Эдийн засаг: Онол, практик, Дугаар 35 (2022)
Economics: Theory and Practice, Volume 35 (2022)
ISSN:261-247X; eISSN: 2617-2488; www.econjournal.mn *Судалгааны өгүүлэл*

Forecasting Time Series: UK inflation

Gantungalag Altansukh

National University of Mongolia

Received: 2021.10.05; Revised: 2022.02.27; Accepted: 2022.05.04

Abstract: Related to the less volatile periods in inflation since 1980s, the forecast performance of naïve models have been improved. In that notion, simple univariate autoregressive model is selected as a benchmark against the proposed models which are multivariate vector autoregressive and vector error correction model specification, these expected to improve forecast accuracy in UK inflation due to the improved set of information. Proposed models are forecasted using h step ahead out-of-sample methodology together with both iterative and direct forecast, and they are compared based on the forecast accuracy measures. I will deal with several issues related to a forecasting to make sure selecting the best model appropriately that perform well most of the times.

Keywords: Forecasting time series, Inflation, UK;
JEL classification: C32, C53.

Department of Economics, Building V, Room 609, Ulaanbaatar, Mongolia. email: a.gantungalag@num.edu.mn;

1 Introduction

One of the key objectives of any country's Central Bank is not only to minimize fluctuations in inflation through its monetary policy to provide the economic stability but also to keep inflation at a low and stable level. However, achieving this objective is difficult without the ability to predict these changes in inflation.

The UK economy has experienced low inflation and stable growth in the periods preceding the 1970s, however there was an unexpected high rate of inflation and fluctuations in economic growth during the 1970s due to monetary policy changes and the influence of a substantial oil price changes (Barnett et al 2010). It started recovering to the stable periods of inflation and growth from the 1980s to this present day. In particular, the change in monetary policy in 1992 related to the first adoption of the inflation targeting[1] for Retail Price Index (RPIX) made the UK economy more stable (Benati 2004). Benati (2004) pointed that the stability of the inflation dynamics after 1992 is the highest since the gold standard. Therefore, the structural change after "Great Inflation[2]" is evident.

Because of the periods of high volatility and several structural changes in the inflation process, researchers tend to propose the highly sophisticated models which include many variables and complicated techniques to deal with problems that may cause forecast failures and to determine which models perform well in the particular time periods and data set, and thus often pay less attention to the naïve forecast models. Stock and Watson (2007) point out that the U.S inflation has become easier to forecast in a sense that declined inflation volatility reduces the risk of large inflation forecast errors, as measured by mean squared forecast error, yet it is harder to forecast in a sense because the standard multivariate models' percentage improvements over univariate models has shrunk. The UK inflation has been experiencing a similar pattern which is consistent with this view (Groen et al 2009). Based on that notion, the value added of candidate models such as models with robust methods to structural breaks to the naïve models may be small and perform comparably most of the time. Perhaps using naïve models for forecasting as a benchmark may be more suited so that one can justify whether the candidate model really improves forecast accuracy rather than doing horse race among the proposed candidate models.

Although there are problems in the simple linear models, for the recent stable inflation periods since mid 1980s the simplest possible benchmark models may show some advantages. For instance, as it is indicated in Greon et al (2009) short run cyclical variations in the most of the major macroeconomic indicators often possibility to be accounted for by a simple autoregressive process and they found those models hard to beat for the UK during the stable periods from 1980s until 2007 based on both rolling window and unweighted average methodologies. For these reasons, naïve univariate models provide a suitable benchmark for evaluating the forecast performance of more complex models.

There are number of good papers that propose advanced models and address main issues for inflation forecasting. While choosing the best forecast model is desirable, it has been hard to rank the best candidate models since the models perform differently in particular time periods and data sets. Perhaps one proposed candidate model improves over another while there can be others that candidate cannot beat which make comparisons difficult. It is common in the recent literature that the researchers propose a variety of models and choose them based on forecast accuracy by comparing with a benchmark model which performs relatively well most of the time so that they can justify their candidate models with higher reliability. Albeit one can choose any model as a candidate depending on what they are emphasizing, there are several common candidate models those are used most of the time; Multivariate linear autoregressive models (Hendry and Castle 2006, 2007), Vector autoregressive models (Hendry and Castle 2006, Evans and Watchel 1993, Clarks and McCracken 2006), Philips curve (Stock and Watson 2007) and Vector error correction models (Hendry and castle 2006, 2007)

[1] The inflation targeting is introduced in 1992 with level of 1%-4% and changed to 2.5% in 1995, 2.5% with a band of 1 in 1997 and 2.0% with a band of 1 in 2003.

[2] This often refers to the high volatile inflation period during 1970s.

against the benchmark Univariate autoregressive models. The question I am raising in this paper is simple-which candidate model has more accuracy against the benchmark univariate autoregression models in UK?

Keeping that question in mind, this paper performs forecasts for UK quarterly inflation using both linear univariate and vector autoregressive specifications, as these models are widely considered in empirical forecasting. This paper also examines whether using a cointegrated system may improve over those two models. Models are compared based on h-step ahead out-of-sample forecasts with the aim of finding the most appropriate specification for forecasting UK inflation.

The remainder of paper is organized as follows. Section 2 discusses key literature in inflation forecasting. Section 3 introduces data set. Methodological approaches are outlined in section 4 and an analysis of the results is made in section 5. Section 6 concludes.

2 Literature review

The literature on inflation forecasting is too large to present here comprehensively but several relevant, key works are worth noting and discussing. The discussion will be built around three main points: the most commonly used forecasting models, variables and methodologies in literatures.

2.1 Inflation forecasting models

There are a lots of disagreement with regard to which inflation model results in superior or smallest error, hence it is hard to rank the forecasting models systematically. Many inflation forecasting models tend to perform well at one point at time and poorly at others.

In macroeconomic application, linear univariate autoregressive models are often considered as naïve models which perform well most of the time. Due to its good performance and simplicity, most papers use it as a benchmark model for forecasting. Good performance of univariate model is supported by many researchers including Marcellino (2006) who studied a large number of inflation models, both linear and nonlinear, in order to make comparisons and found that linear univariate autoregressive models are robust although nonlinear models perform better at some times. Likewise, Meese and Geweke (1984) studied univariate autoregressive models with various data transformations, data periodicity, forecast horizon and seasonality etc which applied to 150 macroeconomic time series and found that simple autoregressive models with lag length selected by AIC generally perform well.

Although the autoregressive form of univariate model is extensively used in practice, in the recent literatures Integrated Moving Average (IMA) form of univariate inflation model rather than autoregressive process is preferred, at least in Stock and Watson (2007), Clarks and McCracken (2006) who studied forecasting in inflation. Stock and Watson (2007) found that the forecast accuracy of low order autoregression deteriorated since mid-1980s when the coefficients of the autoregressive process changed due to the relatively less fluctuations in inflation. Because the moving average coefficient changes inversely with the ratio of the permanent to transitory disturbance term, it increases relating to the recent stable period which cannot be approximated well in the low order autoregression as it was in 1970s. Their study was based on both rolling and recursive methods of autoregressive forecast on the split sample and the IMA (1) model was found to have the lowest mean squared forecast error, hence it is recommended. They also suggested that longer lag is better suited for the less volatile inflation period when the autoregressive form is utilized instead of IMA (1) form due to declining importance of permanent components. By reflecting on those findings, although I choose univariate autoregressive form for sake of simplicity, I perform forecast not only for the whole available sample but also for the more stable sample period from 1992 with longer autoregressive lags to check the robustness of this idea.

The larger alternative of the univariate models is multivariate vector autoregressive model (VAR) that includes other variables in the system. VAR model have been prevalent in the empirical macroeconomic forecasting since it was introduced by Sims (Clements and Hendry 2002) and it is considered as successor of univariate models. Particularly, trivariate VAR system of output, price, and interest rate is widely used in forecasting of US and UK inflation such as in Jacobson et al (2001) and Clark and McCracken (2006). However, the rising issue of instability in a small scale VAR due to the presence of structural breaks has become a central concern, as indicated in Stock and Watson (2001) and Clark and McCracken (2006). But the evidence of instability is mixed. For example, Stock and Watson (2003) have found that coefficients of a system with inflation and output gap to be stable, according to stability tests.

Concerning the issues of the instability and usefulness of small scale VAR model, a well known paper by Clark and McCracken (2006) investigated several statistical methods that may improve the forecast accuracy under the presence of a structural change, focusing on the VAR model system including US output, interest rates and prices. They examined several methods, including different approaches to lag selection, differencing inflation and interest rate, intercept correction, simple averages, Bayesian shrinkage model, and predictive least squares to improve forecast accuracy. As a result, they found that the VAR model with the differenced inflation with compared to the univariate benchmark consistently performed well in terms of the root mean squared forecast error (RMSFE) using real time data. Therefore, based on those issues of instability in VAR system and findings by researchers above, I propose to estimate a VAR system including the output gap, together with differencing methodologies, to deal with issues of system instability and forecast accuracy.

Larger scale VAR models and multivariate single equation models are widely used in the papers by Hendry and Castle (2006, 2007). They chose their variables in the VAR to allow cost-push and demand-pull inflation, and the effect of monetary and external variables to inflation, but the number of variables is restricted due to degrees of freedom (Hendry and Castle 2006). Although I agree with their claim that the increased information set may improve forecast accuracy, I will consider trivariate VAR system of output gap, short run interest rate and inflation as it is common in the rest of the literature and for simplicity. Furthermore, I will perform forecasts based on the VAR system rather than the most general model VARMA because the MA component of VARMA makes the forecasting procedure complicated since MA errors have to be estimated recursively even though the parameters of VARMA model are not assumed to be known (Clements and Hendry 2000, p.123).

However, it should be noted that those univariate autoregression and multivariate VAR models are highly dependent on time series properties of the variables such as order of integration. As a result, those models are vulnerable to structural break which has been extensively discussed in recent papers and is a non negligible issue in UK. For example, Halunga et al (2009) find that the order of integration in UK inflation has changed. According to their study, the whole sample of the UK inflation persistence shifts from I (1) to I (0) at the end of 1981. Splitting the sample at this date and further examination also yield the change in inflation in 1973 from I (0) to I (1). Since the structural break issue is an inseparable part of the inflation forecasting in UK, it is worth to discuss the findings of several influential papers in order to emphasize the importance of robust modelling of structural change, although this paper will not deal with this issue.

Hendry and Castle (2007) examined the forecast performance of UK inflation in order to emphasize the need of modeling the structural breaks in the data by comparing 15 models based on 1-step, 4-step and 8-step forecast errors of quarterly inflation. They found that the models with various transformations which robustify to structural changes have power over the models without robust using single-equation and vector equilibrium correction models that are applied to two sample periods: one which included structural change and the other was a more quiescent period. The period with structural change is 1990q4-1995q3 when the clear break occurred due to the European Exchange rate mechanism and new inflation targeting regime, and the quiescent period covers 1998q3-2003q2 when there is no clear change. The structural break robust transformations are differencing device, rapid updating tools and forecast pooling. Particularly, both under the presence of the structural break and quiescent periods, a pooled forecasting method perform better than any individual forecasting device, especially during the periods of abrupt change. Moreover, this methodology not only dominates others in the volatile periods but also performs well during the quiescent period which implies that the policy makers do not have to switch the policy model in response to a change in regime since this model performs well in both periods.

Another way that we can capture structural break is Hamilton's (1989) state dependent Markov Switching model. Krolzig (2000) studied Markov Switching vector autoregressive process in detail and found that modelling the changing nature of the economy might improve the forecast accuracy rather than considering time-invariant models and traditional model improving methods such as differencing, intercept correction, and multi step estimation. One of the advantages of using the Markov chain is that it produces meaningful forecasts prior to possible changes happening, based on expected duration of regimes which can be calculated by transitional probabilities of regime change which then enable the choice of an appropriate forecasting models (Hamilton 1989). This nonlinear model might be better than linear models since Clements and Sensier (2003), Arghyrou (2005) found evidence that UK inflation has a nonlinear pattern. However, Bessec and Bouabdallah (2005) conclude that the Markov switching models perform better than linear models but the gain over linear models is small and significant only for small horizons, based on a comparison study of models using Mean Absolute error (MAE) and Root mean square error (RMSFE).

However, these advanced models are not a concern of this paper that these could be investigated as an extension of the linear models. Instead non-trivial extensions of univariate and VAR models, forecasting in co-integration system using Vector error correction model (VECM) will be considered against models that account for structural breaks. VECM model may perform better since it mitigates the problems arising from the forecast based on VAR which usually suffer from large number of parameter estimates and hence estimation based on this model will be less precise. Hendry and Clements (1995) emphasized the importance of imposing long run constraints in the system to improve forecast accuracy; it is especially robust in the small sample size on their study of forecast performance for UK M1 money. Several other studies can be mentioned that performed forecasts in co-integrating system and found improved accuracy. For instance, the study by Hendry and Castle (2007) on simple equilibrium correction models and vector equilibrium correction models together with various transformations including differencing, pooling and rapid updating tools which try to account for structural breaks against benchmark univariate time series actually yield more robust forecasts. According to their result, univariate autoregressive models perform extremely well in stable periods but perform badly in the periods where breaks occurred due to lack of concessions to such breaks. However, VECM models with differencing method that removes a bias from the trend in co-integrating equations showed consistently better results.

2.2 General forecasting methodologies in the literature

Although there is a vast literature on methodologies to improve forecasting accuracy, I will briefly discuss the ones which are closely related to my object. Iterative and direct forecasting methods which I will describe in later section are common in the literature and accuracy may depend on choice between the two methods. Hendry and Castle (2006, 2007) conclude that if no structural change occurs, the iterative forecast performs better whereas direct forecast is preferable when there is a change. Particularly, the iterative forecast is very dependent on estimated parameter coefficients; hence it may be a weak estimator for the whole sample when there is a structural break. Unlike iterative forecast, direct forecast is preferable under presence of structural break but forecast accuracy declines in a longer horizon whereas the iterative forecast tends to converge to the long run equilibrium as the horizon increases, with the small variations across the horizons (Hendry and Castle 2007). Thus more break resilient methods should be applied to obtain reliable forecast values.

Another choice is the rolling window forecast which is the forecast based on the most recent constant number of observations that rolls forward over time and drops older observations to mitigate the effects from possible misspecification of model. This rolling window methodology is extensively used in Stock and Watson (2003, 2004) and its usefulness in econometric modelling is supported by Swanson and White (1997), as indicated in the Tashman (2000). However, this method is considered to be an extreme one, hence the less extreme version of this idea, discounted least squares which uses all observations but puts more weight on the recent observations and gradually eliminates the old observation is suggested by Stock and Watson (2003) and Evans (2006). Even though these approaches are very practical and promising to define the appropriate model for forecasting, eliminating the observations regardless of whether they have exhibited the structural change may lead a loss of useful information. However, the rolling regression method is still considered one of the more useful methods for forecasting if the series is long enough and structural change has appeared several times (Tashman 2000).

Moreover, in-sample or out-of sample measures of predictive contents are one issue we should consider here. In sample predictability methodology is useful to define whether selected variables are able to predict the variable being forecasted using full sample regression whereas the out-of-sample methodology is a real time forecast that use the data available up to forecasting period. Although choosing between them is purely based on one's own judgment, researchers generally agree that forecasting method should be out-of-sample methodology rather than assessing in goodness of fit to the past data due to several weaknesses associated with in-sample fit methodology. Tashman (2000) clearly pointed the weaknesses of in-sample fit methodology. Firstly, the in-sample method tends to understate forecast error than the real error because it over fits to the whole sample and is designed to determine the fit for historical data. However, the variation of the past history does not have to persist in the future and future variation might not be a reflection of the past. Secondly, the model and estimation that is selected by in-sample methods might not fit the post sample data. Likewise, the superiority of out-of sample methodology is supported by Clark and McCracken (2003) who concluded that out-of sample methodology is better for the finite sample size. Furthermore, most papers generally use the out of sample methodology e.g. Stock and Watson (2003, 2007), Marcellino (2006), Clark and McCracken (2003). In contrast, the better predictability of in-sample methodology over out-of sample is supported by Inoue and Kilian (2004) based on the study of asymptotic behavior of both the in-sample and out-of-sample. However, I will adopt out-of sample methodology since the out-of sample is favored most of the time, together with both direct and iterative forecasting methods.

2.3 Predictability of variables for inflation forecasting

Non-financial variables for inflation forecasting are common in UK. Hendry (2001) tested a variety of variables that may determine change in UK inflation. Those variables are unemployment as many researchers consider labor market as a primary source of inflation, excess demand for goods and services (output gap), excess money holdings according to the 'Quantity Theory', short and long run interest rate according to the Fisher theory, and unit labor cost and exchange rate. Based on 1% significance level, excess demands for goods and service, short and long interest rate, unit labor cost and exchange rate appeared to be good determinants for inflation whereas neither excess money nor unemployment rate revealed as good indicator. The predictability of the output gap for inflation in US is discussed as well in Stock and Watson (2003) and Hendry (2001), where it was found that the output gap is significantly related to inflation but it leads practical problem of estimating the gap.

Considerable amount of work has been done on inflation forecasting using financial variables, such as asset prices, for the last few decades. It has been an influential variable which is a potential useful predictor in a short and medium term inflation forecasting due to the forward looking behavior. Asset price here can be defined as interest rate spreads, stock returns, bonds, exchange rates, housing and any other financial measures of value of assets. Although which assets should be preferred is a big issue in forecasting, extensively discussed in Stock and Watson (2003), it is not discussed in this paper. Their main idea was that stock prices should be leading indicators for inflation due to three premises in macroeconomic theory: the premise that stock prices reflect the expected present discounted value of future earnings, Irwin Fisher's theory that the nominal interest rate is equal to real interest rate plus expected inflation rate and the notion that temporary high interest rate, following economic slowdown due to the monetary contraction. Hendry (2001) is one of the few researchers who considered role of asset market, specifically the foreign exchange rate, for UK inflation.

Furthermore, the money aggregates which were main predictors during the 1970s and 1980s lost their power due to the large measurement error from ongoing definition changes as new financial instruments are introduced (Stock and Watson 2003).

Based on those discussions around forecasting models, methodologies and predictability of variables, I propose to forecast UK inflation using univariate autoregressive specification, multivariate VAR specification as these are commonly used for forecasting inflation and the forecast based on cointegrated system using VECM model hoping it improves forecast accuracy. The variables in these multivariate models, output gap, short run interest rate and inflation, are selected concerning the study by Hendry (2001) on predictability of variables for UK inflation and various studies on instabilities in a VAR models and forecast accuracy. The proposed methodologies used here are h step ahead out-of-sample methodology together with both iterative and direct forecast which are applied not only to whole sample but also to the relative stable sample period since 1992 to check robustness of the discussions related to structural change, model specification, and well performance of univariate models.

3 Data

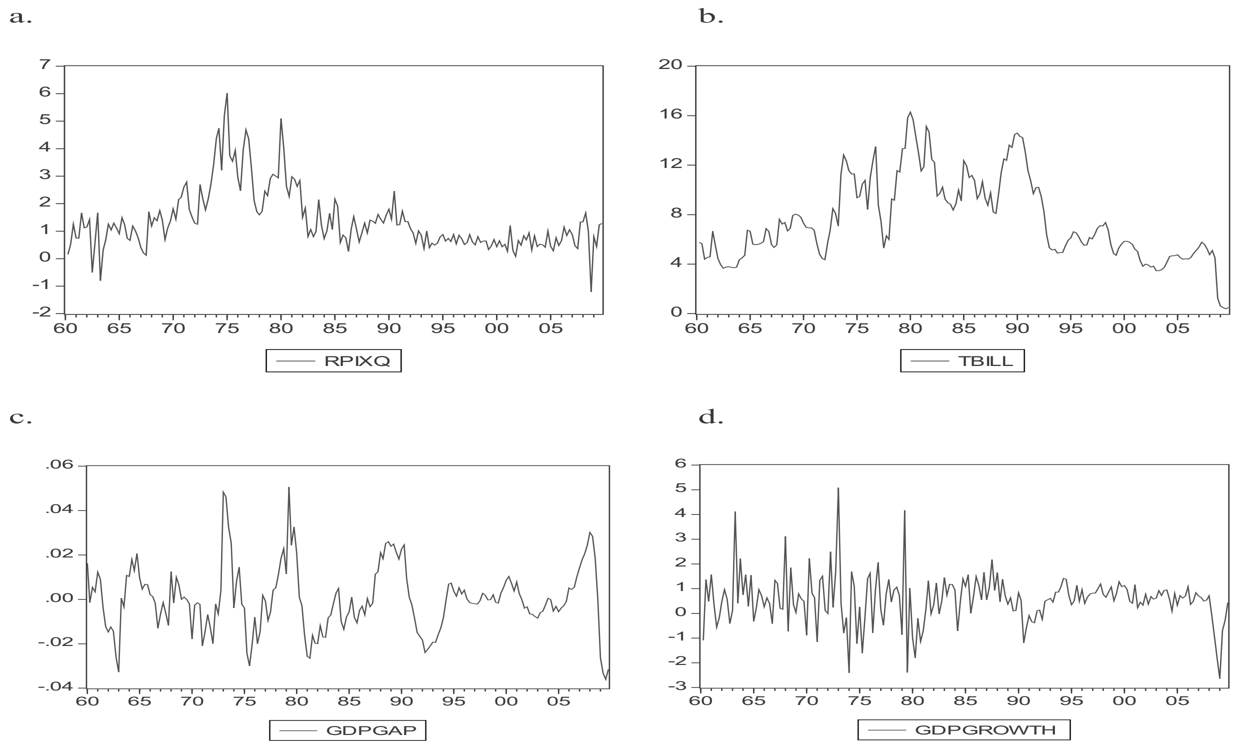
The data set used in the analysis comprises of UK quarterly retail price index for inflation, 3 months Treasury Bills for short run interest rate, and both output growth and output gap for output measure. Data were obtained from two main sources: The Office National Statistics (ONS) and the Bank of England. Retail price index and Interest rate series were available on a monthly basis and aggregated to quarterly observations in a way that the last monthly value of the quarter represents the quarterly values. For the GDP series, the quarterly aggregates were available.

Figure.1a plots UK quarterly inflation series from 1960Q1 to 2009Q4. This whole sample period is divided into two parts: the estimation sample which covers from 1960Q1 to 2004Q4 data and the forecast sample which covers from 2005Q1 to 2009Q4 including 20 observations in total. In order to construct the inflation series, RPIX[3] data which is obtained from ONS starting from 1987 and extended back to 1960 using RPI data from the same source is transformed using several transformations below. The data is shown in Figure 1a after these transformations.

Firstly, RPIX series which were indexed by six different base years originally is rebased based on the latest index year-1987 using chain index method. Then inflation rate is calculated by differencing logged quarterly values of RPIX series and multiplying by a hundred. Secondly, as it is found by Osborn and Sensier (2009) the UK inflation has highly seasonal pattern and this seasonal pattern changes over time. Thus, the seasonal adjustment is done to mitigate this concern. Seasonal adjustment for this series is implemented using X-12 filter in EViews program with default option to remove seasonal oscillations. Thirdly, after seasonal adjustment, three outliers are clearly seen in the plot; 1975Q2, 1979Q3 values and 1990Q2 value relative to the values after around 1983.These abnormal high inflation rates are related to introduction of Value added taxes and Poll taxes in UK as it is pointed in Osborn and Sensier (2009). The outliers in the inflation series were consistent with the outlier periods detected by Bataa et al (2014) who studied change in inflation for G7 countries based on monthly observations. Outliers appearing in the series were taken into account using the

[3] RPIX is a retail price index excluding mortgage interest rate which is introduced in 1975 when mortgage interest rate payment becomes a component of PRI price index. They are similar in the early years after RPIX was first introduces and differ slightly starting from 1987.

Fig. 1: Dynamics of variables



method Stock and Watson (2003) mentioned. Specifically, outliers are replaced with the median of the non-outlier values within three periods on either side of the outlier.

Although the outliers in the estimation period are adjusted, there is one more outlier left in the forecasting period that should not be mitigated in a similar way since it is closely related to forecast accuracy. This unexpected deflation (-1.2%) appears in 2008Q4 but did not affect later periods which might be considered an additive outlier type. If it is the case, according to the study by Chen and Liu (1993), the forecasts from forecasting origin near to this outlier will largely fluctuate due to the carry over effect although the point forecast is not very sensitive with this type of outlier. Therefore, forecasts values which are produced using 2008Q4 observation will distort the forecast accuracy more than the forecasts made for 2008Q4 observation. Thus the forecast accuracy results are shown with these last five observations, as well as without them to check whether the main conclusions are affected.

Figure.1b plots 3 months Treasury Bills (T-Bill) which is a short run interest rate measure in the UK. This is a quarterly data from 1960Q2 to 2009Q4 which is obtained from the Bank of England from post-1975 period and is extended back to 1960s using data obtained from ONS.

In figure 1c and 1d, output gap and output growth are sketched respectively. They are calculated using seasonally adjusted real Gross Domestic Product (GDP). The output growth series is constructed as differencing logged GDP values and multiplied by a hundred. In order to construct output gap, the difference between actual and potential output, the Hodrick-Prescott[4] filter is applied to the logged GDP to calculate potential output. The smoothing parameter λ in this filter is set equal to 1600 as it is widely agreed value for quarterly data (Ravn and Uhlig 2002).

[4] Recently, Kamper, Morley and Wong (2018) shows that Beveridge-Nelson filter produces a more intuitive and reliable estimate of the output gap. This paper can be further improved by employing Berveridge-Nelson filter.

4 Methodology

4.1 Forecasting in Univariate model

This paper focuses on autoregressive models or AR(p) process for forecasting univariate inflation, rather than more general case of autoregressive moving average models (ARMA). Once we get involved with ARMA model, nonlinear techniques has to be considered because of the unobservable behavior of the moving average process. However, there is no loss of generality by considering only AR process in the paper because one of the main issues – integratedness on forecastability - depends on the roots of AR polynomial and can be analyzed without including MA polynomial (Hendry and Clements 2000, p.79). Assume that the inflation in UK is determined by the following AR(p) process,

$$\pi_t = \mu + \sum_{i=1}^{p}(\psi_i \pi_{t-i}) + \varepsilon_t \tag{1}$$

where $\varepsilon_t \sim (0, \delta^2)$, p is a maximum lag length and π_t is a quarterly inflation rate by percentage. For the general case of an AR(p) model, the estimation will be performed using conventional OLS. The parameter estimators by OLS here are though not unbiased but they are consistent, and t and F test statistics are appropriate since our sample size is moderate enough to preventing sample size distortion.

Prior to defining the model specification, the order of integration or stationarity test should be conducted to define data generating process. For that purpose the Augmented Dickey Fuller (ADF) test for unit root is utilized. Since the forecasting methodologies are valid for the stationary variables, we made those variables stationary which were found to be non-stationary by taking differences. After transforming data into the stationary series, the order of specification is defined using both model selection criteria and 'testing down' methodologies and the forecasting based on those two models are compared. Specifically, in the model selection criteria approach, the number of lags is selected by AIC criterion as suggested by Meese and Geweke (1984) and Marcellino et al (2006). Marcellino et al (2006) emphasized the importance of the lag selection choice for forecasting and suggested that the iterative forecast is preferable than direct forecast if lag selection criterion is selected by AIC in one step ahead forecast due to small sample MSFE. Moreover, the reason AIC criterion is preferred rather than BIC criterion is that it usually allows longer lag which is better to account for serial correlations in the error terms. However, the AIC criterion is more conventional and tends to yield over specified lag length. In that concern, true lag order of inflation model in UK might be shorter, the testing down methodology is employed to test whether proposed lag orders by AIC are statistically significant. Particularly, in testing down methodology, the longest lag length is tested sequentially until the null hypothesis of zero coefficients is rejected. As part of model selection procedure, the Lagrange Multiplier test for serial correlation is performed to make sure appropriateness of the selected model. After specifying the model for UK inflation, forecast are implemented based on the estimated parameters of model.

The parameters of equation (1) are estimated and forecasted based on both Iterated and Direct multistep ahead forecasting methods together with pseudo out-of-sample methodology. Pseudo out-of –sample methodology uses only data available through date t, then uses this estimated model in order to forecast for t+h period. This process is repeated throughout a sample to produce sequence of forecasts. Therefore, to perform pseudo out-of-sample forecasting, the historical data is split into two parts: an estimation period and a forecast period. An appropriate number of forecasting periods N is appointed so that the length of forecast period is, at least, equal to forecasting horizon. In our case, the estimation period covers from 1960Q1 to 2004Q4 and the forecast period covers the period between 2005Q1 and 2009Q4 which are twenty observations in total. Moreover, the reduced forecast sample accuracy is calculated as well along with full forecast period in order to see change in accuracy.

4.1.1 Iterative forecasting method

The forecasted values are obtained differently depending on data generating process. For a stationary series or I(0) process, iterated multistep ahead forecasts are found using one step ahead forecast and iterated forward for multistep ahead forecasts. One step ahead AR model for inflation is

$$\pi_{t+1} = \mu + \sum_{i=1}^{p} \psi_i \pi_{t+1-i} + \varepsilon_{t+1} \tag{2}$$

Then the parameters in equation (1) are estimated by conventional OLS and h step ahead forecast of inflation, π_{t+h} is constructed recursively for $h = 1, \ldots,$as,

$$\hat{\pi}^{I}_{t+h|t} = \hat{\mu} + \sum_{i=1}^{p} \hat{\psi}_i \hat{\pi}_{t+h-i|t} \tag{3}$$

The model is estimated over $t = 1, \ldots, T$ and h iterated forecast are calculated for $T + 1, \ldots, T + h$ based on the equation (3). The model is re-estimated over $t = 1, \ldots, T + 1$ and new forecasts are calculated for $T + 2, \ldots, T + h + 1$. This process continues recursively until the forecasts for whole forecast horizon are obtained[5]. For the non stationary series or I(1) series, we need to adjust the forecasted values from equation (3) to obtain forecasts for level series because we transformed the non stationary data to the stationary series by differencing and h step ahead forecasts are made using that stationary series. The forecast adjustment for differenced series is made using the transformation that Marcellino et al (2006) used. The level h step ahead forecast of inflation is then it is calculated by

$$\hat{\pi}^{I}_{t+h|t} = \pi_t + \sum_{i=1}^{h} \triangle\, \hat{\pi}^{I}_{t+i|t} \tag{4}$$

where π_t is I(1) process. For the differenced series, the iterative forecasting methodology produces forecasts that are made for change in every forecast horizon using the predicted values from previous horizon. Therefore, the sum of all forecasted changes throughout proposed horizons must be added to the original series to obtain forecast values for desired horizon.

4.1.2 Direct forecasting method

In direct forecast, the parameters are also estimated by OLS regression but regressors are constant, $\pi_t, \ldots, \pi_{t-p+1}$ and only available information at time T is used. The regression model is given by,

$$\pi_{t+h} = \beta + \sum_{i=1}^{p} \rho_i \pi_{t+1-i} + \varepsilon_{t+h} \tag{5}$$

For stationary series, direct forecasts are found by using the estimated parameters of equation (5) which are estimated by OLS recursively,

$$\hat{\pi}^{D}_{t+h|t} = \hat{\beta} + \sum_{i=1}^{p} \hat{\rho}_i \pi_{t+1-i|t} \tag{6}$$

[5] The recursive methodology of model estimation is utilized in this study due to its common use in the relevant literature. However, the rolling regression which is the forecast based on the most recent constant number of observations that rolls forward over time and drops older observations is also considered to be one of the effective methods for forecasting, especially when there are large number of observations and the series is subject to multiple structural breaks (Tashman, 2000). Although this path is not pursued in this study, the performance of rolling regression can be further assessed as a robustness check to the results presented in Section 5.

The recursive estimation is processed in a same way that we discussed in the iterative forecasting method. If the series being forecasted is differenced, the appropriate transformation to obtain forecast for level series should be made on the forecast values from equation (6) as,

$$\hat{\pi}^{D}_{t+h|t} = \pi_t + \triangle_h \, \hat{\pi}^{D}_{t+h} \tag{7}$$

where π_t is I(1) process. Unlike iterative forecasting method, h step ahead direct forecasting methodology produces the forecast for total quarterly change between forecast origin and forecast horizon that must be added to the original series to derive forecast for level series.

4.1.3 Direct vs Iterative forecast

There is no straight forward answer in the literature that which method is more superior to the empirical matter. One's optimal choice between iterative and direct methodology relates two major issues. Firstly, structural breaks complicate forecasting accuracy. If structural change occurs, an iterative forecast performs badly due to the higher dependence of the estimated parameter coefficients in the model. Unlike iterative forecast, a direct forecast is preferable under presence of structural change but forecast accuracy declines at longer horizons whereas an iterative forecast tends to converge to the long run equilibrium as horizon increases where variations across horizons are small (Hendry and Castle 2007).

Secondly, the accuracy depends heavily on the model selection procedure. For example, an iterated forecast gives more efficient parameter estimates when the one period ahead forecast is correctly specified and its efficiency increases as forecasting horizon gets longer. However, a direct forecast performs better when model is misspecified or if the negative serial correlation exist in the error terms (Marcellino et al 2006, Chevillon and Hendry 2005). Particularly, on one hand, an iterated forecast methodology tends to perform better than a direct forecast if the model has longer lag specification than the true model. On the other hand, if the model is underspecified, the true autoregressive order of model exceeds the number of estimated lags for dependent variable, the direct methodology is preferable. However, in practice the true model and true data generating process is unknown (Allen and Fildes 2005), hence only way we can improve forecast accuracy is to choose closest model based on smallest forecast error.

4.2 Forecasting in multivariate VAR model

Although the forecasting procedure of VAR model is same as in univariate forecasting, building VAR models is more complicated than building univariate models due to several issues related to the variables in the system. Issues are: what variables and in which form should they be included in VAR system of inflation forecasting. Here I will discuss the case where all variables in the VAR are stationary.

The first issue, the predictability of variables, is discussed briefly in the section 2 where we proposed to utilize the VAR system with output gap (Y_t), short run interest rate (R_t) and inflation (π_t)[6]. In the recent literature, output gap is preferred more than output growth due to its good predictability (Hendry 2001) and the provision of stability in the system (Stock and Watson 2003). However, the inclusion of output growth instead of output gap in the trivariate VAR system of inflation is conventional and hence this system should be considered as well. VAR in this paper is composed of three equations: inflation as a function of lagged interest rate, output, and its own lagged variable; interest rate as a function of lagged inflation, output and its own lagged variable; similarly for output

[6] The conventional trivariate VAR system of inflation is employed to facilitate comparison of results in the literature. However, the VAR system can be extended further with inclusion of more variables.

as a function of lagged interest rate and inflation rate and its own lagged value. The output can be either output growth or output gap.

$$\begin{bmatrix} \pi_t \\ Y_t \\ R_t \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} + \sum_{j=1}^{p} \Gamma_j \begin{bmatrix} \pi_{t-j} \\ Y_{t-j} \\ R_{t-j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{\pi t} \\ \varepsilon_{yt} \\ \varepsilon_{rt} \end{bmatrix} \tag{8}$$

Disturbance term is assumed to be vector white noise which implies Ordinary Least Squares (OLS) to each equation gives consistent and efficient parameter estimates (Osborn and Becker 2007).

The second issue, the form of variables in the system, depends on data generation process of that certain variable. The output growth and output gap are generally agreed as I(0) processes. Some researchers believe that the inflation and interest rate series are stationary or I(0) processes that appear to be I(1) due to structural change. For example, under structural change, the null hypothesis of unit root may be under rejected, suggesting that a unit root exists while there is none (Maddala and Kim 1998). In this case, different form of variables systems are forecasted in order to investigate their effect on forecast accuracy. Three types of VAR is forecasted based on model building strategies suggested by Allen and Fildes (2005): the VAR system of all level variables claiming that all variables are stationary at their levels, the VAR with differenced interest rate and inflation together with level output gap based on unit root test result, and the VAR with differenced interest rate and inflation together with output growth to make comparison with output gap performance. Comparison of output gap and output growth is based on the differenced system because forecasting the VAR with all differenced variables is one way of protection against the bias from structural break (Hendry and Clements 1995). In addition to those three models, the fourth VAR model for stable inflation sample period is estimated in order to compare the results from the whole sample period.

After having taking care of these empirical issues, the model specification process is implemented. The number of lagged values to be included in each equation is determined by three different methodologies; both AIC and BIC criteria and the testing down methodology using joint lag exclusion test. AIC and testing down methodologies are performed in a same way that we discussed for the univariate model. The reason I introduced BIC criterion here is for two reasons. Firstly, the forecast accuracy might be suffered from increased number of parameters estimated in the system due to the longer lag selected by AIC whereas BIC tends to give shorter lag length. Secondly, if the maximum lag order is longer than true lag length, BIC provides consistent estimator for the true lag order and the iterative forecast with BIC selection criterion is asymptotically efficient (Marcellino et al 2006).

Unrestricted VAR parameters are estimated via OLS estimation procedure after models are selected. Using those estimated parameters, we apply forecasting techniques as discussed in the univariate model. Specifically, iterative forecasting method together with out-of-sample methodology is employed for VAR forecasting. VAR model for inflation in vector notation is defined by following

$$X_t = \sum_{j=1}^{p} \Gamma_j X_{t-j} + \delta + V_t \tag{9}$$

where X_t is 3x1 vector of time series variables, $V_t \sim IN_n[0, \Omega]$ for $t = 1, 2 \ldots, T$, $\boldsymbol{\delta}$ is 3x1 intercept matrix and $\boldsymbol{\Gamma}$ is 3x3 matrix of coefficient. Then h step ahead iterative forecast is defined as follows

$$\hat{X}_{t+h|t} = \sum_{j=1}^{p} \hat{\Gamma}_j \hat{X}_{t+h-j|t} + \hat{\delta} \tag{10}$$

If inflation in the VAR model is expressed in a differenced term, we transform the forecasted values using the equation (4) to derive the forecast for level.

4.3 Forecasting from co-integrating relationship

So far we built a model depending on whether the data generation process is stationary or non-stationary. However, several non-stationary variables may exist where the linear combination of them yields stationary variable. If this is the case, differencing every non-stationary variable to convert it to the stationary is not correct way for imposing unit roots (Hendry and Clements 2000 p.127). Engle and Yoo (1987) maintain that the forecasting performance of an unrestricted VAR model does not always give promising results and make the puzzle on inclusion of level series or differenced series immense. According to Engle and Yoo (1987), the VAR with differenced variables suffers from model misspecification because many related macroeconomic variables which trend overtime are differenced while the forecast is made using that relationship. The VAR with level variables tends to underestimate parameters near the unit circle that make forecast biased. Therefore, the VAR with level variables could be a poor forecasting model if the unit root is likely to be the case. However, forecasting based on co-integrating relationships gives more better forecast than either VAR with level and VAR with differenced as stated in Allen and Fildes (2005), Engle and Yoo (1987) and Christoffersen and Diebold (1998). The error correction which is attached with both change and level data provides the framework for estimation, forecasting and testing (Engle and Yoo 1987).

The cointegrating relationship is tested on the variables that are integrated with order one using Johansen's co-integration test which consists of two tests: Maximum eigenvalue statistics and trace statistics. Johansen and Juselius suggest that the maximum eigenvalue test may be preferable (Maddala and Kim 1998). However, in practice the Trace statistics is utilized more often than the Maximum eigenvalue because it jointly tests the eigenvalues effectively, whereas the maximum eigenvalue works progressively by testing on a single estimated eigenvalue (Osborn and Becker 2008). The forecasts are made using the results from both statistics and compared.

In order to set up this test, we consider two things. Firstly, we replace output gap variable by logged output so that every variable in the system is an I(1) process. Secondly, we impose the restriction in the system. Since we have observed linear deterministic trend at least in one variable, we allow unrestricted intercept both in co-integrating relationship and VAR, and trend in the co-integrating relation to allow for the co movements in the data series (see Osborn and Becker 2008). Lag length eight is chosen as suggested by Akaike information Criteria. Based on the number of co-integrating vectors, we perform the forecast using error correction model which is derived using equation (9).

$$\triangle X_t = \sum_{j=1}^{p} \Gamma_j \alpha(\beta^{'} \triangle X_{t-j}) + \delta + V_t \tag{11}$$

Where $\alpha\beta^{'} = \Gamma - I_n$ and α and β are nxr of rank $(r < n)$when $X_t \sim$ I(1) and r is co-integrating rank. Then the forecast values are estimated by recursively,

$$\triangle \hat{X}_{t+h|t} = \sum_{j=1}^{p} \hat{\alpha}(\hat{\beta}^{'} \hat{X}_{t+h-j|t}) + \hat{\delta} \tag{12}$$

Although the vector error correction model makes forecasts for differences according to the equation (11), E-views program results the final forecasted values for level,

$$\hat{X}_{t+h|t} = X_t + \sum_{j=1}^{h} \triangle \hat{X}_{t+j|t} \tag{13}$$

4.4 Forecast accuracy measures

Models are compared based on forecast accuracy. Accuracy of forecasts is measured in terms of the bias and efficiency of the forecasts from each forecast model which is captured by Mean Error (ME).

The means absolute error (MAE) and square root mean forecast error (RMSFE) which are closely related to ME errors are reported along with ME since they are convenient to be compared.

$$ME = \frac{\sum_{i=1}^{n} e_{i,j,t+h|t}}{n}$$

$$MAE = \frac{\sum_{i=1}^{n} |e_{i,j,t+h|t}|}{n}$$

$$RMSFE = \sqrt{\frac{\sum_{i=1}^{n} e_{i,j,t+h|t}^2}{n}}$$

If the associated forecast error is,

$$e_{j,t+h|t} = \pi_{t+h} - \hat{\pi}_{j,t+h|t}$$

Where π_{t+h} is actual inflation rate and $\pi_{j,t+h|t}$ is h step ahead forecasted values for model j, n is the number of forecasts $(i = 1, 2, .., n)$ computed for each model j over each forecast horizons.

In practice, the models compared to each other tend to be nested that make comparison difficult. As stated in Clark and West (2006), the additional variables in the larger model usually make noise that fluctuate forecasting accuracy and increase associated MSFE. They suggested adjusting MSFE of models compared to account for this noise. The statistical set up of MSFE comparison is according to Clark and West (2006):

Let's consider model 1 and 2 which are parsimonious and larger variable model respectively. In our case parsimonious models would be the univariate model that is nested in the VAR or VECM. The h step ahead forecasted values of two models are denoted by $\hat{y}_{1t,t+h}$ and $\hat{y}_{2t,t+h}$ with associated forecast errors are $y_{t+h} - \hat{y}_{1t,t+h}$ and $y_{t+h} - \hat{y}_{2t,t+h}$, then we compute following

$$\hat{f}_{t+h} = (y_{t+h} - \hat{y}_{1t,t+h})^2 - [(y_{t+h} - \hat{y}_{2t,t+h})^2 - (\hat{y}_{1t,t+h} - \hat{y}_{2t,t+h})^2]$$

The null hypothesis of equal MSFE for the two models is tested by regressing $\hat{f}_{t+h}$ on a constant and using the resulting t statistic for zero coefficients. The alternative hypothesis is that the smaller MSFE for larger variable model than parsimonious model. If the statistic is greater than $+1.28$ for one sided 0.10 or $+1.645$ for two sided 0.05 test, the alternative hypothesis is rejected. In the regression for the one step ahead forecast errors on a constant, the usual least squares standard error is used because it is consistent under the assumption that the error term is a white noise process. However, for more than one step ahead forecast, the standard errors are made robust with the Newey-West approach using the default option in E-views program to account for autocorrelation in error terms due to the overlapped forecast errors.

5 Empirical Results

5.1 Unit Root test

In order to investigate possible changes in persistence of inflation series over the period, the ADF test is employed over various sample periods. It performs a hypothesis test for the existence of a unit roots against the alternative hypothesis of either stationary or trend stationary time series data. Figure 1a indicates a trend in some periods although there is no clear overall trend observed. However, including a trend where it is not required does not asymptotically affect the test but excluding a trend that should be present results in misspecification. Therefore I performed ADF tests on level data conducted with a constant and trends and on the differenced data with only a constant, as differencing the data removes trend from the series.

The order of Integration in UK inflation has changed several times, as indicated in the study by Halunga et al (2009). According to their study, the whole sample of the UK inflation persistence shifts from I(1) to I(0) at the date 1981 m12. Splitting the sample at this date, further examination also yields that there is a change in inflation in 1973 m12 from I(0) to I(1). The whole sample period between 1960 and 2009 is split as a result of their findings and an ADF test is applied to those sub-samples to determine any a change in inflation. ADF test is also applied to the estimation period between 1960 and 2004 to define order of integration that eventually leads model specification. Table.1 represents the ADF test results.

Tab. 1: ADF test results

	Level				1st Difference				
Periods	Exogenous variables	Lag**	T-Stat	P-value*	Exogenous variables	Lag**	T-Stat	P-Value*	The Order of Integration
1960Q1-2004Q4	C&T	2	-2.6101	0.2764	C	7	-6.2953	0.0000	I (1)
1960Q1-1973Q4	C&T	0	-4.9138	0.0011	C	1	-8.4122	0.0000	I (0)
1974Q1-1981Q4	C&T	1	-3.2589	0.0926	C	1	-6.0513	0.0000	I (1)
1982Q1-2009Q4	C&T	7	-2.5248	0.3158	C	6	-4.4980	0.0004	I (1)

*MacKinnon (1996) one-sided p values, ** The number of lag length is according to the Akaika information criteria

The results show that the order of inflation in UK changed at least one time from I(1) to I(0) at 1973 as it is indicated in Halunga et al (2009). The other change in 1981m12 has not revealed in the ADF test.

Even though the ADF test is sufficient, it is not a powerful test in the presence of structural break (Maddala and Kim 1998). Specifically, under structural change unit root test, the null hypothesis may be under rejected, pointing that a unit root exists while there is none (Maddala and Kim 1998). The obvious structural change we can point from the graph is related to the inflation targeting in 1992. Therefore, ADF test alone is not very indicative. Knowing that defining the order of inflation is not negligible issue, the Phillips Peron (PP) test should be included along with the ADF test to obtain more indicative answer since it shows higher power than ADF test when we are testing a stationary and trend stationary alternative hypothesis (Maddala and Kim 1998). Table.2 represents the PP test results. PP test results indicate several changes in the UK inflation in the past which is consistent with the results found by Halunga et al (2009). These changes in inflation suggest that perhaps the inflation is not an I(1) process throughout whole sample. Furthermore, the order of integration for the modelling period between 1960Q1 to 2004Q4 has revealed different results in two unit root tests that do not give indicative answer while real data generating process is never known in the practice. Therefore, it is more sensible to consider both I(0) and I(1) processes for forecasting purposes and compare them to find the better forecast.

5.2 Univariate forecasting results

The model specification is made according to the method we discussed in the previous section but based on the order of integration: I(0) and I(1) process.

Tab. 2: The Phillips Peron test results

	Exogenous variables	T statistics	P Value*	Lag length**	The order of integration
1960Q1-2004Q4	C&T	-4.0878	0.0079	4	I (0)
1960Q1-1973Q4	C&T	-4.9632	0.0009	3	I (0)
1974Q1-1981Q4	C&T	-3.1212	0.1192	1	I (1)
1982Q1-2009Q4	C&T	-7.6753	0.0000	0	I (0)

*MacKinnon (1996) one-sided p values, ** The number of lag length is according to the Newey West.

Based on the assumption that the inflation is I(1) process, the inflation model is found as AR(7) as it is suggested by AIC criterion and AR(2) when testing down methodology is applied to the AR(7) model. AR(3) is selected for the model for I(0) inflation process. Those models completely accounted for serial correlation[7]. In addition to those three models, the model specified for stable inflation sub sample period from 1992Q1 to 2004Q4 is estimated. Based on those four models, one, two, four and eight step ahead forecasts are made using both iterative and direct forecasting methodologies. The forecast results are reported in terms of the forecast accuracy measures in Table.3 and Table.4, for direct and iterative methodologies respectively.

Tab. 3: Summary of quarterly inflation forecast error based on direct forecast method

		Models			
	Horizon	I (1)-AR(2)	I (0)-AR(3)	Sub-Sample: I(0)-AR(3)	AIC:I(1)-AR(7)
ME	1-step	0,0307	-0,0409	0,0989	0,0192
	2-step	0,0655	-0,0206	0,1083	0,0487
	4-step	0,0918	-0,0207	0,1623	0,0775
	8-step	0,0683	-0,0328	0,1565	0,0571
MAE	1-step	0,5321	0,5199	0,4743	0,5065
	2-step	0,7330	0,5216	0,5001	0,6765
	4-step	0,9326	0,5531	0,4853	0,8656
	8-step	0,6895	0,6506	0,6340	0,6615
RMSFE	1-step	0,7419	0,7208	0,6696	0,7059
	2-step	1,0004	0,7303	0,6902	0,9274
	4-step	1,1876	0,7627	0,6923	1,0919
	8-step	0,9326	0,8625	0,8333	0,8648

Inspection of Table.3 and Table.4 suggests four main findings.

Firstly, iterated AR forecast performs well compared to the direct AR forecast. Most of horizons and model, the forecast error of iterative forecast is lower than direct forecast suggesting that the iterative one step ahead forecast is preferable for both stationary and non stationary processes of UK inflation, which is consistent with the finding by Chevillion and Hendry (2004). Specially, a better performance

[7] Lagrange multiplier test for serial correlation is conducted but not included in the paper to conserve space. It is available upon request from the author.

Tab. 4: Summary of iterative quarterly inflation forecast error

	Horizon	Models I (1)-AR(2)	I (0)-AR(3)	Sub Sample: I(0)-AR(3)	AIC:I(1)-AR(7)
	1-step	0,0307	-0,0409	0,0989	0,0192
ME	2-step	0,0302	-0,0566	0,0856	0,0545
	4-step	0,0510	-0,1073	0,1651	0,0649
	8-step	0,1039	-0,1861	0,1865	0,1317
	1-step	0,5321	0,5199	0,4743	0,5065
MAE	2-step	0,5432	0,5177	0,4834	0,5118
	4-step	0,6359	0,5365	0,4739	0,5548
	8-step	0,5326	0,4926	0,5622	0,5391
	1-step	0,7419	0,7208	0,6696	0,7059
RMSFE	2-step	0,7976	0,7791	0,6725	0,7417
	4-step	0,8426	0,7820	0,6249	0,7404
	8-step	0,7750	0,7602	0,7233	0,7656

of iterative process can be observed in MAE and RMSFE accuracy measure that consistently deliver lower error at every horizons comparing to direct forecast methodology. However, this observation is not robust in the ME accuracy measure. As we mentioned in the previous section, the direct forecasting methodology is supposed to perform better than iterative forecast due to the fact that the several structural breaks have occurred in the past. Perhaps, the efficiency of the iterative forecast is outweighing the robustness of the direct forecast to any structural break. Furthermore, well specified models with no serial correlation in the error term give good results although the bias from structural break in the parameter estimates exists.

Secondly, the forecast errors in model estimated over sub sample period are consistently low compared to the whole sample errors, which are valid for both in iterative and direct forecasts. It can be explained in two ways. First, estimating periods over sub sample reduces the probabilities of break occurring leading reduction of forecast bias substantially. Particularly, lower sample errors in iterative forecast than in direct forecast supports the claim that iterative forecast performs well under no structural change. Second, forecasts based on sub sample estimation period with reduced volatility increases forecast accuracy because the probability of forecasted values strays away from actual values is small.

Thirdly, forecasts based on models with longer lags showed good performance. It is clearly seen in the iterative forecast. For example, RMSFE error is smaller in AR (3) model than AR (2) and AR (7) model improves over these two models. This observation is consistent with the claim by Marcellino et al (2006) who compared a direct and iterated multi step AR methods for macroeconomic time series and found a good performance of the iterative forecast with higher order autoregression, especially the lag length is selected by AIC criterion. The approval of lag selection by AIC criterion is stated in the paper by Meese and Geweke (1984) as well. Although we are not concerned with moving average component, the better performance of AIC criterion model can be justified by better accounting for moving average component that may exist in the whole sample. Longer order AR process relatively well approximates the invertible moving average process in the population rather than shorter order AR process if there is any. Therefore, a longer AR model is preferable for forecasting if we suspect moving average component in the inflation process as it is indicated in the Stock and Watson (2007). Evidently, if we disregard the performance of sub sample model, the best performance comes from iterative forecast with AIC criterion.

Fourthly, most of the models show increasing forecast errors as forecast horizon increases at least at short horizons. It can be caused either way: on the one hand, it may be due to the accumulation of innovation errors and the fall in predictability when the economy moves in a stable way and model is specified correctly (Hendry and Castle 2006). This explanation fits quite well both in direct and iterative forecasts for the stable inflation sub sample period. For example, increased forecast errors in longer horizon during stable periods might be showing decline in predictability. On the other hand, for forecast using full estimation period, it might be due to the structural break in the inflation series where accuracy of direct forecast supposed to decreases as forecast horizon increases and large forecasting bias arises in iterative forecast. However, the decline in forecast accuracy of iterative and direct forecast is observed in one, two and four step ahead forecast horizon but not consistently in eight steps a head forecast errors. For instance, the eight step ahead forecast errors declined both in AR(2) and AR(3) models in iterative forecast and the same happened in AR(2) and AR(7) models in direct forecast which is difficult to explain since the pattern is not consistent. Moreover, the preference on either I(0) or I(1) process for forecasting is not explicit because they are nested each other in iterative and direct forecast.

Probably, one of the causes of fluctuations in forecast accuracy is the outlier in 2008Q4 and its carry over effect to the next periods' forecast. In that sense, the last five observations are excluded from the sample and the forecast accuracy is measured again. The forecast accuracy sharply improves, RMSFE errors decline up to fifty percent when the last five observations are excluded. However, the general patterns and findings remain similar[8].

5.3 Multivariate VAR forecasting results

Three types of VAR systems depending on order of integration plus one more VAR system based on sub sample estimation period, same as univariate model, are utilized used for forecasting using an iterative forecasting methodology. Both whole and reduced forecast period accuracy are calculated and compared.

Model 1: The VAR system of output gap, level interest rate and inflation appears to have a lag length of seven according to the AIC and one according to the BIC criterion. The lag exclusion Wald statistics[9] for joint significance of endogenous variables at both lag length is performed and these appear to be significant.

Model 2: The model specification for the VAR system of output gap, differenced inflation rate and interest rate, is selected by AIC and BIC criteria. A lag length of eight according to the AIC and zero according to the BIC criterion are chosen. The inclusion of all eight lags is supported by the lag exclusion Wald test statistics.

Model 3: Same lag is specified for the VAR system which includes output growth instead of output gap and differenced inflation rate, interest rate according to the AIC criterion, but lag length one is selected by BIC criterion.

Model 4: the model 4 represented here is model 1 which applies for sub sample between 1992Q1 and 2004Q4. Both AIC and BIC criteria gives lag length 2 and it appears significant in lag exclusion Wald test. All models completely account for serial correlation[10] up to 12 quarters. The iterative forecast results based on AIC criterion and BIC criterion are summarized in Table.5 and Table.6 respectively.

The forecast based on the full estimation sample are generally gives similar results to those found in the univariate models although they could not outperform univariate models. To summarize, iterative forecast with the longer lag suggested by AIC criterion performs better in model 2 and

[8] The detailed results are available upon request from the author.
[9] The detailed results are available upon request from the author.
[10] The detailed results are available upon request from the author.

Tab. 5: Summary of forecast errors for VAR models based on AIC criterion

	Horizon	VAR-Models based on AIC criterion Model-1	Model-2	Model-3	Model-4
ME	1-step	-0,0481	0,1204	0,0357	0,2513
	2-step	-0,0970	0,1861	0,0276	0,2283
	4-step	-0,4216	0,1378	-0,1217	0,1909
	8-step	-0,4560	0,1485	-0,0524	0,2086
MAE	1-step	0,5196	0,5847	0,5730	0,5692
	2-step	0,6320	0,6348	0,6357	0,5603
	4-step	0,6279	0,5475	0,6045	0,5244
	8-step	0,5485	0,5833	0,5379	0,5727
RMSFE	1-step	0,8082	0,8154	0,8205	0,7561
	2-step	0,9599	0,9263	0,9511	0,7654
	4-step	0,9603	0,7925	0,9030	0,6802
	8-step	0,8924	0,7953	0,8288	0,7305

model 3 whereas model 1 with BIC criterion improved over those models although the improvement is marginal. Moreover, forecasts on model 4 show highest accuracy compared to other three models in terms of MAE and RMSFE which is consistent with univariate results although the ME errors for model 4 is unusually high. However, if we disregard better stable inflation period sample performance, the best model is VAR with level variables by BIC criterion. As we expected, the relative performance of output gap is better than output growth but only marginally. Finally, we have seen that the forecast accuracy declines as forecast horizon increases in univariate models at least at short horizons (up to 4 step ahead forecast). However, this conclusion cannot be made in any of VAR models whereas we expected to see this pattern at least during the stable inflation period when the model specification is made with higher accuracy.

Although the univariate models generally perform well especially during the stable period inflation, they are nested in the VAR for some models. Perhaps, the bad performance of VAR model is due to the noise from additional variables in the system that increase associated forecast error. Therefore, we adjusted forecast error according to the methodology proposed in section 4. The results are given in the Table.7 where we see whether forecast accuracy in univariate model really is better. The adjusted forecast errors of univariate and multivariate VAR models based on same order of integration are tested and compared. For example based the assumption that inflation is I(0) process, the model 1 in VAR forecast are tested against the iterative AR(3) univariate model (I(0) models). For assumption that inflation is I(1), the model 2 is tested against AR(7) (I(1) models). Additionally, the sub sample model results in univariate are tested against the model 4 (sub sample models). Using resulting statistics, we cannot conclude that the VAR models outperform univariate that is hidden under noise since none of the test results can reject the null hypothesis of equal RMSFE of models. In order to reject null hypothesis, the estimated values should be sufficiently large.

Perhaps disappointing performance of VAR model is due to the outlier in forecasting period. When the forecast sample reduced, the conclusions are changed massively. Table.8 and Table.9 represent the forecast accuracy for reduced forecast sample period which lead following analysis. Firstly, multivariate VAR models outperform the univariate models as we hoped. The improved sets of information do yield better predictability in the system. Secondly, the forecast accuracy in model 4 found to have less accuracy than the forecast accuracy estimated on whole sample period which contradicts with what we expected and found previously. However it may be justified by reduced

Tab. 6: Summary of forecast errors for VAR models based on BIC criterion

		VAR-Models based on BIC criterion			
	Horizon	Model-1	Model-2	Model-3	Model-4
ME	1-step	-0,0105	0,0109	0,0330	0,2513
	2-step	-0,0659	0,0498	0,0089	0,2283
	4-step	-0,2750	0,0617	-0,1117	0,1909
	8-step	-0,4283	0,0444	-0,1227	0,2086
MAE	1-step	0,5169	0,5356	0,5788	0,5692
	2-step	0,5763	0,6277	0,6996	0,5603
	4-step	0,5897	0,7801	0,7790	0,5244
	8-step	0,5663	0,5011	0,5985	0,5727
RMSFE	1-step	0,7525	0,7870	0,8247	0,7561
	2-step	0,8548	0,9086	0,9985	0,7654
	4-step	0,9075	1,0587	1,1235	0,6802
	8-step	0,8502	0,7745	0,8463	0,7305

Tab. 7: Test statistics for adjusted RMSFE

	T-statistics		
Forecast horizon	I(0) Models	I(1) Models	Sub sample Models
1step	-0,740978	-2,044051	0.285412
2step	-1.174138	-1,789537	-1.029622
4step	-0.518303	-2,067206	-2.042052
8step	-0.200532	-1,949697	-0.862147

Rejection value equals greater than $+1.28$ for one sided 0.10, $+1.65$ for two sided 0.05.

impact of output gap on the inflation dynamics relating to the stable inflation regime due to inflation targeting policy. Hendry and Castle (2007) found a small coefficient of the output gap during 1990s sample compared to the sample up to 1990s. Indeed when inflation is relatively low in 1990s, output gap suggests excessive demand driving inflation upward in a system which increases the forecasting bias. Possibly, the difficulty of measuring the output gap may lead the forecast error as well, as indicated in Hendry and Castle (2007). However, better performance of output gap over output growth remained the same though the improvement is small. Thirdly, the model 1 performed better over both longer and shorter horizons compared to the both model 2 and model 3. The reason that differenced VAR model cannot perform well is perhaps related to the model misspecification caused by differencing method that remove trend in the series. Once trend is removed, the relationship between variables that is useful for forecasting in a VAR is cut. If the root of the autocorrelation is close to unity, then the level forecast could have performed badly. However, good performance of model 1 with BIC criterion is evident both before and after forecast sample reduction suggesting that the inflation and interest rate variables may be stationary at level that made them appear to be I(1) processes due to the structural change. Fourthly, the BIC criterion generally leads higher forecast accuracy except in model 2. That may be pointing to noise coming from longer lag models by AIC criterion. Many parameter estimations in longer lag model arises instability in the system and tend to have a poor forecast. Moreover, the efficiency in BIC might suggest overspecified lag length although the forecast is made the assumption that the model is specified correctly. Finally, the forecast accuracy increases in a short horizons but moves not in systematic way in the mid and long horizons suggesting the economic instability or unanticipated shift in the data that fluctuates forecast accuracy up and down.

Based on those comparison statistics, we can conclude that the VAR models perform better than the benchmark univariate autoregressive models due to the improved set of information. Specially, the best VAR models selected by BIC criterion, model 1, improve over forecast accuracy of every single univariate models in terms of RMSFE accuracy measures.

Tab. 8: Reduced forecast sample errors For VAR models based on AIC criterion

		VAR-Models based on AIC criterion			
	Horizon	Model-1	Model-2	Model-3	Model-4
ME	1-step	-0,0844	-0,0353	0,0398	0,1550
	2-step	-0,0676	0,0204	0,1259	0,1913
	4-step	-0,1049	0,1189	0,2278	0,2328
	8-step	-0,0823	0,3798	0,4466	0,2471
MAE	1-step	0,2267	0,3813	0,3846	0,2680
	2-step	0,2638	0,3428	0,3631	0,2891
	4-step	0,2372	0,3435	0,3546	0,2956
	8-step	0,1748	0,3873	0,4466	0,3041
RMSFE	1-step	0,3257	0,4363	0,4360	0,3854
	2-step	0,3968	0,4121	0,4522	0,4120
	4-step	0,3466	0,4147	0,4446	0,4232
	8-step	0,2737	0,4936	0,5573	0,4518

Tab. 9: Reduced forecast sample errors For VAR models based on BIC criterion

		VAR-Models based on BIC criterion			
	Horizon	Model-1	Model-2	Model-3	Model-4
ME	1-step	-0,0579	-0,0047	-0,0460	0,1550
	2-step	-0,0740	0,0784	-0,0051	0,1913
	4-step	-0,0654	0,2231	0,1134	0,2328
	8-step	-0,0533	0,2838	0,3013	0,2471
MAE	1-step	0,2385	0,3542	0,3575	0,2680
	2-step	0,2368	0,4228	0,4081	0,2891
	4-step	0,2132	0,5118	0,4069	0,2956
	8-step	0,1913	0,3885	0,3916	0,3041
RMSFE	1-step	0,3058	0,4282	0,4333	0,3854
	2-step	0,3501	0,5129	0,4812	0,4120
	4-step	0,3203	0,5923	0,4069	0,4232
	8-step	0,2888	0,4684	0,4411	0,4518

5.4 Forecast results in co-integrating relationship

The VAR system of logged real GDP, inflation and interest rate which are known as I(1) processes according to the unit root test over the whole sample period is forecasted using the co-integrating relationship among them. The possible co-integrating relationship is tested based on Johansen's co-integration test using E-Views program with the option D (unrestricted intercept and trend in co-integrating equation and intercept in VAR). The lag length eight (seven for differenced endogenous) is selected by AIC and one (zero for differenced endogenous) is selected by BIC. The co-integration test with eight lag did not yield any co-integrating relationship[11] whereas test with one lag yield one co-integrating equation based on both Trace and Maximum eigenvalue statistics. The co-integration test result is presented in Table.10. Based on one co-integrating relationship among the variables, the iterative forecasting is produced using vector equilibrium correction model (VECM) and forecast accuracy is calculated on both forecasting period before and after last five observations are excluded. The results are given in Table.11 in terms of the forecast error.

Tab. 10: Johansen's co-integration test results based on BIC criterion

Hypothesized number of CEs	Trace/max-eigenvalue statistics	Critical value (0.05)	Probability**
Co-integration rank test (Trace)			
None *	70.79174	42.91525	0.0000
At most 1	25.29554	25.87211	0.0588
At most 2	7.457594	12.51798	0.2992
Co-integration rank test (Max-eigenvalue)			
None *	45.49620	25.82321	0.0000
At most 1	17.83795	19.38704	0.0828
At most 2	7.457594	12.51798	0.2992

* denotes rejection of the hypothesis at the 0.05 level, **MacKinnon-Haug-Michelis (1999) p-values

Since we found one co-integrating relationship among the variables, restricting co-integrating vector in Error correction model is supposed to give better forecast than either model 1 or model 2 (model 3). The expectation was valid based on comparison of forecast error between VAR models and VECM for full forecast period. It improves over most of VAR models except the forecast based on model 4. However, it was weak forecast comparing to the univariate models. It could not beat any of the univariate iterative forecasts.

However, inspecting reduced forecast sample error indicated that the VECM's improvement over VAR models is less robust. Although, reduced forecast sample delivered a good performance that was better than the models 2 and model 3 at short horizons, it could not outperform the most accurate VAR model, model 1. Similarly, it performs better than some of univariate models at short horizons but it could not outperform the best univariate model- iterative forecast with AIC criterion. Therefore, the forecast based on VECM model using cointegrated relationship is not as strong model as expected.

[11] Co-integration test results based on AIC criterion is available from the author upon request.

Tab. 11: Summary of forecast error for models based on co-integrating relationship

	Horizon	full forecast period	reduced forecast period
ME	1-step	0,1228	0,0725
	2-step	0,2212	0,1733
	4-step	0,2937	0,2944
	8-step	0,3679	0,3771
MAE	1-step	0,5106	0,2474
	2-step	0,5946	0,296
	4-step	0,637	0,353
	8-step	0,6503	0,3784
RMSFE	1-step	0,7313	0,3432
	2-step	0,8103	0,414
	4-step	0,8211	0,4857
	8-step	0,8035	0,5714

6 Conclusion

The aim of this empirical work was to investigate the forecasting models for the UK's inflation which improve over the benchmark univariate autoregressive model. For this purpose two candidate models: VAR and VECM are forecasted together with the benckmark univariate autoregressive model. The VAR system of inflation, output gap, and short run interest rate is employed as it is commonly used for inflation forecasting. Due to the improved information set, it is expected to outperform the univariate benchmark. The forecast based on VECM using the cointegrated relationship among the variables in the VAR is performed due to the possibility of several non-stationary variables thus the linear combination of them yields a stationary variable and we hoped that by imposing a long run relationship the forecast accuracy can be improved.

The model specification is made using AIC criterion along with the testing down methodology for the univariate autoregressive models whereas for the VAR and VECM models, both AIC and BIC criterion are employed together with lag exclusion test. These model specification methods are applied to both I(0) and I(1) inflation processes and compared, as unit root tests could not give an indicative answer as to whether the inflation series was stationary at their level or non-stationary. Although there are many methodologies available for forecasting, we used direct and iterative forecast together with out-of-sample forecasting methodology.

Due to the outlier in the forecasting period, sudden deflation in 2008Q4, the forecast made for this observation and its carry over effect to years after this observation fluctuate forecast accuracy considerably. To capture the effect of this outlier observation the forecast accuracy is calculated both including and excluding the last five observations of forecast period and the results are compared. To summarize them, general findings are as follows. Firstly, the iterative forecast performs well compared to the direct forecast based on RMSFE in univariate autoregressive models. Iterative forecasts with longer lags as suggested by AIC criterion was found to be an especially good forecast model which is consistent with results from the study by Marcellino et al (2006), Meese and Geweke (1984) and suggestions that Stock and Watson (2007) made related to the moving average components. The robustness of this result remained same in reduced forecast periods. Therefore, the iterative forecast applied to the model which was selected by the AIC criterion,was found to be the best in the univariate case. The performance of the direct forecast in univariate models is incomparable to the VAR models since it can be applied only to univariate autoregressive models.

The best performing model in the VAR is the VAR with level variables by the BIC criterion although the BIC criterion did not improve over AIC selection criterion for all models. Compared to the benchmark univariate model, the best VAR models could not improve forecast accuracy and it is nested in univariate models when accuracy is calculated for full forecast sample period whereas it is found opposite in reduced forecast sample. According to the results in reduced forecast sample, the improved information set does yield better forecasts. Based on the suspicion that VAR models make noise that make them nested with univariate models, where they actually improved the univariate, the adjusted MSFEs of both models are compared according to the Clark and West (2006). However, the result was not thrilling.

Another contradicting result from full forecast period and reduced forecast period is the performance of the stable period sub sample inflation in the estimation period. Although the full forecasting sample suggested that reduced volatility increases predictability and forecast accuracy, it is not supported in the reduced forecast sample which is the case both in univariate and VAR models. Several explanations were suggested in section 4.

The preference on the order of integration is not very evident. For VAR models, systems with level variables generally perform well as opposed to the differenced variable system which is supported by both results from full and reduced forecast periods and the AIC and BIC criteria at least over short horizons. In univariate benchmark models, it is not conclusive as to whether the I(0) or I(1) is preferred. We can only say longer lags gives better accuracy.

As we found one cointegrating relationship among the variables, imposing this long run relationship in the system, we expected to see improved forecast accuracy according to the assumptions we made throughout this paper. The result is mixed depending on whether the forecast period includes the last five observations. When the full forecast sample's accuracy is measured the VECM yield better forecasts than both VAR with level and differenced variables, although it could not outperform univariate benchmark. In contrast, reduced forecast sample accuracy indicated that the forecast based on the cointegrated system only can outshine VAR with differenced variables in short horizons but is less accurate than the best VAR model with level variables by BIC criterion. Similarly, it could not improve over the best univariate model by AIC criterion although it produced better forecasts compared to some of the other univariate models. Therefore, the forecast based on VECM model was not as strong as expected.

It should be noted that the forecast accuracy is heavily dependent on the outlier observation in the forecasting period as well as the model selection procedure and forecast methodologies. Additionally, it may be possible that extending forecast observations or periods may improve forecast accuracy since 20 observations in total is not sufficient. Therefore, further extensions should be done to improve forecast accuracy.

References

Allen, P.G., and Fildes, R. (2005) Levels, Differences and ECMs-Principles for Improved Econometric Forecasting. *Oxford Bulletin of Economics and Statistics* 67, Supplement

Arghyrou , M., Martin, C., and Milas, C. (2004) Non-linear inflationary Dynamics: Evidence form the UK. *Oxford Economic Papers* 57, pp 61-69

Barnett, A., Groen, J.J., and Mumtaz, H. (2010) Time Varing Inflation Expectation and Economic Fluctuations in the UK: A structural VAR analysis. Bank of England Working Paper No. 392

Bataa, E., Osborn, R.D., Sensier, M., and Dijk, D. (2014) Identifying Changes in Mean, Seasonality, Persistence and Volatility for G7 and Euro Area Inflation. *Oxford Bulletin of Economics and Statistics* 76 (3), 360-388

Benati, L. (2004) Evolving Post-World War II U.K. Economic Performance. *Journal of Money, Credit and Banking* 36, (4) pp 691-717

Bessec, M., and Bouabdallah, O. (2005) What Causes the Forecasting Failure of Markov Switching Models? A Monte Carlo Study. *Studies in Nonlinear Dynamics and Econometrics* 9, (2) article 6

Chen, C., and Liu, L.M. (1993) Forecasting Time Series with Outliers. *Journal of Forecasting* 12,(1) pp 13-35.

Chevellion, G., and Hendry, D.F. (2005) Non-parametric Direct Multi-step estimation for forecasting economic process. *International Journal of Forecasting* 21, pp 201-208

Christoffersen, P.F., and Diebold, F.X., (1998) Co-integration and Long-Horizon Forecasting. *Journal of Business and Economic Statistics* 16, (4) pp 450-458

Clark, T.E., and West, K.D. (2006) Approximately Normal Tests for Equal Predictive Accuracy in Nested models. *Journal of Econometrics* 138, pp 291-311

Clark, T.E., and McCracken, M.W. (2006) Forecasting with Small Macroeconomic VARs in the Presence of Instabilities. Unpublished Booklet: Federal Reserve Board, Washington D.C

Clark, T.E., and McCracken, M.W. (2006) Averaging Forecasts from VARs with Uncertain Instabilities. Unpublished Booklet: The Federal Reserve Bank of Kansas City.

Clements, P.M., and Hendry, D.F. (1995) Macro-Economic Forecasting and Modelling. *The Economic Journal* 105 (431) pp.1001-1013

Clements, M. P., and Sensier, M. (2003) Asymmetric Output-gap Effects in Phillips Curve and Mark-up Pricing Models: Evidence for the US and the UK. *Scottish Journal Of Political Economy* 50, (4) pp 359-374

Clements, M.P., and Hendry, D.F. (2002) A Companion to Economic Forecasting. Oxford: Blackwell

Clements, M.P., and Hendry, D.F. (2000) Forecasting Economic Time series. Cambridge: Cambridge University Press

Engle, F.E., and Yoo, B.F. (1987) Forecasting and Testing in Co-integrated systems. *The Journal of Econometrics* 35, pp143-159

Evans, M., and Watchel, P. (1993) Inflation Regimes and the Sources of Inflation Uncertainty. *Journal of Money, Credit and Banking* 25, (3)

Groen, J.J., Kapetanios, G., and Price, S. (2009) Real Time Evaluation of Inflation Report Forecast for UK. Unpublished booklet: Bank of England etc

Kamber, G., Morley, J., Wong, B. (2018) Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter. *The Review of Economics and Statistics*, 100 (3): 550–566.

Halunga, A.G., Osborn, R.D., and Sensier, M. (2009) Changes in the Order of Integration of US and UK Inflation. *Economics Letters* 102, pp30-32

Hamilton, J. (1989) A New Approach to Econometric Analysis of Non-stationary Time Series and the Business Cycle. *Econometrica* 57, (2) pp 357-384

Hendry, D.F., (2001) Modelling UK inflation, 1875-1991. *Journal of Applied Econometrics* 16, 255-275

Hendry, D. F., and Castle, J.L. (2007) Forecasting UK inflation: the Roles of Structural breaks and Time Dissaggregation. Unpublished Booklet: Oxford University.

Hendry, D.F., and Castle, J.L. (2006) Forecating UK Inflation: Empirical Evidence on Robust Forecasting Device.

Inoue, A., and Kilian, L. (2004) In –sample or Out-of –sample predictability: Which one should be use?. *Econometric Review* 23, (4) pp371-402

Jacobson, T., Jansson, P., Vredin. A„ and Warne, A. (2001) Monetary Policy Analysis and Inflation Targeting in a Small Open Economy. *Journal of Applied Econometrics* 16, pp 487-520

Krolzig, H.M. (2000) Predicting Markov Switching Vector Autoregressive Process. Unpublished Booklet: Nuffield College Oxford.

Maddala, G.S., and Kim, I.M. (1998) Unit Root, Co-integration and Structural change. Cambridge University Press.

Marcellino, M., Stock, J.H., and Watson, M.W. (2006) A Comparison of Direct and Iterated Multi-step AR methods for Forecasting Macroeconomic Time Series. *Journal of Econometrics 135*, (1-2) pp 499-526

Meese, R., and Geweke, J. (1984) A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomic Times Series. *Journal of Business and Economic Statistics* 2, (3) pp 191-200

Osborn, R.D., and Becker, R. (2007, 2008) Co-integration [lecture note] delivered for module ECON60521 Time Series Analysis at University of Manchester.

Osborn, R.D., and Becker, R. (2005, 2006,2007) VAR processes [lecture note] delivered for module ECON60521 Time Series Analysis at University of Manchester.

Osborn, R.D., and Sensier, M. (2009) UK Inflation: Persistence, Seasonality and Monetary Policy. *Scottish*

Journal of Political Economy 56, 24-44.

Ravn, M.O., and Uhlig, H. (2002) On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations. *The Review of Economics and Statistics* 84, (2) pp 371-376

Stock, J.H., and Watson, M.W. (2001) Vector Autoregressions. *Journal of Economic Perspectives* 15, (4) pp101-115

Stock, J.H., and Watson, M.W. (2003) Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature* 41, (3) pp 788-829

Stock, J.H., and Watson, M.W. (2004) Combination Forecasts of Output Growth in a Seven-Country Data Set. *Journal of Forecasting* 23, pp405-430

Stock, J.H., and Watson, M.W. (2007) Why has US Inflation Become Harder to Forecast. *Journal Of Money, Credit and Banking* 39, pp3-33

Tashman, L. (2000) Out-of-sample Test of Forecast Accuracy: An analysis and Review. *International Journal of Forecasting* 16, pp437-450