Data Analytics MSc Dissertation MTH797P, 2019/20

Forecasting Inflation with Time Series Analysis

Yuting Lu, ID 190534035

Supervisor: Dr. Pedro Vergel

A thesis presented for the degree of
Master of Science in *Data Analytics*

School of Mathematical Sciences
Queen Mary University of London

Declaration of original work

This declaration is made on September 10, 2020.

Student's Declaration: I Yuting Lu hereby declare that the work in this thesis is my original work. I have not copied from any other students' work, work of mine submitted elsewhere, or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

Referenced text has been flagged by:

1. Using italic fonts, **and**
2. using quotation marks "...", **and**
3. explicitly mentioning the source in the text.

Abstract

This paper investigates the performance of time series model in data-poor and data-rich environment for predicting inflation rate in the US. The data-poor model includes random walk, sARIMA selected by AIC and BIC. The data-rich model includes ARDL and ARDL with two kinds of penalty term: Lasso and adaptive Lasso. The in-sample data we used for model training starts from 1959M01 to 2005M12. The pseudo out-of-sample period starts from 2006M01 to 2019M12 with three subsamples: before Great Recession, during Great Recession and after Great Recession. We consider both short-term (1, 3 months) and medium to long-term (6, 12 months) forecast. We use two types of inflation measurements: CPI and PCE. Five predictor variables were used for multivariate model. We applied rolling forecast estimation with fixed window to evaluate the model performance. The evaluation metrics we used are RMSE and MAE. Our result confirmed previous literature that Philips curve is episodic. In addition, we confirmed the advantage of using ARIMA model for short-term forecast.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The prediction of economic and financial time series has been challenging due to the unprecedented changes in economic trends. The volatility in market also leads to challenges in forecasting. One important economic series is inflation. It is continuously monitored by the monetary policymakers in order to supervise the economic and control price stability. In January 2012, the Federal Open Market Committee (FOMC) had a meeting and set a 2% target inflation rate as measured by annual change in PCE).[2] Generally, when inflation climbs toward 3% the FED raises the FED funds rate. When inflation falls below 1% the deflationary fears will be sparked and hence the FED would use quantitative easing. The public like investors also keep tracks of inflation to form an understanding of the market and to make investment decisions.

However, over the past two decades the behaviour of inflation has been hard to predict. For example, when the Phillips curve was first proposed, it worked in the way that wages could affect inflation. More specifically, it states that inflation and unemployment rate have an inverse relationship. That is, higher inflation associates with lower unemployment. The idea behind this theory is that when unemployment rate decreases, wages would begin to increase, hence the rising production costs were passed on to consumers.

But the situation has changed as the labour market grew more competitive. As the economy becoming globalised, wage growth is now largely driven by productivity gains, not just labour deficiency. Hence even when unemployment rate falls to new lows the inflation did not occur, see Figure 1.1.

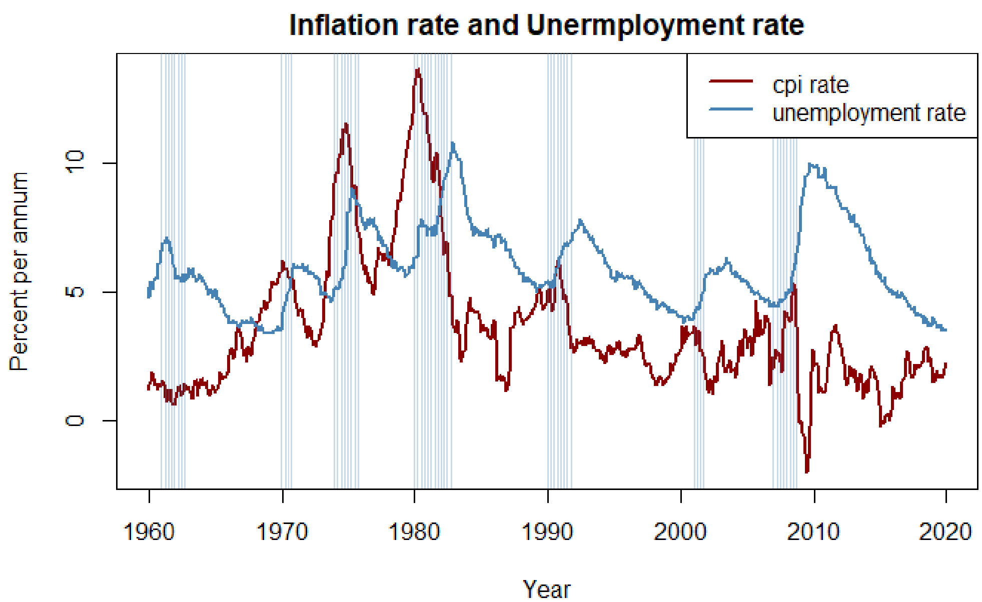


Figure 1.1: Plot of annual inflation rate and unemployment rate. The shaded areas are the recession periods.

This paper investigates the forecasting ability of six models for two measures of inflation: CPI and PCE monthly annual rate for 4 horizons and 4 time periods. The models differ in three ways: hyper-parameter selection, regularisation, and number of predictors (data-poor and data-rich environment) in a rolling window evaluation. As it is well known that forecasting inflation is hard during recession, we have separated the period of Great Recession out to see what model performs the best during that period.

In the remainder of this paper we first present the literature review that outlines the history of inflation prediction in Chapter 2. Then we introduce our data and data source in Chapter 3. Next in the Chapter 4 we present the methodology including the models and evaluation metrics we used. Chapter 5 is where we presents the formation of our experiments including data transformation and model building. Moreover, we present our results in this Chapter. In Chapter 6 we discuss the results, including the strengths and limitations of our study, as well as future work for improving our study. Finally, we have our conclusion in this chapter.

Chapter 2

Literature Review

The evolution of inflation forecasting model ranges from extrapolation to econometric modelling. In this chapter we introduce literature that is related to this paper.

ARIMA model was first developed and described by Box et al. in 1970[3]. Soon it became the most well-known model for forecasting time series in both statistics and economics areas.

An early study by Landsman and Damodaran (1989)[4] in forecasting inflation where they used the univariate ARIMA method showed that, the forecast accuracy of model was improved by ARIMA parameter estimator due to its lower mean squared percentage error. In addition, the study by Stockton and Glassman (1987)[5] showed that ARIMA model can have significant high accuracy in terms of short-term forecasting compares to other more sophisticated models.

However, a study by Meyler, Kenny and Quinn (1998)[6] mentioned that although ARIMA model could provide higher accuracy when forecasting inflation compares to models like vector autoregressive (VAR) and the Bayesian VAR, ARIMA can have bad performance when the data is unstable and have relatively big changes.

For multivariate models, there are many literatures about forecasting inflation using Phillips curve. As we included a form of Phillips curve: ARDL, we also present the literature review of Phillips curve. For example, Stock and Watson (2008)[7], and Banbura et al. (2013)[8] provided extensive literature surveys. Notice that we will directly refer to ARDL/ regression model in the remaining chapters.

Phillips curve was first proposed by AW Phillips, who claimed that there is negative relationship between unemployment rate and inflation.[9] Gordon (1982, 1990)[10] conducted the first studies in this field. In the 1970s, the Phillips curve was amended by adding supply shocks and zero long-run trade-offs. In 1977 Gordon came up with a triangle model of inflation. The triangle model includes three factors that Gordon believes inflation has: built-in inflation (inertia), demand-pull and supply-push factors. This model was helpful in the early 1980s where inflation decreased due to a sharp decrease in unemployment rate.

Stock and Watson (1999)[11] were the first to study the stability of Phillips curve by performing a pseudo out-of-sample experiment on forecasting inflation at one-year horizon. They concluded that for the period between 1970 and 1996, Phillips curve models performs better than their univariate models in predicting four quarter ahead inflation when doing recursive forecasts. Moreover, they found that using various series data in the model is better than single series model.

Atkenson and Ohanian(2001)[12] introduced a simple benchmark model: Random Walk. They showed that the forecast from this model outperform AR model during the period 1984 to 1999. They also showed that the Phillips curve based on NAIRU (using either the unemployment gap or an activity index) could not beat the random walk model. Fisher, Liu, and Zhou (2002)[13] on the other hand used rolling regressions to show that Phillips curve models performs better than Random walk in the period 1977 to 1984 and some period after that. Hence, they concluded that the Phillips curve's performance is episodic. Moreover, factors like sample period, forecasting horizon and inflation measure could affect the model's performance.

Later during 2003 and 2010 some researches were conducted to confirm the above findings. For example, Stock and Watson (2007)[14], Canova (2007)[15], Ang, Bekaert and Wei (2007)[16] extended the AO's analysis with qualifications and confirmed their findings. In addition, the Stock and Watson (2008)[7]and (2010)[17] study provides support for the Fisher et al. (2002)[13] findings.

An important study which also act as an indicator for this paper is the Stock and Watson (2008) study[7]. They used US data ranging from 1953 to 2008 with quarterly frequency to implement a pseudo out-of-sample exercise. Overall, they included 157 models with single forecast and 35 combination forecasts. Five different types of inflation measures were used including all CPI, all PCE, core CPI, core PCE and the GDP deflator. Although some results have relatively weak robustness, main qualitative results of the literature are confirmed in this study. This paper concluded that there is strong correlation between forecasting performance and the sample period as well as phase of the business cycle. Hence a key finding is that the performance of Phillips curve forecasts is episodic.

In the study by Li and Chen (2014)[18], they showed that in an out-of-sample experiment the lasso-based models performs better than dynamic factor models.

In summary, most literature provided evidence that model forecast is episodic and the forecasting performance depends on factors like the time period and features of data sample set.

Chapter 3

Data

We used a large macroeconomic dataset FRED-MD for this project. The dataset is publicly available at the Federal Reserve of St-Louis's website. It contains 128 variables with monthly data from the year 1959 to 2020.
The macroeconomic indicators are separated to eight groups: Output and Income, Labour market, Housing, 'Consumption, orders, and inventories', Money and Credit, Interest and exchange rates, Prices, Stock market.
The details of this dataset can be found on the paper by McCracken and Ng.[19]

3.1 Variables of interest

Consumer Price Index The Consumer Price Index (CPI) is the most used economic indicator for inflation measurement. It is released by the Bureau of Labour Statistics. It is used to measure the average change over time for the price of consumer items, goods and services for daily living. CPI has two main usages: 1. Can be used for social security payment adjustment. 2. Can act as the reference rate for some financial contracts as well as inflation swaps.

To calculate CPI, we first define eight major groups of goods and services: Housing, Apparel, Transportation, Education and Communication, Food and Beverages, Medical Care, Recreation and Other Goods and Services. Then take price changes for each item in the basket of goods and averaging them. In this way we can assess the price changes with the cost of living by investigating the changes in CPI.

Personal Consumption Expenditures The personal consumption expenditures (PCE) is released by the Bureau of Economic Analysis. It contains the actual and imputed expenditures of households, includes durable and non-durable goods and services data. As we mentioned in the introduction chapter, it is used as a indicator measurement for the FED to set inflation goal: 2%.

They are both very important in decision making processes and follows similar trends. However, the CPI tends to have higher inflation index/rate, as shown in Figure 3.1. The difference is that, CPI has used one set of expenditure weights for many years while PCE uses a Fisher price index that utilise current and previous expenditure data.

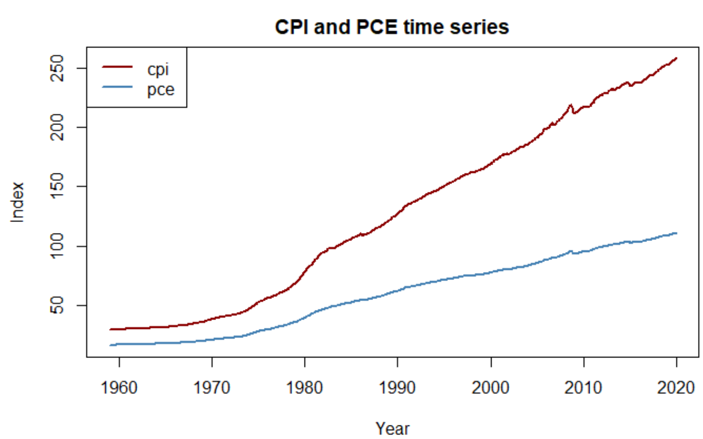


Figure 3.1: CPI and PCE time series plot.

3.2 Predictor Variables

For the data-rich model ARDL, we included five economics time series as the independent variables, they represent the collection of real economic activities.

The predictor variables are: Housing Start (HOUST, from Group: Housing), Industry Production (INDPRO, from Group: Output and Income), Nonfarm payment (PAYEMS, from Group: Labor market), Oil Price (OILPRICEx, from Group: Prices) and 3-month Treasury Bill (TB3MS, from Group: Interest and exchange rates).

Chapter 4

Methodology

In this chapter we introduce the methodology used to select model and tune hyper parameters. In addition, we present the evaluation of model performance.

4.1 Univariate model

For the univariate model we select the random walk (RW) model as our benchmark. RW is a simple form of ARIMA that often predicts economics time series well. It equals to ARIMA (0,1,0). If a more sophisticated model with assumptions cannot beat RW it cannot be used as a proper guide for policy, hence we chose it as the baseline model.

4.1.1 Baseline model: Random Walk

Random Walk is a time series that at each time point the series move randomly away from its current location.[20] It can be written as

$$y_t = y_{t-1} + \varepsilon_t$$

where ε_t is a white noise variable with zero mean and constant variance σ^2.

Random Walk with Drift which can be written as

$$y_t = c + y_{t-1} + \varepsilon_t$$

where c is the average of the changes between consecutive observations. An increase in the value of y_t implies positive c, and y_t will tend to drift upwards.

4.1.2 Autoregressive Integrated Moving Average

The ARIMA model with different parameters were developed and described by Box et al. in 1970[3]. We first introduce some concepts that are useful for understanding ARIMA.

1. Stationarity: A process is said to be stationary if there are no big changes at each time point. There should be no periodic variations. For example, a white noise series is stationary as it does not have trends or seasonality, it has no significant changes at any given time point.

2. Differencing: One way to make a non-stationary time series stationary is to compute the difference between consecutive observations.[20] It can be written as $y'_t = y_t - y_{t-1}$. Sometimes two differencing is needed to make the time series stationary: $y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$.

There is also seasonal differencing, it means the difference between one observation and the last one from the same season. It is written as: ${y_t}' = y_t - y_{t-s}$, where s is the period of seasonal pattern. For example, for monthly data s=12.

3. ARIMA model: An ARIMA(p, d, q) model can be written as

$$y'_t = c + \phi_1 y'_{t-1} + ... + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where y'_t is the differenced series, ε_t is the error term, p = order of the autoregressive part, d = degree of first differencing involved, q = order of the moving average part.[20]

4.1.3 Seasonal ARIMA

Box and Jenkins[3] proposed a variation of ARIMA model for seasonal time series data: sARIMA. A seasonal ARIMA includes additional seasonal terms $(P, D, Q)_s$. It allows seasonal parts to be modelled. In this paper we denotes sARIMA model as: (p, d, q)(P,D,Q)[s].

4.1.4 Parameters Selection

We first choose the order of differencing based on statistical test. That is, to investigate what level of differencing leads to stationary of data. We focus on two main tests, details of them can be found in Rob J Hyndman's book Forecasting: Principles and Practice[20]:

- Augmented Dickey-Fuller (ADF) test: The null hypothesis is that the series is non-stationary.
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test: The null hypothesis is that the data are stationary.

We reject the null hypothesis for p<0.05.

We then choose the hyperparameters p, q, P, Q based on information criteria AIC and BIC, they are defined as follows:

$$AIC = -2log(L) + 2(p + q + k + 1)$$

$$BIC = AIC + [log(T) - 2](p + q + k + 1)$$

where L is the likelihood of the data, $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$.[20] We want to choose the model that minimise the information criteria.

After we selected the hyper-parameters, the values for c, ϕ_t, θ_t will be estimated by the maximum likelihood estimation (MLE) method. MLE find the values of the parameters that maximise the probability of achieving the data that we have observed. That is, to minimise $\sum_{t=1}^{T} \varepsilon_t^2$.

4.2 Multivariate models

For multivariate model we used linear model with dynamic effect: Autoregressive Distributed Lag (ARDL) Model. Moreover, we introduced two types of penalty term that were used with the ARDL model.

4.2.1 Linear Regression model

Time series regression is a special case of classic regression model.[21] It has the form:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_k x_{kt} + \varepsilon_t, \qquad t = 1, ..., T$$

In vector form:

$$y_t = \beta_0 + \mathbf{X_t^T} \beta + \varepsilon_i$$

where $y_t \in \mathbb{R}$ is the dependent variable, $X_t = (x_{1t}, ..., x_{kt})^T \in \mathbb{R}^k$ is the set of predictors, $\varepsilon_i \sim N(0, \sigma^2)$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)^T$ is the set of parameters with β_0 a constant.

To estimate the parameters, we use ordinary least squares (OLS) which is to minimise the sum of squared residuals (SSR):

$$\hat{\boldsymbol{\beta}} = argmin_{\beta_0, ...\beta_k} \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{j=1}^{k} \beta_j x_{jt} \right)^2.$$

4.2.2 LASSO-type Penalised Regression

Least Absolute Shrinkage and Selection Operator (LASSO) was first proposed by Robert Tibshirani in 1996[22]. The aim of using LASSO is to estimate a model that could produce forecasts with small variance as well as identify the set of predictors that explains the dependent variable better.

To obtain LASSO estimates, we minimise the sum of squared residuals subject to a penalty term of L_1 norm of the coefficients:

$$\hat{\boldsymbol{\beta}}^{\boldsymbol{LASSO}} = argmin_{\beta_0,\ldots\beta_k} \left\{ \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{j=1}^{k} \beta_j x_{jt} \right)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \right\}$$

where $\lambda \geq 0$ is a penalty/tuning parameter. When $\lambda = 0$, the LASSO estimates are the same as OLS estimates.

4.2.3 adaLASSO

The adaptive Lasso was propsed by Zou (2006):[23]

$$\hat{\boldsymbol{\beta}}^{\boldsymbol{adaLASSO}} = argmin_{\beta_0,\ldots\beta_k} \left\{ \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{j=1}^{k} \beta_j x_{jt} \right)^2 + \lambda \sum_{j=1}^{k} w_j \, |\beta_j| \right\}$$

where $\mathbf{w}$ is a known weights vector.

The idea behind adaptive Lasso is to assign different weights to different coefficients and hence the weights are more "tailored" to the coefficients.

Notice that the selection of parameter λ is done with k-fold cross validation, see details of CV in section 4.3.

4.2.4 Autoregressive Distributed Lag Model

An ARDL(p,q) model is a linear model with dynamic information. It assumes that a time series Y_t can be represented by a linear function of p lags of the dependent variable and q lags of l additional predictors where $l = 1, ..., k$:

$$\begin{aligned} Y_t = \beta_0 &+ \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} \\ &+ \delta_{11} X_{1,t-1} + \delta_{12} X_{1,t-2} + ... + \delta_{1q} X_{1,t-q} \\ &+ ... \\ &+ \delta_{k1} X_{k,t-1} + \delta_{k2} X_{k,t-2} + ... + \delta_{kq} X_{k,t-q} \\ &+ \epsilon_t \end{aligned} \tag{4.1}$$

It was suggested in a study by Min B. Shresthaa and Guna R. Bhatta[1] that for mixed of stationary and non-stationary variables using ARDL model is recommended. Their suggestion for model selection can be summarised in the image below.

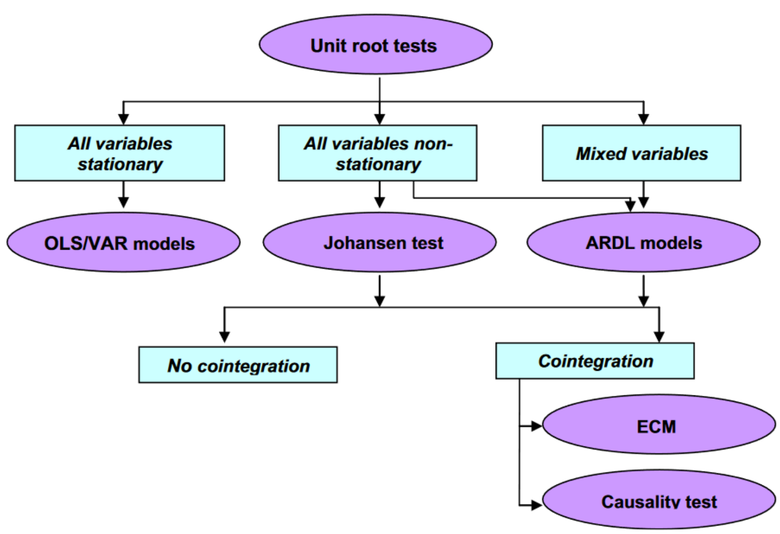


Figure 4.1: Model selection for time series data.[1] where OLS: Ordinary least squares; VAR: Vector autoregressive; ARDL: Autoregressive distributed lags;ECM: Error correction models.

4.3 Cross Validation

A very common statistical technique for variable selection is cross validation (CV). The idea of CV is splitting the sample into two subsets, the in-sample set and out-of-sample set. Then use only the in-sample set to estimate parameters on the model and measure model's performance on the out-of-sample set.

There are mainly two ways to split the sample: exhaustive (leave-n-out) and non-exhaustive (K-fold) CV.[21]

To use leave-n-out CV we set n observations to be the testing set and use the other observations to estimate the parameters of the model. When n=1, we call the leave-one-out CV a time series cross validation method[20]. In this procedure, the series of test sets each contain a single observation. Each corresponding training set contains only observations that occurred up to the first observation of the testing set.

The K-fold CV is different, as it split the original sample into K non-overlapping folds/subsamples. Then one of the subsamples is removed and the model is estimated with the remaining K-1 subsamples. Each time the out-of-sample forecasts are produced with the estimated model. The forecasting error will be kept. This procedure is done for each subsample which leads to K forecasting error.

In this dissertation, we used different types of CV method based on different situation.

First, we used K-fold CV is to select our penalty parameters. The optimal λ is the one that minimises the forecasting error. Next we explain this in a mathematical way. We follow the notations from Peter Fuleky's Macroeconomic Forecasting in the Era of Big Data.[21]

Let $\mathbf{V} \subseteq \{1, ..., T\}$ be the indices of data in the testing set and let $\mathbf{T} \subseteq \{1, ..., T\}$ be the indices of data in the training set. Let $\hat{\boldsymbol{\beta}}_{\boldsymbol{T}}(\lambda)$ be the parameter estimate based on the training data $\mathbf{T}$ using the tuning parameter

λ. For each λ,

$$CV(\lambda, V) = \sum_{t \in V} (y_t - \mathbf{x_t'} \hat{\boldsymbol{\beta}}_{\boldsymbol{T}}(\lambda))^2$$

is the corresponding prediction error on the testing set $\mathbf{V}$.

Let $\mathbf{V} = \{V_1, ..., V_B\}$ be the collection of testing sets with corresponding training sets $\{T_1, ..., T_B\}$. The CV error for the parameter λ is calculated as

$$CV(\lambda) = \sum_{i=1}^{B} CV(\lambda, V_i).$$

We choose the optimal λ where

$$\hat{\lambda} \in argminCV(\lambda).$$

We applied leave-one-out (time series) CV method for measuring model's forecasting ability, this is discussed in chapter 4.4.1.

4.4 Model Evaluation

We designed a pseudo out-of-sample forecasting experiment. The model performance is measured with two evaluation metrics.

4.4.1 Pseudo out-of-sample forecasting

To simulate the case of a real-time forecaster, we used a pseudo out-of-sample forecasting experiment. That is to conduct all model specification and estimation with observation up to date t, make a h-step ahead forecast for date $t + h$. Then move to date $t + 1$ and carry out the same procedure through the sample. Stock and Watson (2008)[7] states that the advantages of using pseudo out-of-sample forecast evaluation includes better capturing model specification uncertainty, instability and estimation uncertainty.

Moreover, we apply a time series CV (as described in chapter 4.3) to avoid look-ahead-bias as this method prevent the use of future observation.

Finally, for the model estimation we used rolling estimation (moving window with fixed number of observations) rather than recursive estimation (expanding window starting from the same observation).

4.4.2 Evaluation Metrics

There are several measures for evaluating the accuracy of forecast, see e.g. [24]. They have different characteristics and can be used for different purposes. We focus on the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE), they are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

where $\hat{y}_t$ is the predicted y.

Both RMSE and MAE are negatively oriented, i.e. the lower the value the higher the prediction accuracy. They are also both on the same scale as the variable of interest. The difference is that RMSE gives more weight to large errors while MAE does not. They are both very important indicators for measuring model performance. For example, lower RMSE might shows that the model is better at predicting extreme cases, while a lower MAE might show that the model is more robust. Hence, we include both of them for comparison.

Chapter 5

Empirical Application and Results

In this chapter we present the uses of our methodology described in chapter 4, as well as the results we achieved.

In our empirical study, we investigate the forecasting ability of different models in different time period with short-term horizons (h=1, 3) and medium to long-term horizons (h=6, 12) for two inflation measures: CPI and PCE. We aim to investigate the difference of model performance in different circumstances.

5.1 Empirical setup

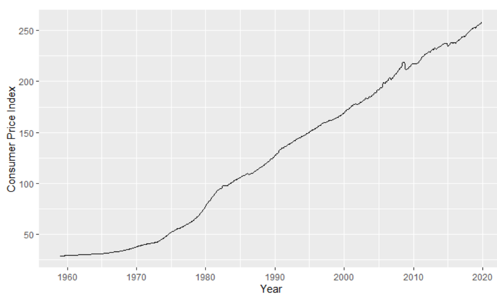
1. Forecast objective: The aim is to forecast inflation level in the US. We transformed the original price index series to monthly year-on-year consumer price inflation following:

$$\pi_t = \Delta_{12} log(CPI_t) = log(CPI_t) - log(CPI_{t-12})$$

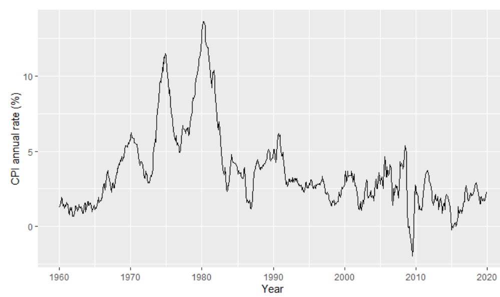
$$\pi_t = \Delta_{12} log(PCE_t) = log(PCE_t) - log(PCE_{t-12}).$$

We chose the year-on-year inflation because the US government uses year-on-year inflation as the target variable. After transformation we lose the first 12 observations, so the transformed series starts from 1960.

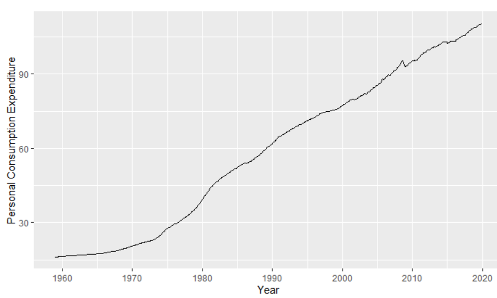
The plots for the original series and transformed (annual monthly rate) series are shown below:



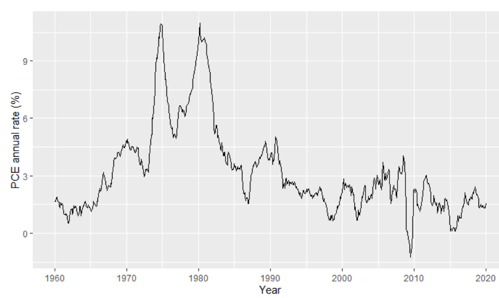
(a) CPI



(b) CPI annual rate



(c) PCE



(d) PCE annual rate

Figure 5.1: Inflation index and their transformation.

2. Predictor variables transformation: Following on the forecast objective with the transformation on CPI, we have also transformed the predictor variables. We follow McCrackenand Ng (2016)[19] in the choice of transformations. The details on the dataset and the series transformation are all in McCracken and Ng (2016)[19]. Moreover, we performed ADF test on the transformed data, all series after transformation passed the ADF test of being stationary. The transformation is summarised on the table below:

Variables	Transformation	Definition
Inflation variables		
CPI	$100\Delta_{12}$ln	Consumer Price Index: Total All Item for the US
PCE	$100\Delta_{12}$ln	Personal Consumption Expenditure: Chain Index
Predictor variables		
TB3M	Δ	3 Month Treasury Bill: Secondary Market Rate
HOUST	ln	Housing Starts: Total New Privately Owner
INDPRO	$100\Delta_{12}$ln	Industry Production Index
PAYEMS	$100\Delta_{12}$ln	Total Nonfarm Payroll
OILPRICEx	$100\Delta_{12}$ln	Crude Oil, spliced WTI and Cushing Spot market price

Table 5.1: Predictor variables transformation. Transformations are: (i) $100\Delta_{12}$=100 $ln(P_t/P_t - 12)$, (ii) Δ: $X_t = P_t - P_{t-1}$, and (iii) $ln : X_t = ln(P_t)$.

3. Forecast horizon: We consider both short-term forecast where h=1, 3 and long-term forecast where h=6, 12.

4. Pseudo-Out-of-Sample periods: The pseudo-out-of-sample period is 2006M1 - 2019M12. The sub-periods are the period before Great Recession (2005M1 - 2007M11), the Great Recession (2007M12 - 2009M06), as well as the period after Great Recession (2009M07 - 2019M12).

The volatility of inflation in each time period is presents in the Table 5.2. It can be seen that for both CPI and PCE, during recession period the volatility is more than twice higher of the other times. This could be the reason why forecasting inflation is very hard for recession period; it is also the reason why we have included sub-samples for model evaluation: to test the model's robustness in more extreme circumstances.

	Range	Volatility_CPI	Volatility_PCE
Full pseudo-out-of-sample	2006M01 - 2019M12	1.3017	0.9966
Before Great Recession	2005M01 - 2007M11	0.8478	0.5587
During Great Recession	2007M12 - 2009M06	2.4359	1.8352
After Great Recession	2009M07 - 2019M12	0.9924	0.7667

Table 5.2: Volatility/ Standard deviation in each time period for CPI and PCE rate.

5.1.1 Building sARIMA model

First, we want to test the stationarity of the training set. We implement the hypothesis test described in Chapter 4.1.4: ADF and KPSS test. The resulted p-value for CPI, CPI with one differencing, PCE, PCE with one differencing are presented in Table 5.3.

	CPI_train	diff(CPI_train)	PCE_train	diff(PCE_train)
ADF test	0.1968	0.01	0.2481	0.01
KPSS test	0.01	0.1	0.01	0.1

Table 5.3: Resulted p-value from adf test and kpss test, reject the null hypothesis when p-value is less than 0.05.

We conclude from Table 5.3 that both series are stationary after one differencing, which indicates d=1 for our ARIMA model.

Next, we utilise the ACF & PACF plot to determine the value of hyper-parameters. The plots for CPI and PCE are shown in Figure 5.2 and Figure 5.3. They have similar patterns; we present our observations from the plots and discuss how we decided for the hyper-parameters.

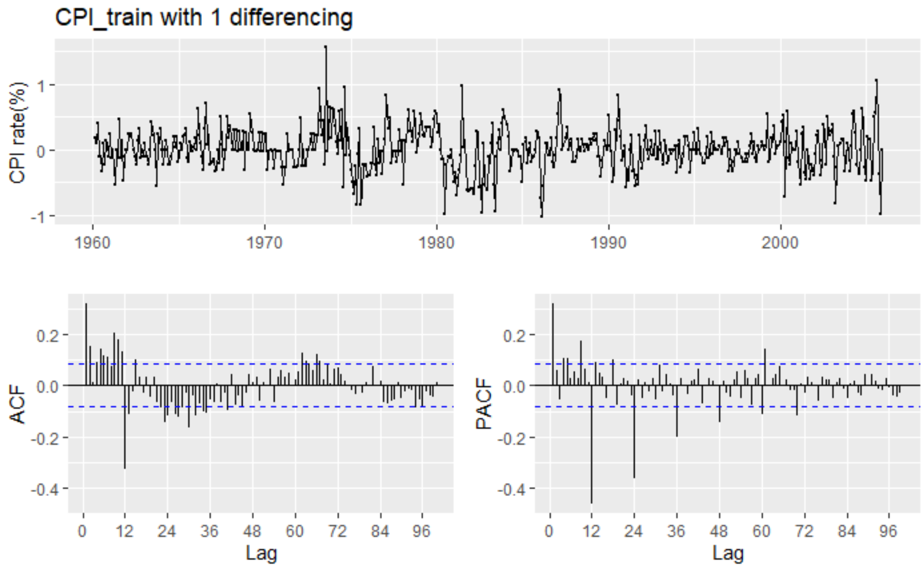


Figure 5.2: CPI training set after one differencing together with its ACF & PACF plot

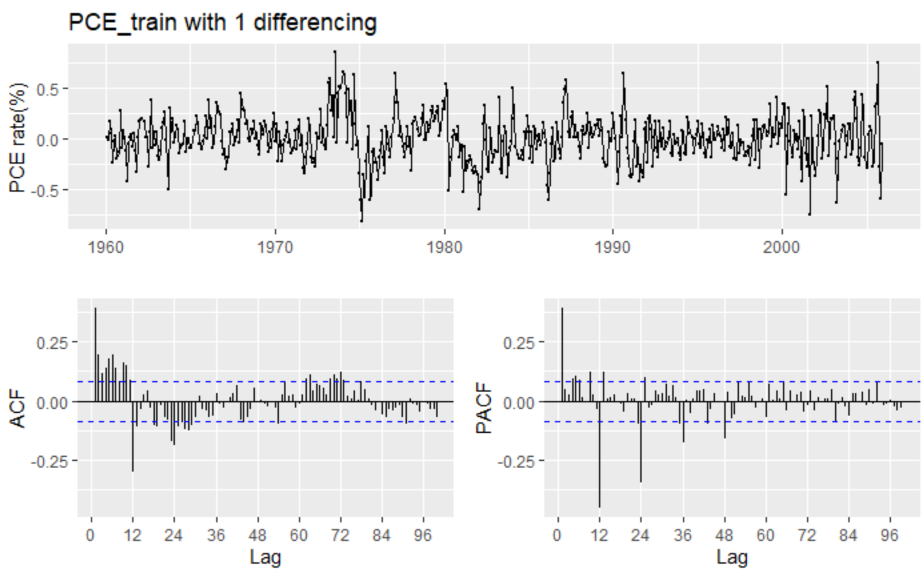


Figure 5.3: PCE training set after one differencing together with its ACF & PACF plot

Observations from the plots:

- The exponential decay in the seasonal lags of the PACF suggests a seasonal MA(1) component.
- The significant spike at lag 2 in the ACF suggests a non-seasonal MA(2) component.

Therefore, we started with a seasonal ARIMA(0,1,2)(0,1,1)[12] model. Based on our observations, we applied grid search and identified two final sARIMA models based on lowest AIC & BIC.

The selected models are

- sARIMA_aic:(1,1,6)(1,1,2)[12]
- sARIMA_bic:(1,1,2)(0,1,2)[12]

Their residual plots are shown in Figure 5.4 and Figure 5.5. Since almost all spikes are within the significance limits and the residuals have normal distribution, the residuals are approximately white noise. Hence these two models can be used for forecasting inflation rate.

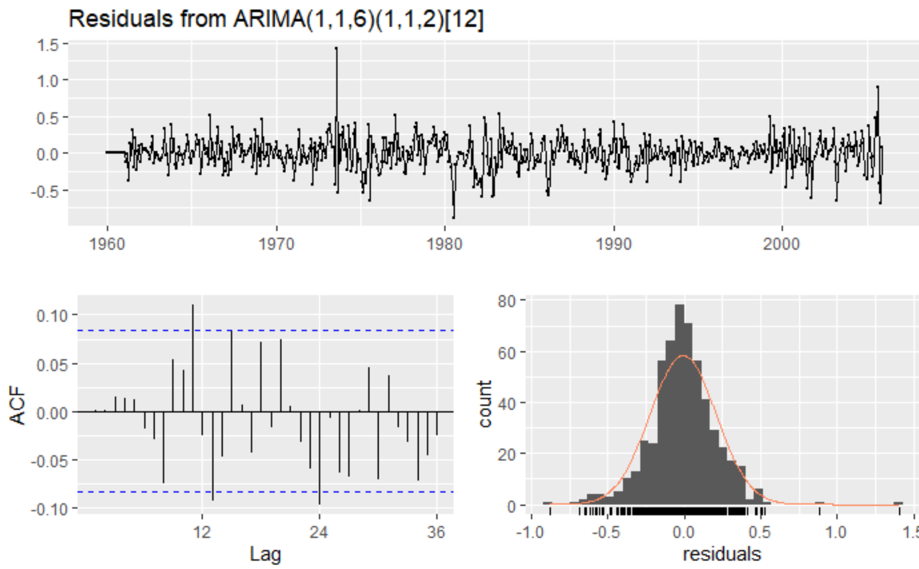


Figure 5.4: The residual plot of model selected by lowest AIC.

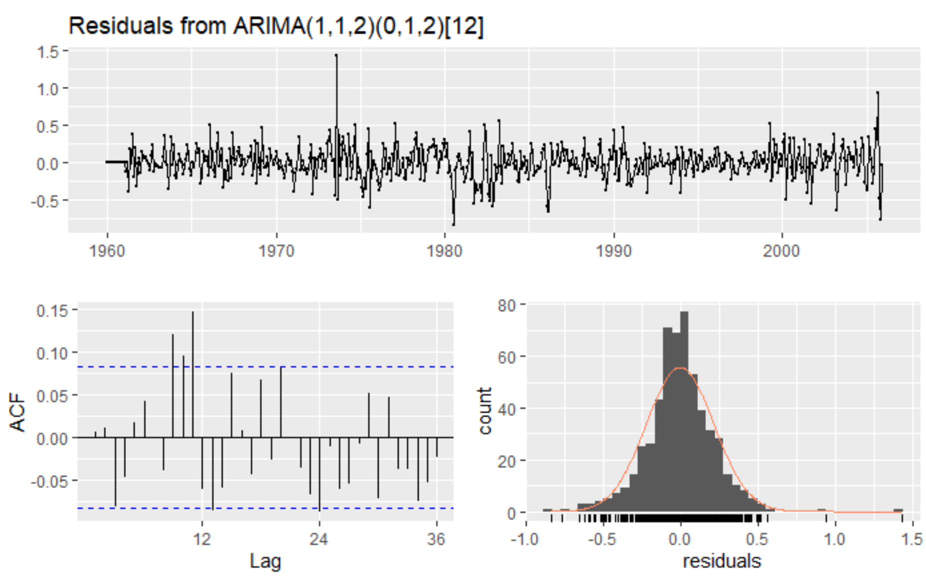


Figure 5.5: The residual plot of model selected by lowest BIC.

5.1.2 Building ARDL model

To build ARDL model, we first selected the lag orders of each variables with AIC, then we perform regression with all of their lags. Finally, we select the significant lags (p-value < 0.05) of these variables. Notice that these steps are all done with the training dataset.

For CPI, the used variables and their lags are: CPI: 1, 3, 9, 12, 13; 3-month treasury bill: 2, 8, 10; Housing start: 2; Industry production: 3; Nonfarm Payroll: 3, 6; Oil price: 2, 7, 12, 13.

For PCE, the used variables and their lags are: PCE: 1, 2, 12, 13; 3-month treasury bill: 1, 7, 10; Housing start: 6; Industry production: 3; Nonfarm Payroll: 3, 4, 6; Oil price: 2, 7, 12, 13.

For the model estimation and forecast evaluation we followed the methodologies described in Chapter 4. For details of pseudo out-of-sample design see Chapter 4.4.1. For details of the penalty terms we used see Chapter 4.2.2.

5.1.3 Computer Implementation

All data processing, visualisation and model building are implemented with packages in R.

Plots are produced with the `ggplot2` R-package. [25] We use `lubridate` for manipulating dates. [26]

The Arima models were built with the `forecast` R-package. [27] For the implementation of ARDL model with penalty term and rolling forecast function we used `LasForecast` R-package. [28]

5.1.4 Empirical Results

The empirical results are summarised in the two tables below. They compare models in 5 dimensions: 1. Two types of model: data-poor and data-rich. 2. Four forecasting horizons. 3. Four pseudo out-of-sample periods. 4. Two evaluation metrics: RMSE and MAE. 5. Two inflation measurements.

Full out-of-sample period: 2006M1 - 2019M12						
CPI	RW	sARIMA_aic	sARIMA_bic	ARDL_ols	ARDL_lasso	ARDL_adalasso
h=1	0.5857	0.2984	0.2884	0.7238	0.7214	0.7441
h=3	1.0841	0.8030	0.7610	0.7693	0.7637	0.7694
h=6	1.5188	1.3896	1.2844	0.7786	0.7700	0.7708
h=12	1.9970	2.2594	2.0373	0.7624	0.7571	0.7464
PCE						
h=1	0.3678	0.2047	0.1973	0.4804	0.4842	0.5297
h=3	0.7475	0.5702	0.5288	0.5060	0.5095	0.5418
h=6	1.1091	1.0378	0.9395	0.5164	0.5186	0.5431
h=12	1.4886	1.7752	1.5521	0.5091	0.5086	0.5375
Before recession: 2006M1 - 2007M11						
CPI	RW	sARIMA_aic	sARIMA_bic	ARDL_ols	ARDL_lasso	ARDL_adalasso
h=1	0.5893	0.3221	0.3249	0.6974	0.7397	0.7591
h=3	1.0984	0.6707	0.7185	0.7053	0.7497	0.7694
h=6	1.2065	0.9962	1.0736	0.7062	0.7467	0.7708
h=12	1.5463	1.5776	1.5860	0.7176	0.7633	0.7850
PCE						
h=1	0.3864	0.2098	0.2170	0.4610374	0.4951	0.5680
h=3	0.7513	0.4788	0.5129	0.4647	0.4997	0.5730
h=6	0.8040	0.6953	0.7458	0.4632	0.4990	0.5726
h=12	1.0293	1.2180	1.2069	0.4690	0.5019	0.5786
During recession: 2007M12 - 2009M6						
CPI	RW	sARIMA_aic	sARIMA_bic	ARDL_ols	ARDL_lasso	ARDL_adalasso
h=1	0.7622	0.5906	0.5698	1.1169	1.1378	1.2355
h=3	1.9676	1.7464	1.7126	1.1352	1.1516	1.2633
h=6	3.1990	3.2255	3.1650	1.1112	1.1324	1.2366
h=12	4.5608	4.6294	4.1800	1.1018	1.1230	1.2241
PCE						
h=1	0.4978	0.3974	0.3758	0.6773	0.7231	0.8981
h=3	1.3662	1.2342	1.1856	0.6910	0.7399	0.9062
h=6	2.3713	2.3518	2.3265	0.6859	0.7344	0.9046
h=12	3.4913	3.7746	3.6413	0.6808	0.7391	0.9007
After recession: 2009M7 - 2019M12						
CPI	RW	sARIMA_aic	sARIMA_bic	ARDL_ols	ARDL_lasso	ARDL_adalasso
h=1	0.4669	0.2147	0.2099	0.7824	0.7844	0.7958
h=3	0.8492	0.5443	0.4989	0.8139	0.8058	0.8099
h=6	1.1824	0.8854	0.7897	0.7913	0.7804	0.7693
h=12	1.2808	1.3813	1.2281	0.7753	0.7665	0.7553
PCE						
h=1	0.3956	0.1547	0.1518	0.4501	0.4393	0.4392
h=3	0.6519	0.3759	0.3576	0.4824	0.4713	0.450
h=6	0.9206	0.6255	0.5982	0.4956	0.4822	0.4492
h=12	1.0038	0.9447	0.9272	0.4847	0.4716	0.4495

Table 5.4: The table shows RMSE of each model with four forecast horizons: 1, 3, 6, 12 in four time periods for two inflation measurements: CPI and PCE. The smallest RMSE for each model at different forecast horizons are shaded.

Full out-of-sample period: 2006M1 - 2019M12						
CPI	RW	sARIMA_aic	sARIMA_bic	ARDL_ols	ARDL_lasso	ARDL_adalasso
h=1	0.4589	0.2117	0.2014	0.5058	0.5050	0.5155
h=3	0.7513	0.5463	0.5209	0.5278	0.5264	0.5258
h=6	1.1010	0.9695	0.8843	0.5290	0.5251	0.5245
h=12	1.5284	1.6737	1.4747	0.5182	0.5118	0.5021
PCE						
h=1	0.2767	0.1452	0.1386	0.3347	0.3311	0.3486
h=3	0.5042	0.3990	0.3669	0.3458	0.3421	0.3554
h=6	0.7760	0.7160	0.6513	0.3488	0.3441	0.3571
h=12	1.1336	1.2613	1.1202	0.3437	0.3361	0.3552
Before recession: 2006M1 - 2007M11						
CPI	RW	sARIMA_aic	sARIMA_bic	ARDL_ols	ARDL_lasso	ARDL_adalasso
h=1	0.4529	0.2490	0.2400	0.4926	0.5200	0.5390
h=3	0.7734	0.5025	0.5441	0.4980	0.5251	0.5429
h=6	0.9881	0.8519	0.9004	0.5002	0.5236	0.5449
h=12	1.4476	1.2384	1.1338	0.5056	0.5312	0.5504
PCE						
h=1	0.2882	0.1588	0.1576	0.3502	0.3626	0.3997
h=3	0.5367	0.3832	0.3796	0.3502	0.3640	0.4021
h=6	0.6713	0.5958	0.6189	0.3471	0.3654	0.4012
h=12	0.9482	0.9321	0.9138	0.3517	0.3640	0.4058
During recession: 2007M12 - 2009M6						
CPI	RW	sARIMA_aic	sARIMA_bic	ARDL_ols	ARDL_lasso	ARDL_adalasso
h=1	0.4554	0.4174	0.4030	0.7824	0.7844	0.7958
h=3	1.308	1.2025	1.2636	0.8139	0.8058	0.8099
h=6	2.5339	2.4652	2.4971	0.7913	0.7804	0.7693
h=12	4.4752	4.4882	4.0163	0.7753	0.7665	0.7558
PCE						
h=1	0.3101	0.2829	0.2683	0.4542	0.4738	0.5516
h=3	0.9696	0.8674	0.8650	0.4619	0.4838	0.5602
h=6	1.8888	1.7702	1.7823	0.4592	0.4732	0.5719
h=12	3.4378	3.679	3.5426	0.4518	0.4722	0.5723
After recession: 2009M7 - 2019M12						
CPI	RW	sARIMA_aic	sARIMA_bic	ARDL_ols	ARDL_lasso	ARDL_adalasso
h=1	0.3649	0.1706	0.1645	0.4670	0.4622	0.4686
h=3	0.5864	0.4361	0.3935	0.4905	0.4852	0.4787
h=6	0.8364	0.7226	0.6249	0.4927	0.4882	0.4819
h=12	1.0535	1.0967	0.9648	0.4776	0.4653	0.4522
PCE						
h=1	0.3255	0.1179	0.1167	0.3148	0.3081	0.3076
h=3	0.4721	0.3015	0.2887	0.3277	0.3202	0.3138
h=6	0.6470	0.5039	0.4845	0.3305	0.3204	0.3135
h=12	0.8191	0.7599	0.7535	0.3231	0.3112	0.3145

Table 5.5: The table shows MAE of each model with four forecast horizons: 1, 3, 6, 12 in four time periods for two inflation measurements: CPI and PCE. The smallest MAE for each model at different forecast horizons are shaded.

Chapter 6

Discussion and Conclusion

In this chapter we discuss the empirical results we obtained in Chapter 5.1.4 as well as our experiments and finally have a conclusion for this paper.

6.1 Discussion

6.1.1 Discussion of the result

Observations from the table for CPI:

- For univariate model, in general as the forecasting period increase, forecasting error also increase.

- For short-term forecast h=1 the model with lowest RMSE and MAE is sARIMA selected by BIC, except for the period before Great Recession where sARIMA_bic has RMSE 0.0028 higher than sARIMA_aic. For short-term forecast h=3, RMSE shows the same result as for h=1 while for in the MAE table, sARIMA_aic has the lower value in both before recession and during recession periods.

- For medium-term forecast h=6, both table reflects the same best model as for h=3. For long-term forecast h=12, one different observation from before is that the benchmark model Random Walk outperform the others for the full out-of-sample period and before recession period.

- For multivariate model, it can be seen that they starts to outperform univariate model from h=3, 6. In addition, the forecasting error does not necessarily increase as forecasting period increase.

- For short-term forecast h=1, RMSE table shows that except for the full out-of-sample period that ARDL_lasso outperform the other models, ARDL_ols is the best model. MAE table has similar results except for after recession period where ARDL_lasso outperforms the others. For short-term forecast h=3, RMSE table shows similar result for as for h=1. Except for the after-recession period that ARDL_lasso performs better than ARDL_ols. For MAE table there exist three different results compares to the case h=1.

- For medium-term forecast h=6, the RMSE table shows same result as for h=3 except for after recession period. While it is the recession period for MAE table. For long-term forecast h=12, RMSE shows same result as for h=6 except for full out-of-sample period while for MAE table there are no exception.

Observations from the table for PCE:

- For univariate model, results of PCE is consistent as of CPI, except for the recession period. For example, for the PCE measurement, in both RMSE and MAE table random walk outperforms the other univariate models when h=12.

- For multivariate model more difference occurred. The results are consistent only for the before recession period.

6.1.2 Discussion of the study

We discuss the strengths and limitations of our study.

Strengths

- We have different scenarios for the model comparison. That is, we included different forecasting horizon, time period, inflation measurement and error measure, to see how models perform under different circumstances.

- We investigated the relationship between information criteria (for univariate model) and penalty term (for multivariate mode) and model performance.

- We implemented time series cross validation for model estimation and model evaluation to simulate the real forecasting process.

Limitations and future work

- For the multivariate model, we did not utilise all predictor variables in the dataset. The accuracy might be higher if we were to use more of them. In addition, having more predictor variables can help us better mimic the real-world scenario. One approach to include "big data" in time series forecasting is to use Principal component analysis (PCA) for dimension reduction.[21]

- Our method of selecting predictor variables for ARDL model might not be very effective in terms of improving accuracy. For example, the inclusion of oil price only had little (in some circumstances none) help in increasing model accuracy even though this variable and its lags appears to be significant in the model. We could include more experiments to test the strength of each variable in improving model performance.

- We could include the recursive forecasting method as well for better comparison between models in future work to make the study more comprehensive.
- Our study considered only linear model. In the future we can include machine learning model which has the non-linearity feature.

6.2 Conclusion

We have the following conclusions based on the results in Chapter 5.1.4 and the discussion we have in Chapter 6.1.

First, our results also confirmed the Stock and Watson (2008)[7] study. That the performance of Phillips curve forecasts is episodic. In fact, the same result can also apply to the sARIMA model.

Second, for short-term forecast h=1, univariate model sARIMA is the model with highest forecasting accuracy. For h=3 the the model with lowest error differ in different time period.

Third, for medium to long-term forecast h=6, 12, both RMSE and MAE indicates a better performance of ARDL model. Especially during recession period, the ARDL model is more stable than sARIMA.

Fourth, the forecasting error for PCE is smaller than CPI, which indicates that it is easier to forecast PCE rate than CPI rate.

In general, inflation is harder to forecast during recession as we can see the forecasting error is highest for that period. In terms of univariate model, the result shows that in some cases for long-term forecast random walk can be useful while sARIMA model selected by bic has a higher frequency of being selected as the best performing model. In terms of multivariate model, we found that penalty term can improve the ARDL model's performance in some cases. However, factors like time period, forecasting horizon and inflation measurement can all affect the result.

Bibliography

[1] Shrestha MB, Bhatta GR. Selecting appropriate methodological framework for time series data analysis. The Journal of Finance and Data Science. 2018;4(2):71–89.

[2] Reserve F. What are the Federal Reserve's objectives in conducting monetary policy?; 2012. Accessed 2020-08-26. Available from: https://www.federalreserve.gov/faqs/money_12848.htm.

[3] Box GE, Jenkins GM, Reinsel G. Time series analysis: forecasting and control Holden-day San Francisco. BoxTime Series Analysis: Forecasting and Control Holden Day1970. 1970;.

[4] Landsman WR, Damodaran A. A comparison of quarterly earnings per share forecasts using James-Stein and unconditional least squares parameter estimators. International Journal of Forecasting. 1989;5(4):491–500.

[5] Stockton DJ, Glassman JE. An evaluation of the forecast performance of alternative models of inflation. The review of economics and statistics. 1987;p. 108–117.

[6] Meyler A, Kenny G, Quinn T. Forecasting Irish inflation using ARIMA models. 1998;.

[7] Stock JH, Watson MW. Phillips curve inflation forecasts. National Bureau of Economic Research; 2008.

[8] Banbura M, Mirza H. Forecasting euro area inflation with the Phillips curve. manuscript, European Central Bank. 2013;.

[9] Phillips AW. The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957. economica. 1958;25(100):283–299.

[10] Gordon RJ. The Phillips curve now and then. National Bureau of Economic Research; 1990.

[11] Stock JH, Watson MW. Forecasting inflation. Journal of Monetary Economics. 1999;44(2):293–335.

[12] Atkeson A, Ohanian LE, et al. Are Phillips curves useful for forecasting inflation? Federal Reserve bank of Minneapolis quarterly review. 2001;25(1):2–11.

[13] Fisher JD, Liu CT, Zhou R. When can we forecast inflation? Economic Perspectives-Federal Reserve Bank of Chicago. 2002;26(1):32–44.

[14] Stock JH, Watson MW. Why has US inflation become harder to forecast? Journal of Money, Credit and banking. 2007;39:3–33.

[15] Canova F. G-7 inflation forecasts: Random walk, Phillips curve or what else? Macroeconomic Dynamics. 2007;11(1):1–30.

[16] Ang A, Bekaert G, Wei M. Do macro variables, asset markets, or surveys forecast inflation better? Journal of monetary Economics. 2007;54(4):1163–1212.

[17] Stock JH, Watson MW. Modeling inflation after the crisis. National Bureau of Economic Research; 2010.

[18] Li J, Chen W. Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. International Journal of Forecasting. 2014;30(4):996–1015.

[19] McCracken MW, Ng S. FRED-MD: A monthly database for macroeconomic research. Journal of Business & Economic Statistics. 2016;34(4):574–589.

[20] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts; 2018.

[21] Fuleky P. Macroeconomic Forecasting in the Era of Big Data: Theory and Practice. vol. 52. Springer; 2019.

[22] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(1):267–288.

[23] Zou H. The adaptive lasso and its oracle properties. Journal of the American statistical association. 2006;101(476):1418–1429.

[24] Shcherbakov MV, Brebels A, Shcherbakova NL, Tyukov AP, Janovsky TA, Kamaev VA. A survey of forecast error measures. World Applied Sciences Journal. 2013;24(24):171–176.

[25] Wickham H. ggplot2: elegant graphics for data analysis. springer; 2016.

[26] Grolemund G, Wickham H, et al. Dates and times made easy with lubridate. Journal of statistical software. 2011;40(3):1–25.

[27] Hyndman RJ, Khandakar Y, et al. Automatic time series for forecasting: the forecast package for R. 6/07. Monash University, Department of Econometrics and Business Statistics ...; 2007.

[28] Gao Z. LasForecast. GitHub; 2019. Accessed 2020-08-10. https://github.com/zhan-gao/LasForecast.