Stochastic Time Series Modeling for Coking Coal Production in India

[1]Dr. T. Jai Sankar, [2]Ms. I. Angel Agnes Mary, [3]Ms. K. Nalini

[1]Assistant Professor, [2]Guest Faculty, [3]Guest Faculty
Department of Statistics
Bharathidasan University, Tiruchirappalli – 620 023, India.

Abstract:
This paper deals with the stochastic time series modelling for Coking Coal (Metallurgical and Non-Metallurgical) production in India during the years from 1981 to 2021. Coking Coal is an essential input for production of Iron and Steel. The largest single use of Coal in the Steel Industry is as a fuel for the blast furnace and for the production of Coal for reduction of iron ore or for injection with the hot blast. This study considers Autoregressive (AR), Moving Average (MA) and ARIMA processes to select the appropriate ARIMA model for Coking Coal production in India. ARIMA (p, d, q) and its components autocorrelation function (ACF), partial autocorrelation function (PACF), root mean square error (RMSE), mean absolute percentage error (MAPE), normalized BIC and ARIMA (1,2,2). Based on the selected model, Coking Coal production in India is projected to decline from 44.79 million tonnes in 2022 to 32.6 million tonnes in 2031.

Index Terms - ARIMA, BIC, Forecasting, MAPE, Coal Production, RMSE.

I. INTRODUCTION:

Coking Coal is a naturally occurring sedimentary rock found in the Earth's crust. These include hard, semi-hard, semi-soft coking coal and powdered coal for injection (PCI). All these apply to different grades of Coal used for making steel. Coking Coal generally contains more carbon, less ash and less moisture than thermal coal, which is used for power generation.

Major part of the export trade is low volatile hard Coking Coal with high swelling index and good fluidity. 66% of Steel production relies on Coal inputs. World Crude Steel production was 1.6 billion tonnes in 2013. About 0.6 ton of Coal produces 1 ton of Steel, which means about 770 kg of Coal is used to produce 1 ton of Steel. Coal India Limited (CIL) marketing division coordinates marketing activities for all its subsidiaries. CIL has set up Regional Sales Offices and Sub-Sales Offices at selected locations in the country to cater the needs of the consuming sectors in various regions. In India's import policy, Coal can be freely imported by the consumers based on their commercial discretion considering their requirements.

Coking Coal is imported by Steel Authority of India Limited (SAIL) and other steel manufacturing units mainly to bridge the gap between demand and domestic availability and improve quality. Coal-based power plants, reinforce plants, captive power plants, sponge iron plants, industrial consumers and coal traders import the non-coking Coal.

II. MATERIAL AND METHODS:

As the aim of the study was to design and development of Stochastic Time Series modelling for Coking Coal production in India, various forecasting techniques were considered for use. ARIMA model, introduced by Box and Jenkins (1976), was frequently applied for discovering the pattern and predicting the future values of the time series data. Box and Pierce (1970) measured the distribution of residual autocorrelations in ARIMA. Akaike (1970) found the stationary time series by an AR (p), where p is finite and bounded by the same integer. Moving Average (MA) models were applied by Slutzky (1973).Jai Sankar et al. (2010) applied ARIMA (1,1,0) model for cattle production and forecast the yearly production of cattle in the Tamil Nadu during the period of 1970 to 2010. Jai Sankar (2011) used a stochastic model approach to fit and forecast fish product export in Tamil Nadu during the period of 1969 to 2008. Jai Sankar et al., (2011) selected ARIMA (1,1,0) model to for Bovine Production Forecasting in Tamil Nadu during the years from 1970 to 2008. Jai Sankar and Prabakaran (2012) applied ARIMA (1,1,0) model to forecast the milk production in Tamil Nadu during the period of 1978 to 2008. Jai Sankar (2014) considered ARIMA (0,1,1) model for designing of a Stochastic Model for Egg Production Forecasting in Tamil Nadu during the years from 1996 to 2008.Jai Sankar and Vijayalakshmi (2017) found ARIMA (1,1,1) model for Ghee production in Tamil Nadu during the years from 1977 to 2008. Jai Sankar et al., (2017) considered ARIMA (0,1,1) model for Rice production in India during the years from 1950 to 2013. Jai Sankar et al. (2017) applied ARIMA (1,1,0) model for Wheat production in India during the years from 1950 to 2013. Jai Sankar and Pushpa (2019) applied ARIMA (2,1,0) model for design and development of time series analysis for *Saccharum officinarum* Production in India during the years from 1950 to 2017. Xiaofan Zhang et al., (2020) used Coal price forecast based on ARIMA model forecasted up to May 2020 to December 2023. Jai Sankar and Pushpa (2022) applied ARIMA (1,1,0) model for *Solanum tuberosum* production in India during the years from 1950 to 2018. Jai Sankar and Pushpa (2022) used ARIMA (0,1,1) model for Bajra (*Pennisetum glaucum*) production in India during the years from 1951 to 2018. Jai Sankar and Pushpa (2022) used ARIMA (0,1,1) model for *Musa paradisiaca Linn* production in India during the years from 1961 to 2019.

The time series when differenced follows both AR and MA models and is known as ARIMA model. Hence, ARIMA model was used in this study, which required a sufficiently large data set and involved four steps: identification, estimation, diagnostic checking and forecasting. Model parameters were estimated to fit the ARIMA models.

Autoregressive process of order (p) is, $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$;

Moving Average process of order (q) is, $Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$; and

The general form of ARIMA model of order (p,d,q) is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where Y_t is Coking Coal production, ε_t 's are independently and normally distributed with zero mean and constant variance σ^2 for t = 1,2,..., n; d is the fraction differenced while interpreting AR and MA and ϕs and θs are coefficients to be estimated.

Trend Fitting**:** The Box-Ljung Q statistics was used to transform the non-stationary data into stationary data and also to check the adequacy for the residuals. For evaluating the adequacy of AR, MA and ARIMA processes, various reliability statistics like R^2, Stationary R^2, RMSE, MAPE, and BIC as suggested by Gideon Schwartz (1978) were used as below:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \right]^{1/2} ; \; MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{(Y_i - \hat{Y}_i)}{Y_i} \right| \; and$$

$$BIC(p,q) = \ln v^*(p,q) + (p+q) [\ln(n) / n]$$

where p and q are the order of AR and MA processes respectively and n is the number of observations in the time series and v* is the estimate of white noise variance σ^2.

III. RESULTS AND DISCUSSION:

In this study, data were collected from the Annual Report (2021-2022) of the Directorate of Economic Survey and Statistics of the Ministry of Coal, Government of India for the period 1981 to 2021 and used to fit the ARIMA model to predict the future Coking Coal production.

Table 1: Actual Coking Coal Production (million tons) in India

Year	Production	*Year*	Production	*Year*	Production
1981	32.6	*1995*	44.2	*2009*	34.8
1982	36.1	*1996*	40.1	*2010*	44.4
1983	37.6	*1997*	40.5	*2011*	49.6
1984	36.3	*1998*	43.5	*2012*	51.7
1985	36.6	*1999*	39.2	*2013*	51.6
1986	35.7	*2000*	33.2	*2014*	56.8
1987	39.5	*2001*	31.1	*2015*	57.5
1988	41.0	*2002*	28.7	*2016*	60.8
1989	42.8	*2003*	30.2	*2017*	61.6
1990	44.4	*2004*	29.4	*2018*	40.2
1991	45.3	*2005*	30.2	*2019*	41.1
1992	46.3	*2006*	31.5	*2020*	52.9
1993	45.3	*2007*	32.1	*2021*	44.8
1994	45.1	*2008*	34.5		

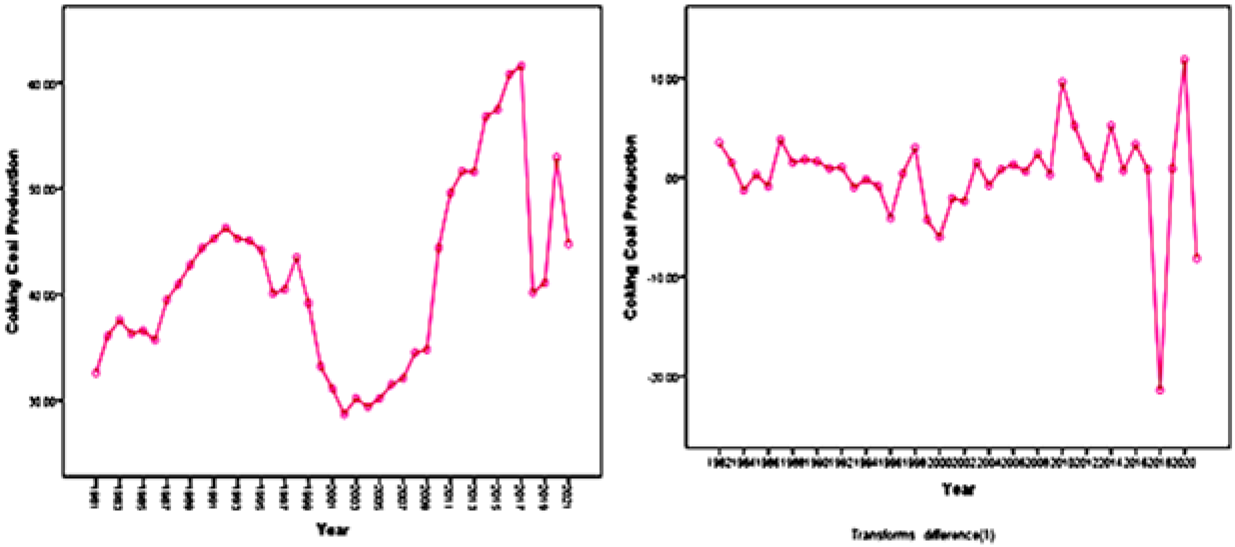

Figure 1. Time plot of Coal Production and First Differencing for Stationarity

Figure 1 depicts that the data were used for time plot of Coal production and first differencing for stationarity, and shows that after first differencing, the time plot is non-stationary.

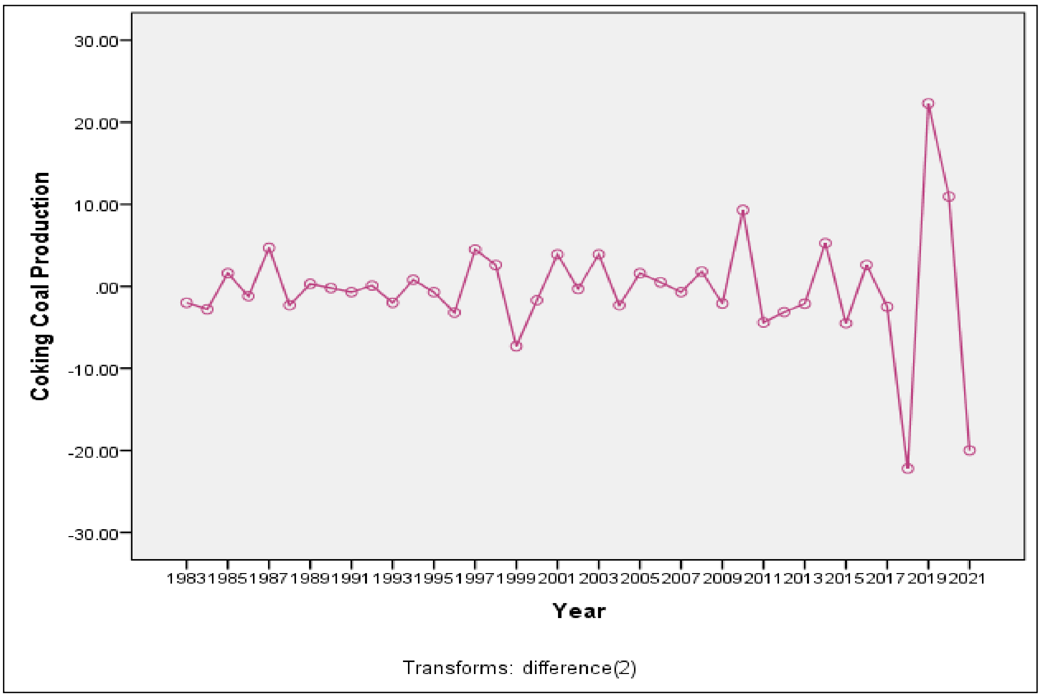

Figure 2. Time plot of Coking Coal Production with Second Differencing

Figure 2 reveals that, again non-stationarity in mean was corrected through second differencing of the data. The newly constructed variable Y_t could now be examined for Stationarity. Since, Y_t was stationary in mean, the next step was to identify the values of p and q. For this, the ACF and PACF of various orders of Y_t were computed and presented in Table 2 and Figure 3.

Table 2: ACF and PACF of Coking Coal Production

Lag	AC		Box-Ljung Statistic		PAC		Lag	AC		Box-Ljung Statistic		PAC	
	Value	Std. Error	Value	Sig.[b]	Value	Std. Error		Value	Std. Error	Value	Sig.[b]	Value	Std. Error
1	-0.297	0.154	3.719	0.054	-0.297	0.160	*13*	0.011	0.128	19.027	0.122	0.039	0.160
2	-0.437	0.152	11.967	0.003	-0.576	0.160	*14*	0.013	0.125	19.038	0.164	-0.057	0.160
3	0.279	0.150	15.425	0.001	-0.159	0.160	*15*	-0.092	0.123	19.602	0.188	-0.125	0.160
4	-0.043	0.148	15.510	0.004	-0.368	0.160	*16*	0.071	0.120	19.949	0.223	-0.108	0.160
5	0.024	0.146	15.536	0.008	-0.046	0.160	*17*	-0.026	0.117	19.999	0.274	-0.197	0.160
6	0.095	0.144	15.976	0.014	-0.005	0.160	*18*	0.027	0.115	20.054	0.330	-0.003	0.160
7	-0.054	0.141	16.122	0.024	0.194	0.160	*19*	0.093	0.112	20.738	0.351	0.094	0.160
8	-0.156	0.139	17.375	0.026	-0.074	0.160	*20*	-0.178	0.109	23.394	0.270	0.041	0.160
9	0.147	0.137	18.522	0.030	0.148	0.160	*21*	-0.042	0.106	23.552	0.315	-0.075	0.160
10	0.035	0.135	18.591	0.046	-0.054	0.160	*22*	0.169	0.103	26.250	0.241	0.011	0.160
11	-0.086	0.132	19.010	0.061	0.080	0.160	*23*	-0.006	0.100	26.253	0.289	-0.074	0.160
12	0.012	0.130	19.019	0.088	-0.117	0.160	*24*	-0.099	0.097	27.297	0.291	-0.030	0.160

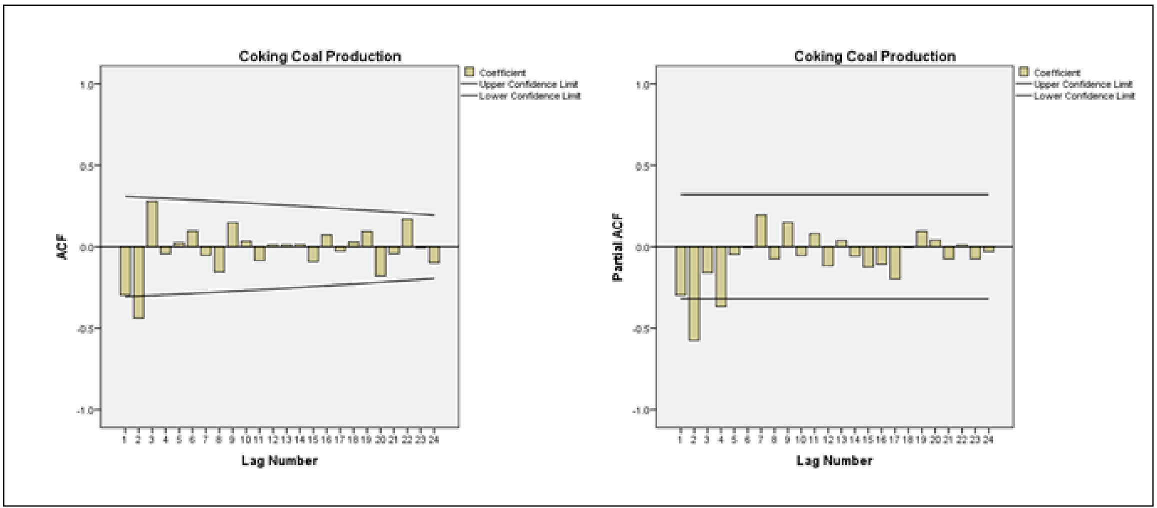

Figure 3. ACF and PACF of differenced data

Table 3: Estimated ARIMA Model Fit Statistics

Fit Statistic	Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC
(1,2,1)	0.462	0.653	5.293	7.378	54.276	3.056	21.819	3.615
(1,2,2)	**0.563**	**0.718**	**4.839**	**7.938**	**43.552**	**3.326**	**17.508**	**3.529**

The ARIMA models are discussed with values differenced twice (d=2) and themodel which had the minimum normalized BIC was chosen. The various ARIMA models and the corresponding normalized BIC values are given in Table 3. The value of normalized BIC ofthe chosen ARIMA was 3.529.

Table 4: Estimated ARIMA Model of Coking Coal Production

		Estimate	SE	t	Sig.
Constant		-0.023	0.088	-0.258	0.798
AR	**Lag 1**	-0.564	0.216	-2.618	0.013
Difference		2			
MA	**Lag 1**	0.002	1.691	0.001	0.999
	Lag 2	0.988	1.577	0.627	0.535

The ACF and PACF of the residuals are given in Figure 4, which also indicated the 'good fit' of the model. Hence, the fitted ARIMA model for the Coking Coal production data was

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$$\mathbf{Y_t = -0.023 - 0.564Y_{t-1} - 0.002\varepsilon_{t-1} - 0.988\varepsilon_{t-2} + \varepsilon_t}$$

Table 5: Residual of ACF and PACF of Coking Coal Production

Lag	ACF	Std. Error	PACF	Std. Error	*Lag*	ACF	Std. Error	PACF	Std. Error
1	-0.130	0.160	-0.130	0.160	*13*	-0.040	0.185	-0.029	0.160
2	-0.180	0.163	-0.201	0.160	*14*	-0.050	0.185	-0.021	0.160
3	0.197	0.168	0.151	0.160	*15*	-0.096	0.185	-0.139	0.160
4	0.066	0.174	0.085	0.160	*16*	0.039	0.186	-0.024	0.160
5	-0.003	0.174	0.085	0.160	*17*	0.026	0.187	-0.005	0.160
6	0.075	0.174	0.088	0.160	*18*	0.052	0.187	0.082	0.160
7	-0.116	0.175	-0.120	0.160	*19*	0.075	0.187	0.100	0.160
8	-0.170	0.177	-0.225	0.160	*20*	-0.132	0.188	-0.146	0.160
9	0.036	0.181	-0.114	0.160	*21*	0.005	0.190	-0.121	0.160
10	-0.049	0.181	-0.109	0.160	*22*	0.139	0.190	-0.069	0.160
11	-0.132	0.182	-0.088	0.160	*23*	0.042	0.193	0.002	0.160
12	-0.051	0.184	-0.054	0.160	*24*	-0.086	0.193	-0.040	0.160

Model parameters were estimated and reported in Table 3 and Table 4. The model verification is concerned with checking the residuals of the model to improve on the chosen ARIMA (p,d,q). This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders, up to 24 lags were computed and the same along with their significance which is tested by Box-Ljung test are provided in Table 5. This proves that the selected ARIMA model is an appropriate model.

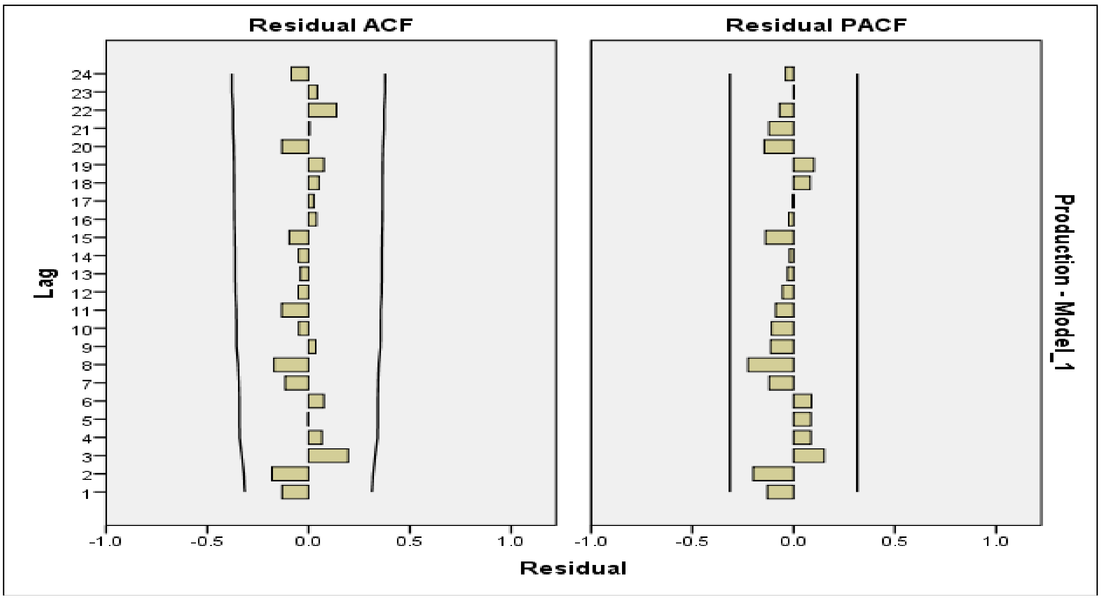

Figure 4. Residuals of ACF and PACF

Forecasting value of Coking Coal production from the year 2022 to 2031 given in Table 6. To assess the forecasting ability of the fitted ARIMA (p,d,q) model, important measures of the sample period forecasts' accuracy were computed. Figure 5shows that the actual and forecasted value of Coking Coal production data with 95% confidence limits.

Table 6: Forecast of Coking Coal Production(million tons)

Year	Forecast	UCL	LCL
2022	41.26	50.95	31.57
2023	42.87	59.83	25.91
2024	41.55	62.32	20.77
2025	41.84	66.62	17.06
2026	41.19	69.21	13.17
2027	41.03	72.2	9.87
2028	40.56	74.63	6.49
2029	40.23	77.03	3.43
2030	39.78	79.26	0.31
2031	39.37	81.32	2.59

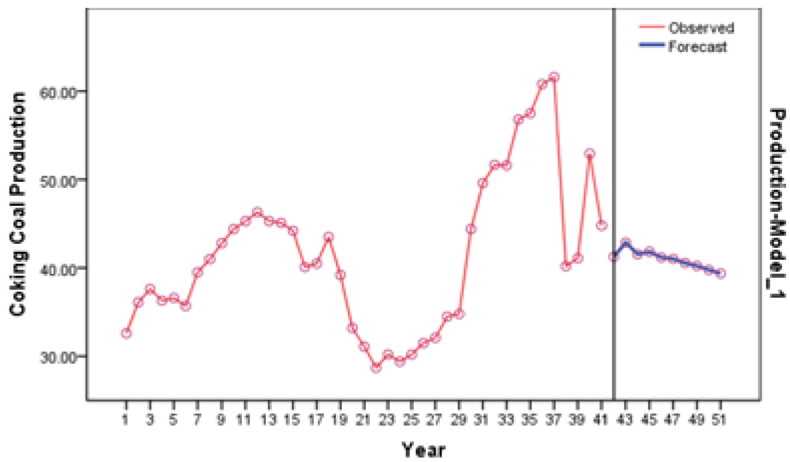

Figure 5. Actual and Estimate of Coking Coal Production

IV. CONCLUSION:

The results showed that, the Coal production would not remain stable throughout the year. The most appropriate ARIMA model for Coking Coal production forecasting of data was found to be ARIMA (1,2,2). From the temporal data, it can be found that forecasted production would decrease from 44.79 million tonnes in 2022 to 32.6 million tonnes in 2031 in India using time series data from 1981 to 2021 on Coking Coal Production. In future, these results will be helpful for taking necessary steps from the government authorities to increase Coking Coal production in India.

REFERENCES:

[1] Akaike, H. (1970). Statistical Predictor Identification. *Annals of Institute of Statistical Mathematics*. 22: 203-270.

[2] Box, G.E.P. and Jenkins, G.M., (1976). Time Series Analysis: Forecasting and Control. San Francisco, Holden-Day, California, USA. https://link.springer.com/chapter/10.1057/9781137291264_6.

[3] Box, G.E.P. and Pierce, D.A., (1970). Distribution of Residual Autocorrelations in ARIMA Models. *J. American Stat. Assoc.*,**65**: 1509-1526. https://www.jstor.org/stable/ 2284333?seq=1.

[4] Jai Sankar, T., Prabakaran, R., Senthamarai Kannan, K. and Suresh, S (2010). Stochastic Modeling for Cattle Production Forecasting. *International Journal of Modern Mathematics and Statistics.* 4(2): 53-57. https://medwelljournals.com/fulltext/? doi=jmmstat.2010.53.57

[5] Jai Sankar, T. (2011). Forecasting Fish Product Export in Tamilnadu - A Stochastic Model Approach. *International Journal of Recent Research in Science and Technology.* 3(7): 104-108. https://updatepublishing.com/journal/index.php/rrst/article/view/ 749/734.

[6] Jai Sankar, T., Senthamarai Kannan, K., Suresh, S. and Prabakaran, R., (2011).Stochastic Modeling for Bovine Production Forecasting. *International Journal of Agricultural and Statistical Sciences*. 7(1): 141-148. https://www.researchgate.net/ publication/289812991_Stochastic_modelling_for_bovine_production_forecasting.

[7] Jai Sankar, T. and Prabakaran, R. (2012). Forecasting Milk Production in Tamilnadu. *International Multidisciplinary Research Journal.* 2(1): 10-15. https://www.academia.edu/62866649/Forecasting_milk_production_in_Tamilnadu?bulk.

[8] Jai Sankar, T. (2014). Design of a Stochastic Forecasting Model for Egg Production. *International Journal of Innovative Science, Engineering & Technology*. 1(6): 319-325. https://www.ijiset.com/v1s6/IJISET_V1_I6_52.pdf.

[9] Jai Sankar, T. and Vijayalakshmi, C. (2017). Implementation of ARIMA Model for Ghee Production in Tamilnadu. *International Journal of Pure and Applied Mathematics.* 113(6): 56-64. https://acadpubl.eu/jsi/2017-113-pp/articles/6/7.pdf.

[10] Jai Sankar, T. , Pushpa, P. and Vijayalakshmi, C. (2017). Stochastic Forecasting Analysis for Rice Production in India. *Far East Journal of Theoretical Statistics.* 53(2): 63-75. http://www.pphmj.com/abstract/10739.htm.

[11] Jai Sankar, T., Pushpa, P. and Vijayalakshmi, C. (2017). Stochastic Time Series Analysis for Wheat (*Triticum aestivum*) Production in India. *International Journal of Pure and Applied Mathematics.* 113(7): 111-119. https://acadpubl.eu/jsi/2017-113-pp/articles/7/14.pdf.

[12] Jai Sankar, T.and Pushpa, P. (2019). Design and Development of Time Series Analysis for *Saccharum officinarum* Production in India. *A Journal of Composition Theory*. 12(9): 203- 211. http://www.jctjournal.com/gallery/29-sep2019.pdf.

[13] Jai Sankar, T.and Pushpa, P. (2022).Design and Development of Stochastic Modelling for *Solanum tuberosum* Production in India. *Journal of Algebraic Statistics.* 13(2): 3353-3361. https://www.publishoa.com/index.php/journal/article/view/1067.

[14] Jai Sankar and Pushpa (2022). Design and Development of ARIMA Model for Bajra (*Pennisetum glaucum*) Production in India. *Journal of Algebraic Statistics.* 13(2): 3512-3523. https://publishoa.com/index.php/journal/article/view/1089.

[15] Jai Sankar and Pushpa (2022). Design and Development of Stochastic Modelling for *Musa paradisiaca Linn* Production in India. *Journal of Algebraic Statistics.* 13(2): 3591-3599. https://publishoa.com/index.php/journal/article/view/1103.

[16] Slutzky, E., (1973). The summation of random causes as the source of cyclic processes. *Econometrica*, **5**: 105-146. https://www.scribd.com/document/388512380.

[17] Schwarz Gideon (1978). Estimating the Dimension of a Model. *Ann. Statist.*, **6(2)**: 461-464. https://doi.org/10.1214/aos/1176344136.

[18] Xiaofan Zhang, Chao Liu and Yuhang Qian (2020). Coal Price Forecast Based on ARIMA Model. *Financial Forum.* **9(4):** 180-184. file:///C:/Users/Dell/Downloads/Coal_Price_Forecast_Based_on_ARIMA_M.