International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Time Series Forecasting of Producer Price Index, using ARIMA

Nikhil Garg[1], Aakash Varshney[2], Adarsh Agrawal[3]

Undergraduate IIT Delhi

Abstract: ***Producer Price Index (PPI) is a key indicator of economic stability of a country. This project aims to forecast the quarterly future PPI of USA using ARIMA Model for the years 2003 2007, using a data set with quarterly PPI data for the years 1960 2002. Based on our analysis, it was interpreted that the ARIMA(1,1,1) was best suited for modeling the future PPI, with maximum log-likelihood of and the minimum AIC of 393. The Ljung Box test reveals that the residuals are free from heteroscedasticity and serial correlation.***

1. Introduction

Producer Price Index (PPI), is a family of indexes that measures the average change in selling prices received by domestic producers of goods and services over time. It is basically a measure of changed of prices from the perspective of the seller. It considers three areas of production: industry-based, commodity-based and commodity-based nal demand-intermediate demand. An index takes the weighted average of the changes across all industries, and reports the results with respect to a base year.

PPI holds many important uses. It is a short term indicator of in ationary trends, it is used as an analytical tool by many businesses and researchers and used by many international organisations such as Eurostat, IMF for economic monitoring of countries and comparison.

In this project, we have used a dataset which provides the quarterly PPI data of USA from the years 1960 - 2002.

2. Literature Review

There are many forecasting models which have been used to predict PPI values, these include gray box model, regression analysis and time series analysis. Prasad S. Bhattacharyay and Dimitrios D. Thomakos in their paper on CPI and PPI prediction, compared various models including VAR (using Philip's concepts on unemployment and in ation rate), ARIMA models, models where the e ects of the exchange rate and import prices are taken into account. Past studies in the literature also use commodity prices as an economic indicator to achieve forecast improvements, though the results are mixed. Combining both aggregate and disaggregate indicators through Bayesian shrinkage procedures, Zellner and Chen (2001) also report higher forecast accuracy

3. Data Used

The data we have used for the paper gives the quarterly PPI of USA from 1960 2002, from the
US Bureau of Labour Statistics. The link:
https://drive.google.com/file/d/0BwogTI8d6EEiM3JlTDdXNEFfRDg/edit

4. Methodology

4.1 Overview

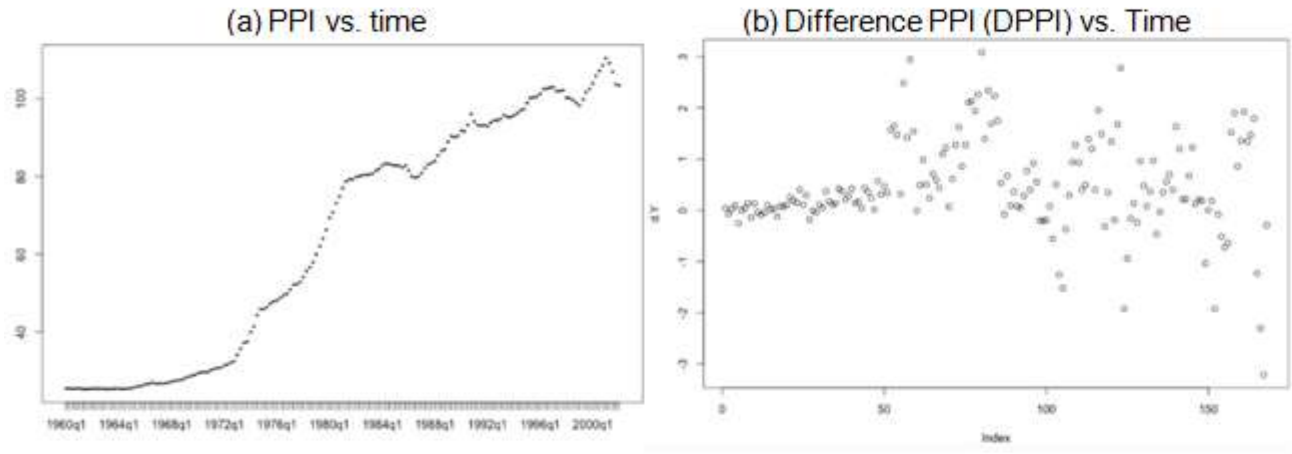
We follow the standard Jenkins Box approach of ARIMA modeling to determine an appropriate model for our forecast. The rst step is to determine the stationarity of the data we have. This is done by the Augmented Dickey Fuller Test (ADF). If stationarity is not achieved we di erence our data n times till we achieve stationarity. Once we have this, we plot the Autocorrelation Function (ACF) and Partial Autocorrelation functions (PACF) for the stationary data we have,and determine the order of AR process and MA process through the number of spikes in the PACF and ACF respectively.

We then compare our obtained model with other models by comparing the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values, and also check the e ciency of our model by plotting the ACF of our residuals and checking whether they are correlated.

4.2 Graphs and their interpretations

Based on the above written methodology, we plot relevant graphs and give our interpretations regarding the same.

4.2.1 Plots of PPI to visualise data



The above two graphs show the plots of PPI vs time and DPPI vs time. As can be visually observed, PPI is positively correlated with time, while DPPI can visually be seen to be have a constant mean and be less dependent on the time. We will verify these visual observations statistically.

4.2.2 Checking for Stationarity of the data

Augmented Dickey Fuller Test for PPI variable

```
> # Augmented Dickey-Fuller test
> adf.test(Y, alternative="stationary")

        Augmented Dickey-Fuller Test

data:  Y
Dickey-Fuller = -1.3857, Lag order = 5, p-value = 0.8327
alternative hypothesis: stationary
```

From the above ADF test on the PPI variable, the p-value is very high and we cannot reject the null hypothesis, and therefore we conclude that the variable is non-stationary. Hence we apply the same test on the DPPI variable to check for its stationarity. Augmented Dickey Fuller Test for DPPI variable

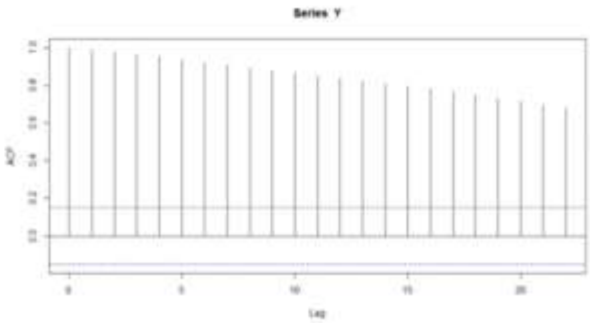
```
> # DF and ADF tests for differenced variable
> adf.test(d.Y, k=0)

        Augmented Dickey-Fuller Test

data:  d.Y
Dickey-Fuller = -6.8398, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(d.Y, k = 0) : p-value smaller than printed p-value
> adf.test(d.Y)

        Augmented Dickey-Fuller Test

data:  d.Y
Dickey-Fuller = -3.2459, Lag order = 5, p-value = 0.08252
alternative hypothesis: stationary
```

The p-value of the ADF test on the DPPI variable is 0.01, and hence we can say with 99% con dence that the null hypothesis (DPPI is non-stationary) can be rejected. We can say that the rst difference is stationary, and estimate the best model by determining the AR and MA parameters.

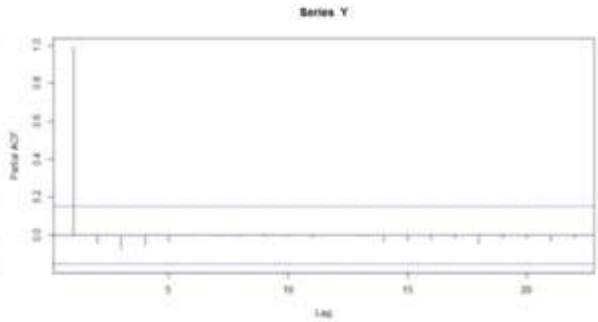
4.2.3 ACF and PCF plots to determine Order of ARIMA Process

PPI

(a) ACF of PPI vs lag

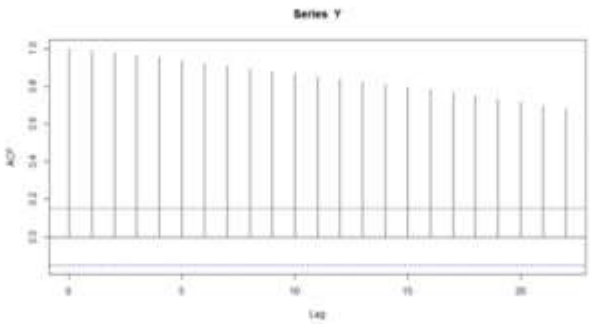


(b) PACF PPI vs lag

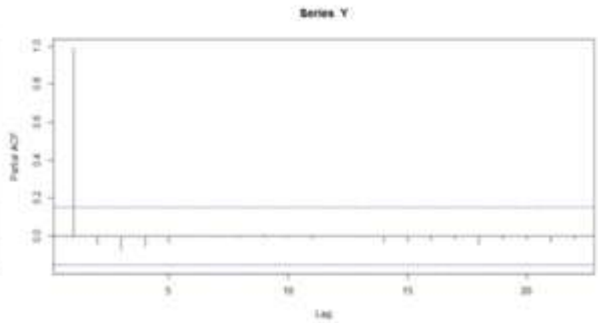


PPI

(a) ACF of PPI vs lag

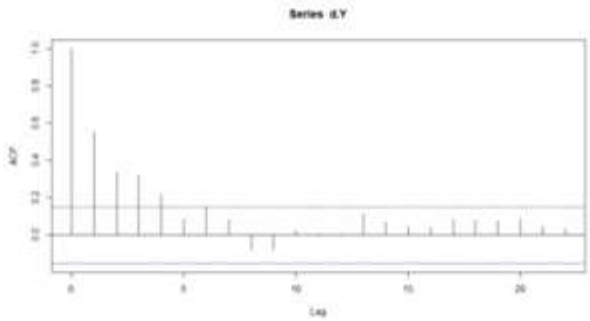


(b) PACF PPI vs lag

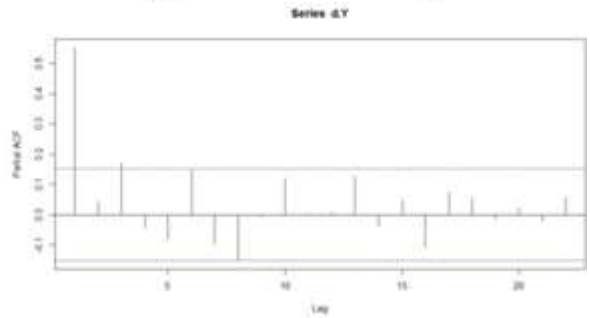


DPPI

(a) ACF of DPPI vs lag

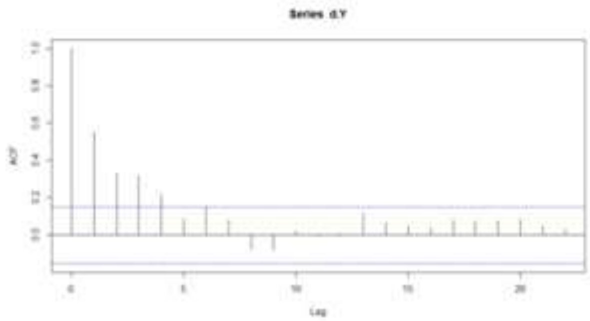


(b) PACF of DPPI vs lag

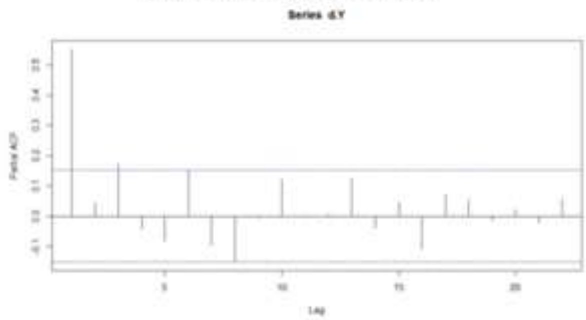


DPPI

(a) ACF of DPPI vs lag



(b) PACF of DPPI vs lag



The MA order can be determined from the number of spikes in ACF plot of DPPI, and it is determined to be 1 indicating a MA(1) process. We will compare our results with other MA orders and check if our estimate is correct. The AR order can be determined from the number of spikes in PACF plot of DPPI, and it is determined to be 1. This indicates an AR(1) process.

Furthermore, analysing the ACF plot of PPI, con rms our previous observation of non-stationarity in the original PPI data, which can be seen from the ACF plot of PPI vs lag. The plot isn't exactly geometrically decreasing and remains slowly decreasing which indicates non-stationarity of data.

4.2.4 Comparison with Di erent ARIMA models and coefficients of equations obtained

```
> # ARIMA(1,0,0) or AR(1)
> arima(Y, order = c(1,0,0))

Call:
arima(x = Y, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.9996     64.522
s.e.  0.0005     38.002

sigma^2 estimated as 1.058:  log likelihood = -248.2,  aic = 502.4
>
> # ARIMA(2,0,0) or AR(2)
> arima(Y, order = c(2,0,0))

Call:
arima(x = Y, order = c(2, 0, 0))

Coefficients:
         ar1      ar2  intercept
      1.6474  -0.6475    230.406
s.e.  0.0003   0.0003        NaN
sigma^2 estimated as 0.6061:  log likelihood = -198.27,  aic = 404.54
Warning messages:
1: In log(s2) : NaNs produced
2: In sqrt(diag(x$var.coef)) : NaNs produced
>
> # ARIMA(0,0,1) or MA(1)
> arima(Y, order = c(0,0,1))
Call:
arima(x = Y, order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      1.0000    64.6863
s.e.  0.0182     2.3345

sigma^2 estimated as 231.6:  log likelihood = -702.48,  aic = 1410.96
>
```

```
> # ARIMA(0,1,1)
> arima(d.Y, order = c(0,0,1))
Call:
arima(x = d.Y, order = c(0, 0, 1))
Coefficients:
         ma1  intercept
      0.4872     0.4654
s.e.  0.0579     0.0908

sigma^2 estimated as 0.6284:  log likelihood = -199.5,  aic = 404.99
>
> # ARIMA(1,1,1)
> arima(d.Y, order = c(1,0,1))

Call:
arima(x = d.Y, order = c(1, 0, 1))

Coefficients:
         ar1      ma1  intercept
      0.7245  -0.2547     0.4397
s.e.  0.1152   0.1682     0.1576

sigma^2 estimated as 0.5783:  log likelihood = -192.59,  aic = 393.17
>
> # ARIMA(1,1,3)
> arima(d.Y, order = c(1,0,3))

Call:
arima(x = d.Y, order = c(1, 0, 3))

Coefficients:
         ar1     ma1      ma2     ma3  intercept
      0.7334  -0.241  -0.1082  0.1217     0.4324
s.e.  0.1242   0.142   0.0970  0.0800     0.1664

sigma^2 estimated as 0.5638:  log likelihood = -190.48,  aic = 392.97
>
> # ARIMA(2,1,3)
> arima(d.Y, order = c(2,0,3))

Call:
arima(x = d.Y, order = c(2, 0, 3))

Coefficients:
         ar1      ar2      ma1     ma2     ma3  intercept
      1.5191  -0.7084  -1.0502  0.2100  0.3179     0.4405
s.e.  0.2253   0.1589   0.2103  0.1314  0.1036     0.1438

sigma^2 estimated as 0.5474:  log likelihood = -188.22,  aic = 390.44
>
```

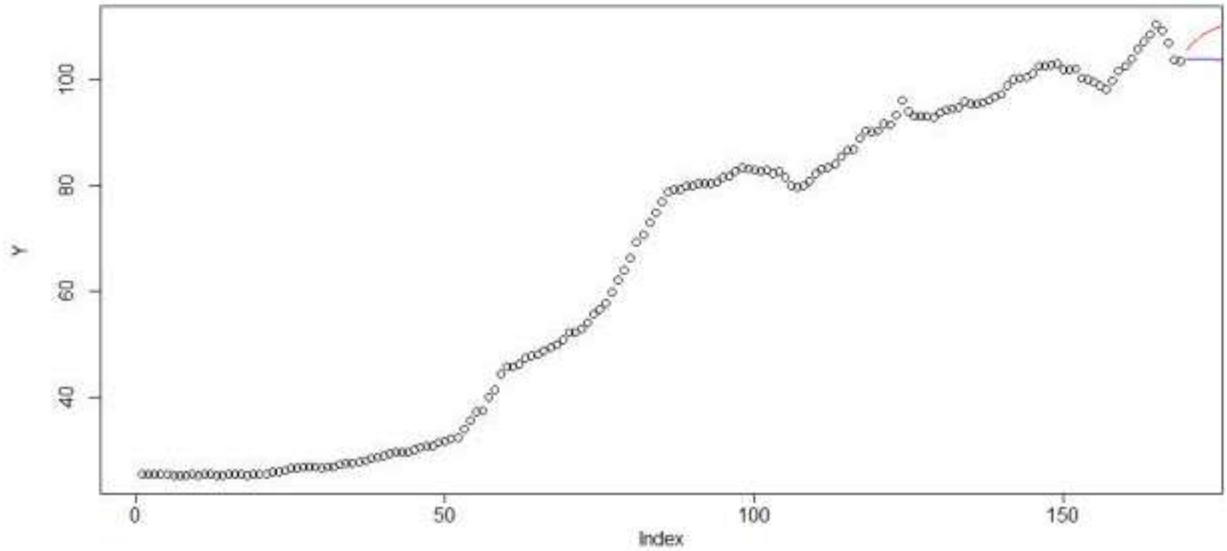
Table 1: ARIMA Models

	ARIMA (1,0,0)	ARIMA (2,0,0) -	ARIMA (0,0,1)	ARIMA (1,0,1)	ARIMA (1,1,0)	ARIMA (0,1,1)	ARIMA (1,1,1)	ARIMA (1,1,3)	ARIMA (2,1,3)
Const	64.37	64.18	64.69	64.67	0.46	0.47	0.43	0.43	0.44
L1.ar	0.999	1.64	-	0.99	0.55	-	0.72	0.73	1.51
L2.ar	-	-0.64	-	-	-	-	-	-	-0.71
L1.ma	-		1	0.53 -		0.48	-0.25	-0.24	-1.05
L2.ma	-	-	-	-	-	-	-	-0.1	0.21
L3.ma	-	-	-	-	-	-	-	0.12	0.32
AIC	502	424	1401	441	412	405	393	392	390
BIC	511	426	1420	543	408	414	406	411	412

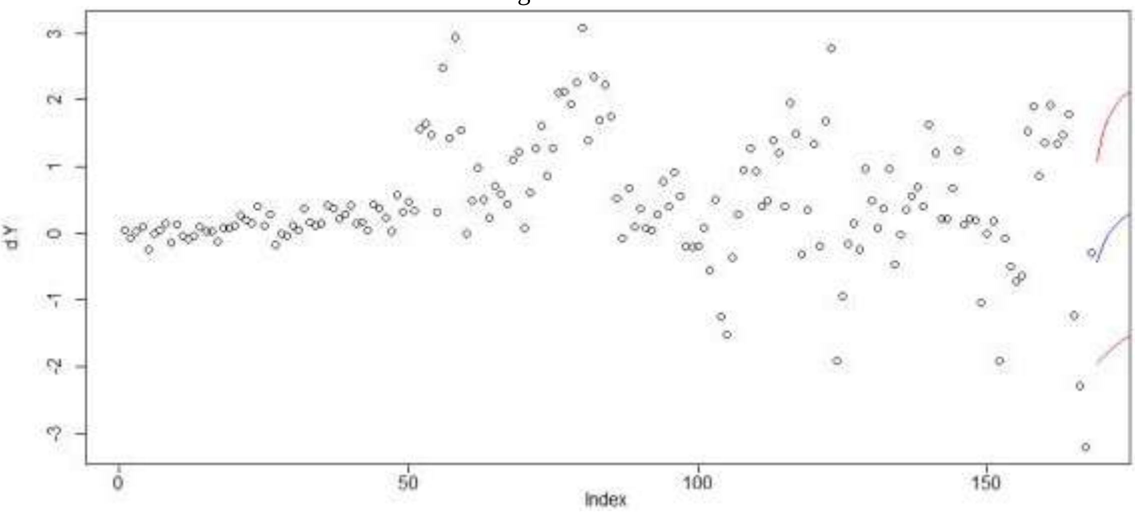
The above table mentions the lag coefficients obtained when we tested the data set with the different ARIMA models mentioned. On comparing the AIC values we obtain the minimum AIC to be associated with ARIMA (1,1,1)

4.2.5 Forecasting

Forecasting of PPI variable



Forecasting of DPPI variable



We have shown the forecasts of PPI and DPPI in the above two graphs with a 5% confidence interval(2.5% both side), on an ARIMA(1,1,1) for the difference variable.

5. Conclusion

In this term paper we have analyzed the Producer Price Index of USA, and forecast the PPI for the future years using a ARIMA Model. We found the rst difference to be stationary, and estimated the most appropriate ARIMA model which comes out to be a ARIMA (1,1,1) model.

The model was verified by comparing with other models and we obtained a minimum AIC value of 393 and the corresponding BIC to be 406 corresponding to our model, indicating that it is better than the other models.

References

[1] D.M.K.N. Seneviratna and Mao Shuhua, Forecasting the Twelve Month Treasury Bill Rates in Sri Lanka: Box Jenkins Approach, IOSR Journal of Economics and Finance (IOSR-JEF) 1(1), 2013, 44 - 47.

[2] M. Al-Shiab, The Predictability of the Amman Stock Exchange using the Univariate Autoregressive Integrated Moving Average (ARIMA) Model, Journal of Economic & Administrative Sciences, 22(2), 2006.

[3] P. Chujai, N. Kerdprasop, and K. Kerdprasop, Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models, Proc. IMECS Conf., Hong Kong, 2013.

[4] S. Nasiru and S. Sarpong, Empirical Approach to Modelling and Forecasting Inflation in Ghana, Current Research Journal of Economics Theory, 4(3), 2012, 83-87

[5] W. Dongdong, "The Consumer Price Index Forecast Based on ARIMA Model," IEEE Computer Society Washington, DC, USA ©2010, vol. 01, pp. 307-310, 14-15 August 2010.

[6] Okasha, K.M., Abu Shanab, M.M.D. (2014): Forecasting Monthly Water Production in Gaza City Using a Seasonal ARIMA Model Scholars Journal of Physics. Mathematics and Statistics, vol. 1, no. 2, pp. 61-67

[7] Olanrewaju, K. O., Olakunle, O. A., Emmanuel, A. O. (2014): Forecast Performance of Multiplicative Seasonal Arima Model: an Application to Naira/ Us Dollar Exchange Rate, American Journal of Applied Mathematics and Statistics, vol. 2, no. 3, pp. 172- 178

[8] Box, G., Jenkins, G. (1970): Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day

[9] "https://www.datascience.com/blog/introduction-to-forecasting-with-arima-in-r-learn-data-science-tutorials" [Online]

[10] "http://fmwww.bc.edu/repec/sce2004/up.8248.1077921898.pdf" [Online]