FORECASTING THE US INFLATION RATE

Faith Hill

Bloti B. Teh

Instructor: Dr. Ran Ji

Faith Hill, Bloti B. Teh, and Huy Nguyen are graduate students at **George Mason University | College of Engineering and Computing | Volgenau School of Engineering | Department of Information Sciences and Technology | Fairfax Campus** ; Data Analytics Engineering | Operations Research (OR)/SYST 538-001| Analytics for Financial Engineering and Econometrics; bteh@gmu.edu; fhill2@gmu.edu; hnguy40@gmu.edu;

Abstract

We forecast price inflation, by examining available datasets from the Bureau of Labor Statistics (Consumer Price Index) with the application of various predictions or forecasting tools available.

We implement the methodology of a Long-Memory Process and Fractional Differencing and FARIMA Modeling and Linear Regression to model and describe the behavior of consumption and checks for residual autocorrelation.

Keywords: Inflation; Economic Uncertainty; Forecasting; Inflation Forecasting

I. Introduction

We define the US inflation rate by year as the percentage of change in products and rendered services from one year to another and it is the natural rise and fall of every thriving economy. "Inflation is hard to forecast." Considering a widespread of empirical evidence, this statement made by Stock and Watson (2003), to date is considered as a stylized fact in the study if macroeconomics. Thus, inflation forecasts are important central bank outputs. Far beyond academia, lies a full-fledged commitment by the Federal Reserve (Fed) to price stability, the development of models that accurately describe the underlying data-generating mechanism of inflation and its relevance to policy makers [1]. When it comes to policy making, we argue that there is a strong relation between economic uncertainty and inflation. More precisely, the data-generating properties of inflation (persistence and volatility) are linked to economic authorities and uncertainties in policy. If a Central Bank targets inflation using interest rates, high policy uncertainties can affect and firm decisions [2]. For example, if there are lingering uncertainties concerning looming interest rates increase, consumers might discount a future hike and cut back on spending and hence inflation is in the lead prior to policies and decisions made [3].

To emphasize the widespread nature of inflation, there is a comprehensive literature covering forecasting and inflation, and therefore we consider the term spread in this research project. Proceeding through this research, we breakdown the term spread into three main components: the expected real change, the expected inflation changes, and consider the term premium.

The empirical findings suggest that only the long-term rates contain valuable information on inflation expectations and why the term spread can forecast inflation is exposited [4]. On the other hand, the effect of a low spread is restrictive on monetary policy. Consequently, when the spread is low on short-term interest rates there is a relatively high long-term interest rates, and invariably, ironic that monetary policy will considerably slow down and inflation will decrease [5]. From the consideration of a univariate time series perspective, there is an unprecedented outcome which attests that US inflation rate is best captured by a long-memory process [6]. History also assumes that US inflation forecasts can be either (1) or (0) [7].

The need for long-memory stationary models is proven to be very imperative. Contrary to ARMA processes where stationary models are used to model stationary time series and they also have only short memories in that their auto-correlation functions decay to zero exponentially fast. Therefore, there exist a $D > 0 \; and \; r > 1 \; such \; that \; p(k) < D|r|^k \; for \; all \; k$. On the contrary, many financial time series appear to have long memory since their ACFs decay at a slow polynomial rate rather than a fast geometric rate, that is $p(k) \sim Dk^{-a} \; for \; some \; D \; and \; a > 0$. A polynomial rate of decay is sometimes called hyperbolic rate [8].

Given a recent evidence, economic policy uncertainty affects in part the patterns of inflation forecasting models in that they do not directly control policy uncertainty and thus often mis specified [9]. In this project we implement Fractional Differencing – one of the most widely used models for stationary, long-memory processes. Furthermore, we maintain a stance that there is a relationship between economic uncertainty and inflation. The principal result of our analysis is that US inflation and consumer inflation expectations have an important shortcoming while also exploring the feasibility of eliciting policy uncertainty to contain some direct information about the distribution of points and an individual's uncertainty about future inflation.

II. Objective

To achieve the goal of the project - forecast price inflation, we examine available datasets from the Bureau of Labor Statistics (Consumer Price Index) with the application of various predictions or forecasting tools available. Then conduct model comparison and identify the best model in forecasting price inflation. We will implement the methodology of a Linear Regression model, Long-Memory Process using Fractional Differencing and $FARIMA\ Modeling$ to model and describe the behavior of consumption and checks for residual autocorrelation.

III. Data

The dataset we explored was taken from US Bureau of Labor Statistics (BLS) – Consumer Price Index (US City Average) spanning the period of 2010 – 2022. The data represents measures of inflation for all CPI items ranging from food and beverages, transportation, housing, apparel, medical care, education and communication, other goods and services, recreation, energy, commodities, services, nondurables, and durables. The data is under the series title: All items in US city average, all urban consumers, chained, not seasonally adjusted – Chained CPI for All Urban Consumers (CCPIU) [10]. Another set of data also incorporated in our project from BLS the gold price daily from NAQDAQ, bond rate, CPI data from the US Bureau of Labor Statistics – Consumer Price Index, SP500 index from an R package (tidyquant) and daily oil price from the U.S. Energy Information Administration (EIA).

IV. Methodology

To model and determine the long-and short-term dependence of variables and interdependencies between inflation and uncertainty and model comparisons using AIC and BIC, we consider a $FARIMA\ Process\ Model(p, d, q); *\ if\ \Delta^d Y_t\ is\ an\ ARMA(p, q) process, then\ Y_t$ a fractionally integrated process of order d or, simply $I(d)$ process.

ARMA processes are widely used to model stationary time series. These stationary ARMA processes only have short memories by auto-correlating function decays to zero exponentially fast – which is expressed as $D > 0\ and\ r < 1\ ; Such\ that\ p(k) < D|r|^k$. However, on the contrary, we implement a long-memory process using a fractional differencing. Many financial time series have a long-memory since their ACFs decay at a relatively slow- polynomial rate, given as: $p(k) \sim Dk^{-\alpha}\ where\ D\ and\ \alpha > 0$. Fractional differencing is one of the most widely used models for stationary, long-memory processes, where the integer values of d is defined in the equation below:

$$\Delta^d = (1-B)^d = \sum_{k=0}^{d} \binom{d}{k} (-B)^k$$

The definition of Δ^d is extended to non-integer values of d with a restriction $d > -1$. The only values needed for d to model the long-memory process must be greater than 1. The function $f(x) = (1-x)^d$ has an infinite Taylor series expansion:

$$(1-x)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-x)^k$$

Since $\binom{d}{k} = 0 \; if \; k > d \; and \; d > -1; is \; an \; integer, when \; d \; is \; an \; integer \; we \; have$:

$$(1-x)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-x)^k = \sum_{k=0}^{\infty} \binom{d}{k} (-x)^k$$

The right side of the equation is the finite binomial expansion for d the nonnegative integer such that the left side of the equation extends the binomial expansion to all $d > -1$. Since $(1-x)^d$ is defined for all $d > -1$. Thus, if $d > -1$, then:

$$\Delta^d Y_t = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k Y_{t-k}$$

A process Y_t is a fractional $ARIMA\ (p, d, q)$ process, also known as $ARFIMA \; or \; FARIMA\ (p, d, q)$ process, if $\Delta^d Y_t$ is an $ARMA\ (p, q)$ process. Thus, Y_t is a fractionally integrated process of order d, or $I(d)$ process, which is also an $ARIMA$ process extended to the non-integer values of d. Normally, $d > 0, with \; d = 0$ which is common in $ARMA$ process, but d could be potentially negative. Note that if $0 < d < \frac{1}{2}$, then we have a long-memory stationary process. If $d > \frac{1}{2}$, then Y_t can be differenced an integral number to become a stationary process with long-memory.

The $FARIMA$ model serves as a control for model uncertainty. This is evident in prediction purposes. It allows us to average over an entire distribution of fractional differencing parameter rather than the use of just one model. Since unit root models are mostly favorited by data, unit root and long-memory models have unique importance for the persistence in the data and incorporating model uncertainty is pivotal for prediction.

The methodology is based on the function $fracdiff()$ in R's $fracdiff$ package. This will fit a $FARIMA\ (p, d, q)$ process. Since there is no R function that will choose the values of p, d, q, we input these values manually. To begin, we choose the value of d using the $fracdiff()$ with $p =$

$q = 0$, which is default values. ***<u>The estimate was …</u>*** and inflation rates were fractionally differenced using the value of d and $auto.arima()$ function applied to the fractionally differenced series. The outcome – BIC select the value $p = q = d = 0$. Where the value of $d = 0$ implies no further differencing applied to the already fractionally differenced series. Fractional differencing was performed using the R function $diffseries()$ and $fracdiff$ package.

As part of our methodology, we implemented the long-memory processes to plot $DiffSqrtCpi$ and along with the corresponding ACF to check for signs of long memory. We executed the following R script:

```
library("fracdiff")
fit.frac = fracdiff(DiffSqrtCpi,nar=0,nma=0)
fit.frac$d
fdiff = diffseries(DiffSqrtCpi,fit.frac$d)
acf(fdiff)
```

Figure 1. R Code to test for Long-memory Process

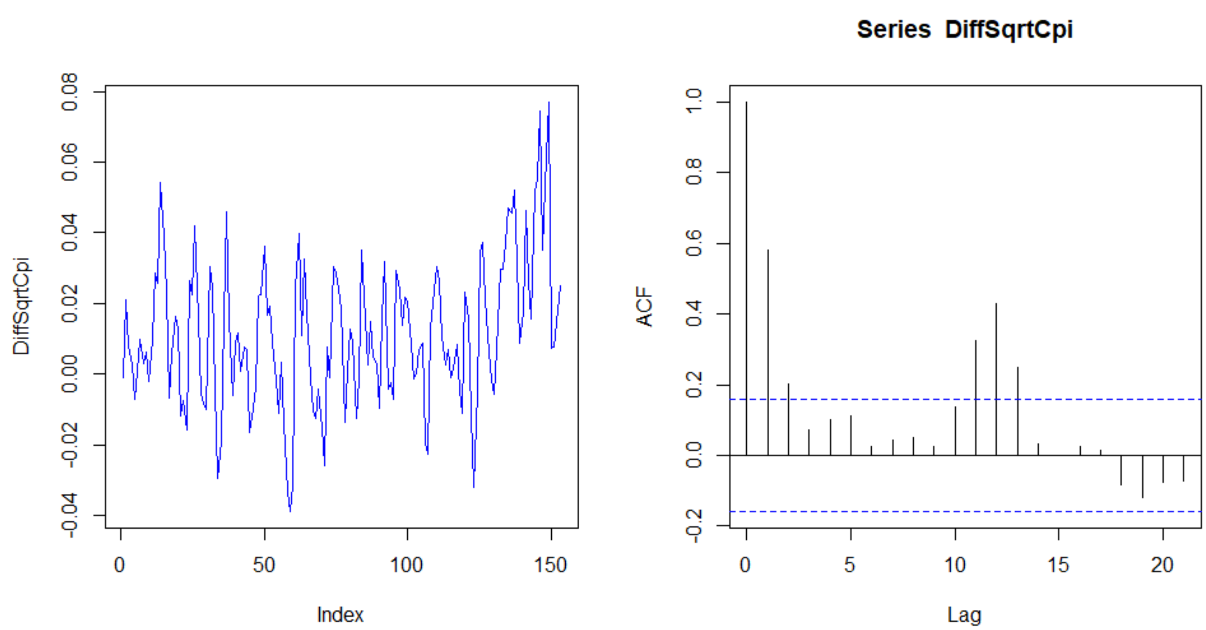

Figure 2. A plot of time series DiffSqrtCpi: A plot of autocorrelation function for time series DiffSqrtCpi

The ACF in Figure 2 decays very slowly as its value is not strictly decreasing. There is visible moment in time where it increases as movements occurs from the left to right, which a characteristic of an $AR(p)$ model or the need for fractional differencing – i.e., a long-term memory process. Applying the code $fracdiff$ to estimate the of fractional differencing, we applied the output $d = 0.4340421$ and plotted the ACF for Figure 3.

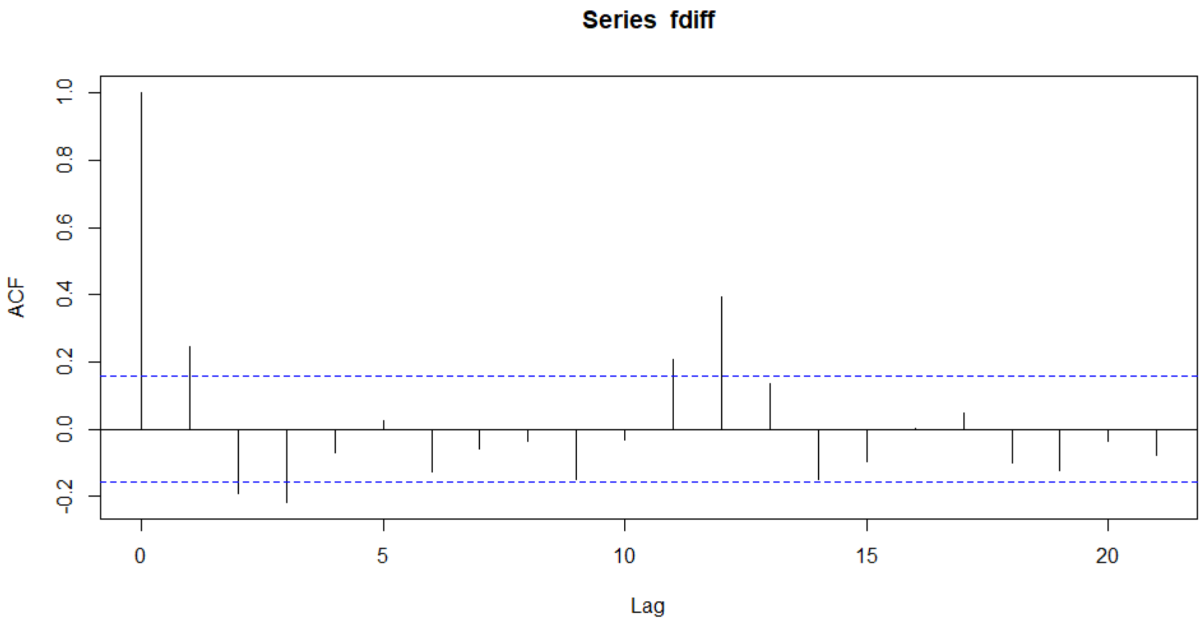

Figure 3. ACF of fractionally differenced time series

The function $lm()$ has been implemented in this project to fit linear models and perform regression, single stratum analysis of variance, and the analysis of covariance to data frames in R Language. Models with different variables, from one variable to all five variables, use the AIC function to determine the best model for the linear regression to produce the best model fit a parsimonious one.

V. Forecasting

We generated the behavior of $ccpiu\_value$, to determine the types of differencing, seasonal, nonseasonal, or both to recommend a fitting seasonal ARIMA model to the CCPIU dataset. We observe that the time series is not is not stationary and thus, the need to take differences to make it stationary.

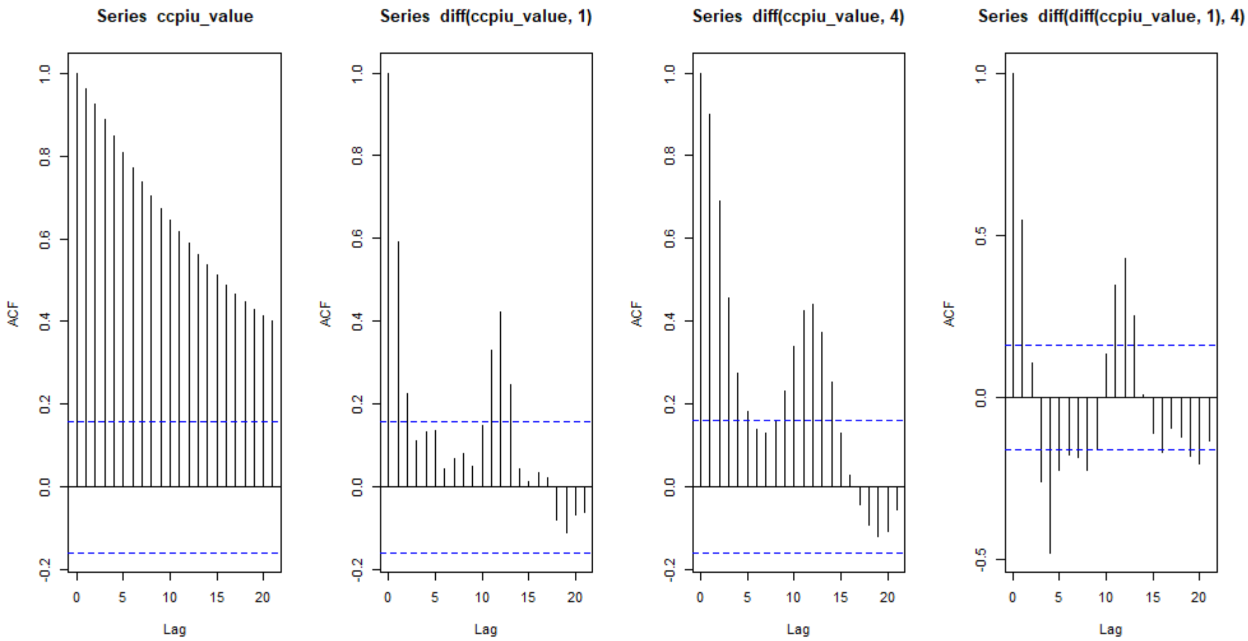

Figure 4. Plots of autocorrelation functions for the ccpiu_values time series.

In Figure 1 and 2 above we plot the autocorrelations functions (ACF) for various differences of the $ccpiu\_value$.

- The leftmost plot is the ACF for the untransformed $ccpiu\_value$ data.
- The left-middle plot is the ACF for the first difference of the $ccpiu\_value$ data.
- The right-middle plot is the ACF for the seasonal difference (of order four) of the $ccpiu\_value$ data.
- The rightmost plot is the ACF for the combined nonseasonal and seasonal difference
- (Of order four) of the $ccpiu\_value$ data.

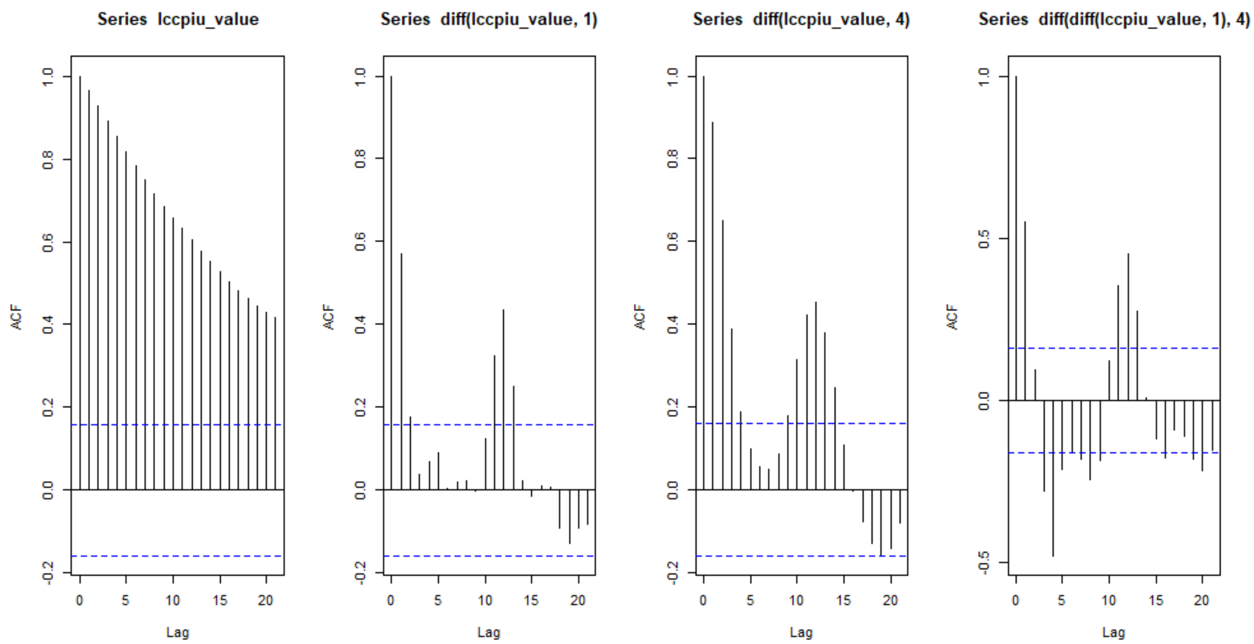

Figure 5. log (ccpiu_value).

We also determined an ARIMA model that provides a good fit to log ($ccpiu\_value$). Taking a logarithmic transform of the data results in an ACF that has similarities with Figure 1 above. In either case the combination of one seasonal difference and nonseasonal difference can be modeled with a seasonal moving average (MA) model with a single coefficient.

Fitting the chosen model to both the raw ($ccpiu\_value$) time series and logarithm of $ccpiu\_value$ we compare for the best Ljung-Box test on each model residuals showing the results in Figure 3 below.

```
> ccpiu_value_ts_model = arima( ccpiu_value, order=c(0,1,0),
+                                 seasonal=list(order=c(0,1,1), period=4) )
> Box.test( residuals(ccpiu_value_ts_model), lag=10, type="Ljung" )

        Box-Ljung test

data:  residuals(ccpiu_value_ts_model)
X-squared = 69.62, df = 10, p-value = 5.249e-11

>
> log_ccpiu_value_ts_model = arima( lccpiu_value, order=c(0,1,0),
+                                     seasonal=list(order=c(0,1,1), period=4) )
> Box.test( residuals(log_ccpiu_value_ts_model), lag=10, type="Ljung" )

        Box-Ljung test

data:  residuals(log_ccpiu_value_ts_model)
X-squared = 64.764, df = 10, p-value = 4.497e-10
```

Figure 6. Residuals & P-values

Intuitively, larger p-values in the Box.text indicates that the time series has a resemblance of independent events. Therefore, the logarithm of $ccpiu\_value$ has a larger p-value and represents a better model fit producing more independent residuals.

Checking the ACF of the residuals along with the corresponding model, we observe the model with the larger p-value is the best model fit. However, the model is not reflective of statistically significant autocorrelations.

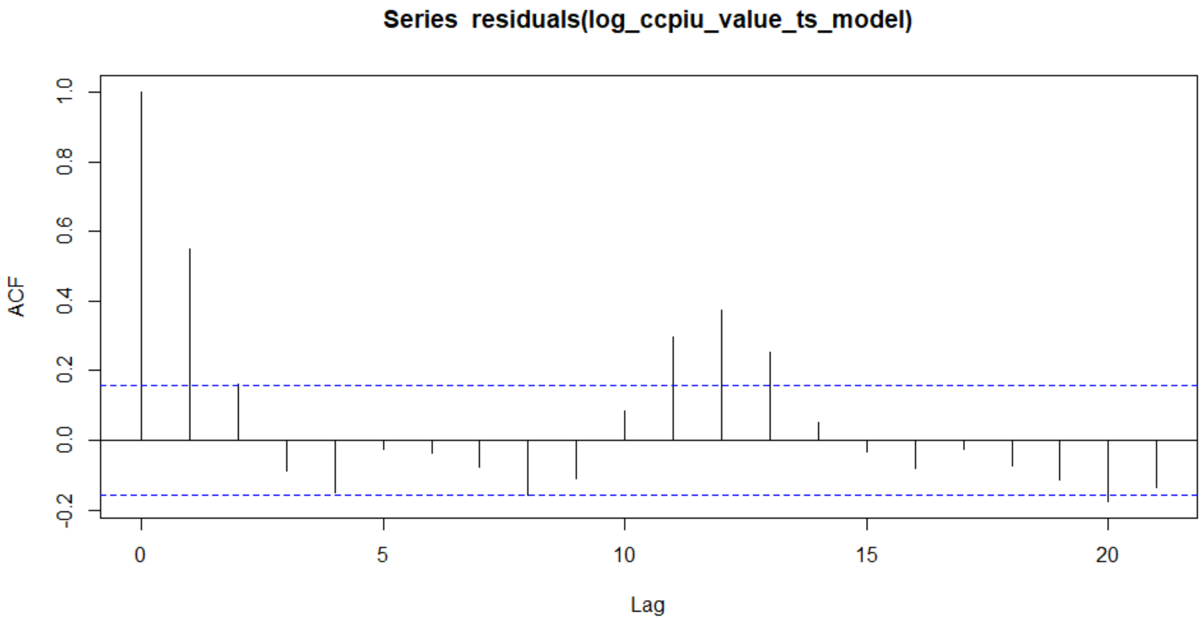

Figure 7. ACF

VI. Forecasting Results

After several iterations in the linear regression model, we determine that the model with two predicting variables – oil price and CPI is the best predicting model to determine inflation rate comparable to the rest of the regression models developed.

```
> fit7<-lm(Inflation_diff ~ oil_diff + CPI_diff)
> summary(fit7)

Call:
lm(formula = Inflation_diff ~ oil_diff + CPI_diff)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0160490 -0.0019686  0.0000553  0.0020448  0.0193254

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.741e-04  2.613e-04  -2.197    0.029 *
oil_diff     2.066e-04  4.637e-05   4.456 1.29e-05 ***
CPI_diff     1.364e-03  1.976e-04   6.900 4.79e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.00371 on 235 degrees of freedom
Multiple R-squared:  0.4231,	Adjusted R-squared:  0.4182
F-statistic: 86.17 on 2 and 235 DF,  p-value: < 2.2e-16
```

Figure 8. Summary of Linear regression model.

The p-value of both the predicting variables in the summary from Figure 4 are significantly less than 0.05, indicating that the output variables have a statistically significant impact on inflation rate. Also, the adjusted R-squared value of 0.4182 implies a 41.82% chance of variability in inflation can be attributed to the association between the inflation and the predictor variables oil price and CPI.

$$inflation\_diff = 4.637e^{-5} * (oil\_diff) + 1.976e^{-4} * (CPI\_diff)$$

In our seasonal model, we plot the predictions (the solid lines) and two-sigma errors bars (the dashed lines) on predictions and forecasting for the next eight quarters. The red curves are for the consumption data and green curves for the logarithm model.

We noticed the log model predicts larger numbers with a larger confidence interval.

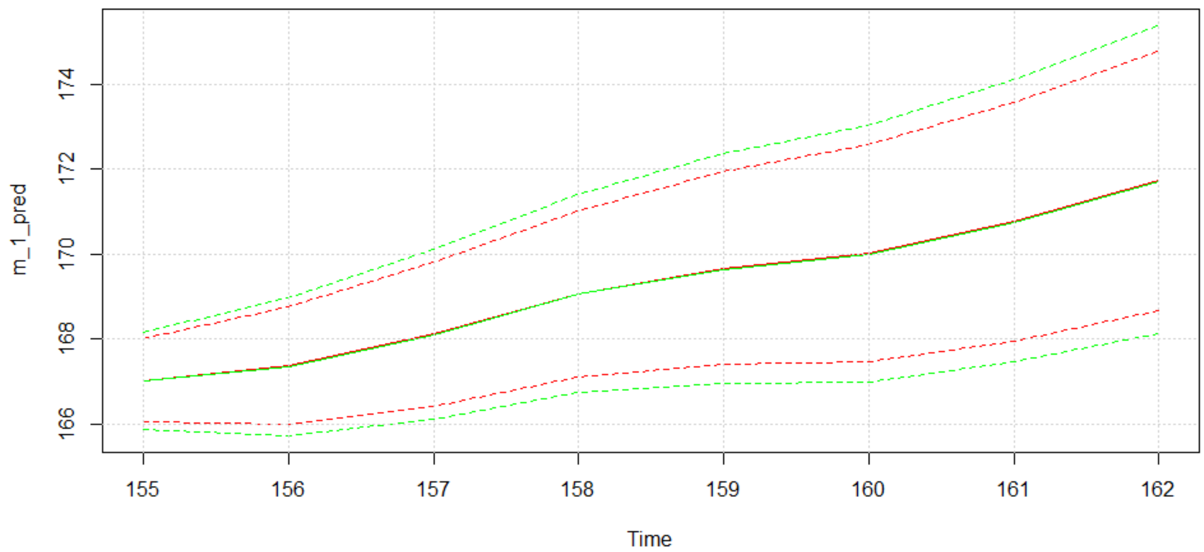

Figure 9. Forecasting Model

VII. Check for Model Robustness

To test for model robustness, on the ARIMA model, the model determined by $auto.arima$ translates as the same model suggested from the acf plots. Thus, we compare the values of AIC, and BIC.

Running the $auto.arima$ produces the output below.

```
> auto.arima( lccpiu_value, ic="bic" )
Series: lccpiu_value
ARIMA(0,2,2)

Coefficients:
          ma1      ma2
      -0.3657  -0.5241
s.e.   0.0652   0.0638

sigma^2 = 8.253e-06:  log likelihood = 677.05
AIC=-1348.11   AICc=-1347.95   BIC=-1339.04
```

Figure 10. AIC & BIC values

Checking for robustness on the linear model, we implemented multiple trials to determine the best model to predict inflation, and thus the most parsimonious. The models are tested using AIC in R studio in the below output.

```
> #One-Variable Model
> AIC(lm(Inflation_diff ~ CPI_diff))
[1] -1966.351
> AIC(lm(Inflation_diff ~ SP500_return))
[1] -1875.339
> AIC(lm(Inflation_diff ~ gold_diff))
[1] -1890.218
> AIC(lm(Inflation_diff ~ bond_diff))
[1] -1857.484
> AIC(lm(Inflation_diff ~ oil_diff))
[1] -1941.75

> # Two-Variable Model
> AIC(lm(Inflation_diff ~ CPI_diff + SP500_return))
[1] -1970.547
> AIC(lm(Inflation_diff ~ gold_diff + SP500_return))
[1] -1905.019
> AIC(lm(Inflation_diff ~ gold_diff + CPI_diff))
[1] -1965.756
> AIC(lm(Inflation_diff ~ oil_diff + CPI_diff))
[1] -1983.656
> AIC(lm(Inflation_diff ~ oil_diff + gold_diff))
[1] -1950.614
> AIC(lm(Inflation_diff ~ oil_diff + SP500_return))
[1] -1940.632

> # Three-Variable Model
> AIC(lm(Inflation_diff ~ CPI_diff + oil_diff + SP500_return))
[1] -1982.743
> AIC(lm(Inflation_diff ~ gold_diff + oil_diff + SP500_return))
[1] -1949.854
> AIC(lm(Inflation_diff ~ gold_diff + CPI_diff + SP500_return))
[1] -1969.551

> # Four-Variable Model
> AIC(lm(Inflation_diff ~ CPI_diff + oil_diff + SP500_return + gold_diff))
[1] -1981.427
> AIC(lm(Inflation_diff ~ gold_diff + oil_diff + SP500_return + bond_diff))
[1] -1948.759
> AIC(lm(Inflation_diff ~ gold_diff + CPI_diff + SP500_return + bond_diff))
[1] -1974.602
```

> # **Five-Variable Model**
> AIC(lm(Inflation_diff ~ gold_diff + CPI_diff + SP500_return + bond_diff + oil_diff))
[1] -1981.307

In the five variable model considered, the AIC score is lower than two-variable model of oil price and CPI. Thus, it was not selected to be the best predicting model.

VIII. Conclusions

In conclusion for our linear regression model AIC test indicated that model with 2 predicting variables oil_diff and CPI_diff is the best to predict the $Inflation\_diff$.

For the long-memory models in Figure 2 and seasonal ARIMA models in Figure 9, the plot predictions (the solid lines) and two-sigma error bars (the dashed lines) on these predictions for the next eight quarters. The red curves are for the consumption data, and the green curves are for the logarithm model. We notice that the log model predicts larger numbers with a larger confidence interval.

IX. Appendices: Code References

We demonstrated how to predict US inflation using three different models: Linear Regression model, Long-memory Process, ARIMA seasonal models.

Firstly, a linear regression model AIC test indicates that model with 2 predicting variables oil_diff and CPI_diff is the best to predict the $Inflation\_diff$.

Secondly, a long-memory and seasonal ARIMA models plot predictions for the next eight quarters. A significant observation on the log model predicts larger numbers with a larger confidence interval.

Details found at: https://github.com/fhill09/Predicting-US-Inflation-Rate

Presentation Slides found at: https://prezi.com/view/ApW4XDdRPUmVggDeFnkx/

References

[1] V. Plakandaras, P. Gogas, T. Papadimitriou and R. Gupta, "The Informational Content of the Term Spread in Forecasting the US Inflation Rate: A Nonlinear Approach," Journal of Forecasting, J. Forecast., 2017.

[2] M. Balcilar, R. Gupta and C. Jooste, "Long memory, economic policy uncertainty and forecasting US inflation: A Bayesian VARFIMA Approach," 2017.

[3] N. Bloom, "Econometrica," in The Impact of Uncertainty Shocks, 2009, p. 623–685.

[4] M. F. Estrella A, "The predictive power of the term structure of interest rates in Europe and the United States: implications for the European Central Bank. European Economic Review.," European Economic Review 41(7): 1375–1401, p. 41(7): 1375–1401, 1997.

[5] W. M. Stock JH, "Phillips Curve Inflation Forecasts," NBER Working Paper No. 14322., 2008.

[6] C. a. L. A. G. A. Barros, "Inflation Forecasting in Angola; A Fractional Development Review," vol. Review 25, 2013.

[7] Y. a. A. P. P.-L. Lovcha, "A Fractionally Integrated Approach to Monetary Policy and Inflation Dynamics," 2013.

[8] D. Ruppert and D. S. Matteson, Statistics and Data Analysis for Financial Engineering, New York: Springer, 2015.

[9] V. Colombo and P. M. a. E. O. Jones, "Economic Policy Uncertainty in the US: Does it matter for the Euro Area?; The Time-Varying Correlation between Uncertainty, Output and Inflation: Evidence from DCC-GARCH Model," Economics Letters, no. 121; 118, 2013.

[10] Y. J. F. Y. Z. Rui Hou, "ResearchGate," 26 May 2022. [Online]. Available: https://www.researchgate.net/publication/360865283_Short_time_asymptotics_for_American_maximum_options_with_a_dividend-paying_asset. [Accessed 09 September 2022].