (2004). This process consists of fitting the restricted vector autoregressive (VAR) model specified as follows:

$$\Delta\pi_t = a_0 + \sum_{i=1}^{p_1} a_i \Delta\pi_{t-i} + e_{1,t} \tag{3.10}$$

$$output\_gap_t = b_0 + \sum_{i=1}^{p_2} b_i \Delta\pi_{t-i} + \sum_{i=1}^{p_3} c_i output\_gap_{t-i} + e_{2,t} \tag{3.11}$$

$$M3_t = d_0 + \sum_{i=1}^{p_4} d_i \Delta\pi_{t-1} + \sum_{i=1}^{p_5} f_i output\_gap_{t-i} + \sum_{i=1}^{p_6} g_i M3_{t-i} + e_{3,t} \tag{3.12}$$

$$ex\_rate_t = h_0 + \sum_{i=1}^{p_7} h_i \Delta\pi_{t-i} + \sum_{i=1}^{p_7} l_i output\_gap_{t-i} + \sum_{i=1}^{p_9} m_i ex\_rate_{t-i} + e_{4,t} \tag{3.13}$$

where the disturbance vector $e_t = (e_{1,t}, e_{2,t}, e_{3,t}, e_{3,t})'$ is independent and identically distributed with covariance Σ. Rapach and Weber (2004) describe this process in the following steps:

1. First start by determining the lag orders $(p_1, p_2, \cdots, p_9)$ to be used in estimation of equations (3.10) through (3.13), in this case the Schwartz Information Criteria (SIC) was used to select lag orders and the lags were selected from zero to the maximum lag of six for each variable.
2. The second step is to estimate the equations using the full sample of observations via OLS and then compute the residuals $\{\hat{e}_t = (\hat{e}_{1,t}, \hat{e}_{2,t}, \hat{e}_{3,t}, \hat{e}_{4,t})'\}_{t=1}^{T}$. As in many applications including Enders (2010), Rapach and Weber (2004) generate the series of disturbances for the pseudo-sample by drawing randomly with replacement from the OLS residuals a pseudo-sample of disturbances containing $\{\hat{e}_t^* = (\hat{e}_{1,t}^*, \hat{e}_{2,t}^*, \hat{e}_{3,t}^*, \hat{e}_{4,t}^*)'\}_{t=1}^{T}$. Due to the presence of a lagged dependent variable, that is, the initial condition for lags of the dependent variable (say $\Delta\pi_1^*$ through $\Delta\pi_{p_1}^*$) are selected by random draw

from the actual $\{\Delta\pi_t\}$, it is advisable to construct a pseudo-sample residuals with $T+50$ elements to avoid initial condition problems. Hence, the pseudo-series of disturbance terms was generated (drawn with replacement) with an additional 50 disturbance from the OLS residuals to give $\{\hat{e}_t^*\}_{t=1}^{T+50}$.

3. The third step is to use the pseudo-series of disturbances to generate $\{\Delta\pi_t^*, output\_gap_t^*, M3_t^*, ex\_rate_t^*\}_{t=1}^{T+50}$ using the estimated values of the coefficients in equations (3.10) through (3.13) taken as fixed and setting the initial lagged observations for $\Delta\pi_t$, $output\_gap_t$, $M3_t$ and ex_rate_t to zero. Then discard the first $50-p$ observations, where $p=\max\{p_1,\cdots,p_9\}$, to remain with the pseudo-sample of $T+p$ observations, which corresponds to the sample size of the original sample. These constructed pseudo-samples are then used to calculate the forecast test statistics.
4. The last step is to repeat this previous steps as many times as possible (this report used 500 times as in Rapach and Weber (2004)) in order to get the empirical distribution for each of the described test statistics.

4. Results and discussion

4.1. Data description and their transformation

The data set that is used for this research is of a quarterly frequency over a period 1997:1-2012:4 and is from the central bank of Rwanda (BNR). The variables of interest are *CPI,* nominal gross domestic product (NGDP), money supply (M3) and the exchange rate, and are obtained from the central bank of Rwanda (BNR). All variables were transformed using natural logarithms and differenced once for stationarity where necessary, denoted by L and D respectively (Table 1). The inflation variable is derived from the CPI, while the output gap is generated from real GDP using the Hodrick-Prescott (HP) filter. Real GDP was obtained by calculating the ratio of NGDP to CPI

$$\ln(real\_GDP) = \ln(NGDP / CPI) = \ln(NGDP) - \ln(CPI).$$

The stationarity of the HP-generated output gap is obvious since the cyclic components from the de-trended series are always stationary. Figure 1 shows the estimate of potential output and the output gap.

Table 1: Summary of variables and their corresponding transformation

Variable	Description	Transformation
CPI	Consumer price index	DLCPI=pi
M3	Money supply	DLM3
EXRATE	Exchange rate	DLEX
NGDP	Nominal Growth Domestic Product	Output gap

To test the stationarity of the variables, each series was tested using the augmented Dick-Fuller (ADF) test. The null hypothesis of a unit root was rejected at the 1 per cent level in all cases. Table 2 summarizes the results of the ADF unit root tests for the inflation, money supply and exchange rate variables (appropriately transformed). In the first panel of Table 2 inflation (the log difference of CPI) is stationary. However, many studies find that inflation has a unit root (e.g. for the USA, Stock and Watson (1999, 2003, and 2008); Clark and

McCracken (2005, 2006) find that the inflation rate is $I(1)$ and use the change in inflation). Since there is little evidence on this in the literature for Rwanda, Stock and Watson's (2003) suggestion that the sensitivity of the results is checked by running both inflation and the change in inflation as the dependent variable is followed. The comparison of the results from this approach is reported in Tables 3 and 4.

Table 2: Unit root test

Null Hypothesis: PI has a unit root
Exogenous: Constant
Lag Length: 4 (Automatic - based on AIC, maxlag=6)

		t-Statistic
Augmented Dickey-Fuller test statistic		-4.043043
Test critical values:	1% level	-3.540198
	5% level	-2.909206
	10% level	-2.592215

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: DLM3 has a unit root
Exogenous: Constant
Lag Length: 4 (Automatic - based on SIC, maxlag=6)

		t-Statistic
Augmented Dickey-Fuller test statistic		-3.637475
Test critical values:	1% level	-3.548208
	5% level	-2.912631
	10% level	-2.594027

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: DLEX has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic
Augmented Dickey-Fuller test statistic		-3.933914
Test critical values:	1% level	-3.540198
	5% level	-2.909206
	10% level	-2.592215

*MacKinnon (1996) one-sided p-values.

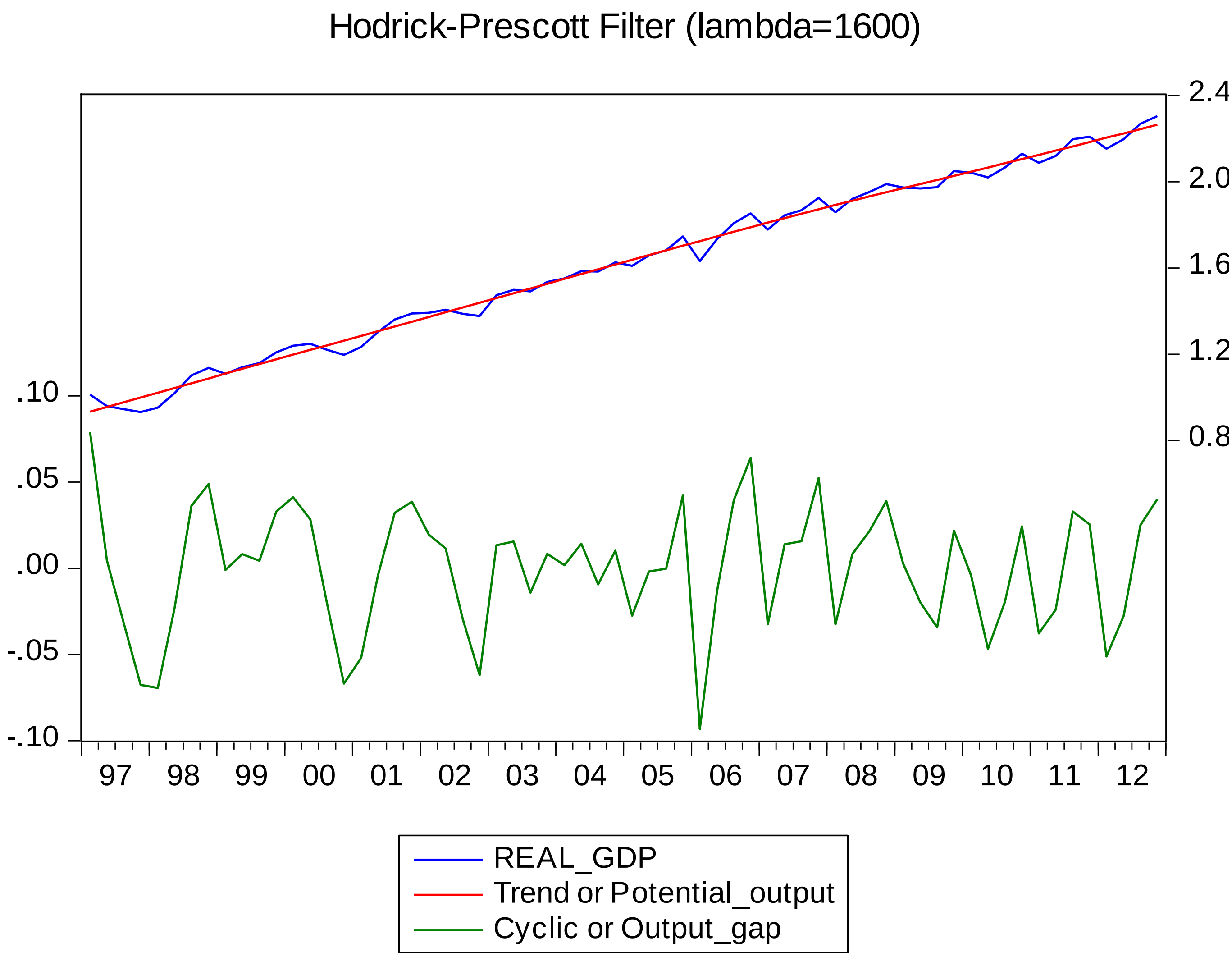

Figure 1: HP filtering for the Output gap

4.2. Empirical results

As discussed in section 3.3.2 on pseudo out-of-sample forecasting, the sample is split into two subsamples: one for in-sample estimation and another for forecast evaluation. Two observations were dropped from sample to adjust for differenced data and the lag lengths were selected according to the information criteria by setting the maximum lag of six. For a small sample size like in this study, the split of subsamples is crucial. To split the subsamples for in-sample estimation and out-of-sample forecasting, the approach by Clark and McCracken (2005) of taking roughly 20% of the full sample for the out-of-sample forecasting is adopted. This was deemed reasonable, given the relatively small sample of quarterly data available for Rwanda. The in-sample data used to estimate the model and produce the first forecast run from 1997:Q3 +h-1 through 2009:Q4. Out–of–sample forecasts were produced over the period 2010:Q1 to 2012:Q4, at different horizons (in quarters). The forecast horizons considered are 1, 2 and 4 periods ahead. Thus the forecasting sample includes a total of 12 one–step ahead forecasts, the first of which is generated from a model estimated with 50 "in-sample" observations before lag adjustment. More generally, for h-step ahead forecasts, the forecasting sample has a total of 12-h+1 forecasts. To generate the empirical results the Gauss programming language was used. The codes published by Rapach and Weber (2004) were used with some adjustments, and are gratefully acknowledged.

The recursive in-sample estimates were used to forecast inflation out-of-sample. The lag structure of each model was selected using the SIC across different forecast horizons using the full sample. Lag values for q1 were selected from a range of zero to the maximum of six lags, and to ensure that each predictor appears in the model, q2, q3, and q4 were selected from a range of one to six lags. The pseudo out-of-sample forecasting performance of each model is summarized in Tables 3 and 4. The tabular summary reports lags of the ADL models, the in-sample Wald statistic, the MSFE, mean absolute forecast error (MAFE), the Relative MSFE, and the four out-of-sample test statistics (MSE-T, MSE-F, ENC-T and ENC-NEW). In order to see how well the Phillips curves forecast Rwandan inflation over the period, a graph that illustrates the out-of-sample forecast performance is reported at each forecast horizon.

As mentioned before the reported results discuss the use of both the inflation level (reported in Table 3) and the change in inflation (reported in Table 4) as the dependent variable. Starting with the results in Table 3, our more likely case given the results of the unit root test in Table 2, the relative MSFEs indicate that the forecasts of the competing models outperform the benchmark at the 1- and 2-step ahead horizon (except for the augmented Phillips curve with money supply at the h=2 horizon). But at the h=4 horizon (1-year ahead forecasts) none outperform the benchmark.

In Table 4, using the change in inflation as the dependent variable confirms the robustness of the findings in Table 3, in that the number of relative MSFEs that are less than one is 6 in Table 4 against 5 in Table 3. (also, good performance in the relative MSFE is observed at the h=4 horizon in Table 4).

The MSE-T and MSE-F statistics are used to test the null hypothesis that the out-of-sample MSFE from the competing forecasts is equal to the out-of-sample MSFE from the AR benchmark forecasts against one sided (upper-tail) alternative hypothesis that the out-of-sample MSFE from the competing forecasts is lower than the out-of-sample MSFE from the benchmark forecasts. The ENC-T and ENC-NEW statistics are used to test the null hypothesis that the out-of-sample forecasts from AR benchmark encompass the out-of-sample forecast from competing forecasts against the one-sided (upper-tail) hypothesis that the out-of-sample forecasts from benchmark forecasts do not encompass the out-of-sample forecasts from competing forecasts. However, as discussed earlier, all four out-of-sample test do not have the asymptotic distribution for nested models, hence the bootstrap technique was applied to generate the p-values for the four out-of sample and the in-sample Wald tests. The p-values are given in parentheses where the significance was assessed at the 10% level and it is indicated in bold.

Comparing the results from Tables 3 and 4 based on these statistical tests, the p-values in Table 3 generated using bootstrap methods, generally support the rejection of equal forecasts between competing and benchmark models; and the rejection of forecast encompassing (i.e. of the null that the AR benchmark forecasts encompasses the competing forecasts). The rejections of these null hypotheses strengthen the results of the relative MSFE. Table 3 shows that the MSE-F tests support the results of the relative MSFE more than MSE-T, while both ENC-T and ENC-NEW are supportive, except for ENC-NEW for the augmented Phillips curve using money at h=2.

Similarly, in Table 4, although the relative MSFEs are generally less than one, there is some evidence that the statistical tests do not support some of these results. For instance, the relative MSFE is less than one for the Phillips curve forecasts made using the output gap at h=1 and h=2 horizons and for the augmented Phillips curve including the exchange rate (i.e. inflation regressed on the output gap and the exchange rate) at h=4. However none of the four statistical tests is significantly different from zero based on a p-value at the 10% level. This supports the usefulness of statistical tests, as pointed out by Stock and Watson (2003). An observable result from the comparison of the statistical tests in Table 4, is that, the MSE-T is less supportive of the relative MSFE compared to MSE-F, and also that the ENC-T is less supportive of relative MSFE results than ENC-NEW.

It is worth highlighting that the reported results in Table 3 are consistent with the unit root test (Table 2) suggesting the use of inflation in levels as the dependent variable. However, to check the robustness of these results, the case where the log differences of inflation are used as the dependent variable, is reported in Table 4. The results reported in Tables 3 and 4 show that the analysis is robust to this specification issue, in the sense that forecasts from the generalised Phillips curve models are able to outperform the benchmark forecasts at times irrespective of the specification. On the basis of the unit root test, and in the absence of new information that supports the alternative option, the specification using inflation in levels is preferred, and more weight should be given to these results.

The Wald test is the in-sample F-test using the full sample to test the null hypothesis that the variable Granger-causes inflation. This test is convenient for in-sample forecasting to assess whether a variable added to an AR model has predictive content for forecasting inflation. Thus, the reported Wald test statistic in this study examines the predictive content of added variables such as the output gap and/or M3 and the exchange rate in the ADL model using the full sample in order to see whether there is any improvement. This test was applied to each of competing models. For the Phillips curve, the null hypothesis for this test sets all coefficients on the output gap to zero. If the null is rejected then the output gap has in-sample predictive content with respect to future inflation, otherwise it adds no additional information to that from the AR model. Similarly, in the case of the augmented Phillips curves, the Wald test was applied by setting the coefficients on M3 or the exchange rate to zero and seeing whether the added variable has adds more information to that contained in the standard Phillips curve to predict inflation.

The results of the Wald tests are reported in Tables 3 and 4 where the p-value given in parentheses was obtained through a bootstrap procedure. Based on reported p-values, the Wald tests show that for h=1 and 2, the output gap in the standard Phillips and the money supply in the augmented Phillips curves for h=2 are significantly different from zero at the 10% level in Table 3 whereas in Table 4 none of the variables is significant different from zero. This rejection of the null hypothesis from the Wald test in Table 3 implies that both variables (output and M3) are relevant in sample in Rwanda which agrees with the results from the pseudo out-of-sample forecasts.

It has to be noted that the small sample size is probably an issue for both the estimation and the evaluation of the forecasts. Therefore, it is not erroneous to conclude that some dissimilarity of results between in-sample and out-of-sample forecasts based on Wald tests may due to the effects of the small sample. This can be observed in the graphical representations of the out-of-sample estimates that show little evidence of mimicking actual data as the horizon increases. Indeed this issue may not be only due to sample size but also to some missing information not included in models (model misspecification).

Table 3: Pseudo out-of-sample test results, using inflation in levels

Horizon (h):	**1 Quarter Ahead**		**2 Quarter Ahead**		**4 Quarter Ahead**	
Out-of-sample period	**2010:1-2012:4**		**2010:1-2012:4**		**2010:1-2012:4**	
MSFE_AR	67.25		27.79		13.13	
MAFE_AR	6.49		4.077		2.76	
Lags						
q1	6		6		6	
q2	1		1		1	
q3	1		1		1	
q4	2		1		1	
Phillips Curve						
wald	**9.46**	(0.02)	**5.12**	(0.09)	2.18	(0.25)
MSFE	46.18		23.06		18.65	
MAFE	5.58		3.91		3.27	
Rel MSE	0.68		0.83		1.42	
MSE-T	**1.53**	(0.04)	0.42	(0.25)	-1.11	(0.68)
MSE-F	**5.47**	(0.00)	**2.45**	(0.02)	-3.55	(0.98)
ENC-T	**3.04**	(0.00)	1.64	(0.11)	0.16	(0.45)
ENC-NEW	**5.86**	(0.00)	**5.34**	(0.00)	0.22	(0.18)
Augmented Phillips Curve using M3						
wald	2.19	(0.18)	**4.60**	(0.07)	1.32	(0.33)
MSFE	61.52		32.41		18.09	
MAFE	6.18		4.79		3.46	
Rel MSE	0.91		1.16		1.37	
MSE-T	0.68	(0.13)	-0.53	(0.46)	-1.56	(0.75)
MSE-F	**1.12**	(0.08)	-1.71	(0.82)	-3.29	(0.98)
ENC-T	**2.43**	(0.00)	1.18	(0.18)	-0.41	(0.55)
ENC-NEW	**3.20**	(0.00)	**1.65**	(0.05)	-0.37	(0.83)
Augmented Phillips Curve using the exchange rate						
wald	2.51	(0.16)	0.45	(0.65)	2.33	(0.35)
MSFE	56.61		21.85		16.49	
MAFE	6.16		3.82		3.21	
Rel MSE	0.84		0.78		1.25	
MSE-T	**1.03**	(0.09)	0.57	(0.20)	-0.90	(0.51)
MSE-F	**2.25**	(0.03)	**3.25**	(0.04)	-2.44	(0.82)
ENC-T	**2.90**	(0.00)	1.69	(0.14)	0.24	(0.42)
ENC-NEW	**4.22**	(0.00)	**5.74**	(0.01)	0.34	(0.30)

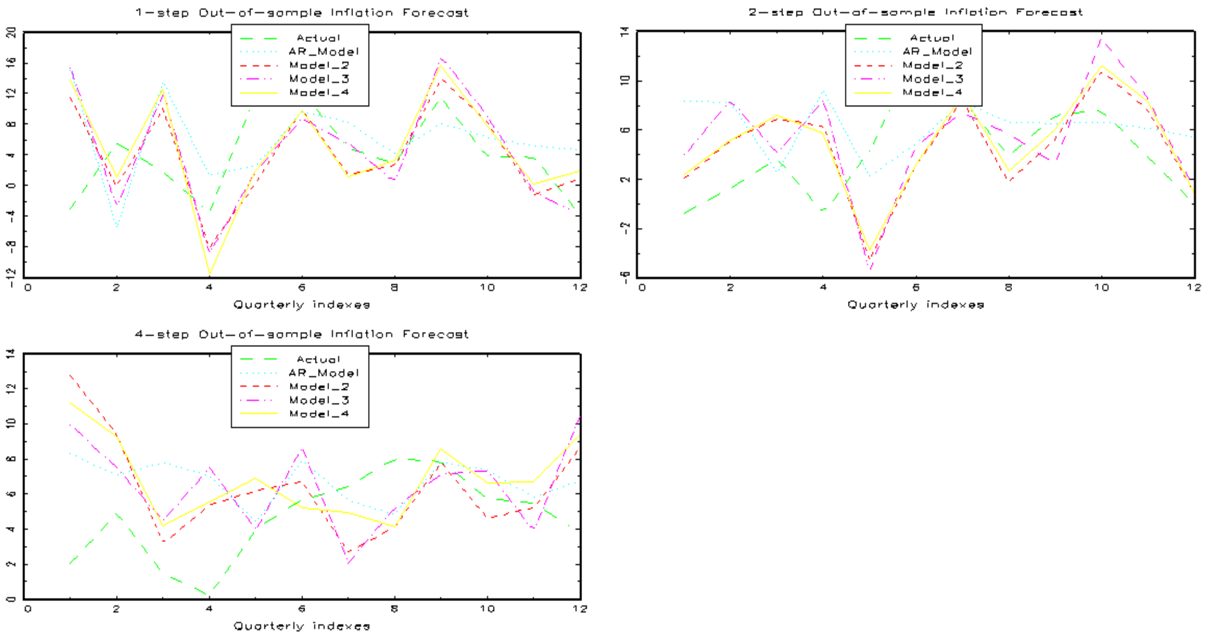

Figure 2: The out-of-sample fit of competing forecasts relative to observed data using inflation in levels.

Table 4: Pseudo out-of-sample test results, using change in inflation

Horizon (h):	1 Quarter Ahead		2 Quarter Ahead		4 Quarter Ahead	
Out-of-sample period	**2010:1-2012:4**		**2010:1-2012:4**		**2010:1-2012:4**	
MSFE_AR	42.27		22.93		18.44	
MAFE_AR	5.15		3.39		3.12	
Lags						
q1	5		5		5	
q2	1		1		1	
q3	1		1		1	
q4	1		1		1	
Phillips Curve						
wald	0.22	(0.77)	0.13	(0.83)	0.00	(0.98)
MSFE	40.81		22.90		11.90	
MAFE	5.17		3.51		2.74	
Rel MSE	0.96		0.99		0.64	
MSE-T	0.29	(0.26)	0.03	(0.33)	1.15	(0.18)
MSE-F	0.43	(0.19)	0.01	(0.33)	**4.94**	(0.01)
ENC-T	0.49	(0.33)	0.12	(0.45)	1.17	(0.28)
ENC-NEW	0.36	(0.25)	0.03	(0.46)	**3.02**	(0.05)
Augmented Phillips Curve including M3						
wald	0.09	(0.81)	0.17	(0.70)	0.06	(0.86)
MSFE	37.93		23.62		9.50	
MAFE	4.70		3.47		2.49	
Rel MSE	0.89		1.03		0.51	
MSE-T	**1.17**	(0.07)	-0.47	(0.40)	1.19	(0.15)
MSE-F	**1.37**	(0.09)	-0.32	(0.38)	**8.47**	(0.01)
ENC-T	1.39	(0.10)	0.06	(0.49)	1.21	(0.27)
ENC-NEW	0.89	(0.22)	0.01	(0.50)	**5.74**	(0.03)
Augmented Phillips Curve including Exchange rate						
wald	0.22	(0.62)	0.03	(0.90)	0.68	(0.65)
MSFE	43.54		24.53		15.32	
MAFE	5.27		3.54		3.15	
Rel MSE	1.03		1.07		0.83	
MSE-T	-0.33	(0.39)	-2.13	(0.85)	0.91	(0.18)
MSE-F	-0.35	(0.37)	-0.71	(0.38)	1.83	(0.12)
ENC-T	-0.10	(0.54)	-2.02	(0.91)	1.02	(0.32)
ENC-NEW	-0.05	(0.52)	-0.33	(0.62)	1.14	(0.23)

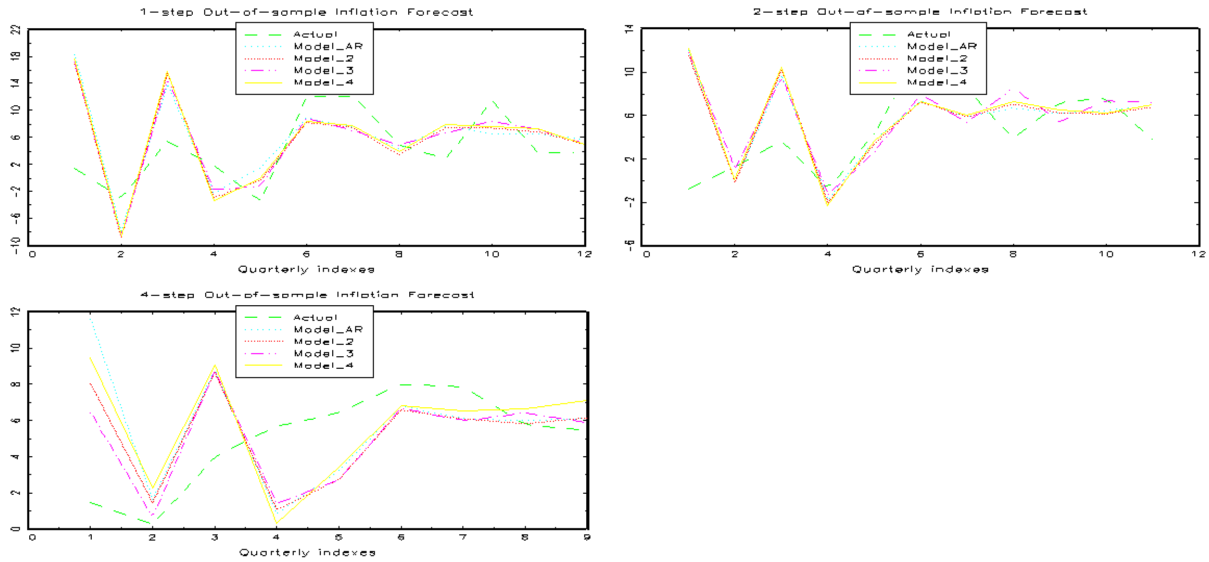

Figure 3: The out-of-sample fit of competing forecasts relative to observed data using change in inflation.

The forecast horizon was limited to 4 periods due to the sample size here, but it should be noted that evidence for other countries suggests that the transmission of policy changes to inflation takes roughly 2 years (8 quarters). To consider the transmission mechanism over 2 years, the next results focus on forecasts over a horizon of h=8 quarters. However, to achieve these results in the present study, the size of the out-of-sample set of observations will have to be increased. But this will however imply a reduction in the in-sample set of observations. Obviously this reduction in the number of in-sample observations affect the in-sample estimation as noted earlier.

The results reported in Tables 5 and 6 also use inflation in levels and the change in inflation, respectively. To accommodate the minimum out-of-sample number of observations which can host 8 quarters of forecasts, 32 percent of available observations were used (equivalent to 20 observations). The results in Table 5 show no evidence that the forecasts from the Phillips curve models outperform the benchmark (the MSFEs exceed one in all cases at all horizons). Similarly results are reported in Table 6. This lack of forecast performance may of course be partly due to the effects of small sample size and the related decisions regarding the number of out-of-sample observations.

Table 5: Pseudo out-of-sample test results, using inflation in levels for 8 Quarters

Horizon (h):	1 Quarter Ahead	2 Quarter Ahead	4 Quarter Ahead	8 Quarter Ahead
Out-of-sample period	2008:1-2012:4	2008:1-2012:4	2008:1-2012:4	2008:1-2012:4
MSFE_AR	84.61	56.41	36.23	20.11
MAFE_AR	6.91	5.31	4.37	3.84
Lags				
q1	6	6	6	6
q2	1	1	1	1
q3	1	1	1	1
q4	1	1	2	1
Phillps Curve				
wald	**9.46** (0.01)	**5.12** (0.07)	2.18 (0.13)	0.25 (0.73)
MSFE	98.05	67.53	46.32	20.35
MAFE	7.94	6.26	5.42	3.91
Rel MSE	1.15	1.19	1.27	1.01
MSE-T	-0.43 (0.40)	-0.47 (0.41)	-1.73 (0.89)	-0.25 (0.42)
MSE-F	-2.74 (0.94)	-3.29 (0.94)	-4.35 (0.99)	-0.23 (0.54)
ENC-T	**1.44** (0.06)	0.86 (0.19)	0.47 (0.35)	0.47 (0.39)
ENC-NEW	**2.90** (0.01)	**2.04** (0.03)	0.46 (0.10)	0.20 (0.23)
Augmented Phillips Curve using M3				
wald	2.19 (0.20)	**4.60** (0.05)	1.32 (0.30)	1.38 (0.46)
MSFE	92.53	61.67	45.64	22.63
MAFE	7.82	6.45	5.46	4.13
Rel MSE	1.09	1.09	1.26	1.12
MSE-T	-0.36 (0.34)	-0.30 (0.32)	-1.76 (0.81)	-3.33 (0.86)
MSE-F	-1.71 (0.65)	-1.70 (0.67)	-4.12 (0.94)	-2.22 (0.73)
ENC-T	**1.41** (0.07)	0.77 (0.22)	0.08 (0.43)	-2.67 (0.88)
ENC-NEW	**2.94** (0.03)	**1.95** (0.05)	0.04 (0.43)	-0.85 (0.84)
Augmented Phillips Curve using the exchange rate				
wald	0.03 (0.88)	0.45 (0.60)	2.87 (0.33)	5.55 (0.27)
MSFE	103.88	67.11	42.54	20.33
MAFE	8.12	6.22	5.27	3.88
Rel MSE	1.22	1.19	1.17	1.01
MSE-T	-0.61 (0.45)	-0.48 (0.38)	-0.89 (0.51)	-0.08 (0.31)
MSE-F	-3.71 (0.90)	-3.19 (0.75)	-2.96 (0.73)	-0.22 (0.33)
ENC-T	1.23 (0.10)	0.84 (0.22)	0.90 (0.29)	1.88 (0.24)
ENC-NEW	**2.20** (0.06)	1.89 (0.10)	2.36 (0.11)	3.53 (0.10)

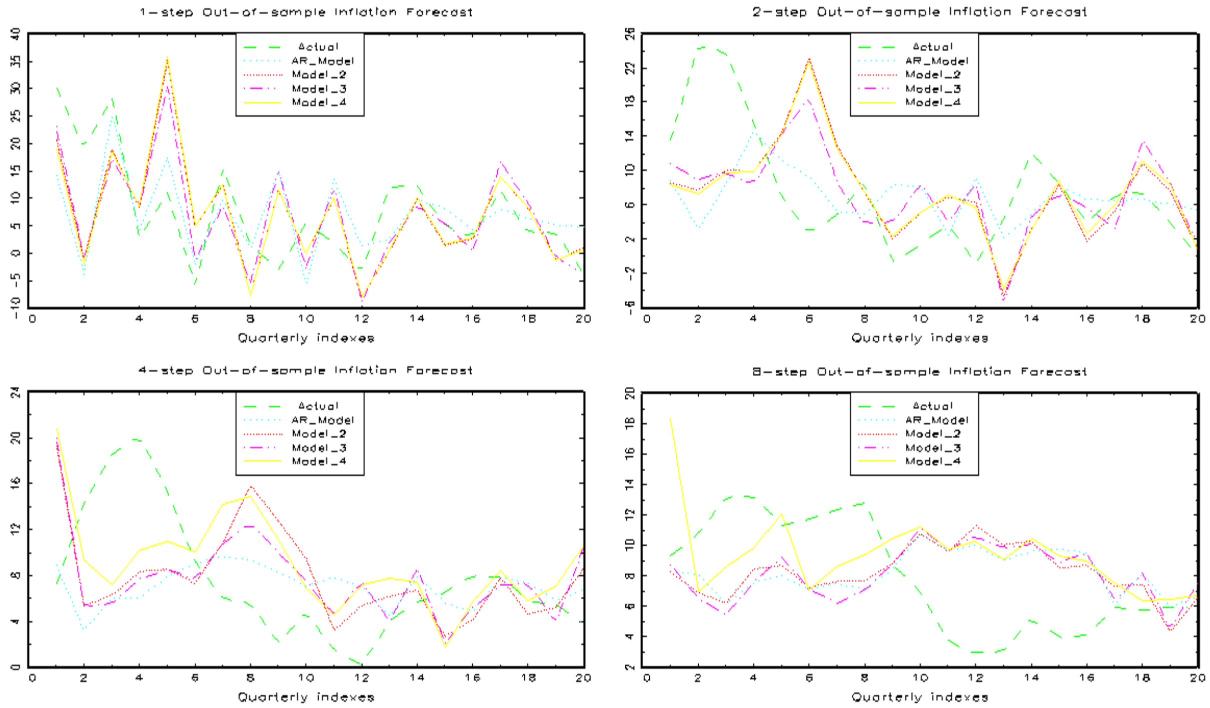

Figure 4: The out-of-sample fit of competing forecasts relative to observed data using inflation in levels for 8 Quarters.

Table 6: Pseudo out-of-sample test results, using change in inflation for 8 Quarters

Horizon (h):	1 Quarter Ahead		2 Quarter Ahead		4 Quarter Ahead		8 Quarter Ahead	
Out-of-sample period	2010:1-2012:4		2010:1-2012:4		2010:1-2012:4		2010:1-2012:4	
MSFE_AR	53.07		29.47		51.66		68.37	
MAFE_AR	5.89		4.49		5.77		6.91	
Lags								
q1	1		1		3		6	
q2	4		1		6		1	
q3	1		2		2		2	
q4	1		1		1		1	
Phillps Curve								
wald	**24.57**	(0.00)	3.20	(0.37)	**43.50**	(0.02)	1.87	0.61
MSFE	162.38		34.10		53.78		66.09	
MAFE	9.33		4.70		6.13		6.88	
Rel MSE	3.06		1.15		1.04		0.96	
MSE-T	-2.26	(0.89)	-1.08	(0.56)	-0.19	(0.40)	1.09	(0.18)
MSE-F	-13.46	(0.98)	-2.58	(0.69)	-0.67	(0.58)	0.45	(0.20)
ENC-T	-1.50	(0.89)	-0.49	(0.59)	2.26	(0.08)	1.23	(0.32)
ENC-NEW	-3.05	(0.99)	-0.55	(0.73)	**4.85**	(0.03)	0.28	(0.39)
Augmented Phillips Curve using M3								
wald	0.08	(0.82)	1.97	(0.37)	1.67 (0.42)		0.35	(0.73)
MSFE	159.77		38.15		107.29		122.42	
MAFE	9.22		4.73		8.12		9.49	
Rel MSE	3.01		1.29		2.07		1.79	
MSE-T	-2.08	(0.80)	-0.94	(0.47)	-1.77	(0.75)	-3.74	(0.86)
MSE-F	-13.35	(0.97)	-4.32	(0.69)	-8.81	(0.92)	-5.74	(0.79)
ENC-T	-1.32	(0.85)	-0.08	(0.52)	-0.02	(0.56)	-3.04	(0.92)
ENC-NEW	-2.83	(0.98)	-0.18	(0.57)	-0.02	(0.57)	-1.32	(0.89)
Augmented Phillips Curve using the exchange rate								
wald	0.474	(0.50)	0.29	(0.70)	0.01	(0.96)	0.37	(0.77)
MSFE	157.84		33.08		57.88		67.68	
MAFE	9.35		4.48		6.41		6.91	
Rel MSE	2.97		1.12		1.12		0.99	
MSE-T	-2.41	(0.89)	-0.81	(0.36)	-0.55	(0.32)	0.30	(0.16)
MSE-F	-13.27	(0.96)	-2.07	(0.41)	-1.83	(0.41)	0.13	(0.19)
ENC-T	-1.55	(0.91)	-0.20	(0.51)	2.14	(0.07)	0.59	(0.43)
ENC-NEW	-3.01	(0.97)	-0.25	(0.54)	**4.75**	(0.05)	0.13	(0.49)

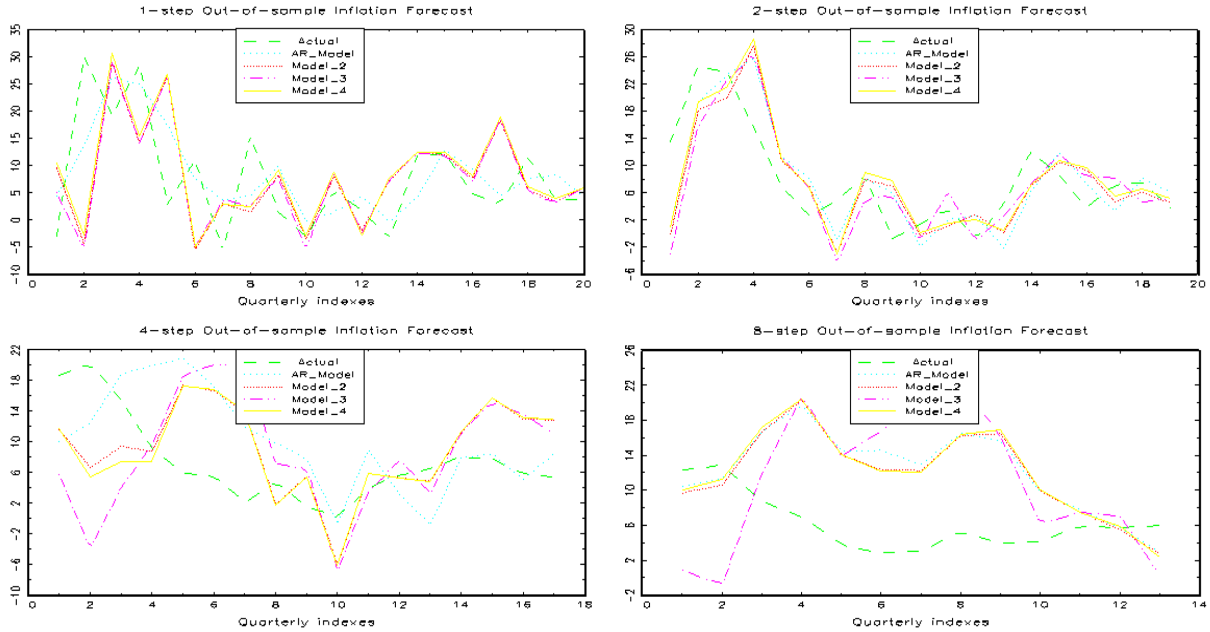

Figure 5: The out-of-sample fit of competing forecasts relative to observed data using change in inflation for 8 Quarters.

5. Conclusion and recommendations

This study evaluates the forecasts of inflation in Rwanda using single equation Phillips curve models. The evaluation was based on out-of-sample forecasts of competing Phillips curve forecasts relative to AR benchmark forecasts. The relative MSFEs of competing forecasts to AR model were computed and formal tests for the out-of-sample predictive content of competing forecasts were undertaken (that is MSE-T, MSE-F, ENC-T and ENC-NEW). The study also compared forecasts of inflation obtained using both inflation in levels (the preferred specification given the results of unit root testing) and change in inflation as the dependent variable. The results are robust to this choice.

The study finds that competing Phillips curve forecasts generally outperform the AR benchmark forecasts. However this may depend to some extent on the sample period and forecast horizon selected, highlighting the challenge of limited data for Rwanda. For forecasts using inflation in levels (Table 3), all variables were found to be good predictors of inflation in Rwanda at h=1, 2, except M3 at h=2. The output gap appears to be an important variable based on MAFEs. Therefore, the researcher strongly recommends that Rwandan economic policymakers take the output gap, money supply and exchange rate (in semi-structural Phillips curve-type models) into consideration in their modelling and forecasting of inflation in order to enhance monetary policy implementation. Of course, there is an open window for further research into models and variables that could help predict inflation in Rwanda.

References

Ang, A., Bekaert, G. and Wei, M. (2007) Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, vol. 54, pp. 1163–1212.

Atkeson, A. and Ohanian, E.L. (2001) Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 25, no. 1, pp. 2–11.

Ben Taieb, S. and Hyndman, R. (2014) Boosting multi-step autoregressive forecasts, *International Conference on Machine Learning (ICML)*, (forthcoming).

Berg, A., Charry, L., Portillo, L. and Vlcek, R. (2013) The Monetary Transmission Mechanism in the Tropics: A Narrative Approach, IMF working paper, WP/13/197.

Bhansali, R.J. (1996) Asymptotically efficient autoregressive model selection for multistep prediction, *ann. inst, statist. Math*, vol. 48, no. 3, pp. 577- 602.

Blinder, S.A. (1997) Is there a Core of Practical Macroeconomics That We Should all Believe? *The American Economic Review*, vol. 87, no. 2, pp. 240-243.

Bowdler, C. and Jansen, S. E. (2004) *A markup model of inflation for the euro area*, European Central Bank Working paper series, no. 306.

Chevillon, B. (2007) Direct multi-step estimation and forecasting, *Journal of Economic Surveys*, vol. 21, no. 4, pp. 746–785.

Chhibber, A. (1991) *Africa's Rising Inflation: Causes, Consequences and Cures*, World Bank Working paper, no. 0577.

Clark, T.E and McCracken, M.W. (2006) The Predictive Content of the Output Gap for Inflation: Resolving In-Sample and Out-of-Sample Evidence, *Journal of Money, Credit and Banking*, Vol. 38, No. 5, pp. 1127-1148.

Clark, T.E. and McCracken, M.W. (2001) Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics*, vol. 105, pp. 85–110.

Clark, T.E. and McCracken, M.W. (2005) Evaluating direct multistep forecasts, *Econometric Reviews*, vol. 24, no. 4, pp. 369-404.

Claus, I. (2000) *Is the output gap a useful indicator of inflation?*, Reserve Bank of New Zealand Working paper, DP2000/05.

Clausen, B. and Clausen, R.J. (2010) *Simulating Inflation Forecasting in Real-Time: How Useful Is a Simple Phillips Curve in Germany, the UK, and the US?* IMF Working Paper, WP/10/52, 1 February.

Cogley, T. and Nason, J.M. (1995) Effects of the Hodrick-Prescott filter on trend and difference stationary time series: Implications for business cycle research, *Journal of Economic Dynamics and Control*, vol. 19, pp. 253-278.

de Brouwer, G. and Ericsson, N.R. (1998) Modeling Inflation in Australia, *Journal of Business & Economic Statistics*, vol. 16, no. 4, pp. 433-449.

Dickey, D.A. and Fuller, W.A. (1979) Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, vol. 74, pp. 423-31.

Diebold, F.X. (2012) *Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the use and abuse of Diebold-Marian test*, NBER Working Paper, no. 18391.

Diebold, F.X. and Mariano, R.S. (1995) Comparing predictive accuracy, *Journal of Business and Economic Statistics*, vol. 13, no. 3, pp. 253–263.

Enders, W. (2010) *Applied Econometric Time Series*. 3rd ed. New York: Wiley, 544 pages.

Fedec, A. (2012) *Rwanda Inflation Rate*, [online], available: http://www.tradingeconomics.com/rwanda/inflation-cpi , [20 March 2012]

Fisher, I. (1926) A statistical relationship between unemployment and price changes, *International Labor Review* 13 (June): 785–92. Reprinted 1973. I discovered the Phillips curve. *Journal of Political Economy*, vol. 81, no. 2, pp. 496-502.

Gruen, D., Pagan, A. and Thompson, C. (1999) The Phillips curve in Australia, *Journal of Monetary Economics*, vol. 44, pp. 223-258.

Hamzacebi, C., Akay, D. and Kutay, F. (2009) Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting, *Expert Systems with Applications*, vol. 36, pp. 3839–3844.

Harvey, A. (2011) Modelling the Phillips curve with unobserved components, *Applied Financial Economics*, vol. 21, no. 12, pp. 7-17.

Harvey, A.C. and Jaeger, A. (1993) Detrending, Stylized Facts and the Business Cycle, *Journal of Applied Econometrics*, vol. 8, no. 3, pp. 231-247.

Harvey, D.I., Leybourne, S.J. and Newbold, P. (1998) Tests for Forecast Encompassing, *Journal of Business & Economic Statistics*, vol. 16, no. 2, pp. 254-259.

Hodrick, R. and Prescott, C.E. (1997) Post war business cycles: An empirical investigation, *Journal of Money, Credit and Banking*, vol. 29, no. 1, pp. 1-16.

Ibarra, R. (2012) Do disaggregated CPI data improve the accuracy of inflation forecasts?, *Economic Modelling* vol. 29, pp. 1305–1313.

Inoue, A. and Kilian, L. (2005) In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?, *Econometric Reviews*, vol. 23, no. 4, pp. 371-402.

Kang, I. (2003) Multi-period forecasting using different models for different horizons: an application to U.S. economic time series data, *International Journal of Forecasting*, vol. 19, pp. 387–400.

Kilian, L. (1999) Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions? *Journal of applied econometrics*, vol. 14, pp. 491-510.

Marcellinoa, M., Stock, J.H. and Watson, M.W. (2006) A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series, *Journal of Econometrics*, vol. 135, pp. 499–526.

Matheson, T.D. (2008) Phillips curve forecasting in a small open economy, *Economics Letters*, vol. 98, pp. 161–166.

McCracken, M.W. (2007) Asymptotics for out of sample tests of Granger causality, *Journal of Econometrics*, vol. 140, pp. 719–752.

MINECOFIN. (2009, September) *Rwandan economy registered robust growth of 9.4% during the 2011/12 whilst inflation remains the lowest in the region*, [online], available: http://www.minecofin.gov.rw/node/621, [25 March 2013]

Mise, E., Kim, T. and Newbold, P. (2005) On suboptimality of the Hodrick–Prescott filter at time series endpoints, *Journal of Macroeconomics*, vol. 27, pp. 53–67.

Moser, G., Rumler, F. and Scharler, J. (2007) Forecasting Austrian inflation, *Economic Modelling*, vol. 24, pp. 470-480.

National Bank of Rwanda (2012) *Monetary policy framework*, [online], available: http://www.bnr.rw/index.php?id=180 [1 March 2014]

Newey, W.K. and West, K.D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, vol. 55, no. 3, pp. 703-708.

NISR. (2013, February) *Consumer Price Index (CPI)*, [online], available: http://statistics.gov.rw/publications/consumer-price-index-cpi-february-2013, [15 March 2013]

Orphanides, A. and van Norden, S. (2005) The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time, *Journal of money, credit, and banking*, vol. 37, no. 3, pp. 583-601.

Phillips, A.W. (1958) The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957, *Economica, New series*, vol. 25, no. 100, pp. 283–299.

Proietti, T. (2011) Direct and iterated multistep AR methods for difference stationary processes, *International Journal of Forecasting*, vol. 27, pp. 266–280.

Rapach, D. and Weber, E.C. (2004) Financial variables and the simulated out-of-sample forecastability of U.S. output growth since 1985: An encompassing approach, *Economic Inquiry*, Vol. 42, No. 4, pp. 717–738.

Sayinzoga, A. and Simson, R. (2006) Monetary policy in Rwanda: A cointegration analysis, *South African Journal of Economics*, vol. 74, no. 1, pp. 65-79.

South African Reserve Bank (March, 2001) *Monetary Policy Review*, [online], available: https://www.resbank.co.za/Lists/News%20and%20Publications/Attachments/2487/mprmarch.pdf , [27 March 2014]

Stock, J.H. and Watson, M.W. (2011) *Introduction to econometrics*, 3rd ed., Pearson Education, 827 pages.

Stock, J.H. and Watson, M.W. (1999) Forecasting inflation, *Journal of Monetary Economics*, vol. 44, no. 2, pp. 293–335.

Stock, J.H. and Watson, M.W. (2003) Forecasting Output and Inflation: The Role of Asset Prices, *Journal of Economic Literature*, vol. 41, pp. 788–829.

Stock, J.H. and Watson, M.W. (2007) *Why has U.S. inflation become harder to forecast?*, NBER Working Paper no. 12324.

Stock, J.H. and Watson, M.W. (2008) *Phillips Curve Inflation Forecasts*, NBER Working Paper no. 14322.

Svensson, L.E.O. (1997) Inflation forecast targeting: Implementing and monitoring inflation targets, *European Economic Review*, vol. 41, pp. 1111-1146.

Tsay, R. (2008) *Out-of-Sample Forecasts*, [online], available: http://faculty.chicagobooth.edu/ruey.tsay/teaching/bs41202/sp2008/lec4a-08.pdf, [10 July 2013]

West, K.D. (1996) Asymptotic inference about predictive ability, *Econometrica*, vol. 64, no. 5, pp. 1067-1084.

West, K.D. (2001) Tests for Forecast Encompassin g WhenForecasts Depend on Estimated Regression Parameters, *Journal of Business & Economic Statistics*, vol. 19, no. 1, pp. 29-33.

Appendix A

The Hodrick-Prescott filter decomposes a series into a trend and a stationary component. Consider a series y_t where $t = 1, \cdots, T$ and you want to decompose y_t into a trend μ_t and a stationary component $y_t - \mu_t$. The Hodrick-Prescott filter uses the following sum of squares: $L = \frac{1}{T}\sum_{t=1}^{T}(y_t - \mu_t)^2 + \frac{\lambda}{T}\sum_{t=2}^{T}[(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})]^2$ where λ is a constant and T is the number of usable observations. The problem is to select μ_t (the control or instrument) so as to minimize the sum of squares L. The 1st term minimises the variance of y_t around μ_t ; the 2nd term is a penalty for variation in the second difference of the trend component. The sensitivity of the trend to fluctuations is adjusted by modifying λ. If $\lambda = 0$, then the 2nd term will vanish and the minimization require $y_t = \mu_t$ hence the trend is equal to itself. As $\lambda \to \infty$ the minimization of L is obtained when $(\mu_{t+1} - \mu_t) = (\mu_t - \mu_{t-1})$, so for $\lambda \to \infty$ it follows that μ_t is a linear trend (constant growth rate) (Enders, 2010). In many applications λ is set as follow: $\lambda = 1600$ for quarterly data; $\lambda = 100$ for annual data; and $\lambda = 14400$ for monthly data (Hodrick and Prescott, 1997).

Appendix B

Dickey-Fuller (DF) and Augmented Dickey-fuller (ADF) tests

Generally, economic time series appear to be trended variable when plotted. A trended series is non-stationary, and to make it stationary it needs to be de-trended. A unit root test is a pre-test before the modelling process and it is used to test the hypothesis that there is a unit root (stochastic trend in the series) against the alternative that there is no unit root (no trend) (Stock and Watson, 2011). This section aim to discuss the unit root tests such as, Dickey-Fuller (Dickey and Fuller, 1979) and its extension known as Augmented Dickey-fuller (ADF) test used in this report.

Consider an AR(1) model: $y_t = \phi_1 y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim IID(0, \sigma^2)$. Subtracting both side by y_{t-1} yield $\Delta y_t = \Phi y_{t-1} + \varepsilon_t$ (B.1)

where $\Phi = \phi_1 - 1$. The idea is to test whether the series contain a unit root, that is, $\phi_1 = 1$. The null hypothesis in DF test is

$$H_0 : \Phi = 0 \Rightarrow \Delta y_t = \varepsilon_t \text{ i.e. } y_t \text{ is } I(1)$$

The alternative is $H_1 : \Phi < 0$ which is chosen to maximize the power of the test in the likely direction of departure from the null. Equation (B.1) is estimate using OLS in order to get the estimated value of Φ and its standard error, then the t-statistic on estimated Φ is compared to the critical value from Dickey-Fuller table. The decision on whether to reject or fail to reject the null hypothesis is guided by critical value. When t-statistic is more negative than the relevant critical value implies a rejection of the null hypothesis that, y_t is $I(1)$, in the direction of the one-sided alternative that it is I (0). The DF tests do not have a normal distribution under the null, even in large samples.

Two others equation models considered by Dickey and Fuller (1979) to test for the presence of unit root are:

$$\Delta y_t = \mu + \Phi y_{t-1} + \varepsilon_t \tag{B.2}$$

$$\Delta y_t = \mu + \Phi y_{t-1} + \beta t + \varepsilon_t \tag{B.3}$$

Note that the equation (B.1) is a pure random walk model; (B.2) is a random walk with drift; and (B.3) is the random with drift and deterministic trend. The process for testing the presence of unit root in models (B.2) and (B.3) is the same as in (B.1), but each has its own appropriate table for critical value that depends on the regression and sample size (Dickey and Fuller, 1979). Indeed, for equation model (B.3), the hypotheses for test differ to others. In (B.3) the null hypothesis is

$H_0 : (\mu, \beta, \Phi) = (\mu, 0, 0)$ a random with drift

And the alternatives are:

$$H_{11} : (\mu, \beta, \Phi) = (\mu, \beta, \Phi)$$
$$H_{12} : (\mu, \beta, \Phi) = (\mu, \beta, 0)$$
$$H_{13} : (\mu, \beta, \Phi) = (\mu, 0, \Phi)$$

The statistics were labeled τ for model (B.1), τ_μ for model (B.2), and τ_τ, for model (B.3).

However, in case of high order of AR model, the DF test discussed above of AR(1) models is not suitable since the error term is autocorrelated, hence further tests such as augmented Dickey-Fuller for parametric is more convenient.

The ADF test is like DF test but with addition of lags to control the serial correlation. To illustrate this, let assume that the true process is generated by AR(2) model, that is

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \qquad \text{(B.4)}$$

where $\varepsilon_t \sim IID(0, \sigma^2)$, to manipulate this, add and subtract $\phi_2 y_{t-1}$ in the right hand side of (B.4) and subtract both side of (B.4) by y_{t-1}, it yield

$$\Delta y_t = \mu + \Phi y_{t-1} + \alpha_1 \Delta y_{t-1} + \varepsilon_t \qquad \text{(B.5)}$$

where $\Phi = \phi_1 + \phi_2 - 1$ and $\alpha_1 = -\phi_2$. It is clear that the term Δy_{t-1} is augmented to DF models discuss in previous section. Thus, the generalize form of AR(p) for ADF test is

$$\Delta y_t = \mu + \Phi y_{t-1} + \sum_{j=2}^{p} \alpha_j \Delta y_{t-j+1} + \varepsilon_t \qquad \text{(B.6)}$$

Where $\Phi = \phi_1 + \phi_2 + \cdots + \phi_p - 1 = -(1 - \sum_{j=1}^{p} \phi_j)$ and $\alpha_j = -\sum_{i=j}^{p} \phi_i$. If $\sum \phi_j = 1$, $\Phi = 0$, then (B.6) has a unit root.

The estimated of ADF statistics are obtained in a similar way as in (B.1) but using various information criteria, to identify the optimal lags length for ADF equation. Information criteria are discussed in Appendix C.

Appendix C

Model selection: information criteria

The lag length for the considered model may be determined using model selection criteria. The idea is to minimize a function of information criteria of the form: $IC(p) = \ln \sigma_p^2 + C_T$ over $p = 0,1,\cdots, p_{\max}$ where $p_{\max}$ is the maximum lag order the practitioner deems acceptable, T is sample size and σ_p^2 is the estimated regression error variance of the model. In general, σ_p^2 decreases as more lags are included. C_T is a penalty term which is increased as more lags are included. The penalty term differs depending to the information criteria. Two most commonly used model selection criteria are Akaike information criteria (AIC) and Schwartz Bayes information criteria (SIC). For AIC the penalty term is $C_T = \frac{2p}{T}$ while for SIC the penalty term is given by $C_T = \frac{p \ln T}{T}$.