

Ms_AJEBA_122223.docx

by Reviewer

Submission date: 13-Aug-2024 03:50PM (UTC+0700)

Submission ID: 2431452267

File name: 119214_ Reviewer _Ms_AJEBA_122223_213087_702961652.docx (157.06K)

Word count: 5844

Character count: 31038

The Box-Jenkins approach: forecasting of organic sorghum production in Chhattisgarh, India

Abstract

Chhattisgarh has taken important steps towards promoting millets cultivation and improving the livelihood of farmers. To increase millets production in Chhattisgarh, the state government launched the Millet Mission in September 2021. This mission has been started with a view to make Chhattisgarh the 'millet hub of India'. The study was conducted on forecasting of sorghum crops in Chhattisgarh India using historical data on currently cultivated area, production and yield of sorghum crops. The time series data was collected from 2001 to 2023, and analysis of the study was carried out using path analysis and Box Jenkins ARIMA model; and among various 20 models the best and suitable ARIMA (2, 1, 2), (3, 1, 3), and (2, 1, 2) model was selected based on AIC, BIC, MAPE, RMSE, MAE. With the help of the selected appropriate model, the area, production and yield of sorghum cultivation in Chhattisgarh was forecasted for the year 2024 to 2030. But marvellous, stochastic and fluctuating trend was observed in sorghum production and yield over the forecast period.

Keywords: Econometric Modeling; Forecasting; SCAPY; AIC; BIC; AFC; MAPE;

JEL Code: C01, C22, C51, C52, C53

1. Introduction

Agriculture is the most important livelihood strategies in India, with two thirds of the county's workforce depend on farming. Organic farming can be seen as an approach to agriculture, where the aim is to create integrated environmentally and economically sustainable agricultural production systems (Krishna *et al.*, 2023). Sorghum is the major staple food of millions of rural poor in arid and semi-arid regions of the world. It is the second cheapest source of energy and micronutrients after pearl millet, and majority of population in the central India depends on sorghum for their dietary and energy requirements (Rao *et al.*, 2006). Because of its drought adaptation capability, sorghum is a preferred crop in tropical, warmer and semi-arid regions of the world with high temperature and water-stress conditions. In India, the crop is grown in both rainy (June–October) and post-rainy (November–February) seasons. In spite of its multiple uses as food, feed, fodder and bio-fuel, the area under grain sorghum has drastically declined (Mishra *et al.*, 2017). Sorghum is second larger millet crops, when compared with other crops in respect to the source of energy which is sorghum has pearl millet (361 Kcal/100 g), (349 Kcal/100 g), and maize with (325 Kcal/100 g). Sorghum has carbohydrate content of 67.5 g/100 g; with 56 to 65% starch content,

20 to 22% of amylase, 2.6-2.8% sucrose and fiber (1.2 g/100 g). It is a good source of vitamin and minerals (Pearl Millet News, 2018; Saravand et al., 2022).

Time series forecasting is an important statistical analysis technique used as a basis for manual and automatic planning in many application domains (Gooijer and Hyndman, 2006). Forecasts are calculated using mathematical models that capture a parameterized relationship between past and future values to express behaviour and characteristics of a historic time series. The parameters of these forecast models are estimated on a training data set to fit the specifics of the time series by minimizing the forecast error. Time series data collected in many situations are hierarchical in structure. These datasets generally contain information in clusters which can be combined into another series of interest. Here, the time series are aggregated along the hierarchy based on dimensional attributes such as location (Hyndman et al., 2011; Athanasopoulos et al. 2009).

India shares around 20% of the world's sorghum area and is the fourth largest producer of this cereal crop. The cultivation area of sorghum in India was more than 16 million ha in 1981, but has gradually decreased to 6.3 million ha in 2012. Production of this cereal has also faced a decline from 12 million tonnes to 6 million tonnes during this period. But marvellous, yield of sorghum, on the contrary, has climbed up from 7.3 tonnes/ha to only 9.5 tonnes/ha in the same tenure (Pal and Paul, 2016). Sorghum (Jowar) production in India stood at 4803.38 thousand tonnes from the cultivated area of 5024.45 thousand hectares in the year 2017-18, and in the of 2019-20 sorghum production 4772.01 thousand tonnes from the cultivated area of 4823.76 thousand hectares. While Sorghum production in Chhattisgarh 5.32 thousand tonnes from the cultivated area of 3.59 thousand hectares in the year 2017-18, and in the of 2019-20 sorghum production 4.07 thousand tonnes from the cultivated area of 2.90 thousand hectares (IMR, 2023; DES, 2023).

In this present study researchers' focus on forecasting of the sorghum production in Chhattisgarh State. Sorghum (Jowar) is produced in almost all the areas of Chhattisgarh state. But it's mostly cultivated area in Surguja division- Balarampur, Koriya, Surajpur, Jashpur and Sarguja (Ambikapur) district, Bastar division-Sukma, Bijapur, Bastar (Jagdulpur), Kanker, Kondagaon and Narayanpur district, Raipur division-Gariyaband, Dhamtari and Mahasamund district, Durg division- Rajnagaonv, Kabirdham, Balod and in Bilaspur division- Korba, Bilaspur district of the state. There is immense potential for increase in the production of sorghum in Chhattisgarh. To increase millets production in Chhattisgarh, the state government launched the Millet Mission in September 2021. This mission has been started with a view to make Chhattisgarh the 'millet hub of India'. This mission has not only increased the income of farmers in forest and tribal areas, but has also increased the prominence of the state (Babu et al., 2023). However, millet remains an important crop for the state's food security and cultural heritage. Encouraging millet cultivation and consumption in Chhattisgarh can not only provide nutritional benefits to the population, but also contribute to sustainable agricultural development and food security.

2. Literature Review

Many research works have been done by researchers in the past on sorghum production, the review of which is as follows. (Saravand et al., 2022) was conducted Pearl millet crop in Gujarat and India with the use of historical data on an area, production and yield of Pearl millet crop. The data was collected for 20 years from the year 1999-2000 to 2018-2019, and analysis was carried out using Compound Growth Rate, path analysis and Box Jenkins' ARIMA model. The best selected ARIMA model was (0, 0, 6) and (0, 0, 5) for Gujarat and India respectively. (Lama et al., 2020) was studies of price index of Ragi and used structural break analysis. The volatile ragi price index series were modelled and forecasted used of GARCH model and its asymmetric extensions. The results indicated improvement in modelling and forecasting performance of the models after incorporation of the policy interventions. (Bellundagi et al., 2016) conducted research on ragi production in Karnataka, and different linear and nonlinear growth models were explored. The forecasting results showed that, even though there was a deceleration in area, the production of ragi was increasing due to increase in productivity in the future time. (Vijay and Mishra, 2018) Studies was pearl millet production in Karnataka, and used ARIMA and ANN models, and ARIMA (0, 1, 1) model was selected for forecasting of the future value from 2011 to 2014. Thus, following researchers was conducted research on millet production, i.e. (Tripathi et al., 2013) was research on pearl millet production and productivity, (Prabhu et al., 2022) was conducted research on forecasting minor millet in India, (Sathish et al., 2022) was studies on trend analysis of minor millet in India, (Sankar and Pushpa, 2023) was studies on forecasting of millets production in India. (Nireesha et al., 2016) was conducted research on pearl millet production in Andhra Pradesh, India, and also some of the investigated works were i.e., Kour et al., 2017; Das et al., 2019; Dharamraja et al., 2019; Patra and Mahapatra, 2020; Chandra, 2023; Gandhi et al., 2023; and Chandra, 2024.

We have documented a detailed literature on time series analysis and prediction the various data series from 1950-51 to 2022-23. But researchers' carried study on forecasting of pearl millet production, millet (Ragi) prices, minor millet production, tea production, groundnut production and coffee production, and moreover study related to Karnataka, Gujarat, Odisha and Andhra Pradesh; and no study has been found regarding sorghum production in Chhattisgarh. Thus, it's a gap and motivated us to undertake study on time series modeling and forecasting of sorghum production in Chhattisgarh, India.

3. Methods and Methodology

3.1. Data Collection

We used time series data from 2001 to 2023 for the research study, time series data was compiled of official website of Indian Institute of Millet Research (IIMR) <https://www.milletstats.com/apy-stats/>, and official website of Directorate of Economics and Statistics, DA&FW, Govt. of India, website <https://desagri.gov.in/statistics-type/normal-estimates/>.

3.2. Econometric Models

To select the best fitted ARIMA model, several statistical tools are being applied, viz., AIC (Akaike, 1974; Priya et al., 2015; Chandra, 2023; Chandra, 2024), BIC/SIC (Schwarz, 1978), MAPE (Lewis, 1982; Armstrong and Collopy, 1992; Goodwin and Lawton, 1999; Ren and Glasure, 2009; Morcno et al., 2013), Ljung-Box test (Box and Jenkins; 1978; Saravand et al., 2022), RMSE (Draxler, 2014), and MAE (Reddy, 2000; Gujarati et al. 2015; Tofallis, 2015; Chandra and Brahme, 2023, Chandra et al. 2023), and thus, the formulation of the models are given bellow:

AIC written as follow:

$$AIC = \{n(1 + \log 2\pi) + n \log \sigma^2 + 2m\} \dots\dots\dots (1)$$

AIC = (-2log L + 2m); where: m= p + q. L = Likelihood function and -2log L = approximately equal to $\{n(1 + \log 2\pi) + n \log \sigma^2\}$, where: σ^2 = the model MSE.

$$BIC = \log\left(\frac{RSS}{n}\right) + \frac{k}{n} \log n \dots\dots\dots (2)$$

Where, "RSS" = the residual sum of squares; k = the number of coefficients estimated, i.e., 1 + p + q + P + Q; and n = the number of observations,

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{X_i - \hat{X}_i}{X_i} \right| \dots\dots\dots (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X}_i)^2}{n}} \dots\dots\dots (4)$$

$$MAE = \sum_{i=1}^n \left| \frac{X_i - \bar{X}_i}{X_i} \right| \dots\dots\dots (5)$$

$$Q = n(n+2) \sum_{i=1}^k \frac{r_i^2}{n-i} \dots\dots\dots (6)$$

Where, n = the number of observations, r^2 = value of i^{th} the number of observations.

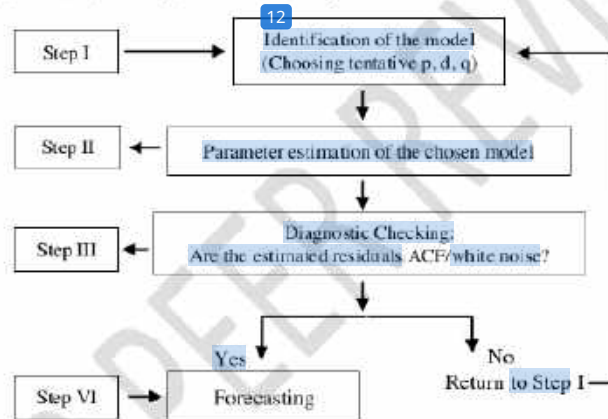
3.3. The Box-Jenkins Approach

Step I- In the first step of the study, the time series data selected for the study was collected from a reliable source, and after which a graphical presentation of the series has seen, which shows whether the series has showed a trend or not. After this, stationarity of the series has checked at the level, and if there has not stationarity at the category level, then stationarity has checked at the first difference, and if there has no stationarity of the series at the first difference, then this process continues till this continues until the series becomes stationary. By the way, most of the series becomes stationary at the first difference. After the series has stationary, the correlogram has seen, with the help of which the model has selected

(Gujarati et al. 2015; Chandra et al. 2023; Chandra, 2023, Chandra, 2024). Figure 1 shows the Box-Jenkins methodology consist of following four steps.

Step II- In the first step, the equation of the selected model (p, d, q) is derived, and then the equation is created by writing the parameters of the selected model with their given values (Gujarati et al. 2015; Chandra, 2023, Chandra, 2024).

Step III- In the third step of the study, a diagnostic check of the residuals of the selected model (p, d, q) was carried out in the second step, in which the autocorrelation (ACF and PACF) Ljung-Box test, WNH, and JB test was done of the selected model. If all the tests after diagnostic testing of the residuals of the selected model (p, d, q) were found to be significant, then the forecasting process of the selected model is done, and if all the tests after diagnostic test of the residuals of were not found to be significant, then all the process is started again from the first phase of the study (Gujarati et al. 2015; Chandra et al. 2023; Ashoka et al., 2014; Chandra, 2023, Chandra, 2024).



Source: (Gujarati et al., 2015; Chandra, 2023; Chandra, 2024)

Figure 1 shows the Box-Jenkins methodology consist of following four steps.

Step IV- In the fourth step of the study, if after diagnostic testing residuals of the model selected in the third step, all the tests are found significant, and follow the Gauss-Markov theorem (Hellin, 2013). Thus, the model constructed is terms the best linear unbiased estimator “BLUE”. Thereafter the forecasting process of the selected model has completed, and then reporting of the model is done (Gujarati et al., 2015; Chandra et al. 2023; Chandra, 2023; Chandra, 2024).

3.4. Equation for ARIMA Models

ARIMA is a linear regression model for time series predicting, and it uses its own lags as predictors. Any 'non-seasonal' time series that exhibits patterns and is not a random white noise can be modelled with ARIMA models. An ARIMA model is characterised by 3 terms: p, d, q (Box and Jenkins, 1978; Gujarati et al. 2015; Chandra, 2023; Chandra and Brahme, 2023; Chandra et al. 2023; Chandra, 2024).

10

Where: p = the order of the AR term, q = the order of the MA term, and d = order of differencing required to make the series stationary (I).

11

Of course, it is quite likely that Y has characteristics of both AR and MA and is therefore ARMA. Thus, if Y follows an ARMA (3, 4) process, it can be written as (Chandra et al. 2023; Chandra, 2024):

18

$$Y_t = \theta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \beta_3 u_{t-3} + \beta_4 u_{t-4} + u_t \dots \dots \dots (7)$$

Where: θ = constant, α = coefficient of AR terms, β = coefficient of MA terms, u = white noise error terms

3.5. Model Identification for Sorghum cultivation area, production and yield

In generally, a non-stationary series is made stationary after differencing ' d times', and is said to be integrated of order ' d ', denoted by $I(d)$. If, the original series is stationary $d=0$, and then the ARIMA model transform into an ARMA model. The time series data used for the present study, i.e., SA, SP, and SY. The series SA, SP, and SY were become stationary after the 1st order differencing. Since, there is no need for further differencing the series, and it is necessary to adopt $d=1$ (first difference) for ARIMA (p, d, q) model. We have checked the correlogram after first difference, and the level in time series (given figure 5). Since, there was no significant spikes of ACF and PACF residuals of the selected ARIMA and ARMA models, to get the appropriate numbers for ' p ' (in AR) and ' q ' (in MA) in the models, and thereafter we were checked white noise in the correlogram after first difference in time series (given figure 5). Since, there was no significant spikes of ACF and PACF residuals of the selected ARIMA and ARMA models, and therefore there was no need for further consideration of any more AR (p) and MA (q). The models convince all the norms (comparatively lowest value of AIC, comparatively low values of BIC, and MAPE, MAE and RMSE). Therefore, these models have been considered to be the best predictive models that have been used to forecast future values of time series, such as DSCA, DSP and DSY. Table 2 shows that the best-fitting ARIMA model with parameters is selected, and Table 3 provides the estimation results of different parameters of AR (p) and MA (q) of ARIMA model for area, production and yield. Using these values, the best-fit ARIMA (p, d, q), models for predicting time series DSA, DSP and DSY were identified. Therefore, the prediction equations for the models can be written as follows. The equations for SCA (8), SP (9), and SY (10) respectively (Chandra, 2023, Chandra, 2024).

$$Y_t = \theta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \beta_1 u_{t-1} + \beta_2 u_{t-2} + u_t \dots \dots \dots (8)$$

$$Y_t = \theta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \beta_3 u_{t-3} + u_t \dots \dots \dots (9)$$

$$Y_t = \theta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \beta_1 u_{t-1} + \beta_2 u_{t-2} + u_t \dots \dots \dots (10)$$

4. Result Analysis

4.1. Stationary test (ADF test)

28

The results of Argument Dickey-Fuller (Dickey-Fuller, 1979) unit root test at level and 1st order difference given in table 1. Before differencing the time series SCA, SP and SY we performed the

stationary test at the level, but the p-values at the level were insignificant. Therefore, series SA, SP and SY were statically not significant. So, it's not stationary. After that we go through 1st differenced, and 1st order difference series DSA, DSP and DSY calculated t-statistics value was respectively = 5.814, 8.122 and 6.787 and p-value were respectively = 0.0006, 0.0000, and 0.0001 which was smaller than critical values at 1%, 5% and 10% level of significance. Hence, we fail to accept the null hypothesis for unit root. It means that the series DSCA, DSP and DSY was not containing the unit root, and thus it is stationary. Figure 4 part (a), part (b), and part (c) represent the plot of correlogram (ACF and PACF) of the stationary series DSCA, DSP and DSY for lags 1 to 12 at the 1st order difference. Figure 2 (a) Scatter plot and (b) Quantiles graphs of Sorghum Area, production and yield in Chhattisgarh. Figure 3 Representation of time series plot of (a) Original and (b) Stationary series.

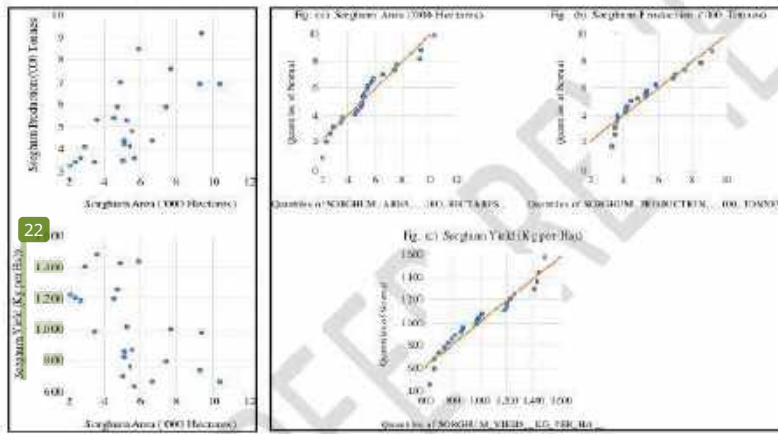


Fig. (a)

Fig. (b)

Figure 2 (a) Scatter plot and (b) Quantiles graphs of Sorghum Area, production and yield in Chhattisgarh

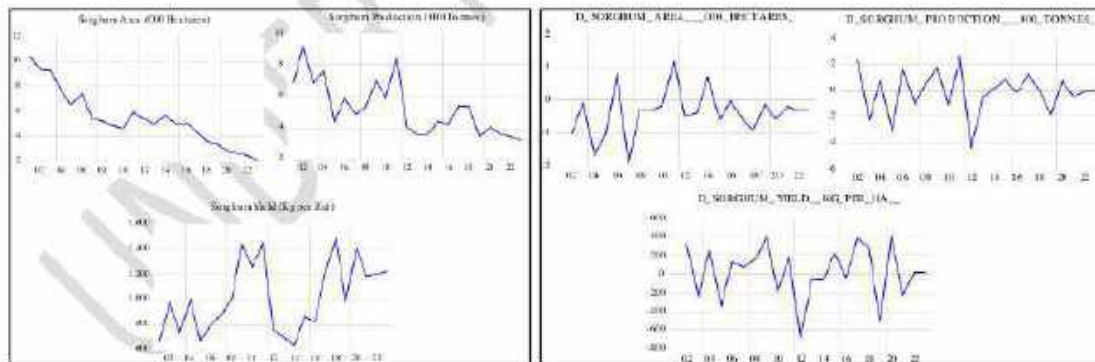


Fig. (a)

Fig. (b)

Figure 3 Representation of time series plot of Original and Stationary series

	Prob.	0.006	0.000	0.000	-	0.999	0.999	-	
(P)	C	-0.1720	0.840	-0.187	-0.496	-2.924	2.924	-0.999	-28.369
	Std. Error	0.0069	0.495	0.508	0.323	52.465	58.543	38.412	
	Prob.	0.0000	0.111	0.718	0.146	0.956	0.961	0.979	
(Y)	C	6.8737	0.828	-0.446	-	-0.436	0.999	-	5.167
	Std. Error	0.0909	0.307	0.256	-	632.03	2901.3	-	
	Prob.	0.0000	0.015	0.100	-	0.999	0.999	-	

Source: Authors' calculation Using EView12

Table 4 Results of the Ljung-Box test and JB test (Normality test)

Variable	Models	leg	Q-Stat.	P-value	Result (Ljung-Box)	Jarque-Bera (p-value)	Result (J-B test)	
Sorghum	(A)	(2, 1, 2)	12	13.525	0.095	Insignificant	0.615	Accepted
	(P)	(3, 1, 3)	12	7.1537	0.307	Insignificant	0.613	Accepted
	(Y)	(2, 1, 2)	12	11.878	0.157	Insignificant	0.512	Accepted

Source: Authors' calculation Using EView12

Based on the estimation results of ARIMA (2, 1, 2), (3, 1, 3), and (2, 1, 2) models (Intercept and coefficients given in table 3) respectively, and the functional form of the time series forecasting models may be presented as follows (Eq. 8, 9, and 10) according to given in table 3:

✦ Model for Sorghum Cultivation Area (SCA)-

$$Y_t = -0.068 - 1.493Y_{t-1} - 0.964Y_{t-2} + 1.695u_{t-1} + 1.000u_{t-2} + u_t$$

✦ Model for Sorghum Production (SP)-

$$Y_t = -0.1720 + 0.840Y_{t-1} - 0.187Y_{t-2} - 0.496Y_{t-3} - 2.924u_{t-1} + 2.924u_{t-2} - 0.999u_{t-3} + u_t$$

✦ Model for Sorghum Yield (SY)-

$$Y_t = 6.8737 + 0.828Y_{t-1} - 0.446Y_{t-2} - 0.436u_{t-1} + 0.999u_{t-2} + u_t$$

4.3. Forecasting Result

This research study is based on annual amount of the sorghum cultivation area, production and yield, and covering the period of 2001 to 2030 (30 observations); of which 23 observations ranging from 2001 to 2023 were historical data and 7 observations ranging the period of 2024 to 2030 was forecasted amount of sorghum cultivation area (SCA), production, and yield. In table 5 shows the forecasting results of ARIMA (2, 1, 2), (3, 1, 3), and (2, 1, 2) for sorghum produce area, production, and yield; and ARIMA (2, 1, 2), (3, 1, 3), and (2, 1, 2) models for SPA, SP and SY which was observed as the best suitable model for predicting the future amount of sorghum area, sorghum production, and sorghum yield respectively; and we have estimated that the yearly amount of SPA, SP and SY achieved in the year 2023-24 from 2.1593 (cultivation area '000' hectare), 2.1363 (production '000 Tonnes), and 1104.379 (yield kg/hectare) respectively to 1.3592 (cultivation area '000' hectare), 1.9874 (production '000 Tonnes), and

ORIGINALITY REPORT

28%

SIMILARITY INDEX

27%

INTERNET SOURCES

13%

PUBLICATIONS

10%

STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	7%
2	www.thepharmajournal.com Internet Source	5%
3	epubs.icar.org.in Internet Source	4%
4	mafiadoc.com Internet Source	2%
5	eurchembull.com Internet Source	1%
6	journalajeba.com Internet Source	1%
7	link.springer.com Internet Source	1%
8	krishi.icar.gov.in Internet Source	1%
9	ndpublisher.in Internet Source	1%

10	www.mdpi.com Internet Source	1 %
11	archive.org Internet Source	1 %
12	Submitted to Nottingham Trent University Student Paper	1 %
13	myresearchjournals.com Internet Source	<1 %
14	Nobin Chandra Paul, Dhananjay D. Nangare. "Trend Analysis of Area, Production and Productivity of Nutri-Cereals (Pearl Millet and Sorghum) in Maharashtra, India: Navigating Challenges for Food Security and the Way Out", National Academy Science Letters, 2024 Publication	<1 %
15	repositorio.tec.mx Internet Source	<1 %
16	Submitted to Kiriri Women's University of Science and Technology Student Paper	<1 %
17	cpbis.gatech.edu Internet Source	<1 %
18	etd.uum.edu.my Internet Source	<1 %

- | | | |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 19 | B S YASHAVANTH, K N SINGH, AMRIT KUMAR PAUL, RANJIT KUMAR PAUL. "Forecasting prices of coffee seeds using Vector Autoregressive Time Series Model", The Indian Journal of Agricultural Sciences, 2017
Publication | <1 % |
| 20 | www.nlp.ir.org
Internet Source | <1 % |
| 21 | www.semanticscholar.org
Internet Source | <1 % |
| 22 | aces.nmsu.edu
Internet Source | <1 % |
| 23 | orbi.umons.ac.be
Internet Source | <1 % |
| 24 | www.aseestant.ceon.rs
Internet Source | <1 % |
| 25 | www.ijcmass.com
Internet Source | <1 % |
| 26 | 1library.net
Internet Source | <1 % |
| 27 | Hamid R. Seddighi. "Introductory Econometrics - A Practical Approach", Routledge, 2013
Publication | <1 % |
| 28 | dokumen.pub
Internet Source | <1 % |

<1 %

29

udsspace.uds.edu.gh

Internet Source

<1 %

30

www.tdx.cat

Internet Source

<1 %

31

"Partial Identification in Econometrics and Related Topics", Springer Science and Business Media LLC, 2024

Publication

<1 %

32

Energy Time Series Forecasting, 2015.

Publication

<1 %

33

eerstekamer.nl

Internet Source

<1 %

34

irc.kdu.ac.lk

Internet Source

<1 %

35

journals.ansfoundation.org

Internet Source

<1 %

36

krishikosh.egranth.ac.in

Internet Source

<1 %

Exclude quotes On

Exclude matches < 1 words

Exclude bibliography On