Full Research Article

Area, Production and Productivity of Wheat (*Triticum aestivum*) in Gujarat State: Forecasting by Using ARIMA Models

L. Netajit Singh[1#*], V. B. Darji[2], Y. Santosh Singh[3]

[1]Anand Argicultural University, Anand, Gujarat, (388 110), India
[1#]*Presently* N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat (369 450), India
[2]Anand Agricultural University, Anand, Gujarat (388 110), India
[3]Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal, India (741 252), India

Abstract

The present investigation was carried out to fit Autoregressive Moving Average (ARIMA) models to arrive at a methodology that can precisely explain the fluctuations of area, production and productivity for wheat crop data in Gujarat state after checking the stationary condition. The data from year 1960−61 to 2012−13 were used for model fitting and forecasting five years ahead from the year 2012−13. The ARIMA models with different p, d and q were judged on the basis of auto correlation function (ACF) and partial auto correlation function (PACF) at various lags. Among different fitted ARIMA models, the final models were selected on the basis of significant autoregressive and moving average term, Akaike's Information Criterion (AIC), Schwartz-Bayesian Criterion (SBC), test of normality (Shapiro-Wilk test) and randomness of residual's (Run test) distribution. Among the ARIMA models, ARIMA (0, 1, 1) family model was found suitable to forecast the pattern of wheat area and production and ARIMA (1, 1, 0) was found suitable for forecasting of wheat productivity trend of Gujarat State. Forecasted values showed an increasing pattern in area, production and productivity of wheat in Gujarat State and predicted values for area, production and productivity of wheat in the year 2017−18 are 12989.2 hundred ha, 40296.9 thousand t and 3148.42 kg ha^{-1} respectively.

1. Introduction

Wheat (*Triticum aestivum*) is the most important food grain crop of the World. In Gujarat, wheat is grown over an area of 1.05 mha with the production of 3.13 mt. and productivity of 2986 kg ha^{-1} (Anonymous, 2013). Gujarat accounted for 1.75% of the total area and 1.32% of the total production of wheat in the country. Mehsana, Banaskantha, Rajkot and Kheda districts in the valleys of the Sabarmati and Mahi rivers are the main producers which together contribute about 55% of the state's production of wheat. Others include Ahmedabad, Sabarkantha, Bharuch and Bhavnagar districts where 6 to 10% of the cropped area is devoted to wheat cultivation.

Wheat is the second most important cereal crop in India after rice and it is severely affected with abiotic factors e.g. Rainfall, humidity and other environmental factors and biotic stresses such as diseases and pest infestation which also indirectly depends upon environment. There are several statistical tools available to predict/forecast the wheat production with the help of assessing the environmental influence on yield.

ARIMA model is an extrapolation method for forecasting and like any other such method, it requires only the historical time series data on the variables under forecasting. Among the extrapolation methods, this is one of the most sophisticated method, as it incorporates the future of all such methods, does not require the investigator to choose initial values of any variables and values of the various parameters a priori. It is robust to handle any data pattern. As one would expect this is quite a difficult model to develop and apply as it involves transformation of the variable, identification of the model, estimation through nonlinear method, verification of the model and derivation of the forecasts (Gupta, 1993).

This paper applies Autoregressive Integrated Moving Average (ARIMA) forecasting model, the most popular and widely used forecasting models for uni-variate time series data. Although it is applied across various functional areas, it's application is very limited in agriculture, mainly due to

unavailability of required data and also due to the fact that agricultural product depends typically on monsoon and other factors, which the model failed to incorporate. In this context, it is worth mentioning, few applications of ARIMA model for forecasting agriculture product. Applying ARIMA model Hossian et al. (2006) forecasted three different varieties of pulse prices namely motor, mash and mung in Bangladesh with monthly data from Jan 1998 to Dec 2000; Wankhade et al. (2010) forecasted pigeon pea production in India with annual data from 1950−1951 to 2007−2008; Mandal (2005) forecasted sugarcane production in India; Khin et al. (2008) forecasted natural rubber price in World market; Rachana et al. (2010), used ARIMA models to forecast pigeon pea production in India. Badmus and Ariyo (2011), forecasted area of cultivation and production of maize in Nigeria using ARIMA model. They estimated ARIMA (1, 1, 1) and ARIMA (2, 1, 2) for cultivation area and production respectively. September 2004. With these exceptions, there is paucity of studies regarding applications of ARIMA model for forecasting area, production and productivity of agricultural products. The main aim of this study is to forecast area, production and productivity of wheat in Gujarat. Forecasting of cultivated land will help the government to make policy on available land used and further food capacity. Forecasting values of production and productivity also important for implementing proper management to face an eventual food deficit in the state of Gujarat.

2. Materials and Methods

The time series data on area, production and productivity of wheat crop for the period 1960−61 to 2012−13 obtained from Directorate of Agriculture, Gujarat state, Gandhinagar were used to fit the ARIMA models.

In regression model, the parameters β's are assumed to be constant over the time. In the forecasting models the errors ε_t's within time period (t=l, 2, 3, . . ., n) are assumed to be uncorrelated i.e. the observations Y_t's are uncorrelated. However, this assumption is rarely met in practice. Usually serial correlations in the observations often exist in time-series data. The statistical concept of autocorrelation was used to measure the relationships between the value of Z at time t (i.e., Y_t) and Y at earlier time periods (i.e., Y_{t-1}, Y_{t-2} ...). The algebraic forms of Autoregressive (AR) and Moving average (MA) processes are:

2.1.1. Autoregressive (AR) process

$$Z_t = C + \varphi_1 Y_{t-1} + a_t \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(1)$$

Where Z_t=Time sequenced random variable
C=Constant term related to mean (μ) such that $C = \mu(1-\varphi_1)$
φ_1=relationship of Y_t with Y_{t-1}

a_t=a random shock element at time t
Similarly, the MA (q) model is again the generalizations of moving average model may be specified as.

2.1.2. Moving average (MA) process

$$Z_t = C - \theta_1 a_{t-1} + a_t \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(2)$$

Where C=Constant term related to mean μ and

θ=relation of at with a_{t-1}

Combining both the model is called ARIMA model, which has general form

$$Z_t = C + \varphi_1 Y_{t-1} + \theta_1 a_{t-1} + a_t \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots(3)$$

2.2. Fitting of box-jenkins ARIMA models

Box-Jenkins time-series models i.e. ARIMA (p, d, q) is known as "Univariate Box-Jenkins technique" (Box and Jenkins, 1976) ARIMA model is an algebraic statement telling how observations on a variable are statistically related to past observation.

This model amalgamates three types of process, viz., Autoregressive of order p; differencing to make a series stationary of degree d and moving average of order q. This method applied only to a stationary time series data. When the data is non-stationary then it has to be brought into stationary by the method of differencing.

2.3. Test for stationarity

The stationarity requirement ensures that one can obtain useful estimates of the mean, variance and ACF from a sample. The stationarity condition of a series was tested by examining the

1. The change of mean and variance over time.

2. The coefficients of AR and MA process i.e. in case of AR (1) and MA (1) process it should be $|\varphi_1|<1$ and $|\theta|<1$.

3. The estimated ACF values which should be tails-off towards zero rapidly.

The significance of autocorrelation was tested by t-test. The standard error of autocorrelation (Bartlett, 1946) was calculated as under

$$s(r_k) = \left(1 + 2\sum_{j=1}^{k-1} r_j^2\right)^{1/2} n^{-1/2} \quad \dots\dots\dots\dots\dots\dots\dots\dots(4)$$

$$t_{rk} = \frac{r_k - \rho_k}{S(r_k)} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(5)$$

k=1, 2, 3,…

The significant value of "t" indicates the presence of autocorrelation. The process of time series modelling involves transformation of data in order to achieve stationary, followed by identification of appropriate models, estimation of parameters, validation of models and finally for prediction. The complete description of these process and steps of time series modelling is clearly explained below.

2.4. ARIMA modelling consists of three operational steps

Identification, estimation and diagnostics checking

2.4.1. Identification

Identification involves the techniques to determine the values of p, q and d. The values are determined by using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). For any ARIMA (p, d, q) process, the theoretical PACF has non-zero partial autocorrelations at lags 1, 2, ..., p and has zero partial autocorrelations at all lags, while the theoretical ACF has non zero autocorrelation at lags 1, 2, …, q and zero autocorrelations at all lags. The nonzero lags of the sample PACF and ACF are tentatively accepted as the p and q parameters. For a non-stationary series, the data is differenced to make the series stationary. The number of times the series is differenced determines the order of d. Thus, for a stationary data d =0 and ARIMA (p, d, q) can be written as ARMA (p, q).

2.4.2. Estimation

The main approaches for fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation which was estimated by using SPSS (version 17) software

2.4.3. Diagnostic checking

The best model was selected on the basis of minimum values of Schwartz-Bayesian Information Criterion (SBC), Akaike Information Criterion (AIC) and Root Mean Square Error (RMSE). Residuals were tested by run test and for randomness by Shapiro–Wilk test for normality and The Ljung and Box for independent were used.

2.5. Test for normality of the residual (Shapiro-Wilk, 1965)

The Shapiro-Wilk statistic was used to test the normality of residuals,

The required test statistics W was defined as

$$W=\frac{S^2}{b}$$ Where $S^2=\sum a(k)\left[e(n+1-k)-e(k)\right]$ ....(6)

The parameter k takes the values

$$K=\begin{cases} 1, 2, 3, 4, ...n/2 \\ 1, 2, 3, 4, ...(n-1)/2 \end{cases}$$

and $b=\sum_{i=1}^{n} (e_i-\bar{e})^2$

2.6. Test for independence of errors (Chi-square test)

$$Q=n(n+2)\sum_{i=1}^{k} (n-k)^{-1}r_i^2(\hat{a})$$ ....(7)

Where, n is the number of observations. The statistic Q approximately follows a χ^2 distribution with (K-m) degrees of freedom, where K is the number of residual autocorrelation and m is the number of parameters estimated in the ARIMA model.

3. Results and Discussion

ARIMA models were fitted for wheat area, production and productivity of Gujarat state as a whole. In fitting of Univariate Box-Jerikins (UBJ) ARIMA models, the autocorrelation up to 10 lags were worked out. If the spikes did not sharply tails-off towards zero and if the visual inspection of the realization indicates that the mean, variance and autocorrelation were not constant over time then the series was considered as non-stationary. Therefore, the new variable X_t was constructed by taking difference of one (i.e. d=1) to make the series stationary.

3.1. Fitting trend on wheat area in Gujarat state by using ARIMA models

The ACF (γ_k) of the transformed variables were tails off toward zero with cut- off third and fourth spikes and PACF (φ_{kk}) of the transformed variables tails off toward zero with cut- off third spike. This suggested that the algebraic family of ARIMA on *p*=0, 1, 2, 3, d=1 and q=0, 1, 2, 3, 4 can be used. The different models among the different value of p and q were fitted. Among the models, those models having lower value of AIC and SBC are given in Table 1. Among these models, ARIMA (1, 1, 0) and ARIMA (0, 1, 1) had significant AR (φ) coefficients and MA (θ) coefficient. The assumptions of residuals *i.e.* normality and independence of residuals were tested by Shapiro-Wilk test and Box-Ljung (Q) test indicated that ARIMA (1, 1, 0) and ARIMA (0, 1, 1) models satisfied the assumptions of normality and independence residuals but ARIMA (0,1,1) model had comparatively lower value of AIC, SBC and RMSE. So, ARIMA (0, 1, 1) model was found suitable to explain the trend of wheat area in Gujarat state and it is given in the Figure 1.

Table 1: Fitted ARIMA models for wheat area in Gujarat state

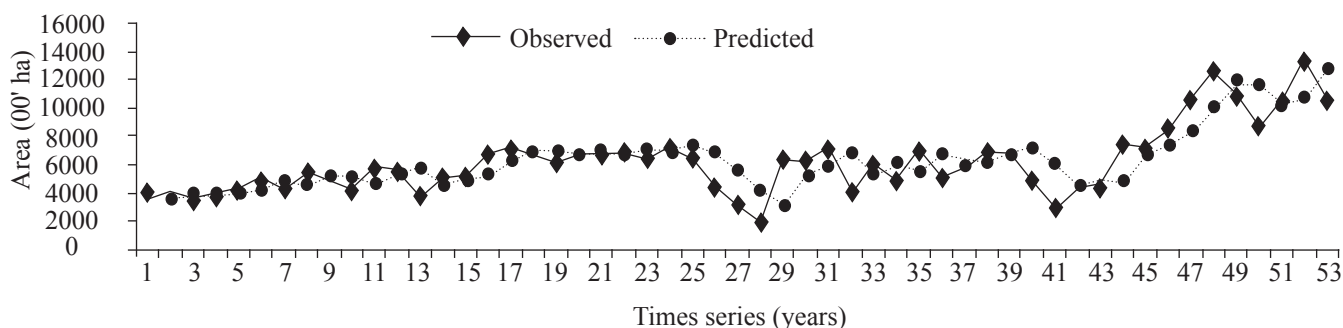| ARIMA	AIC	SBC	AR(φ)	MA(θ)	CONS	RMSE	SW-TEST	BLQ- TEST
(0, 1, 1)	915.88	919.79	-	0.413**	-19.20	1592.02	0.97	26.93
(1, 1, 1)	914.97	920.38	-0.61**	1.00	-51.89	1532.88	0.97	24.59
(1, 1, 0)	917.36	921.26	-0.26**	-	-2.70	1620.07	0.98	32.61
(2, 1, 0)	918.92	927.77	-0.28,-0.11	-	-5.23	1627.99	0.98	29.91
(2, 1, 1)	916.74	924.55	0.58**, 0.06	0.99	63.47	1547.08	0.97	24.18

*p<0.05 and **p<0.01

Figure 1: Trend in wheat area based on ARIMA (0,1,1) model in Gujarat state

3.2. *Fitting trend on wheat production in Gujarat state by using ARIMA models*

The series was made stationary by taking differences of one (i.e. d=1). The value of p and q were identified using ACF and PACF coefficients of various order of X_t. The ACF (γ_k) of transformed variables were dumping-off towards zero with cut-off initial, third and fourth spike and the PACF (φ_{kk}) also cut-off at first and third lag. This suggested that the algebraic family of ARIMA on p=0, 1, 2, 3, d=1 and q=0, 1, 2, 3, 4 can be used. The different models with these different values of p and q were fitted. Among the models, those models having lower value of AIC and SBC are given in Table 2. From the fitted models, ARIMA (0, 1, 1) and ARIMA (1, 1, 0) model had significant MA (θ) and AR (φ) coefficient term of which ARIMA (0, 1, 1) model had lower values of AIC, SBC and RMSE. The assumptions of residuals (i.e. normality and independence of residuals) were tested by Shapiro- Wilk test and Box-Ljung (Q) test indicated that ARIMA (0, 1, 1) satisfied all the assumptions. Therefore, ARIMA (0, 1,1) model were

found suitable to explain the trend of wheat production of Gujarat state and it is given in the figure 2.

3.3. *Fitting trend on wheat productivity in Gujarat state by using ARIMA models*

The series was made stationary, by taking differences of one (i.e. d=1). The autocorrelation function (ACF) and partial autocorrelation function of various order of X_t suggested to use the value of p=0, 1,…4, d=1 and q=0, 1,…., 6. The ACF (γ_k) of the transformed variable were damping off towards zero with cut-off at four lags and the PACF (φ_{kk}) also cut-off at first and fourth lags which are shown in fig. 3. The different models among the different value of p and q were fitted. Among the models, those model having lower value of AIC and SBC are given in Table 3. ARIMA (1, 1, 0) model had comparatively lower value of AIC and SBC with significant AR (φ) coefficients. This model also satisfied the assumptions of residuals i.e. normality and independence. So, ARIMA (1, 1, 0) model was found satisfactory to explained the trend with

Table 2: Fitted ARIMA models for wheat production in Gujarat state

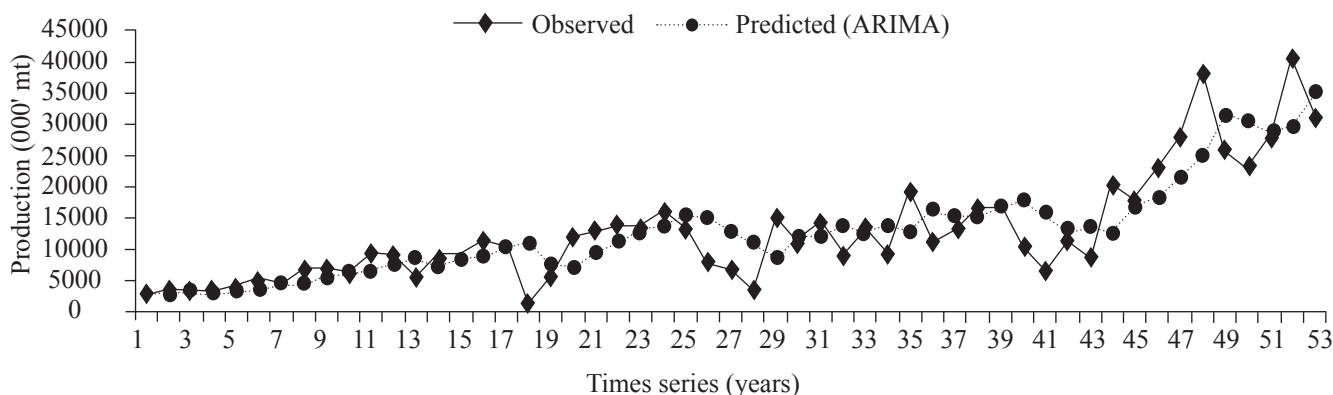| ARIMA	AIC	SBC	AR (φ)	MA (θ)	Cons	RMSE	SW-Test	BLQ- Test
(0,1,1)	1035.36	1093.27	-	0.60**	-126.69	4957.71	0.97	22.13
(1,1,1)	1036.11	1041.97	0.15*	1.0	-74.98	4804.49	0.96	24.44
(1,1,0)	1038.57	1042.47	-0.36**	-	-109.48	5175.43	0.96	30.09*

*p<0.05 and **p<0.01

Figure 2: Trend in wheat production based on ARIMA (0, 1, 1) model in Gujarat state

satisfying all the assumptions regarding to the residuals for the productivity of the wheat in Gujarat state and it is given in the figure 3.

Finally five year ahead forecast was made for area, production and productivity by using ARIMA (0, 1, 1), (0, 1, 1) and (1, 1, 0) models respectively. Table 4, show the forecast values for area, production and productivity at the 95% confidence limit. From table 4 Forecast for the 2013−14 for area was 11705.38 hundred ha with a 95% confidence limit of (8506.91,

14903.85) hundred ha. In same forecast for production and productivity was 34976.17 thousand mt with a 95% confidence limit of (25014.69, 44937.66) thousand mt and 3050.72 kg ha^{-1} with a 95% confidences limit of (2597.32, 3504.12) kg ha^{-1} respectively. For the year 2017−18, the forecast for area, production and productivity was 12989.2 hundred ha, 40296.9 thousand t and 3148.42 kg ha^{-1} with a 95% confidences limit of (8058.2, 17920.2) hundred ha, (27612.56, 52981.24) thousand t and (2439.19, 3857.65) kg ha^{-1} respectively.

Table 3: Fitted ARIMA models for wheat productivity in Gujarat state

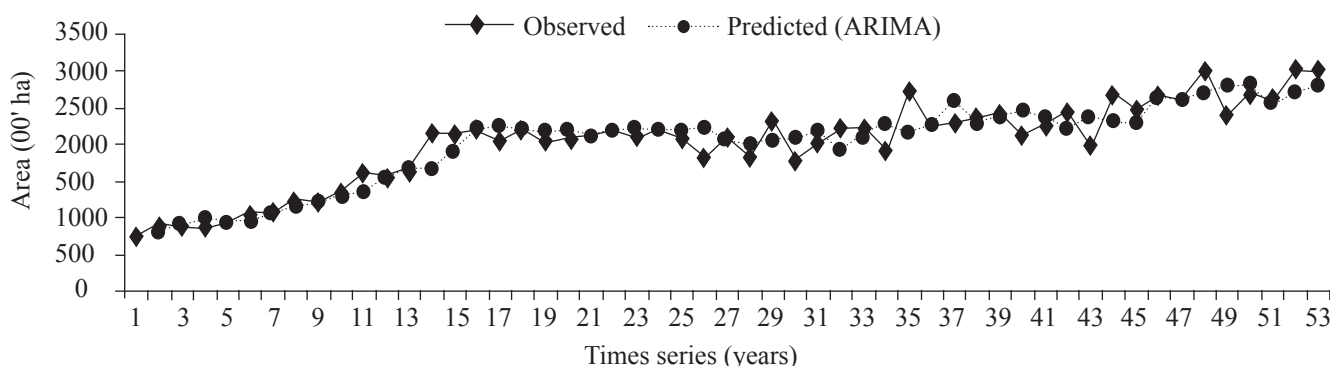| ARIMA	AIC	SBC	AR (φ)	MA (θ)	Cons	RMSE	SW-Test	BLQ- Test
(1, 1, 0)	712.31	718.16	-0.63**	-	57.14	225.79	0.98	27.13
(1, 1, 1)	712.74	716.64	-0.46**	0.32	60.04	221.91	0.98	17.97
(2, 1, 0)	712.18	718.04	-0.78*, -0.23	-	58.77	211.83	0.98	18.49

$^*p< 0.05; ^{**}p<0.01$

Figure 3: Trend in wheat productivity based on ARIMA (1, 1, 0) model in Gujarat state

Table 4: Forecast for area, production and productivity of wheat in Gujarat State by using selected models

Years	Area (000' ha)	Production (000'MT)	Productivity (kg ha^{-1})
2013−14	11705.4±3198.4	34976.17±9961.4	3050.72±453.4
2014−15	12017.3±3708.3	36267.67±10707.3	3056.16±483.4
2015−16	12335.2±4156.0	37584.96±11404.47	3098.09±595.49
2016−17	12659.2±4560.0	38928.04±12061.4	3116.18±639.95
2017−18	12989.2±4931.0	40296.9±12684.34	3148.42±709.23

4. Conclusion

The Box-Jenkins approach was used to model and forecast area, production and productivity of wheat crop in Gujarat. Our forecast showed an increasing pattern in area, production and productivity. ARIMA models are suitable only for short term prediction and hence this is needed to update them every year for achieving more valid forecast.

5. Acknowledgement

6. References

Badmus, M.A., Ariyo, O.S., 2011. Forecasting cultivated areas and production of maize in Nigeria using ARIMA model, Asian Journal of Agricultural Sciences 3(3), 171−176.

Bartlett, M.S., 1946. On the theoretical specification of sampling properties of autocorrelated time series, Journal of the Royal Statistical Society, B 8, 27.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis, Forecasting and Control, Second Edition: Holden Day.

Gupta, G.S., 1993. ARIMA model for forecasting on tea production in India, Indian Economic Journal 41(2), 88−110.

Hossain, M.Z., Q.A. Samad and M.Z. Ali, 2006. ARIMA model and forecasting with three types of pulse prices in Bangladesh: A case study. International Journal of Social Economics 33, 344–353.

Khin, A.A., C.F.C. Eddie, Shamsundin, M.N., Mohamed,

Z.A., 2008. Natural rubber price forecasting in the World market, agricultural sustainability through participate global extension, June 15−19. University Putra Malaysia, Kuala Lumpur, Malaysia.

Mandal, B.N., 2005. Forecasting sugarcane productions in India with ARIMA model. Inter Stat, October, 2005.

Rachana, W., Suvarna M., Sonal, G., 2010. Use of ARIMA models for forecasting pigeon pea production in India, International Review of Business Finance 2(1), 97−107.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples), Biometrika 52, 591−611.

Wankhade, R., Mahalle, S., Gajbhiye, S., Bodade, V.M., 2010. Use of the ARIMA model for forecasting pigeon pea production in India. International Review of Business Finance, 2, 97–102.

Wheat Scenario, 2013. A Snippet, Directorate of Wheat Research Karnal, Haryana (India). Available from http://www.dwr.res.in/sites/default/files/wheatscenarioe-newsletis-1-13.pdf.