FORECASTING INFLATION – AN EMPIRICAL STUDY ON PREDICTIVE POWER OF VARIOUS TIME-SERIES MODELS

by

Peng Zhao

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Economics

Charlotte

2015

Approved by

____
Dr. Craig A. Depken, II

____
Dr. Azhal Iqbal

____
Dr. Matthew R. Metzgar

ABSTRACT

PENG ZHAO. An empirical study on the power of different models for predicting quarterly inflation rate. (Under the direction of DR. CRAIG A. DEPKEN, II)

In this thesis, I investigate the background and causes of major inflation in recent history and empirically study forecasting of future quarterly inflation rates for three typical countries and regions: the United States, the United Kingdom, and the Eurozone. In particular, I empirically investigate the predictive power of four commonly used econometric models: the AR model, the ADL model, the ARIMA model, and the VAR model.

I compare each model's forecasting accuracy by calculating the corresponding RMSFE (Root Mean Squared Forecast Error) of pseudo-out-of-sample forecasting for each country or region. The model that exhibits the smallest RMSFE is my preferred model.

The results suggest that, for each country or region in my dataset, the ARIMA model significantly outperforms the other three models. By determining ARIMA as the most preferable model, I use the ARIMA model to forecast the future (two-year ahead) quarterly inflation rate.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Forecasting inflation is crucial and complicated, and different methods have been studied by economists. The objective of this thesis is to construct four commonly used time-series models, study and compare the predictive power of these models, and then forecast future inflation for the years 2015 and 2016.

Before discussing forecasting inflation, it is important to have a general idea of what inflation is and why it is important. Inflation is defined as a sustained increase in the price level of goods and services within an economy. Why do we care about inflation? Households and consumers care about inflation because, as a general point of view held by many, they believe that prices of goods and services rise as a result of inflation (i.e. lowering purchasing power), and consequently, their standards of living suffer. Investors care about inflation because they believe that inflation will erode their market returns and gains. Governments and economists care about inflation because inflation is tightly related to the overall performance of the entire economy. Inflation and its forecasting have been one of the major concerns of both governments and economists.

What determines fluctuations in inflation, and what is the difference between inflation today and inflation in history? Economic theory posits that there are three basic factors that determine inflation: inflation expectation, output gap (whether actual output is above or below potential output), and supply shocks.

Two basic theories about inflation expectations are rational expectations and adaptive expectations (see Ball, 2008). Rational expectations refers to the hypothesis that expectations are the best possible forecasts based on all public information and news.

For example, if a central bank is going to carry out a new monetary policy, the public believes that this new plan will be carried out and adjust their inflation expectations. Adaptive expectations refers to the assumption that expected inflation next period is equal to the inflation of the previous period.

According to some economists, inflation expectations play an important role in explaining why inflation in recent years has exhibited less variability, becoming more stable than it used to be. For example, Mishkin (2007) argues that in the United States inflation rose in the 1970s during the Great Inflation and has since declined to a much lower level. An intuitive way of thinking about this rise and fall in inflation persistence is that it resulted from an un-anchoring of trend inflation during the period of the Great Inflation, and a re-anchoring in recent years. Economists believe that this is because the fact that the Federal Reserve policy has succeeded in better anchoring inflation expectations.

There are two versions of the Phillips Curve. One version of the Phillips curve captures the short-run relationship between inflation and output: inflation is equal to expected inflation if output is at potential, and higher output raises inflation. This is also known as the Output Phillips Curve (Ball, 2008).

Supply shocks refer to sudden events that cause changes in firms' production costs which in turn result in a change in the rate of inflation. It is also tightly correlated to the fluctuations of inflation.

Throughout recent history, the fluctuations of inflation have changed over time. Now we take a glance at the historic inflations of United States, United Kingdom, and

Eurozone separately, and study why they move in a particular direction during one period of time, and then move in a different direction during another

The U.S. experienced severe and costly inflation during the 1970's when Richard Nixon was the president. After World War II, the United States financed massive aid programs in Europe, known as the European Recovery Program (or Marshall Plan) (see Silvia, Iqbal, House, and Nelson, 2014). As a result, the United States was able to export more commodities to Europe and enhance its profits abroad while other European countries that suffered from the War made use of funds from the United States to rebuild infrastructure.

However, after years of oversupplied money growth, problems arose between the late 1960s and early 1970s. As the current account deficit increased in the United States, the Bretton Woods system suddenly started to look vulnerable—a run on gold was almost inevitable as confidence in the oversupplied dollar on the global economy plummeted (see Silvia, Iqbal, House, and Nelson, 2014). In addition, what is even worse in this period is that oil shocks struck the U.S. economy. As a result, the inflation of the United States increased dramatically to double-digits. Many people blamed the increase in inflation during this period to an increase in oil prices. It is true that as one of the three major factors that affect inflation, supply shocks play a role. The oil crisis in the 1970s is regarded as an important type of supply shock. However, the Federal Reserve's expansionary policy is another important reason.

During the 1970s, not only did inflation greatly increase, but unemployment increased as well. That is why this period is also known as stagflation. Throughout U.S.

history, the supply shocks' effects on overall inflation have become smaller in recent years. During the 2000's, as oil prices steadily increased, the inflation rate fluctuated only between 2% and 4% per annum. We can see that, as a major type of supply shock, oil price does not affect inflation as it once did.

For the United Kingdom, through 1970 to 2014 we see a pattern that is to some extent similar to that of the United States. That is, during the 1970s and 1980s, the inflation was high, and then, the variability of inflation started to become smaller. In addition, due to the global financial crisis and the consequent recession, the inflation during 2008 and 2009 plummeted.

However, the explanation to the once high inflation of United Kingdom is not exactly the same. During the period from 1970 to 1973, the United Kingdom experienced a period of rapid economic growth. This period is known as the Barber boom when Anthony Barber was the Chancellor of the Exchequer. By 1973, as we can see in Figure 8, inflation in the UK exceeded 20%.

There are several reasons to explain why this happened. One important reason is a combination of the inflationary budget (or an expansionary fiscal policy) delivered by Barber in 1972 and the 1973 oil crisis that followed the Yom Kippur War. A fiscal policy is defined as a government's choices of taxes and spending. Barber's expansionary fiscal policy (such as a tax cut) raised people's after-tax incomes, and consequently, a higher public demand and consumption. In addition, a growth of credit in the 1970s also contributed to higher inflation.

For the Eurozone, since it was not established until the late 1990s, we inevitably have fewer observations. Overall speaking, the Euro system seems to have experienced less fluctuation than the United States and the United Kingdom except for the inflation plummet most likely due to global financial crisis during the year of 2008 and 2009.

Significant inflation fluctuations such as those in the 1970s for either the United States or the United Kingdom did not happen in the Eurozone. One important reason was inflation targeting which brought unprecedented transparency and reliability in monetary policy. As stated by Svensson (2000): "During the 1990s an increasing number of central banks have adopted inflation targeting, which due to its logical and transparent design and apparent success so far has become a focus of interest and a natural frame of reference."

Svensson (2000) also points out that "Inflation targeting is characterized by, first, an explicit numerical inflation target. The inflation target is pursued in the medium run, with due concern for avoiding real instability, for instance, in the output-gap; that is, inflation targeting is "flexible" rather than "strict". Second, due to the unavoidable lags in the effects of instruments on inflation, the decision framework is in practice "inflation-forecast targeting."

Though the theoretical explanations for how the inflation rate fluctuates is complicated, both the public and governments are interested in forecasting inflation. In this thesis, I focus on evaluating the performance of various commonly used time-series models (definitions and detailed information of these models are summarized in a later section) in forecasting the future quarterly inflation rate. In particular, I choose AR (1) as the benchmark model, and I also include the ARIMA model, the ADL model, and the VAR model. The key idea behind the variables included in these models is the traditional Phillips Curve which I will explain in detail in a later section.

This thesis is organized into five sections: I. Introduction, II. Inflation Forecasting – a Literature Review, III. Econometric Methodology, IV. Data and Variables, IV. Results, V. Conclusion and Discussion. The Introduction section generally describes the background and causes of inflation in recent years, and in addition, my topic. In the Literature Review section I carefully study and examine economic studies and summarize their contributions on inflation and its forecasting. In section of Data and Variables, I explain my dataset, variables, and how I use them in my analysis. In the Econometric Methodology section, I explain in detail my econometric methodology steps. In the Results and Interpretations section, I show my empirical results and explain how I interpret these results. In the Conclusion and Discussion section, I draw my conclusions and give my comments on them.

INFLATION FORECASTING – A LITERATURE REVIEW

In recent years, two periods of inflation in United States are special. One is the period of the 1970s, which is also known as the Great Inflation. The CPI once reached a double-digit of approximately 15% per annum. The other is the inflation plummet between 2008 and 2009 when a financial crisis happened. Several studies address the history and causes of the Great Inflation and the inflation during financial crisis.

Meltzer (2005) conducts a thorough analysis of the Great Inflation. He believes that political decision-making played an important role in the form of inflation, and he argues that continuation of inflation depends on political choices, analytic errors, and the entrenched belief that inflation will continue. According to Meltzer (2005), economists' explanations of the Great Inflation mainly fall into three groups: theoretical errors, misinformation, and a neglect of money growth as an important factor of inflation.

Economists are also very interested in the relationship between financial crises and inflation dynamics. Figures at the end of this thesis show the historic inflation rate and forecasted two-year ahead inflation for the United States, the United Kingdom, and the Eurozone. If focusing on the historic inflation of U.S. only, it is not difficult to find that during the year of 2008 and 2009, when there was global financial crisis, the inflation rate experienced an obvious plummet.

During the period of global financial crises, economists try to explain why the inflation rate falls dramatically during deep recession. Gilchrist, Schoenle, Sim, and

Zakrajsek (2012) investigate the effect of financial conditions on price-setting behavior. The key findings of their research, according to their empirical analysis, are that when a financial crisis is severe, firms with relatively higher liquidity ratios in the non-durable sector (weak balance sheets) tend to increase prices by approximately 20 percentage points, while firms with relatively lower liquidity ratios in the non-durable sector (strong balance sheets) tend to lower their prices by only about 10 percentage points. No substantial inflation differential for high and low liquidity ratio firms is detected in the durable goods. Their model with financial distortions suggests that during the 2008-2009 financial crisis, there is a substantial attenuation of price dynamics relative to a model without financial distortions.

To forecast inflation rate is difficult, and many previous studies have been dedicated to finding a better model. In forecasting the inflation rate, a lot of research has been performed to compare predictive power of various models. One famous theory about the inflation rate is the Phillips Curve which suggests that there is a historical inverse relationship between the inflation rate and the unemployment rate. The Phillips Curve is widely used in forecasting inflation. The Phillips Curve in different forms has been an important component of various macroeconomic models for many years. However, economists hold different opinions as to whether the Phillips Curve is useful in forecasting inflation.

Blinder (1997) points out that the empirical Phillips curve has worked amazingly well for decades and it should have a prominent place in a core model. In addition, he also argues that Phillips Curves should continue to play such an important role since these curves summarize empirical relationships critical for policymaking.

Another paper (see Rumler and Valderrama, 2008) that supports the Phillips Curve as an indicator of inflation and empirically compares New Keynesian Phillips Curve with a simple autoregressive model and both Bayesian and conventional VAR models. Their findings indicate that for longer horizons (more than 3 months, according to the authors), New Keynesian Phillips Curve delivers relatively more accurate forecasts of inflation in Austria compared to the other three time series models, however, for very short forecast horizons, the New Keynesian Phillips Curve is outperformed by the other time series models.

On the other hand, some economists do not agree with the idea that Phillips Curve is a useful indicator in forecasting inflation. Stock and Watson (2008) suggest that the performance of the Phillips curve forecasts is episodic: there are times, such as the late 1990s, when Phillips curve forecasts improved upon univariate forecasts, but there are other times (such as the mid-1990s) when a forecaster would have been better off using a univariate forecast. Some economists believe that Phillips curve fails to explain the increased variance of inflation during 1970's.

Atkeson and Ohanian (2000) point out that a historical Phillips Curve should change as the economic environment changes. They empirically find that during the period between 1959 and 1969 there exists an obvious downward sloping relationship between unemployment and inflation. However, after 1970 (until 1999 in their dataset), the negative correlated correlation disappears.

Atkeson and Ohanian (2000) also argue that the view in Blinder (1997) is mistaken.

The findings indicate that during the period between mid-1980s and 1990s, economists did not find a version of the Phillips curve that made more accurate inflation forecasts than those from a naive model that presumed inflation over the next four quarters would be equal to inflation over the last four quarters.

Russell and Banerjee, (2007) find that the effectiveness of the Phillips Curve is related to the mean rates of inflation. A key finding of their research indicates that the trade-off between inflation and the rate of unemployment (negative relationship between inflation and unemployment) in the short-run worsens as the mean rate of inflation increases. That is, as the mean rates of inflation increases, the trend line becomes steeper, which suggests that the short-run negative relationship between inflation and unemployment diminishes. This means that when the mean rate of inflation is low, a flat short-run Phillips Curve suggests that the temptation for the monetary authority to carry out expansionary macroeconomic policy is great.

Another famous study on changes in inflation is Mishkin (2007). Several key stylized facts about the dynamics of inflation are given by Mishkin. First, inflation persistence has declined; second, the Phillips curve has flattened; and third, inflation has become less responsive to other shocks. Mishkin believes that this change should be considered as an anchoring of inflation expectations as a result of better monetary policy. Inflation is more likely to fluctuate around a trend level that is determined by where long-run expectations have settled. For forecasting inflation, this means that determining where inflation expectations may be anchored should be the first priority.

The second finding about the dynamics of inflation is that the slope of the Phillips

Curve has flattened since the 1980s. Said differently, the inflation rate is now becoming less responsive to the unemployment gap. Mishkin interprets this phenomenon as something that has both its plus side and negative side. For the plus side, this means that an overheating economy is likely to generate a smaller increase in inflation. On the negative side, it is more costly to compensate for a particular increase in inflation.

Third, as previous mentioned, inflation has become less responsive to other shocks. The oil crises in 1970s caused high volatility of inflation while in recent years inflation is less responsive to oil price fluctuations.

As mentioned previously, economists observe that the fluctuations of inflation tend to become more stable. Said differently, inflation exhibits a smaller variability. It is believed that this is due to inflation expectations. Why does this happen? Economists believe that this is because of inflation expectations. There are some interesting findings as to the form of inflation expectations.

For example, one study (see Blanchflower and MacCoille, 2009) uses data from surveys taken in the UK to examine how individuals form their expectations of future inflation. They point out that the MPC (Monetary Policy Committee) at the Bank of England has an explicit mandate to maintain CPI at a target of 2%, and thus, central banks have an incentive to understand how inflation expectations are formed. They empirically find that inflation expectations are positively correlated with age and education level. Specifically, they find evidence that inflation expectations rise with age, but the more highly educated and home owners are more likely to have lower inflation expectations. In addition, those who are well-educated are more likely to be accurate in their estimates of official inflation one-year ahead, and have less backward-looking expectations.

As we can see, economists have uncovered a lot of interesting findings as to the inflation before and recently. Now, I will discuss in detail how I empirically study forecasting inflation.

ECONOMETRIC METHODOLOGY

Stationarity and DF-GLS unit root test:

The first step is to examine whether the series is stationary or nonstationary. Before testing whether these series are stationary, we need to understand what a stationary (or non-stationary) process is, and why it is important. A stationary stochastic process is defined as a process whose mean and variance are constant over time and the value of the covariance between the two time periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed. Said differently, if a time series is stationary, its mean, variance and autocovariance remain the same no matter at what point they are measured, i.e., they are time invariant (see Gujarati and Porter, 2008)). Similarly, a time series is non-stationary if its mean is time variant i.e. change over time or if its variance is time variant, or both.

The next question is why do we care if a series is stationary? Does stationarity affect forecasting? The answer is yes. Stock and Watson (2007) point out that we assume the dependent variable and regressors are stationary. Our conventional hypothesis test (t, F, and R-square), and models such as AR, ADL, ARIMA, and VAR assume the series data is stationary. If we have a constant mean and variance, we have a constant distribution and we can calculate confidence intervals. If a time series is not stationary, then the mean and variance change over time. Thus the conventional hypothesis tests, confidence intervals and forecasts are not reliable. Therefore, it is important to select the appropriate method to test stationarity.

There are two major types of non-stationarity: trend and break. Here the focus is on trend. Trend can be sub-grouped as deterministic trend and stochastic trend. A deterministic trend is defined as a nonrandom function of time. A typical example would be linear in time. To deal with deterministic trend, detrending (regression on time) is an effective solution.

A stochastic trend is random and varies over time. For example, a series might exhibit a sustained period of increase, and then a sustained period of decrease. A typical example of a series with stochastic trend is random walk. To make a nonstationary series with stochastic trend stationary, transform the series from its level form to its 1^{st} (or higher) order of difference form. If a series with a stochastic trend is stationary after taking the 1^{st} order of difference, then this series is integrated of order 1 (This will be discussed in ARIMA model specifications).

Overall speaking, there are four types of nonstationary process (one that is only related with deterministic trend, and the other three that are related with stochastic trend). The four processes can be mathematically written as follows:

1. Deterministic trend: $Y_t = \beta_0 + \beta_1 * t + u_t$;
2. Pure random walk: $Y_t = Y_{t-1} + u_t$;
3. Random walk with drift: $Y_t = \beta_0 + Y_{t-1} + u_t$;
4. Random walk with drift and deterministic trend: $Y_t = \beta_0 + \beta_1 * t + Y_{t-1} + u_t$;

where Y_t is the dependent variable, t is time, u_t is the white noise error term.

Most of the time, what we face is a stochastic trend under non-stationarity. That is to say, situations described in equation 2 to equation 4. For pure random walk (equation

2), the mean is constant, but the variance is not (this can be done by simply taking expectation and variance on both sides, and some classic linear regression models assumptions). For random walk with drift (equation 3), neither the mean nor variance is constant (this can also be similarly proved). That is to say, based on the definition of stationary, random walk, no matter with or without drift, is nonstationary. For random walk with drift and deterministic trend (equation 4), it contains two components: random walk with drift and a deterministic trend, and it is nonstationary as well.

One thing to notice is that for the random walk models mentioned above, an interesting characteristic of them is that the coefficient of Y_{t-1} is 1 (this is known as a "unit root"). Said differently, if the coefficient of Y_t is 1 (random walk), then we face the problem of a "unit root" (i.e. a problem of non-stationarity). This is why random walk, non-stationarity and unit root can be considered as the same meaning in this case.

There are various methods to test stationarity (e.g. DF test, ADF test), however, I use the DF-GLS test because it is both appropriate and the most powerful test. Before I proceed to discuss DF-GLS test, I will discuss some of the earlier non-stationarity test. So we can have a better understanding why I choose DF-GLS test.

The original Dickey-Fuller (DF) test: 5. $Y_t = \beta_0 + \beta_1 * Y_{t-1} + u_t$

The null hypothesis is that H_0: $\beta_1 = 1$ i.e. there is a unit root (non-stationarity), while the alternative hypothesis is that H_1: $\beta_1 < 1$ i.e. there is no unit root (stationarity). Subtracting Y_{t-1} from both sides yields a modified version of DF test which is written as follows:

The modified Dickey-Fuller (DF) test: 6. $\Delta Y_t = \beta_0 + \delta * Y_{t-1} + u_t$

where $\delta = (\beta_1 - 1)$ in equation 5

The null hypothesis is that H_0: $\delta = 0$ i.e. there is a unit root (non-stationarity), while the alternative hypothesis is that H_1: $\delta < 0$ i.e. there is no unit root (stationarity).

The limitation of the DF-test is that it applies only to AR (1) model. As a matter of fact, in many cases, a higher order of autoregressive model is more appropriate. Thus, a more generalized version of DF-test is developed. It is known as Augmented Dickey-Fuller (ADF) test:

The Augmented Dickey-Fuller (ADF) test: 7. $\Delta Y_t = \beta_0 + \delta * Y_{t-1} + \theta_1 * Y_{t-2} + \ldots + \theta_p * Y_{t-p} + u_t,$ the lag length can be chosen by information criterion i.e. AIC or BIC. Similarly, the null hypothesis is that H_0: $\delta = 0$ i.e. there is a unit root (non-stationarity), while the alternative hypothesis is that H_1: $\delta < 0$ i.e. there is no unit root (stationarity).

When we evaluate a statistical test, two concepts are very important. One is the size of a test, and the other is the power of a test. The size of a test is defined as the probability that the test incorrectly rejects the null hypothesis when it is true, while the power of a test is defined as the probability that the test correctly rejects the null hypothesis when the alternative is true (see Stock and Watson, 2006).

One major issue with DF test or ADF test is the power of these tests. That is to say, the probability that tests such as DF test or ADF test correctly rejects the null hypothesis that there is a unit root is low. To put it simply, these tests tend to report a unit root when actually there is none. For example, based on the DF test and ADF test described above, suppose $\beta_1 \approx 1$, (the coefficient of $Y_{t-1,}$ or alternatively $\delta \approx 0$), it is likely that these tests report that the series is nonstationary i.e. there is a unit root. This is why I choose a test that has a higher power of test (see Gujarati and Porter, 2008).

Specifically, I apply the DF-GLS unit root test to check whether the data are stationary. Like I mentioned above, the reason why I apply DF-GLS rather than earlier unit root test such as ADF test is that DF-GLS exhibits higher power than the ADF test, thus the DF-GLS test is more likely to reject the null hypothesis of a unit root, i.e., a more powerful test is better able to distinguish between a unit AR root and a root that is large but less than 1(see Stock and Watson, 2006). The null hypothesis is $H_{0:}$ the series has a random walk trend, and the alternative hypothesis is $H_{1:}$ the series is stationary around a linear time trend.

Before I proceed to explain my empirical models, I will first make a brief summary and comparison of the different types of models. In model base forecasting, we have two extremes. One extreme is that all models have no additional predictors and they only use lags of the dependent variable. Another extreme is that they have hundreds of variables. Table 1 summarizes these two extremes including different forecasting techniques.

In the last row of Table 1 (AR and ARIMA), these are models based on pure statistical theory and have only dependent variables and their lags. These models are used for short term forecasts. The macro model based on pure economic theory and has hundreds of variables, but it needs out of sample values. The ADL model has one dependent variable, but we can have two or more predictors and it based on pure economic theory. In exponential smoothing, we add time trend, season, and cycle. The VAR model is used for long term forecasts, it has two or more dependent variables and right hand side variables. Therefore, the methods discussed from row 1 to row 2 are for long-run forecasting, while methods discussed from row 3 to row 5 are for short-run forecasting.

Now that I have explained characteristics of various forecasting methods, I can now proceed to introduce models that I am going to use in detail.

AR (1) Model:

The first question is what is an autoregressive process? Stock and Watson explain "a regression model that relates a time series variable to its past values" (Stock and Watson, 2008). In particular, the 1^{st} order autoregressive model (the AR (1) model) is a special case of the p^{th} order autoregressive model. Specifically, an AR (1) model represents Y_t as a linear function of only 1 of its lagged values. Mathematically, it can be written as follows:

8. AR (1): $Y_t = \beta_0 + \beta_1 * Y_{t-1} + u_t$

where Y_t is the dependent variable, Y_{t-1} is the lag one of dependent variable, u_t is the error term, and in addition, $E\ (u_t \mid Y_{t-1}) = 0$.

ADL (p, q) Model:

The differences between an autoregressive model and an autoregressive distributed lag model is that the ADL model adds additional lags of predictors, and it is not just the lag of dependent variable. The autoregressive distributed lag model with p lags of Y_t and q lags of X_t is denoted as ADL(p, q) model. An ADL model contains lags of the dependent variable (the autoregressive component) and a distributed lag of a single additional predictor. The reason why the ADL model includes additional predictors is because additional regressors may help improve forecast accuracy. Mathematically, an ADL (p, q) model can be written as follows:

9. ADL (p, q): $Y_t = \beta_0 + \beta_1 * Y_{t-1} + \ldots + \beta_p * Y_{t-p} + \delta_0 * X_t + \delta_1 * X_{t-1} + \ldots + \delta_q * X_{t-q} + u_t$

Where Y_t is the dependent variable, Y_{t-1} ... $Y_{t-p,}$ $X_{t\ldots}X_{t-q}$ are regressors, u_t is the error term. In addition, $E (u_t \mid Y_{t-1,} Y_{t-2},\ldots, X_{t-1}, \ldots X_{t-2},\ldots) = 0$.

Different from the order of the benchmark model of AR (1), the order of ADL (p, q) model is not pre-determined. That is to say, we need to determine the order of p and q before we can proceed to forecast with this model.

For the AR (p), VAR (p) and ADL (p, q) models, we need to select correct lag order although I set the autoregressive model with only lag one as benchmark. I will discuss several different approaches in selecting p and q, but the best way is the information criterion which I will discuss in detail later.

Stock and Watson (2008) point out that when we select p and q, we need to balance the marginal benefit of including more lags with the marginal cost of uncertainty. The marginal benefit is a better fit and more reliable result, and the marginal cost is uncertainty and large forecast error. For example, if we have only one lag, we may have missed some potential information that may be helpful and the result is not reliable. If the lag is too high, we will have a large forecast error or the issue of collinearity.

The first approach is to use R-square and adjusted R-square. That is to say, selecting the model with highest R-square. However, I am not going to use R-square because R-square will increase when I add more variables.

The second approach is the F-statistic. The idea is to I start a model with several lags and perform an hypothesis test to check whether the last lag is statistically different from 0. For example, I start with an AR (6) model and test whether the coefficient on the 6th lag is significant at the 5% level. If not, I drop it and start to test the 5th lag. The problem is that we may produce a model with too many lags.

In practice, to select the proper lag order, an information criteria is required. In econometric theory, there are two commonly used information criteria. The first is the Bayes Information Criterion (BIC), also known as the Schwarz Information Criterion (SIC). Mathematically, BIC is defined as follows:

10. BIC (K) = ln (SSR (K)/T) + K*(lnT/T)

where K is the number of coefficients in the regression functions (including the intercept term) that minimizes BIC (K), SSR (K) is the sum of squared residuals, T is the number of observations, and K*(lnT/T) is the penalty factor.

Another information criteria is called the Akaike Information Criterion (AIC). Mathematically, the AIC is written as follows:

11. AIC (K) = ln (SSR (K)/T) + K*(2/T).

Similarly, K is the number of coefficients in the regression (including the intercept term) that minimizes AIC (K), SSR (K) is the sum of squared residuals, T is the number of observations, and K*(2/T) is the penalty factor.

When we add more lags, the sum of residuals will decrease (Stock and Watson, 2007). In contrast, the penalty factor will increase when we add more lags. Therefore, the BIC/AIC trades off these two factors, and the order is given by the smallest BIC/AIC value.

In practice, when we try to determine the order of p and q, the results given by BIC and AIC are not always consistent. Now this raises a question, which one should we choose? The order given by BIC or the one given by AIC? Why?

When trying to determine the order of p and q, BIC is more reliable than AIC, thus we pick the value chosen by BIC rather than the one chosen by AIC. The reason is that if we take a look the formula of BIC and AIC, we can see that the difference lies in the second term of their formulas. According to Stock and Watson (2006), the second term in the AIC is not large enough to ensure that the correct order is chosen, even in large samples, so the AIC estimator of is not consistent. As a result, in large samples, the AIC will overestimate the order with nonzero probability.

ARIMA (p, d, q) Model:

There are three components in an autoregressive integrated moving average ARIMA (p, d, q) process. The first component is simply a p^{th} autoregressive process which I have previously mentioned. The second component is the integrated part. If the series is stationary without taking any difference form, then this series is considered as integrated of order 0 (i.e. d = 0), and if the series is stationary when taking 1^{st} difference form, then this series is integrated of order 1 (i.e. d = 1), and so on. The third component is the moving average process. As a matter of fact, the AR process is not the only mechanism that may have generated the dependent variable of Y_t. A moving process is defined as a linear combination of white noise error term. Mathematically:

12. $Y_t = \delta_0 * u_t + \delta_1 * u_{t-1} + \ldots + \delta_q * u_{t-q}$,

where u_t is the error term, q is the order of the error term.

Combing these three components, we have the ARIMA (p, d, q) model

13. ARIMA (p, d, q): $Y_t = \theta + \beta_1 * Y_{t-1} + \ldots + \beta_p * Y_{t-p} + \delta_0 * u_t + \delta_1 * u_{t-1} + \ldots + \delta_q * u_{t-q}$

where Y_t is integrated of order d.

The first step to build an ARIMA model is to check stationarity. I apply the DF-GLS test, since it is the most powerful test. We only look at single mean case for DF-GLS test. If the p-value in single mean case is less than .05, we reject the null hypothesis test and conclude the time series is stationary.

The value of p and q cannot be pre-determined. In addition, BIC or AIC do not work neither. Theoretically, one classical method of determining p and q within ARIMA model is the Box-Jenkins (BJ) Methodology. In particular, determining p and q based on autocorrelation function (ACF) and partial autocorrelation function (PACF). Since the patterns of ACF and PACF of actual data is not clear to identify whether they follow a typical pattern, this method is not very effective. Just as Gujarati and Porter (2008) point out: "In practice we do not observe the theoretical ACFs and PACFs and rely on their sample counterparts, the estimated ACFS and PACFs will not match exactly their theoretical counterparts." In particular, to find the value of p, q, and d for ARIMA model, I use the SCAN (Smallest Canonical Correlation) method.

VAR (p) Model:

Vector Autoregressive Model (VAR(p)) model: suppose there are a total of k variables (including dependent variable), then a VAR model is defined as a set of k time series regressions, in which the regressors are lagged values of all k series. If there are k variables in total, then there are k regression functions. For simplicity, in the case of only two variables, a VAR (p) model can be mathematically written as:

13. $Y_t = \beta_{10} + \beta_{11} * Y_{t-1} + \ldots + \beta_{1p} * Y_{t-p} + \delta_{11} * X_{t-1} + \ldots + \delta_{1p} * X_{t-p} + u_{1t}$

14. $X_t = \beta_{20} + \beta_{21} * Y_{t-1} + \ldots + \beta_{2p} * Y_{t-p} + \delta_{21} * X_{t-1} + \ldots + \delta_{2p} * X_{t-p} + u_{2t}$

where Y_t is the dependent variable, X_t is the independent variable, u_t is the error term.

Similarly with ADL (p, q) model, the order of VAR model is determined by BIC Some advantages of VAR models are: first, we do not need to worry about which variables are endogenous and which variables are exogenous; all variables within VAR model are endogenous. Second, VAR models outperforms many complicated simultaneous-equation models (see Gujarati and Porter, 2008)

Now that we have explained the four models, the next question is: which model should we choose to forecast the two-year ahead inflation? Is there a criteria to determine which model is the most preferable one?

Pseudo out-of-sample forecasting and RMSFE:

In particular, I use pseudo out-of-sample forecasting and RMSFE to determine which model outperforms the other three models. Before I proceed, I explain what are pseudo out-of-sample forecasting and RMSFE and why they are important.

Pseudo out-of-sample forecasting is simply simulating real-time performance of a forecasting model. The purpose of pseudo out-of-sample forecasting is to evaluate the performance of a model. Take the U.S. inflation forecasting that I am going to perform later as an example: The actual sample of U.S inflation at hand ranges from 1970 to 2014, and what I do is simply pretend that the data starts in 1970 and then ends at the end of 2004. Thus, the period between 2005 and 2014 can now be regarded as (pseudo) out-of-sample (even they are actually in sample). By doing that, I use the sample data from 1970 to 2005 with my models to (pseudo) forecast the inflation during the period of 2005 to 2014. The reason for doing this is that I have actual inflation data during 2005 to 2014, thus I can evaluate the forecasting accuracy of each model during that period and find out which model performs best (most preferred model).

Now that we have a test to evaluate the performance of each model, the next question is: what is criteria of determining which model is best? Is there a criteria of evaluating forecasting accuracy? The answer is also yes – RMSFE. Root Mean Squared Forecast Error (RMSFE) is defined as the size of the forecast error, that is, of the magnitude of a typical mistake made using a forecasting model. (Stock and Watson, 2006). There are two sources of RMSFE error, one is from the error term because it is unknown and the other is from the estimated coefficient of $\hat{\beta}_0$ and $\hat{\beta}_1$. Mathematically, the formula of RMSFE is written as follows:

$$15.\ \text{RMSFE} = \{E\,[(Y_{t+1} - \hat{Y}_{t+1|t})^2]\}^{1/2}$$

Where Y_{t+1} is the actual value at period t+1, $\hat{Y}_{t+1|t}$ is the forecasting value at period t+1. Obviously, $(Y_{t+1} - \hat{Y}_{t+1|t})$ is merely the squared forecast error. That is to say, RMSFE is simply the root mean of the average of squared forecast error. Because RMSFE measures the magnitude of forecasting error, the model that yields the smallest RMSFE should be the most preferable model which can then be used to perform (true) out-of-sample forecasting of the next two years.

DATA AND VARIABLES

The raw data were collected from various sources. Detailed information as to the sources of each variable is summarized in tables at the end of this section. All the variables in the dataset were transformed from level form to a year over year form. The range of the data for the United States and the United Kingdom is from 1970 to 2014.

As mentioned previously in the literature section, the historical Phillips Curve should change as the economic environment changes (see Atkeson and Ohanian, 2000). The negative relationship between inflation and unemployment is not constant. After 1970, the negative relationship between unemployment and inflation gradually disappears. Thus, I include the period after 1970 while excluding the period before 1970.

For the Eurozone, the available data are from 1996 to 2014 because the Eurozone was established in the late 1990s. Table 2 through table 4 show the detailed information on variables and data sources.

In this section, I explain in detail why I include the chosen variables in my models and how the data I use are transformed or calculated. Real GDP (real Gross Domestic Product) represents the total aggregate output of an economy adjusted for inflation. The relationship between output and inflation is captured by output Phillips Curve.

According to the output Phillips Curve, when an economy enters a boom, a typical firm produces more than usual. As production increases, the firm's marginal cost increases as well as it was reaching productive capacity. At the same time, unemployment falls when an economy is in a boom, which makes workers more aggressive in pushing for wage increases. Large wage increases further raise marginal costs. To maximize its profits, a firm would inevitably adjust prices with the increase in marginal cost. That is to say, as a general rule, an economic boom would result in an increase in inflation since firms would raise their prices to adjust for an increase in marginal cost and a decrease in unemployment. On the other hand, a reverse result would come from an economic recession.

The second variable that I include is the unemployment rate. The unemployment Phillips Curve suggests that inflation and unemployment are negatively correlated. However, this inverse relationship starts to break down when employment gets very low, or near full employment. Extremely low unemployment rates have proven to be more costly than valuable because an economy operating at or near full employment will cause two important things to happen. First, aggregate demand for goods and services will increase faster than supply, which causes prices to rise. Second, firms will have to raise wages as a result of the workers' pushing up the wage rate. Eventually, this increase will be passed on to consumers in the form of higher prices as the company looks to maximize profits.

Over time, the growth in GDP causes inflation, and inflation further causes even hyperinflation. Once this process is in place, it can quickly become a vicious circle that keeps repeating. This is because in a world where inflation is increasing, people will spend more money because they know that it will be less valuable in the future (See Ball 2008)). This causes further increases in GDP in the short term, bringing about further price increases. As a matter of fact, these are lessons that most advanced countries in the world have learned through their past experience. In the U.S., we must go back more than 30 years to find a prolonged period of stagflation, which was only remedied by going through a painful period of high unemployment and lost production due to severe recession.

The stock market index and the Treasury bond yield are also tightly related to inflation. For stock market investors, annual growth in the GDP is important. If overall economic output is declining or merely holding steady, most companies will not be able to increase their profits, which is the primary driving force of stock performance. However, too much nominal GDP growth is also dangerous, as it will most likely result in an increase in inflation, which erodes stock market gains by making money (and future profits) less valuable.

For treasury yield, a lower interest rate stimulates investment which is an important component of economic growth. That is to say, a lower interest rate tends to stimulate inflation as well because inflation and economic growth are usually positive correlated. The key idea behind choosing unemployment rate and real GDP is the Phillips Curve. An output Phillips Curve captures the positive relationship between inflation and output, while an unemployment Phillips Curve captures the negative relationship between inflation and unemployment. In addition, government Treasury bond (interest rate) is inverse correlated with real GDP, thus, inverse correlated with inflation.

In the data used here, the inflation rate of the United States is measured with the PCE (Personal Consumption Expenditure) deflator, while the inflation rate of both the United Kingdom and the Eurozone is measured with changes in the CPI (Consumption Price Index). Although both the PCE and the CPI are effective and commonly used indicators of inflation, there are several differences between them. The most important reason why I choose the PCE deflator rather than changes in the CPI as a measurement of the inflation rate for the United States is that, according to (Moyer, 2006), for the PCE deflator, compositions of expenditures change from quarter to quarter. For the CPI index, however, the composition of the market basket remains fixed. In addition, PCE deflator is a United States-wide indicator of the average increase in prices for all domestic personal consumption.

The raw data for the PCE deflator is benchmarked to a base of 2009 = 100. The data is then transformed to a year over year basis. To illustrate, suppose I need to calculate the quarterly value of the PCE deflator for the first quarter of 1980, then:

$$16.\ PCE_{1980q1} = (PCE_{1980q1} – PCE_{1979q1})/PCE_{1979q1}$$

where PCE_{1980q1} is the PCE for the year of 1980 quarter 1, and the PCE_{1979q1} is the PCE for the year of 1979 quarter 1. CPI index and real GDP are calculated from their corresponding raw data in a similar way. That is, they are transformed from their level forms to their corresponding quarterly year over year forms.

RESULTS AND INTERPRETATIONS

Table 5 shows the results of DF-GLS test for each country. Each of the variables in the data is either in the form of growth rate or level and thus does not exhibit any linear deterministic trend. The null hypothesis of DF-GLS test is that H_0: the series has a random walk trend (i.e. the series is nonstationary), while the alternative hypothesis of DF-GLS test is that H_1: the series is stationary around a linear time trend (i.e. the series is stationary).

From left to right in Table 5, column 1 to 3 show the variables (detailed description of these variables can be found in table 2 to table 4), DF-GLS statistics, and p-value for the United States, respectively. Column 4 to 6 and column 7 to 9 show the counterparts for the United Kingdom and the Eurozone, respectively.

For the United States, the p-value of the DF-GLS test of each variable (PCE, UR, RGDP, INT, and SP500) are less than 0.0001. This indicates that the null hypothesis of the DF-GLS test that the series is nonstationary can be rejected at a significance level of less than 0.01%. This means that each of the variables in Table 5 for the United States is stationary.

Similarly, for United Kingdom, the p-value of DF-GLS test of each of the variable (CPI, UR, RGDP, INT, and FTSE100) is also less than 0.0001. This indicates that the null hypothesis of the DF-GLS test that the series is nonstationary can also be rejected at a significance level of less than 0.01%. This tells us that each of the variables in Table 5 for United Kingdom are also stationary.

At last, for the Eurozone, the p-value of DF-GLS test of each variable (CPI, UR, RGDP, INT, and STOX) is less than 0.0001. This means that the null hypothesis of DF-GLS test that the series is nonstationary can also be rejected at a significance level of less than 0.01%. This suggests that each of the variables in Table 5 for the Eurozone is stationary as well.

The results of the DF-GLS test in Table 5 suggest that all of the variables for each country or region in the data are highly statistically significant and stationary.

Table 12 to Table 14 show the multiple regression estimates for the United States, the United Kingdom, and the Eurozone, respectively. The time range of these regression in Table 12 is from 1970 to 2004, whereas the actual sample ranges from 1970 to 2014. The periods after 2004 (i.e. 2005 to 2014) are used for pseudo out-of-sample forecasting (see the Econometric Methodology section for a detailed description). A similar approach is applied to both the United Kingdom and the Eurozone.

Table 12 shows the multiple regression estimates for the United States. The dependent variable is PCE_t (inflation at current period t). From left to right, the first column includes all the variables of different order of lags that will be used in the models. The second to fifth columns show the regression estimates for each model (with coefficients over the corresponding standard error in parentheses).The second column shows the results of the AR (1) regression, the coefficient of PCE_{t-1} is 0.9804 which is significant at the 1% level. In addition, the RMSFE of the benchmark AR (1) model is 0.705%.

The third column shows the ADL regression results. In particular, we can see from the table that this is actually an ADL (2, 1) model. The order of lag of the ADL model is chosen based on the BIC (see the Econometric Methodology section for detailed explanation). The smallest BIC value equals to -10.824 (see Table 6) of the ADL model for the United States is reached when PCE is lagged two periods and UR is lagged one period. All the regressors are significant at 1% level. Based on Phillips Curve, inflation and unemployment are negatively correlated. The coefficient of the unemployment rate in the current period t are negative and significant, which is consistent with what is suggested by Phillips Curve. However, the coefficient of unemployment in the previous period is not negative even as it is significant. The RMSFE of the ADL (2, 1) is 0.743%.

The fourth column shows the result of ARIMA regression results. The value of d is zero because the series is stationary based on DF-GLS test. The value of p and q are chosen by the SCAN method which suggests that p = 1 and q = 5. That is to say, the ARIMA model in this case is an ARIMA (1, 0, 5). In addition, only the coefficient of PCE_{t-1} is statistically significant at the 1% level, however for the moving average process of ARIMA model, none of the coefficients for u_{t-1}, u_{t-2}, u_{t-3}, u_{t-4}, and u_{t-5} are statistical significant. In addition, the RMSFE of ARIMA model is only 0.532%.

The fifth column shows the VAR regression results. Similar with the ADL model, the order of lag of the VAR model is also chosen by the BIC, equal to -32.846 (see Table 7). The smallest BIC is reached when the lag of order is one. Thus, this is a VAR (1) model. The coefficients of PCE_{t-1} and $RGDP_{t-1}$ are statistically significant at the 1% level, and the coefficient of INT_{t-1} is significant only at the 10% level. In addition, the

RMSFE of the VAR (1) model is 0.736%.

Overall, for the United States, the multiple regression estimates in Table 12 indicate that an ARIMA (1, 0, 5) outperforms the other three models since it yields the smallest RMSFE in pseudo out-of-sample forecasting, thus it is the most preferable model.

Table 13 shows the multiple regression estimates for the United Kingdom. In this case, the dependent variable is CPI_t (inflation at current period t). From left to right, the first column includes all the variables of different order of lags that are used in the models. The second to fifth columns show the regression estimates for each model (with coefficients over the corresponding standard error in parentheses). The second column shows the AR(1) regression results, the coefficient of PCE_{t-1} is 0.974 which is significant at 1% level. In addition, the RMSFE of the benchmark AR (1) model is 0.573%.

The third column shows the ADL regression results. In particular, we can see from the table that this is actually an ADL(2, 0) model. The lag order of the ADL model is chosen based on BIC (see Econometric Methodology for a detailed explanation). The smallest BIC value, equal to -8.734 (see Table 8) for the ADL model for the United Kingdom is reached when CPI is lagged two periods and UR is lagged zero periods. All the regressors except UR_t are significant at 1% level. UR_t is negative but only significant at 10% level. In addition The RMSFE of the ADL (2, 0) is 0.606%.

The fourth column shows the ARIMA regression results. The value of d is zero because the series is stationary based on DF-GLS test. The value of p and q are chosen by the SCAN method, which suggests that p = 2 and q = 3; the ARIMA model in this case is an ARIMA(2, 0, 3) model. In addition, the coefficient of PCE_{t-1} and PCE_{t-2} are statistical significant at 1% level, however for the moving average process of ARIMA model, still neither one of the coefficients for $u_{t-1,}$ u_{t-2}, and u_{t-3}, are statistical significant. In addition, the RMSFE of ARIMA model is only 0.541%.

The fifth column shows the VAR regression results. Similar to the ADL model, the lag order of the VAR model is also chosen by BIC, equal to -29.6629 (see Table 9). The smallest BIC is reached when the order of lag is two. Thus, this is a VAR (2) model. The coefficients of CPI_{t-1} and CPI_{t-2}, and $RGDP_{t-1,}$ $INT_{t-1,}$ and $FTSE100_{t-2}$ are significant at 1% level, while $RGDP_{t-2}$ is statistically significant at the 5% level. However, the RMSFE of this VAR(2) model is high, which is 1.091%.

For the United Kingdom, the multiple regression estimates in Table 13 indicates that the ARIMA(2, 0, 3) outperforms the other three models since it yields the smallest RMSFE in pseudo out-of-sample forecasting, thus ARIMA is still the most preferable model.

Table 14 shows the multiple regression estimates for the Eurozone. In this case, the dependent variable is CPI_t (inflation at current period t). From left to right, the first column includes all the variables of different lag order that are used in the models. The second through the fifth columns show the regression estimates for each model (with coefficients over the corresponding standard error in parentheses). The second column shows the AR(1) regression results, the coefficient of PCE_{t-1} is 0.790% which is significant at 1% level. In addition, the RMSFE of the benchmark AR (1) model is 0.545%.

The third column shows the regression results of an ADL(1,0) model. The lag order of the ADL model is chosen based on BIC (see the Econometric Methodology for a detailed explanation). The smallest BIC value, equal to -11.4777 (see Table 10) for ADL model for the Eurozone is reached when the CPI is lagged one period and the UR is lagged zero periods. All the regressors are significant at 1% level. In addition, the RMSFE of the ADL (1, 0) is 0.545%.

The fourth column shows the result of ARIMA regression results. The value of d is zero because the series is stationary based on the DF-GLS test. The value of p and q are chosen by the SCAN method, which suggests that p = 0 and q = 3, in this case the model is an ARIMA (0, 0, 3) (or equivalently, an MA (3) process). In addition, the coefficients of u_{t-1}, u_{t-2}, and u_{t-3} are statistically significant at the 1% level. The RMSFE of ARIMA model is only 0.474%.

The fifth column shows the VAR regression results. Similar with the ADL model, the lag order of the VAR model is also chosen by an BIC, equal to -37.316 (see Table 11). The smallest BIC is reached when the lag order is one. Thus, this is a VAR(1) model. Only the coefficients of CPI_{t-1} and UR_{t-1} are significant at the 1% level, while the other regressors are not significant. The RMSFE of this VAR(1) model is high, which is 0.697%.

For the Eurozone, the multiple regression estimates in Table 14 indicate that the ARIMA(0, 0, 3) outperforms the other three models since it yields the smallest RMSFE in pseudo out-of-sample forecasting. Thus once again, ARIMA is the most preferable model.

To summarize, Table 12 through Table 14 present multiple regression results for each country or region. These regression results evaluate the performance of each of the four models. Which model performs the best? The answer is consistent for the United States, United Kingdom, and the Eurozone – ARIMA model outperforms the other three models in terms of the forecasting accuracy (RMSFE). Thus the ARIMA model is my most preferable model to proceed to (true) out-of-sample forecasting for the year of 2015 and 2016.

Figure 1 through Figure 3 plot the pseudo out-of-sample forecasted inflation (dashed line) against the actual inflation (solid line) for the United States, the United Kingdom, and the Eurozone, respectively. For the United States and the United Kingdom, the period used for pseudo out-of-sample forecasting is from 2005 to 2014, while for the Eurozone the period used for pseudo out-of-sample forecasting is from 2008 to 2014.

Figure 4 through Figure 7 plot and compare the two-year ahead (true) out-of-sample forecasting with the AR(1) model, the ADL(2,1) model, the ARIMA(1,0,5) model, and the VAR(1) model respectively. The AR(1) model suggests that the United States will experience a slow decaying of inflation through 2017, while the ADL(2, 1) model and the VAR(1) model suggest that the United States will experience an increase in inflation in the future. The ARIMA(1, 0, 5) model suggests that the United States will experience a decrease in inflation through 2015, quarter 3, and then an increase in inflation through 2016 quarter 4.

Figure 8 through Figure 11 plot and compare the two-year ahead (true) out-of-sample forecasting with the the AR (1) model, the ADL (2, 0) model, the ARIMA (2, 0, 3) model, and the VAR (2) model respectively. the AR (1) model suggests that the United Kingdom will experience a slow decaying of inflation during the next two years, while the ADL (2, 1) model and the VAR (1) model suggest that the United Kingdom will experience an increase in inflation in the future, but the VAR seems to indicate a slower increase in inflation. The ARIMA (2, 0, 3) model suggests that the United Kingdom will experience a decrease in inflation and then an increase in inflation.

Figure 12 to Figure 15 plot and compare the two-year ahead (true) out-of-sample forecasting with the AR (1) model, the ADL (1, 0) model, the ARIMA (0, 0, 3) model, and the VAR (1) model, respectively. The AR (1) model suggests that the Eurozone will experience a slow decaying of inflation through 2017, while the ADL (1, 0) model and the VAR (1) model suggest that the Eurozone will experience an increase in inflation through 2017. The ARIMA(0, 0, 3) model suggests that the Eurozone will experience an increase in inflation through 2015, quarter 4 first, and then, an approximately constant level of inflation through 2017.

CONCLUSION AND DISCUSSION

The results suggest that the ADL model and the VAR model do not perform as well as the ARIMA model or even the AR (1) model that I choose as the benchmark model. This conclusion is to some extent consistent with other studies.

Stockton and Glassman (1987) evaluate the forecast performance of three models of inflation: rational expectation models, expectations-augmented Phillips curves, and monetarist models. Their conclusions are to some extent similar: a simple ARIMA model of inflation produces such a respectable forecast relative to the theoretically based specifications, and theory yields only small dividends in terms of improving the ability to predict the course of inflation accurately. Their findings are also consistent with what is shown here: with this dataset and model specifications, the results indicate that the ARIMA model out-performs the other three time series models and provides the highest forecasting accuracy.

Ang, Bekaert, and Wei (2005) also conduct a comprehensive analysis of different models and investigate the predictive power of these models. They find that the best time-series model for forecasting future inflation is the simple ARMA(1, 1) model. In terms of RMSFE, as is stated by the authors, the ARMA (1, 1) model yields the smallest forecast error. However, they also find that, for CPI measures, survey based measures perform better than the ARMA(1, 1) model, while for PCE measures, the ARIMA model does a better job at forecasting PCE inflation. In addition, they also find that ARMA model performs a much better forecasting than Phillips-Curve-based regressions.

My conclusions are summarized as follows:

1. In pseudo out-of-sample forecasting, ARIMA model out-performs the other three time-series models in terms of forecasting accuracy i.e., the RMSFE.

2. Models based on economic theory do not perform as well as models that only contain lags of the dependent variable. The variables included in ADL model and VAR model are based on economic theories, the most important one is the Phillips Curve. This is especially obvious for the results of the VAR model of the United Kingdom and the Eurozone.

3. According to the ARIMA Results of the two-year ahead forecasting (2015 quarter 1 through 2016 quarter 4), the United States will experience a decrease in inflation during the first two quarters of 2015, and then a continuous increase in inflation through the fourth quarter of 2016.

4. The United Kingdom will experience a decrease in inflation during the first three quarters of 2015, and then a continuous increase in inflation through the fourth quarter of 2016.

5. The Eurozone will experience a continuous increase in inflation through 2015, and then maintained at an approximately constant level through 2016.

REFERENCES

Ang, Andrew; Bekaert, Geert; and Wei, Min, 2005 "Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?" *Journal of Monetary Economics,* Vol.54, pp. 1163-pp.1212.

Atkeson, Andrew; and Ohanian, Lee E., 2001, "Are Phillips Curve Useful for Forecasting Inflation?" , *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol.25, pp.2 – pp.11.

Ball, Laurence M., 2008, *Money, Banking, and Financial Markets*, 1st edition, Worth Publishers.

Barr, David G.; and Campbell, John Y., 1997, "Inflation, Real Interests, and the Bond Market: a Study of UK Nominal and Index-Linked Government Bond Prices", *Journal of Monetary Economics*, Vol.39, pp.361-pp.383.

Bhattarai, Saroj; Won Lee, Jae; and Yong Park, Woong, 2012, "Inflation Dynamics: The Role of Public Debt and Policy Regimes", *Federal Reserve Bank of Dallas Globalization and Monetary Policy Institute Working Paper No.124*.

Blanchflower, David G.; and MacCoille, Conall, 2009, "The Formation of Inflation Expectations: an Empirical Analysis for the UK", *NBER Working Paper 15388*.

Blinder, Alan S., 1997 "Economic Policy and the Great Stagflation", *New York: Academic Press*.

Gujarati, Damodar N.; Porter, Dawn C., 2008, *Basic Econometrics*, 5th edition, McGraw Hill.

Fisher, Jonas D.M.; Chin, Te Liu; and Zhou, Rui Lin, 2002, "When can we forecast inflation?", *Federal Reserve Bank of Chicago Economic Perspective, 1Q/2002,* pp.30-pp.42.

Forni, Mario; Hallin, Marc; Lippi, Marco; and Reichlin, Lucrezia, 2002, "Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area?" *International Macroeconomics discussion paper No. 3146*.

Gilchrist, Simon; Schoenle, Raphael S.; Sim, Jae W.; and Zakrajsek, Egon, 2012. "Inflation Dynamics during the Financial Crisis," *Finance and Economics Discussion Series 2015-012*.

Stockton, David J.; Glassman, James E, 1987, "An Evaluation of the Forecast

Performance of Alternative Models of Inflation", *The Review of Economics and Statistics*, Vol.69, No.1, pp.108-pp.117.

Greene, William H., 2001, *Econometric Analysis*. 7th edition, Prentice Hall.

Meltzer, Allan H., "Origins of the Great Inflation", 2005, *Federal Reserve Bank of St. Louis Review, March/April 2005, 87(2, Part2),* pp.145-pp.75.

Meyer, Brent H.; and Pasaogullari, Mehmet, "Simple Ways to Forecast Inflation: What Works Best?" *Federal Reserve Bank of Cleveland Economic Commentary Number 2010-17*.

Mishkin, Frederic, 2007, "Inflation Dynamics", *International Finance*, Vol.10, pp.317-pp.334.

Moyer, Brian C., 2006, "Comparing Price Measures – CPI, PPI, and PCE", *National Association for Business Economics 2006 Washington Economic Policy Conference*.

Rumler, Fabio; and Valderrama, Maria Teresa, 2008, "Comparing the New Keynesian Phillips Curve series models to Forecast Inflation", *North American Journal of Economics and Finance*, Vol.21, pp.126-pp.144.

Russell, Bill; and Banerjee, Anindya, 2007, "The Long-run Phillips Curve and Non-stationary Inflation", *Journal of Macroeconomics*, Vol.30, pp.1792-pp.1815
Mankiw, Gregory N., "U.S. Monetary Policy during the 1990s", NBER Working Paper 8471.

Sekine, Toshitaka, 2001, "Modeling and Forecasting Inflation in Japan", *IMF Working Paper No. 01/82*.

Silvia, John E.; Iqbal, Azhar; House, Sarah Watt; Nelson, Erik, 2014, "The Great Inflation: Once in a Lifetime?" *Wells Fargo Securities Economics Group Special Commentary*.

Stock, James; and Watson, Mark, 2006, Introduction to Econometrics, 2nd edition, Pearson Education.

Stock, James; and Watson, Mark, 1999, "Forecasting Inflation", Journal of Monetary Economics, Vol.44 (2), pp.293 – pp.335.

Stock, James; and Watson, Mark, 2001, "Vector Autoregressions", *Journal of Economic Perspectives*, Vol.15, No.4, fall 2001, pp.101-pp.115.

Table 1: Comparison of different types of models

Macro model	Long-run forecast, but we need hundreds of variables and out of sample data are needed.
VAR model	Two or more dependent variables and two or more right hand side variables.
Exponential smoothing	Pure statistical predictor. Add time trend and season cycle.
ADL model	One dependent variable and two or more predictors. Out of sample value are needed.
AR ARIMA	Only lags of dependent variable are needed.

Table 2: Data and Variable descriptions for the United States

Variables	Descriptions	Sources
PCE	Quarterly PCE deflator calculated on the year over year basis	*Federal Reserve Bank of St. Louis*
UR	Quarterly unemployment rate	*U.S. Bureau of Labor Statistics*
RGDP	Quarterly real GDP calculated on the year over year basis	*Federal Reserve Bank of St. Louis*
INT	Quarterly ten-year government Treasury bond yield	*Federal Reserve Bank of St. Louis*
SP500	Quarterly S&P 500 index calculated on the year over year basis	*Federal Reserve Bank of St. Louis*

Table 3: Data and Variable descriptions for the United Kingdom

Variables	Descriptions	Sources
CPI	Quarterly CPI calculated on the year over year basis	*OECD Official Website*
UR	Quarterly unemployment rate	*Federal Reserve Bank of St. Louis*
RGDP	Quarterly real GDP calculated on the year over year basis	*Federal Reserve Bank of St. Louis*
INT	Quarterly ten-year government Treasury bond yield	*Federal Reserve Bank of St. Louis*
FTSE100	Quarterly FTSE 100 index calculated on the year over year basis	*Bank of England*

Table 4: Data and Variable descriptions for the Eurozone

Variables	Descriptions	Sources
CPI	Quarterly CPI calculated on the year over year basis	*OECD Official Website*
UR	Quarterly unemployment rate	*ECB (European Central Bank)*
RGDP	Quarterly real GDP calculated on the year over year basis	*ECB (European Central Bank)*
INT	Quarterly ten-year government Treasury bond yield	*ECB (European Central Bank)*
STOX	Quarterly Euro Stoxx index calculated on the year over year basis	*ECB (European Central Bank)*

Table 5: DF-GLS test results for each country or region:

United States			United Kingdom			Eurozone		
Variable	DF-GLS	P-value	Variable	DF-GLS	P-value	Variable	DF-GLS	P-value
PCE	-1.3344	<0.0001	CPI	-1.9904	<0.0001	CPI	-4.0273	<0.0001
UR	-2.1762	<0.0001	UR	-1.5373	<0.0001	UR	-1.7128	<0.0001
RGDP	-2.6757	<0.0001	RGDP	-3.2753	<0.0001	RGDP	-3.2172	<0.0001
INT	-1.144	<0.0001	INT	-0.6547	<0.0001	INT	-0.5193	<0.0001
SP500	-2.2992	<0.0001	FTSE100	-2.2616	<0.0001	STOX	-2.1238	<0.0001

Table 6: ADL model lag selection with BIC for the United States

	UR_t	UR_{t-1}	UR_{t-2}	UR_{t-3}
PCE_t	-7.3177915	-7.3801063	-7.3365877	-7.2925786
PCE_{t-1}	-10.548668	-10.559506	-10.564147	-10.521622
PCE_{t-2}	-10.798801	-10.824492	-10.797986	-10.774642
PCE_{t-3}	-10.758485	-10.784839	-10.760158	-10.750561

Table 7: VAR model lag selection with BIC for the United States

lag	AIC	BIC
0	-21.4935	-21.3864
1	-33.4887	-32.8462*
2	-33.8246	-32.6467
3	-33.8755	-32.1622
4	-33.9356*	-31.6868

Table 8: ADL model lag selection with BIC for the United Kingdom:

	UR_t	UR_{t-1}	UR_{t-2}	UR_{t-3}
CPI_t	-5.6870092	-5.9674544	-5.937413	-5.8979948
CPI_{t-1}	-8.5089465	-8.4793137	-8.4843976	-8.4510807
CPI_{t-2}	-8.7337869	-8.6980834	-8.6906753	-8.6469819
CPI_{t-3}	-8.6902799	-8.6543737	-8.6489144	-8.6133939

Table 9: VAR model lag selection with BIC for the United Kingdom

lag	AIC	BIC
0	-18.3521	-18.245
1	-29.9321	-29.2896
2	-30.8408	-29.6629*
3	-30.7912	-29.0778
4	-31.0023*	-28.7535

Table 10: ADL model lag selection with BIC for the Eurozone:

	UR_t	UR_{t-1}	UR_{t-2}	UR_{t-3}
CPI_t	-10.774658	-10.930201	-10.972539	-10.968282
CPI_{t-1}	-11.477742	-11.399886	-11.368413	-11.282627
CPI_{t-2}	-11.445593	-11.363845	-11.311232	-11.218914
CPI_{t-3}	-11.358834	-11.275627	-11.220233	-11.140889

Table 11: VAR model lag selection with BIC for the Eurozone:

lag	AIC	BIC
0	-29.5866	-29.3838
1	-38.5321	-37.3156*
2	-39.2658	-37.0356
3	-39.5955	-36.3516
4	-39.9283*	-35.6706

Table 12: Regression Results for United States (1970-2004)

Dependent Variable: PCE_t

	AR	ADL	ARIMA	VAR
PCE_{t-1}	0.9804^{***} (0.017)	1.5273^{***} (0.073)	0.9299^{***} (0.034)	1.0093^{***} (0.022)
PCE_{t-2}		-0.5265^{***} (0.076)		
UR_t		-0.4026^{***} (0.119)		
UR_{t-1}		0.3288^{***} (0.113)		-0.0364 (0.029)
$RGDP_{t-1}$				0.0652^{***} (0.022)
INT_{t-1}				-0.0346^{*} (0.020)
$SP500_{t-1}$				-0.0015 (0.003)
U_{t-1}			0.7346 (16.574)	
U_{t-2}			0.7095 (381.300)	
U_{t-3}			0.7922 (326.112)	
U_{t-4}			-0.2226 (88.704)	
U_{t-5}			-0.0398 (15.780)	
Intercept	0.0007 (0.0009)	0.0046^{**} (0.002)	0.0420^{***} (0.042)	0.0024 (0.002)
RMSFE	0.705%	0.743%	0.532%	0.736%

Standard errors in parentheses

* $p<0.1$ ** $p<0.05$ *** $p<0.01$

Table 13: Regression Results for United Kingdom (1970-2004)

Dependent Variable: CPI_t

	AR	ADL	ARIMA	VAR
CPI_{t-1}	0.9742^{***} (0.020)	1.4316^{***} (0.076)	0.5047^{***} (0.074)	1.1875^{***} (0.077)
CPI_{t-2}		-0.4805^{***} (0.075)	0.3405^{***} (0.076)	-0.3686^{***} (0.075)
UR_t		-0.0737^{*} (0.038)		
UR_{t-1}				-0.4952 (0.451)
UR_{t-2}				0.2997 (0.454)
$RGDP_{t-1}$				0.2055^{***} (0.070)
$RGDP_{t-2}$				-0.1472^{**} (0.067)
INT_{t-1}				0.3899^{***} (0.128)
INT_{t-2}				-0.0482 (0.138)
$FTSE100_{t-1}$				0.0052 (0.007)
$FTSE100_{t-2}$				-0.0227^{***} (0.007)
U_{t-1}			1.0205 (277.052)	
U_{t-2}			0.9339 (5.675)	
U_{t-3}			0.9134 (253.046)	
Intercept	0.0015 (0.002)	0.0074^{**} (0.003)	0.0624^{*} (0.033)	-0.0091^{**} (0.004)
RMSFE	0.573%	0.606%	0.541%	1.091%

Standard errors in parentheses

* $p<0.1$ ** $p<0.05$ ***$p<0.01$

Table 14: Regression Results for Eurozone (1996-2007)

Dependent Variable: CPI_t

	AR	ADL	ARIMA	VAR
CPI_{t-1}	0.7898*** (0.100)	0.6748*** (0.099)		0.6061*** (0.112)
UR_t		-0.1451*** (0.048)		
UR_{t-1}				-0.2620*** (0.083)
$RGDP_{t-1}$				-0.0765 (0.070)
INT_{t-1}				0.0635 (0.051)
$STOX_{t-1}$				0.0047 (0.004)
U_{t-1}			0.6543*** (0.183)	
U_{t-2}			1.0130*** (0.221)	
U_{t-3}			0.9020*** (0.213)	
Intercept	0.0042** (0.002)	0.0197 (0.048)	0.1998*** (0.001)	0.0300*** (0.009)
RMSFE	0.545%	0.545%	0.474%	0.697%

Standard errors in parentheses

* p<0.1 ** p<0.05 ***p<0.01

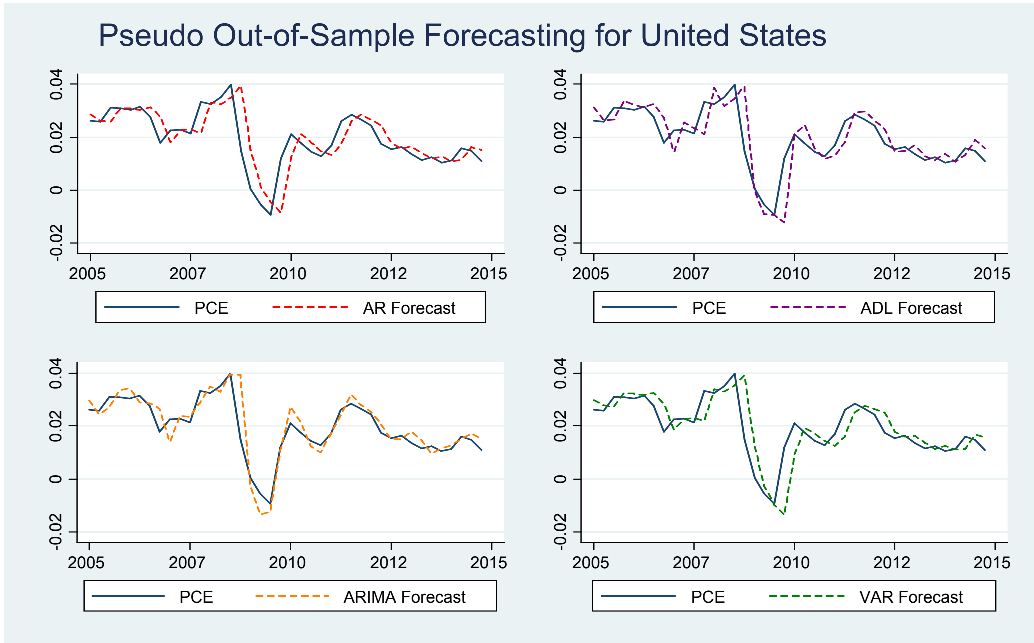

Figure 1: Comparison of out-of-sample forecasting for United States (2005-2014)

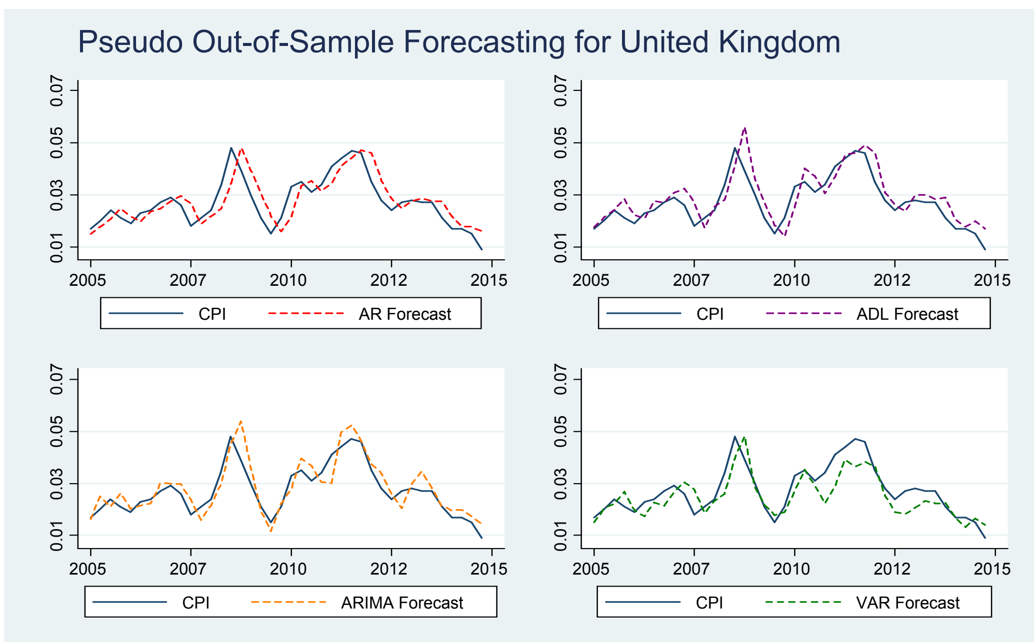

Figure 2: Comparison of out-of-sample forecasting for United Kingdom (2005-2014)

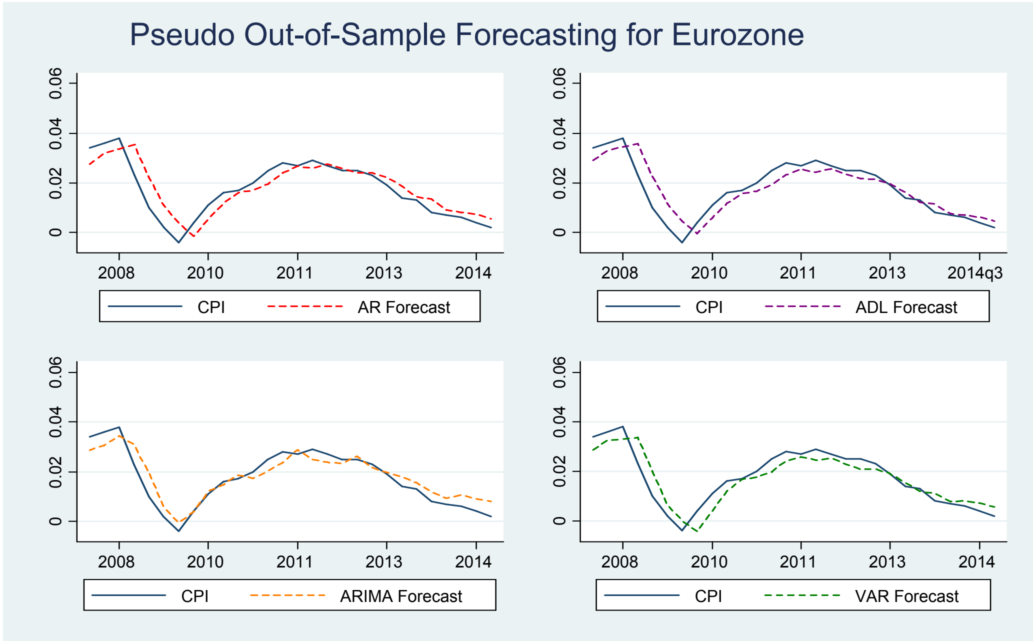

Figure 3: Comparison of out-of-sample forecasting for Eurozone (2008-2014)

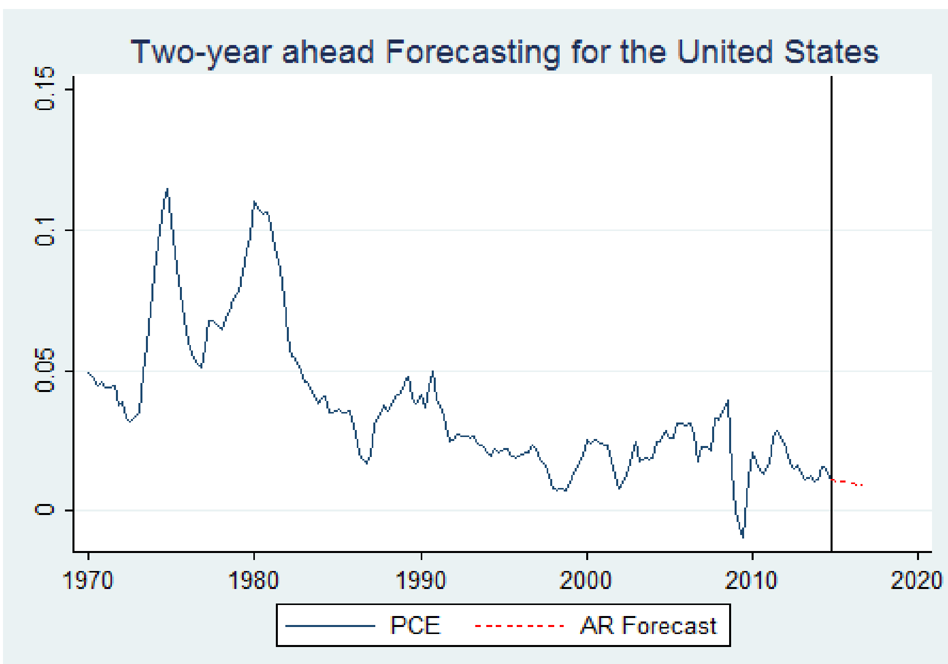

Figure 4: United States Forecasted Inflation with AR Model (2015-2016)

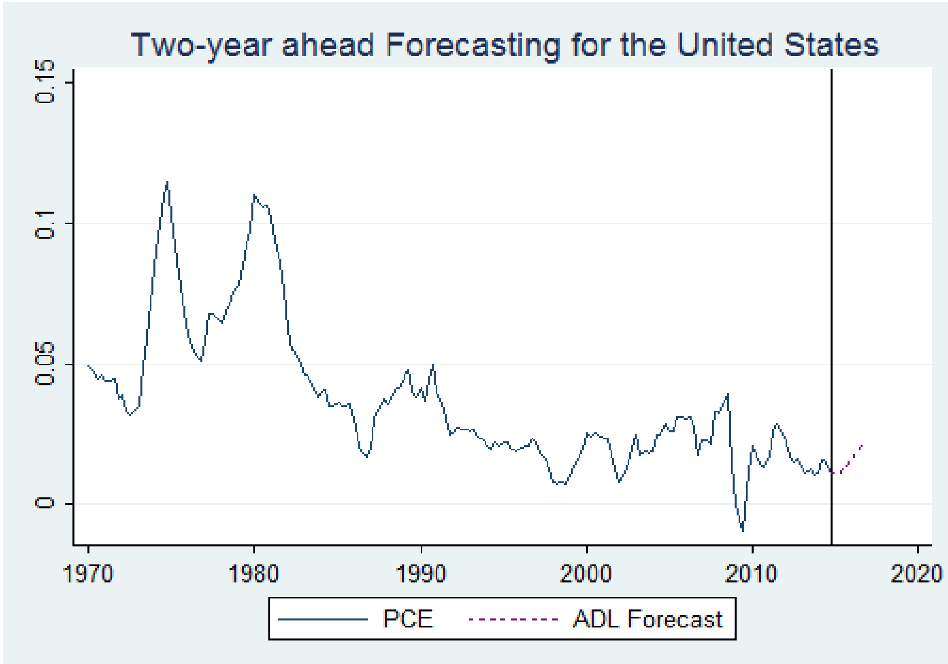

Figure 5: United States Forecasted Inflation with ADL Model (2015-2016)

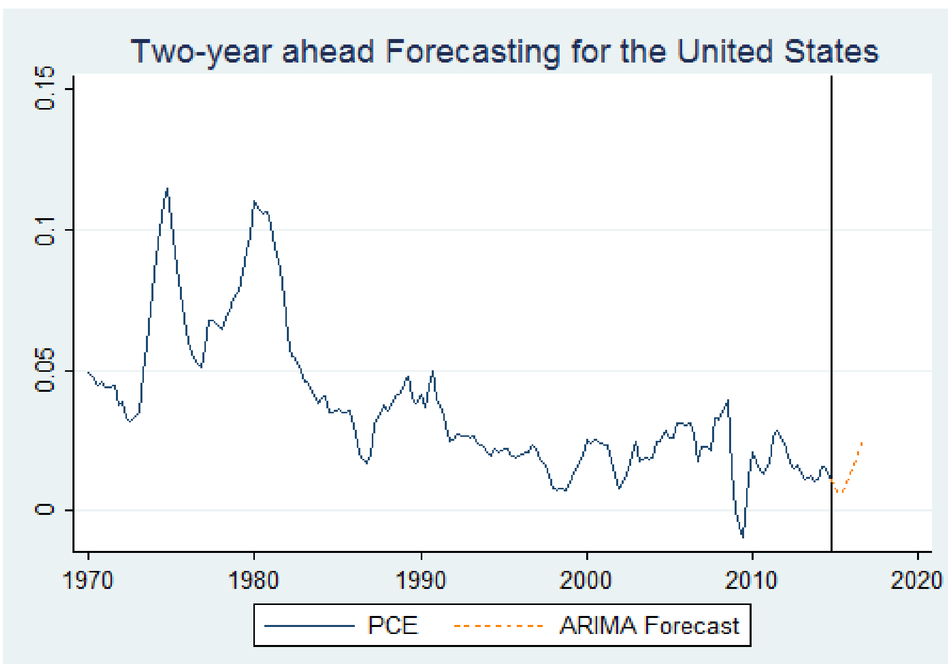

Figure 6: United States Forecasted Inflation with ARIMA Model (2015-2016)

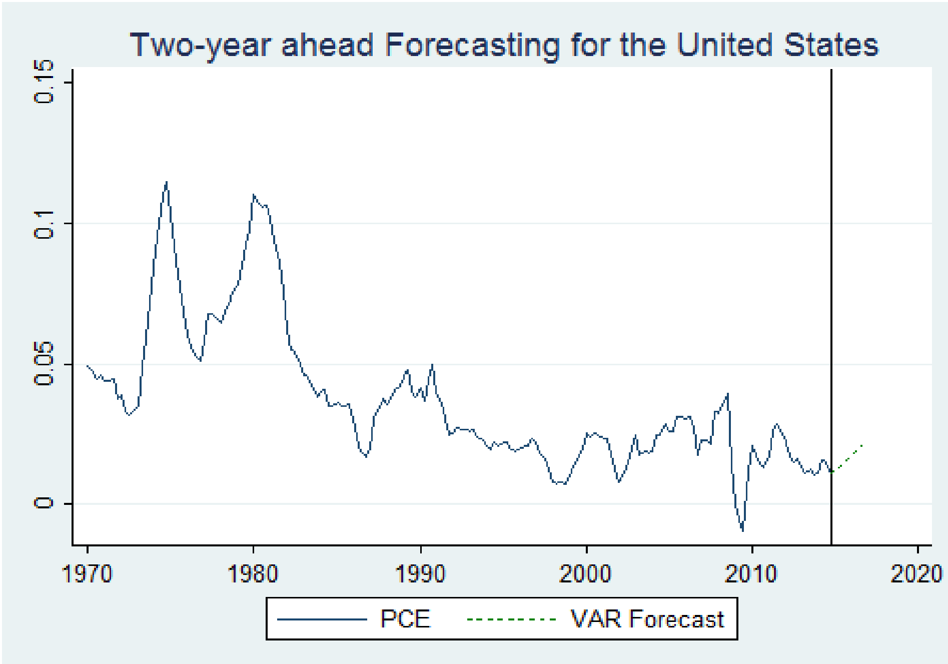

Figure 7: United States Forecasted Inflation with VAR Model (2015-2016)

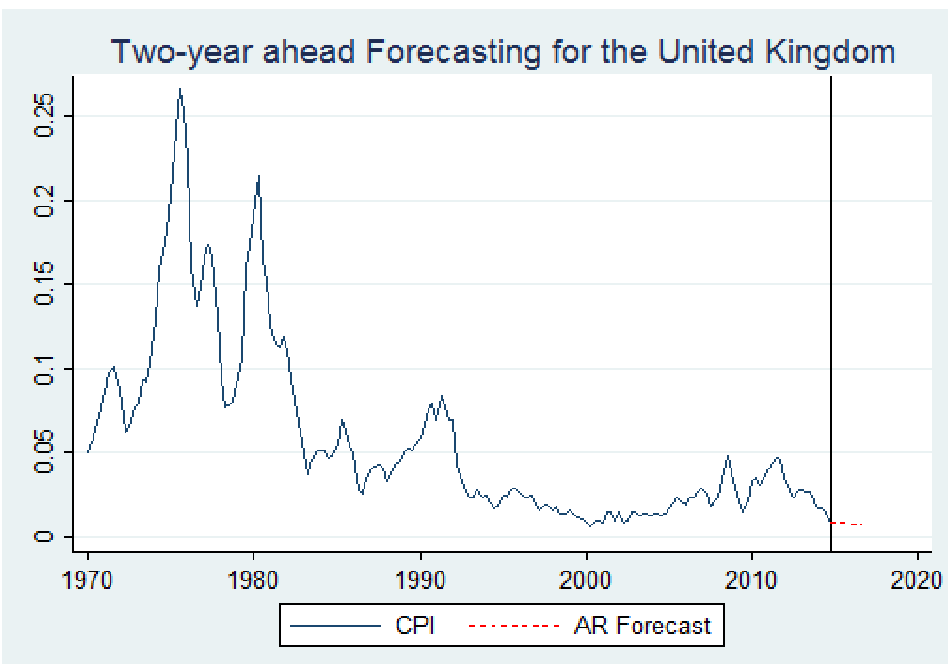

Figure 8: United Kingdom Forecasted Inflation with AR Model (2015-2016)

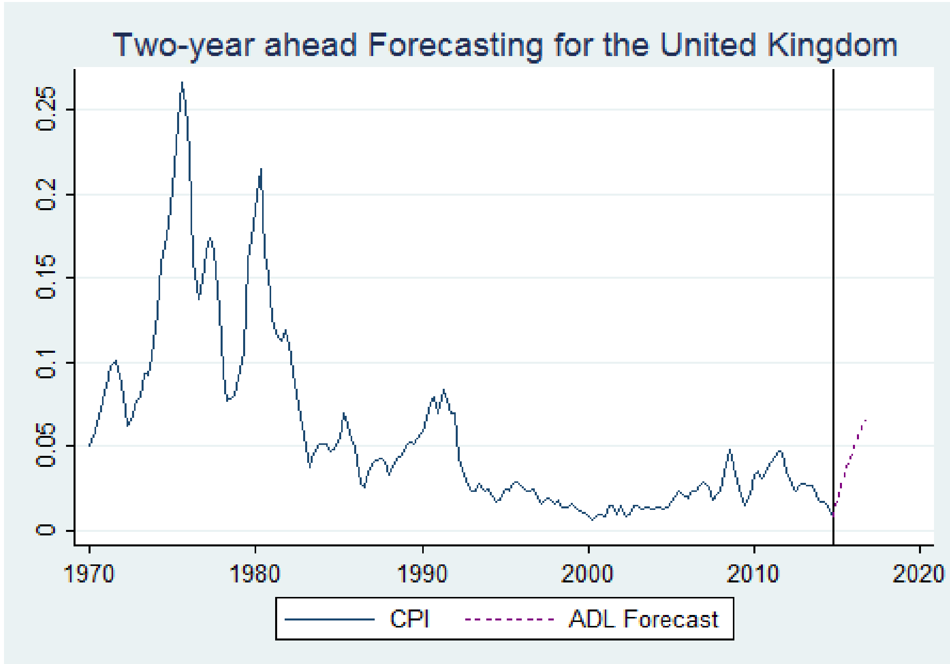

Figure 9: United Kingdom Forecasted Inflation with ADL Model (2015-2016)

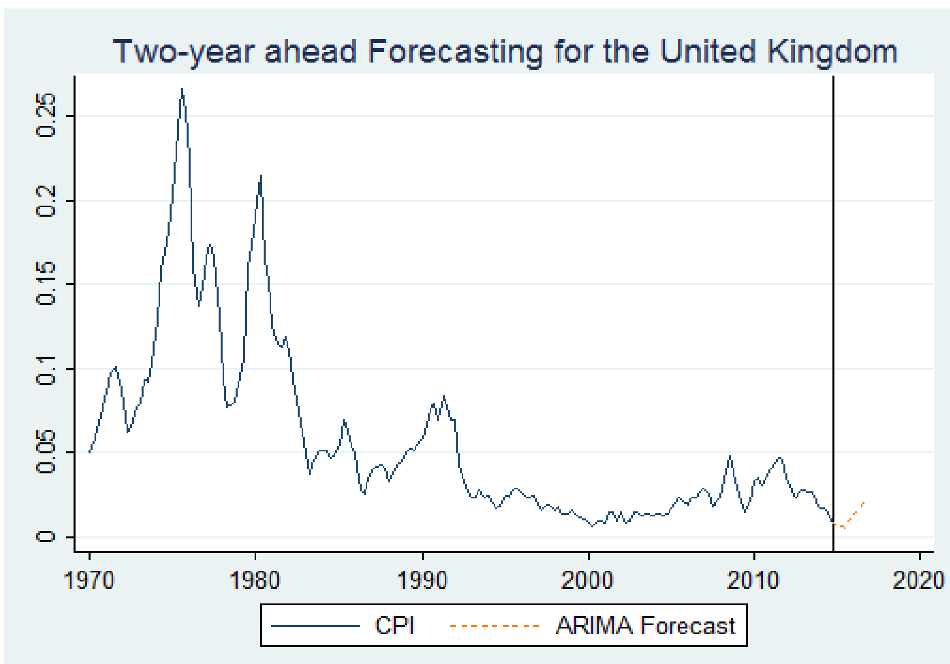

Figure 10: United Kingdom Forecasted Inflation with ARIMA Model (2015-2016)

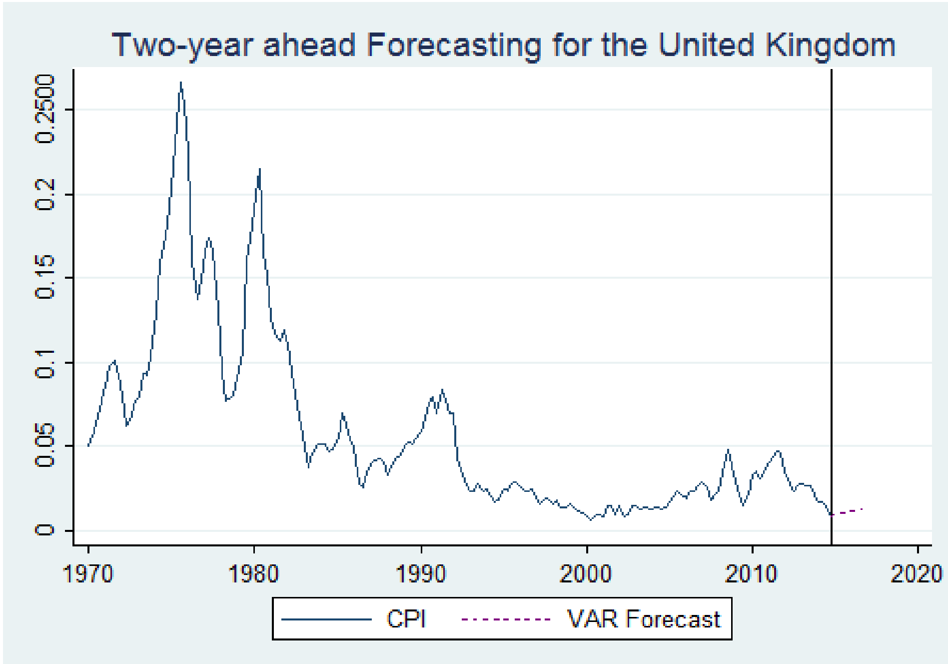

Figure 11: United Kingdom Forecasted Inflation with VAR Model (2015-2016)

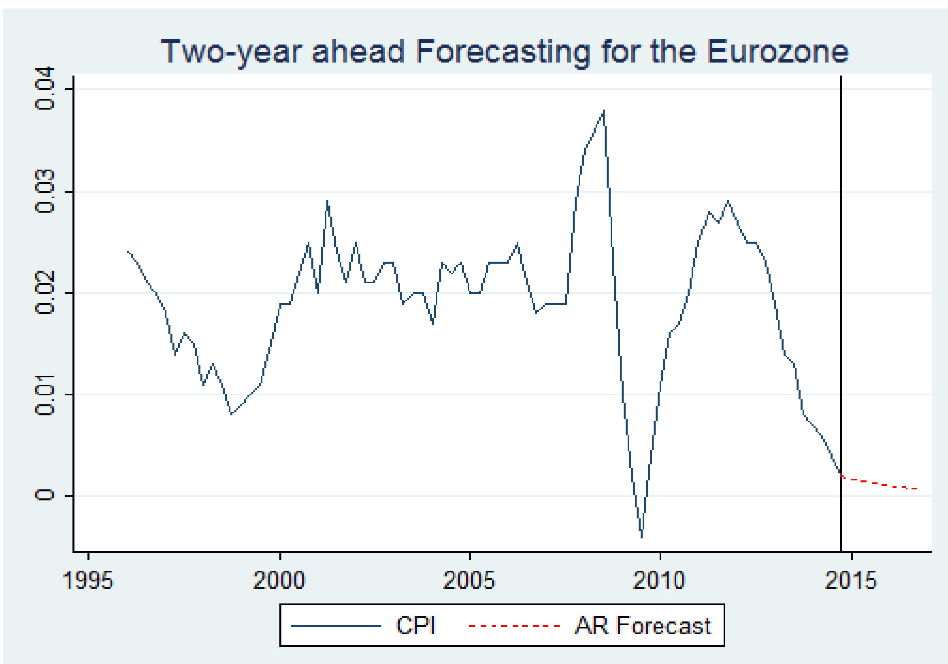

Figure 12: Eurozone Forecasted Inflation with AR Model (2015-2016)

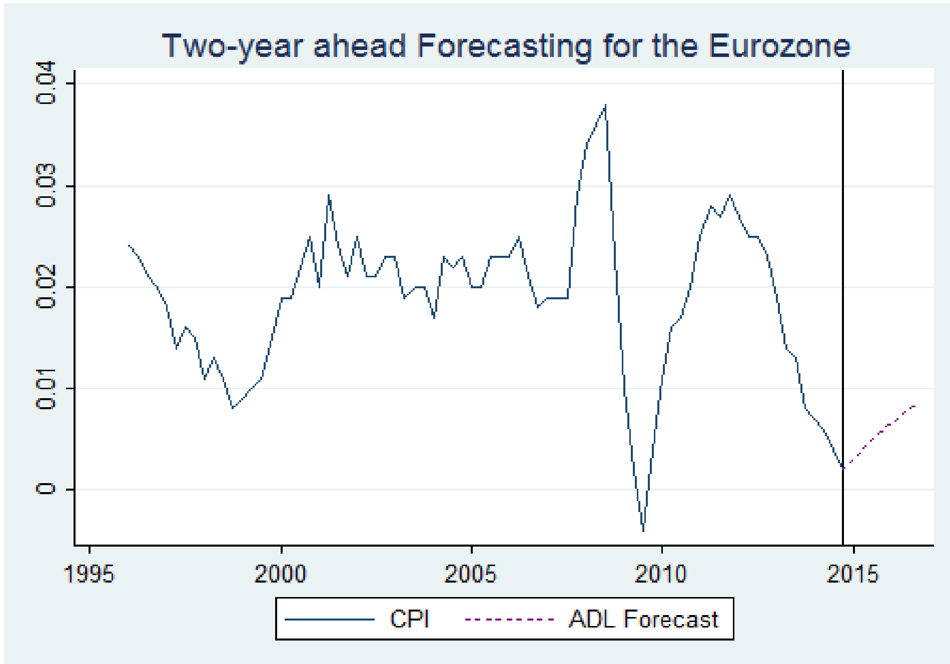

Figure 13: Eurozone Forecasted Inflation with ADL Model (2015-2016)

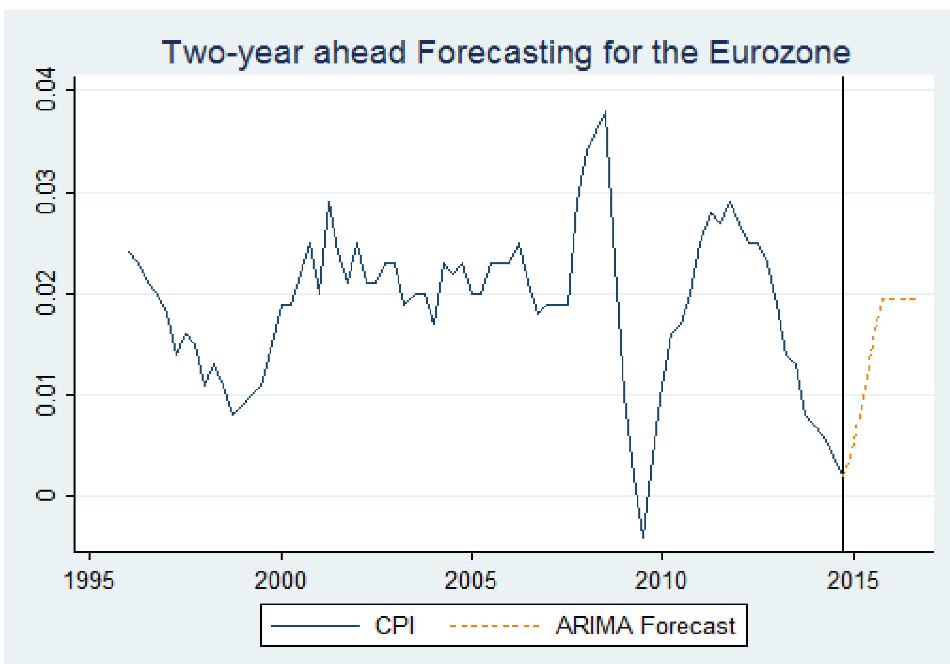

Figure 14: Eurozone Forecasted Inflation with ARIMA Model (2015-2016)

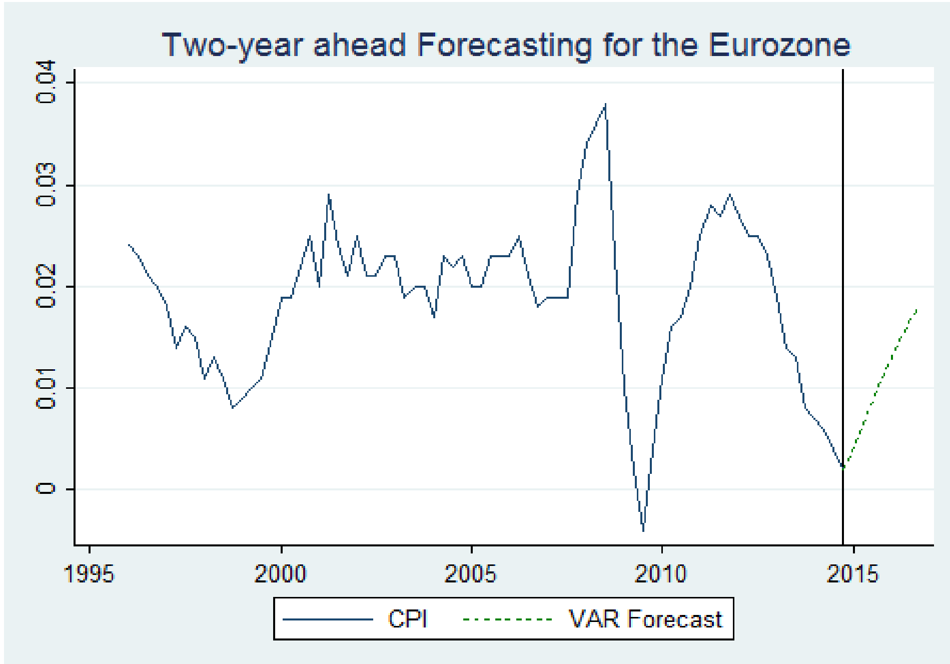

Figure 15: Eurozone Forecasted Inflation with VAR Model (2015-2016)