FORECASTING THE US INFLATION RATE

SYST538/OR538-001
Analytics for Financial Engineering & Econometrics

Team Members:
1. Bloti B. Teh
2. Faith Hill
3. Huy Nguyen

Professor: Dr. Ran Ji

Date: December 12, 2022

PROJECT OBJECTIVE

To forecast inflation rate, we examine available datasets from the Bureau of Labor Statistics (Consumer Price Index) with the application of various predictions or forecasting tools available.

We will 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.

Goal

To conduct model comparison and identify the best model in forecasting price inflation.

"When you think the market can't possibly go any higher (or lower), it almost invariably will, and whenever you think the market "must" go in one direction, nine times out of ten it will go in the opposite direction. Be forever skeptical of thinking that you know what the market is going to do." – William Gallaher

Methodology

- A process Y_t is a fractional ARIMA (p,d,q) process, also known as ARFIMA or FARIMA (p,d,q) if if $\Delta^d Y_t$ is an $ARMA\ (p,q)$ process.
- The FARIMA model serves as a control for model uncertainty. This is evident in prediction purposes.
- The methodology is based on the function fracdiff() in R's fracdiff package. This will fit a FARIMA (p,d,q) process.

$$FARIMA\ Process\ Model(p,d,q);*\ if\ \Delta^d Y_t\ is\ an\ ARMA(p,q) process, then\ Y_t$$

- When there are many potential predictor variables, often we wish to find a subset of them that provide a parsimonious regression model.
- Our next model is based on a linear regression model.

Long-memory Process

Linear Regression Model

FARIMA/ARIMA Process Model

- Performed a data exploratory analysis on the dataset
- Renamed some column names
- The methodology is based on the function fracdiff() in R's fracdiff package
- This will fit a FARIMA (p,d,q) process.
- Fractional differencing was performed using the R function diffseries() and fracdiff package.

```
> summary(df)
   ï..label          ccpiu_value
 Length:154         Min.   :125.0
 Class :character   1st Qu.:133.4
 Mode  :character   Median :137.0
                    Mean   :139.2
                    3rd Qu.:144.2
                    Max.   :166.4
```

```
> describe_vars(df)
            colNum         var nonmissing missing distinct numeric
ï..label         1    ï..label        154       0      154   FALSE
ccpiu_value      2 ccpiu_value        154       0      149   FALSE
```

Summary of dataset -
Chained CPI for All Urban Consumers_2010-2022

Data Processing

- **Performed data exploratory analysis on raw data using Python**
- **Processed commodities used and stocks and bonds taking monthly average**

CPI for All Urban Consumers (CPI-U)
Original Data Value

Series Id:	CUSR0000SA0
Seasonally Adjusted	
Series Title:	All items in U.S. city average, all urban consumers,
Area:	U.S. city average
Item:	All items
Base Period:	1982-84=100
Years:	1982 to 2022

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	HALF1	HALF2
1982	94.40	94.70	94.70	95.00	95.90	97.00	97.50	97.70	97.70	98.10	98.00	97.70		
1983	97.90	98.00	98.10	98.80	99.20	99.40	99.80	100.10	100.40	100.80	101.10	101.40		
1984	102.10	102.60	102.90	103.30	103.50	103.70	104.10	104.40	104.70	105.10	105.30	105.50		
1985	105.70	106.30	106.80	107.00	107.20	107.50	107.70	107.90	108.10	108.50	109.00	109.50		
1986	109.90	109.70	109.10	108.70	109.00	109.40	109.50	109.60	110.00	110.20	110.40	110.80		
1987	111.40	111.80	112.20	112.70	113.00	113.50	113.80	114.30	114.70	115.00	115.40	115.60		
1988	116.00	116.20	116.50	117.20	117.50	118.00	118.50	119.00	119.50	119.90	120.30	120.70		
1989	121.20	121.60	122.20	123.10	123.70	124.10	124.50	124.50	124.80	125.40	125.90	126.30		
1990	127.50	128.00	128.60	128.90	129.10	129.90	130.50	131.60	132.50	133.40	133.70	134.20		
1991	134.70	134.80	134.80	135.10	135.60	136.00	136.20	136.60	137.00	137.20	137.80	138.20		
1992	138.30	138.60	139.10	139.40	139.70	140.10	140.50	140.80	141.10	141.70	142.10	142.30		
1993	142.80	143.10	143.30	143.80	144.20	144.30	144.50	144.80	145.00	145.60	146.00	146.30		
1994	146.30	146.70	147.10	147.20	147.50	147.90	148.40	149.00	149.30	149.40	149.80	150.10		
1995	150.50	150.90	151.20	151.80	152.10	152.40	152.60	152.90	153.10	153.50	153.70	153.90		
1996	154.70	155.00	155.50	156.10	156.40	156.70	157.00	157.20	157.70	158.20	158.70	159.10		
1997	159.40	159.70	159.80	159.90	159.90	160.20	160.40	160.80	161.20	161.50	161.70	161.80		
1998	162.00	162.00	162.00	162.20	162.60	162.80	163.20	163.40	163.50	163.90	164.10	164.40		
1999	164.70	164.70	164.80	165.90	166.00	166.00	166.70	167.10	167.80	168.10	168.40	168.80		
2000	169.30	170.00	171.00	170.90	171.20	172.20	172.70	172.70	173.60	173.90	174.20	174.60		
2001	175.60	176.00	176.10	176.40	177.30	177.70	177.40	177.40	178.10	177.60	177.50	177.40		
											181.5	181.8		
											185.0	185.5		
											191.7	191.7		

	date	year	month	CPI.value	Inflation_rate	gold.avg	bond.rate	adjusted	Oil_price
1	1999-01-01	1999	1	164.700	0.01671	287.2250	5.157895	1248.7747	12.51474
2	1999-02-01	1999	2	164.700	0.01606	287.4112	5.365263	1246.5821	12.01368
3	1999-03-01	1999	3	164.800	0.01726	286.1022	5.580435	1281.6639	14.67652
4	1999-04-01	1999	4	165.900	0.02277	282.6187	5.547727	1334.7567	17.31238
5	1999-05-01	1999	5	166.000	0.02088	276.6868	5.805500	1332.0740	17.71850
6	1999-06-01	1999	6	166.000	0.01963	261.3580	6.041818	1322.5527	17.92318
7	1999-07-01	1999	7	166.700	0.02145	256.1386	5.983333	1380.9900	20.10333
8	1999-08-01	1999	8	167.100	0.02264	256.8143	6.068636	1327.4887	21.27864
9	1999-09-01	1999	9	167.800	0.02628	264.6068	6.071429	1318.1719	23.79667

date	Value
1/1/1982	94.4
2/1/1982	94.7
3/1/1982	94.7
4/1/1982	95
5/1/1982	95.9
6/1/1982	97
7/1/1982	97.5
8/1/1982	97.7
9/1/1982	97.7
10/1/1982	98.1
11/1/1982	98
12/1/1982	97.7
1/1/1983	97.9
2/1/1983	98
3/1/1983	98.1
4/1/1983	98.8
5/1/1983	99.2
6/1/1983	99.4
7/1/1983	99.8
8/1/1983	100.1
9/1/1983	100.4
10/1/1983	100.8
11/1/1983	101.1

- **ARIMA modeling is one of the most popular approaches to time series forecasting.**

- **The methodology here is to use ARIMA models to describe autocorrelations in data**

- **The 'auto.arima()' function in 'R' is used to build ARIMA models by using a variation**

- **It combines unit root tests, minimisation of the AICc, and MLE to obtain an ARIMA model**

```
> summary(df)
  ï..label           ccpiu_value
Length:154         Min.   :125.0
Class :character   1st Qu.:133.4
Mode  :character   Median :137.0
                   Mean   :139.2
                   3rd Qu.:144.2
                   Max.   :166.4
```

```
> describe_vars(df)
            colNum         var nonmissing missing distinct numeric
ï..label         1    ï..label        154       0      154   FALSE
ccpiu_value      2 ccpiu_value        154       0      149   FALSE
```

Summary of dataset -
Chained CPI for All Urban Consumers_2010-2022

Data Set

- The dataset used for both ARIMA and Long-memory process models were pulled from the website of Bureau of Labor Statistics (BLS)
- It comprises of time period from 2010 and 2022
- The third dataset used on the Linear regression model includes commodities such as oil price, gold price, and stock and bonds are incorporated in the dataset.
- Data collected for bond rate is over 30 years and that of gold from 1999
- SP500 data from tidyquant package
    - CPI data from BLS.gov

Forecasting - *Linear Regression*

Forecasting - *ARIMA Model*

Forecasting Results - *Long-memory Process*

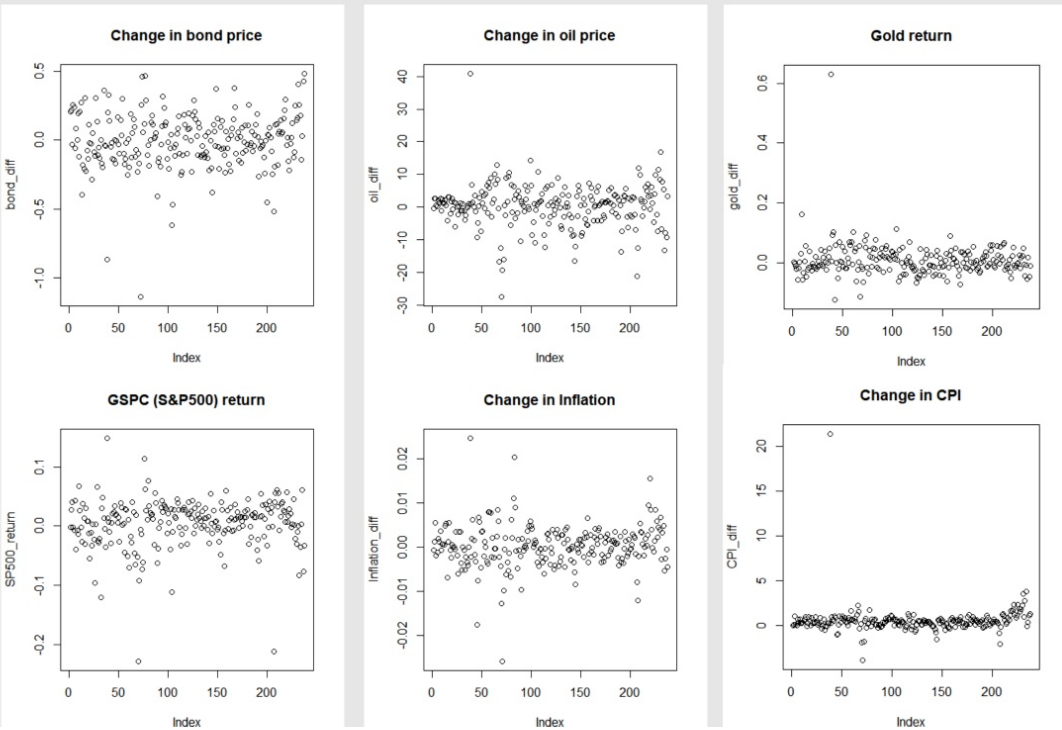
Model Robustness

Conclusion

Models Forecasting

Forecasting performance on Linear regression in scatter plots:

- **Change in bond price**
- **Change in oil price**
- **Gold price return**
- **Change in CPI**
- **CSPC (S&P 500) return**
- **Change in inflation**



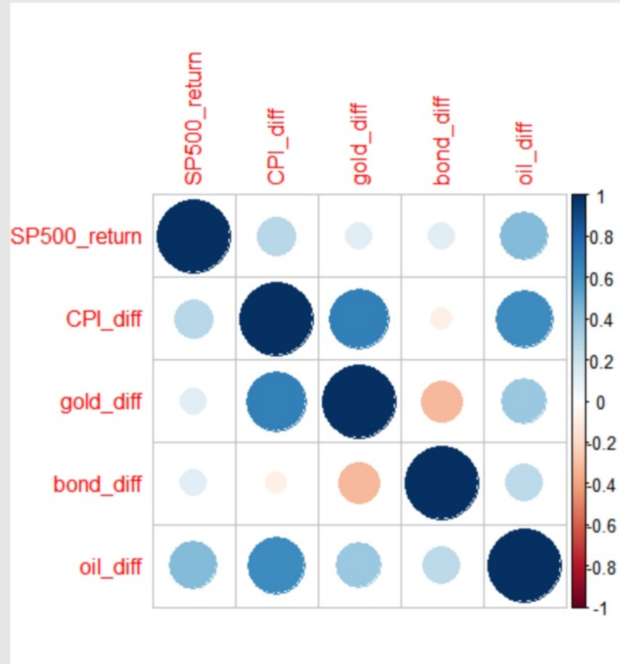


Figure 2.

- **There exists a correlation among the predicting variables**
- **On the contrary there exists no strong correlation among these variables**

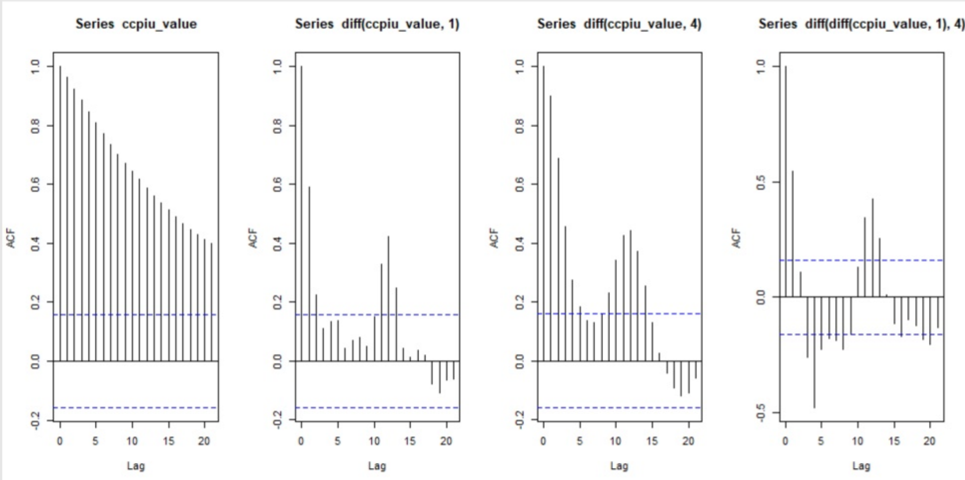


Figure 3.

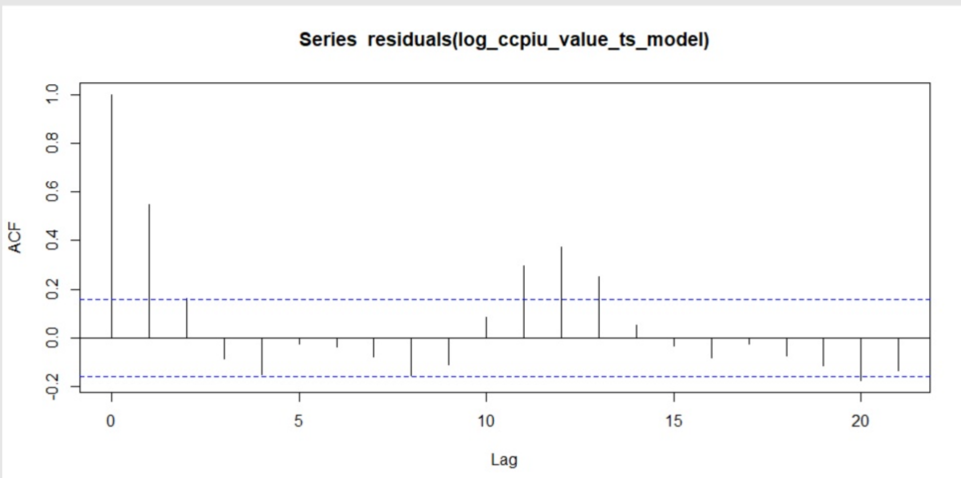


Figure 4.

Behavior of ccpiu_value:

- **In Figure 3 we plot the autocorrelation functions (ACF) for various differences of the**
- **ccpiu_value time series data**
- **The leftmost plot is the ACF for the untransformed consumption data**
- **The left-middle plot is the ACF for the first difference of the consumption data**
- **The right-middle plot is the ACF for the seasonal difference (of order four) of the consumption data**
- **The rightmost plot is the ACF for the combined nonseasonal and seasonal difference (of order four) of the consumption data.**
- **From figure 4, it is intuitively indicative that the time series looks like it is a sequence of independent events.**
- **With large p-values we argue that logarithmic model has a better fit - showing more independent residuals.**

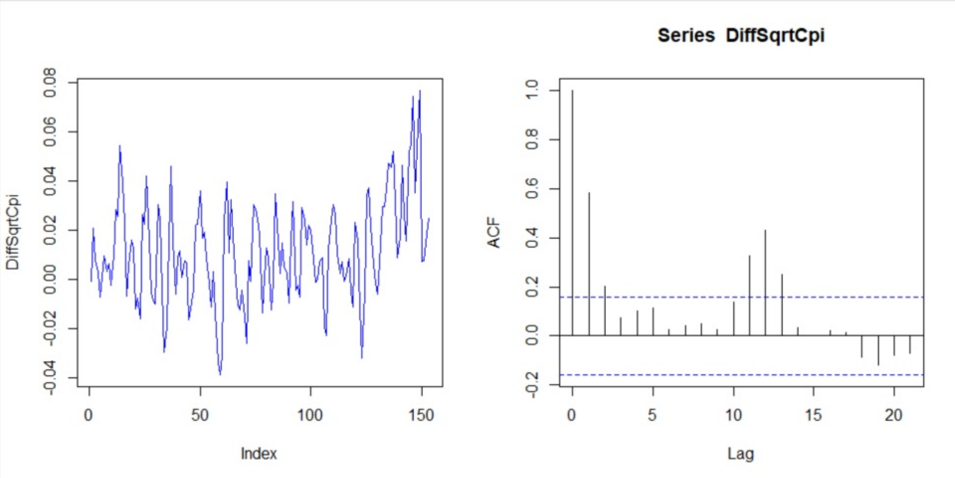


Figure 5. A plot of autocorrelation function 1for Time Series - DiffSqrtCpi.

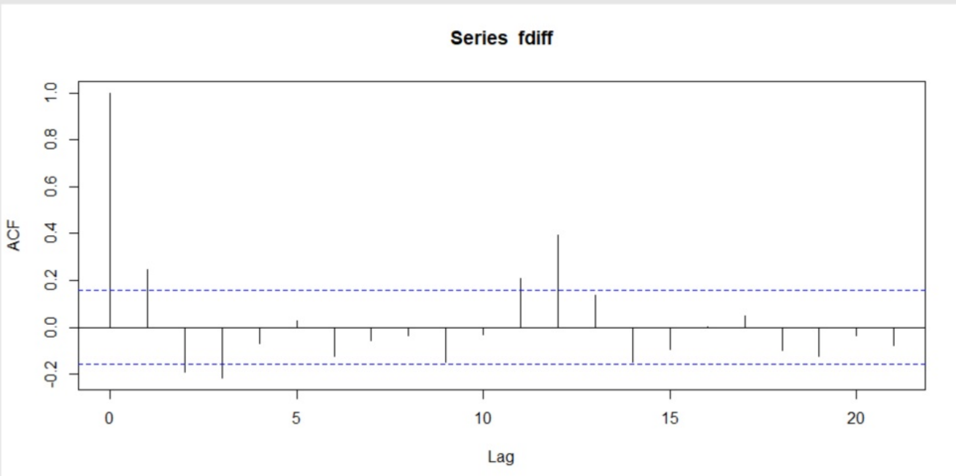


Figure 6.

Behavior of ccpiu_value in - Long-memory Process:

- **In Figure 5, time series and the autocorrelation function (ACF) for the suggested transformations of the Consumer Price Index (CPI)**
- **ACF decays very slowly and its value is not strictly decreasing (there are periods where it increases as we move from left to right).**
- **Applying the code fracdiff to estimate the amount of fractional differencing we get d = 0.4134047**

Model Robustness on Linear Regression & Long-memory processes

Conducted several to tests for model robustness using multiple variables:

- **One variable**
- **Two variables**
- **Three variables**
- **Three-Variable Model**
  AIC(lm(Inflation_diff ~ CPI_diff + oil_diff + SP500_return)) -1982.743
  AIC(lm(Inflation_diff ~ gold_diff + oil_diff + SP500_return))
  AIC(lm(Inflation_diff ~ gold_diff + CPI_diff + SP500_return))
- **Four-variables**
  AIC(lm(Inflation_diff ~ CPI_diff + oil_diff + SP500_return + gold_diff)) -1981.427
  AIC(lm(Inflation_diff ~ gold_diff + oil_diff + SP500_return + bond_diff))
  AIC(lm(Inflation_diff ~ gold_diff + CPI_diff + SP500_return + bond_diff))
- **Five-variables**
  AIC(lm(Inflation_diff ~ gold_diff + CPI_diff + SP500_return + bond_diff + oil_diff)) -1981.307

- **To test for model robustness in long-memory process, we compare the models selected using AIC and BIC.**

- **The ACF for the first difference of the fdiff time series reveals a very significant spike at lag one. However, forcing auto.arima to not consider any differences the result shows that has lower AIC and BIC metrics and should be the preferable model.**

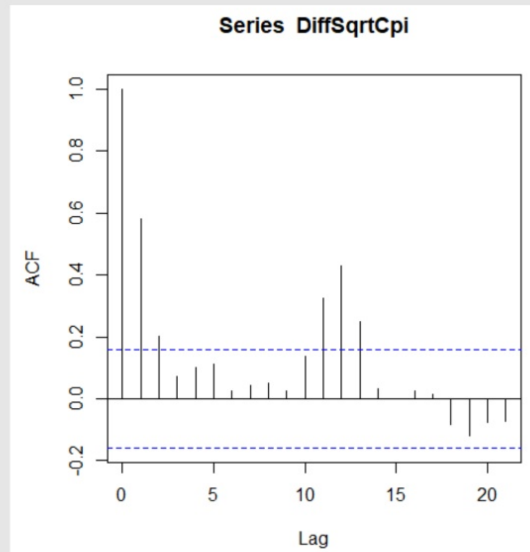


Figure 7.

Conclusions

In conclusion for our linear regression remodel 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 and ARIMA models, in Figure 8, 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 noticed that the log model predicts larger numbers with a larger confidence interval.

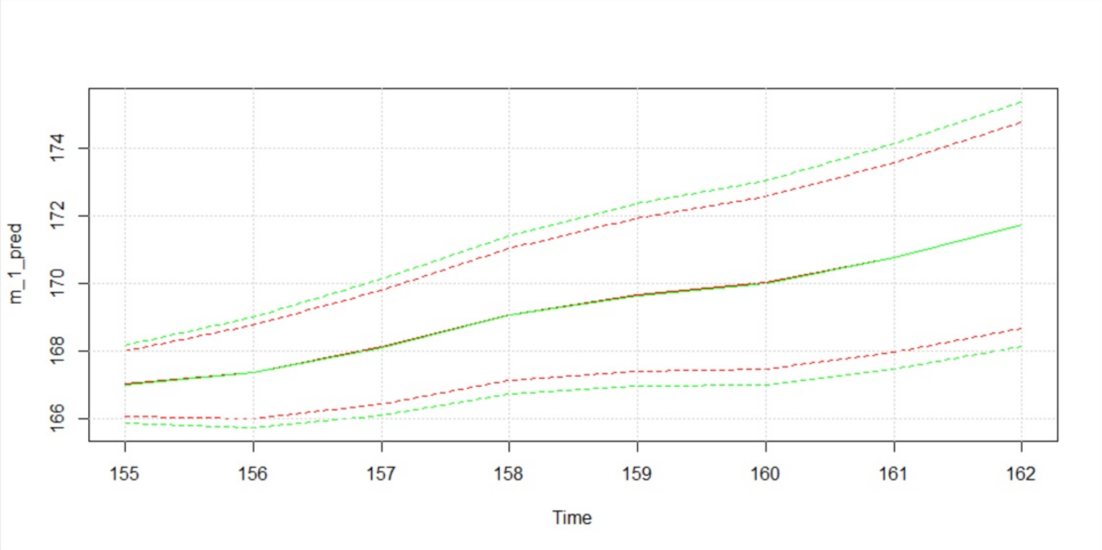


Figure 8. Prediction for eight quarters

FINALLY . . .

THE