Fundamental and Applied Agriculture
Vol. 6(4), pp. 367–382: 2021
doi: 10.5455/faa.129347

AGRICULTURAL ECONOMICS | ORIGINAL ARTICLE

An investigation of vegetable production in Iran by using time series models

Zeynab Latifi [1*], **Hossein Shabanali Fami** [2]

[1]Department of Mechanical Engineering, Faculty of Engineering, Shohadaye Hoveizeh Campus of Technology, Shahid Chamran University of Ahvaz, Susangerd, Iran
[2]Department of Agricultural Development and Management, Faculty of Agricultural Economics and Development, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran

ARTICLE INFORMATION

Article History
Submitted: 03 Oct 2021
Accepted: 08 Dec 2021
First online: 30 Dec 2021

Academic Editor
Mohammad Ataur Rahman
marahman@bau.edu.bd

**Corresponding Author*
Zeynab Latifi
z.latifi@scu.ac.ir

OPEN ACCESS

ABSTRACT

Nowadays, agriculture, particularly vegetable production, plays a vital role in the lives of the people of the Islamic Republic of Iran. The proper forecast of harvesting of vegetable production is essential for the agriculture ministry; so that they can make decisions about storage, import or export, etc. In this paper, the prediction of vegetable production in Iran is done in different time series models such as simple random walk, random walk with drift, linear trend, quadratic trend, simple moving average, simple exponential smoothing, double exponential smoothing, exponential trend, s-curve trend, and Auto-Regressive Integrated Moving Average models. Depending on the availability of the required data, a set of six different groups of vegetables (cucumbers and gherkins, eggplants, garlic, onions, pumpkins, squash and gourds, and tomatoes) has been studied for empirical analysis. The two set data from 1961-62 to 2019-20, and from 1990-91 to 2019-20 has been used to forecast the vegetable production for the next eight years from 2020-21. The total data are divided into training data and testing data. The best models were selected based on the lowest RMSE, minimum values of Akaike Information Criteria, and Schwarz Bayesian Information Criteria. The model diagnosis was performed using Ljung-Box, Runs above and below the median, and Runs up and down tests on ACF and PACF in residuals. For garlic, the quadratic model was selected as the best model; whereas for the rest of the vegetable groups the ARIMA model was determined as the optimum model. For the next eight years, suitable models were forecasted from 2020-21 to 2027-28. Based on the findings, the forecasted production for cucumbers and gherkins is 540.23 to 3021.31, for eggplants is 615.819 to 496.993, for garlic, is 59.1889 to 61.4415, for onions is 2204.69 to 2436.1, for pumpkins, squash, and gourds are 196.068 to 211.065, and for tomatoes is 4399.59 to -4486.66 in 1000 tonnes from 2020-21 to 2027-28. The forecasted data indicate an increase in the production of cucumbers and gherkins, garlic, onions, pumpkins, squash, and gourds; and a decrease in the amount of production of eggplants, and tomatoes in the future. The results can be valuable for the government, and decision-makers to have a suitable plan for crop productions.

Keywords: Agriculture production, AIC, ARIMA, Quadratic model, RMSE

Cite this article: Latifi Z, Fami HS. 2021. An investigation of vegetable production in Iran by using time series models. Fundamental and Applied Agriculture 6(4): 367–382. doi: 10.5455/faa.129347

1 Introduction

Vegetables are rich in water, vitamins, and vital minerals such as fiber, folate, beta-carotene, phytochemicals, etc. these all are necessary for the maintenance of body health; besides, they reduce the risk of chronic diseases (Ulger et al., 2018). There are currently 1000 known edible vegetables, out of which 350 are grown on 60 million hectares of agricultural lands around the world, and annually about one billion tonnes of vegetables are produced (FAO, 2021). According to the FAO, China, with an annual production of 590.676 million tonnes, is the world's largest producer of vegetables; India with an annual production of 132.027 million tonnes has the second rank, the United States with a production of 29.999 million tonnes has the third rank, and Turkey with a production of 25.339 million tonnes is in the next ranks in this regard (Fig. 1). Iran is also being ranked as 11th producer, and produces about 11.827 million tonnes of vegetables in the world, which is equivalent to 1.1% of the total vegetables produced in the world (FAO, 2021).

Iran is one of the largest producers and exporters of vegetables in the Middle East. The diverse climate of different regions has made Iran the fifth largest producer of vegetables; in 2019, out of 383,000 hectares of land under vegetable cultivation in the country, 11.827 million tonnes were harvested (FAO, 2021); In other words, Iran is the favorable condition for growing vegetable, in addition to meet the country's needs in all seasons. So it would be possible to export a variety of vegetable products to other countries, especially the Persian Gulf countries (MoAJ, 2021). The high variety of vegetables and the existence of different crops has made it possible to have cultivation programs four times a year due to the Ministry of Agriculture has implemented many policies and programs such as: reducing or stabilizing the area under cultivation; increasing yield with proper nutrition, and combating spoilage agents; developing seedling cultivation; developing new irrigation methods; developing conservation agriculture, low plowing and rotation and increase of soil organic carbon; developing areas under greenhouse cultivation to reduce water consumption and producing valuable products out of season, etc. (Pakravan and Gilanpour, 2013). Therefore, Forecasting vegetable products are necessary for fair price to farmers, the agricultural industry, and particularly governments; so that they can make the necessary planning based on estimates of domestic agricultural products. Forecasting vegetable availability and population needs might represent an essential role in developing a framework for achieving a sustainable solution to future food insecurity challenges (Vågsholm et al., 2020). To achieve valid predictions, and results, statistics play an important role (Killeen, 2018). Many statistical and economic models have been developed to forecast various topics, including agricultural products (Hanke and Wichern, 2008; Latifi and Shabanali Fami, 2021). Amin et al. (2014) used various time series models for wheat production in Pakistan, and the best model i.e. ARIMA (1,2,2) was selected to forecast the data till 2060. They have noticed that compared to 2010, in 2060 wheat production would become double. To evaluate the trend of vegetable production in terms of area and production in the feeder zones of Chennai city in India, Arivarasi and Ganesan (2015) used time series analysis. The area and production of vegetables in the selected zones were forecasted by the ARIMA model. The forecasted models showed a decreasing trend in both cultivated area and production of vegetables in zone 1; however, in zone 2 an increasing trend was found in cultivated areas whereas decreasing trend was found in the vegetable production for the period 2011-12 to 2014-15. Khayati (2015) used two types of time series models to forecast the major vegetable crops in Tunisia i.e. smoothing, and stochastic models. The results indicated that the Holt model was the best model for potatoes, artichoke, and pepper; and ARIMA and Winters models were appropriate for tomatoes, and onions, respectively. Based on the results, unlike other products, the productivity of potatoes and tomatoes was expected to increase in the future. Rahman et al. (2016) studied the changes from a rice-based cropping system to a shrimp-based cropping system in the coastal area of Bangladesh and their impact on soil environment. Their results showed increasing in the salinity level with changes in the availability of nutrients in the soil. Fauziyanti et al. (2020) with the exponential analysis model tried to forecast food crop productions and food crop consumptions; besides, they tried to find a gap analysis projection between production and consumption of food crops towards 2015-2021 in Bali province in Indonesia. The results showed that the Prediction of food crop production in the Bali province has decreased; while the forecasted food consumption increase in line with the forecasted population in the future. Maghrebi et al. (2020) by using Mann-Kendall and Sen's slope estimator methods investigated the changes in Iran's agricultural production from 1981 to 2013. Results showed that the agricultural development in Iran was not consistent with natural water availability changes across time and space. Despite the decreasing water availability, agricultural production in Iran has increased over the mentioned period. In this study, to find the best models to forecast vegetable production groups in Iran, the software Statgraphics was used. The results will be helpful for farmers, researchers, decision-makers, and the government; that way they can decide on the management of storage, transport, and distribution. On the other hand, accurate prediction plays a vital role in reducing food instability and price determination.

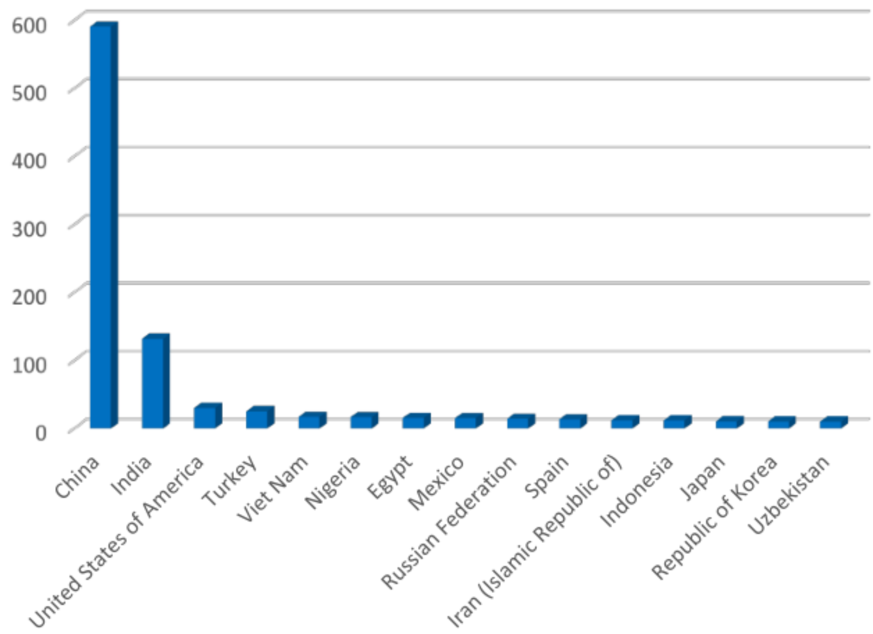

Figure 1. Ranking in vegetables primary production (million tonnes) in 2019 FAO (2021)

2 Materials and Methods

In this research, vegetable products in Iran such as cucumbers, and gherkins, eggplants (Aborigines), garlic, onions, pumpkins, squash, and gourds, and tomatoes have been considered. For garlic and eggplants, the production data have been used from 1990-91 to 2019-20, and for the rest of the vegetable groups, depending on the availability of the required data, the time series data from 1961-62 to 2019-20 have been used (FAO 2021). The total data are divided into training data and testing data. The training data are from 1990-91 to 2015-16, for eggplants, and garlic, and it is from 1961-62 to 2015-16 for the rest of the vegetable groups. Besides, for all the vegetable groups the testing data are from 2016-17 to 2019-20. To select the suitable model for forecasting vegetable production data. Several time series models were fitted to the data as earlier mentioned in previous section.

2.1 Models

2.1.1 Simple random walk model

A random walk model is a non-stationary stochastic time series model. It's defined as a process where the current value of a variable is composed of the past value plus an error term defined as a white noise (ε_t). Suppose ε_t is a variable with average zero and variance (σ^2). Then Z_t is been a random walk

$$Z_t = Z_{t-1} + \varepsilon_t \quad (1)$$

$$Z_t - Z_{t-1} = \varepsilon_t = \Delta Z_t$$

The random walk model is simple to use, and it can easily handle flows around complicated boundaries. The method conserves the total circulation. However, in the random walk model, the computed solutions are noisy due to the statistical errors. In flow control studies, the statistical errors could mask the effects of varying the control parameters. The statistical errors can also cause symmetric flows to turn asymmetric erroneously. To reduce the statistical errors requires a very large number of vortices.

2.1.2 Random walk model with drift

Equation 1 is adjusted as equation 2:

$$Z_t = \alpha + Z_{t-1} + \varepsilon_t \quad (2)$$

$$\Delta Z_t = Z_t - Z_{t-1} = \alpha + \varepsilon_t$$

Suppose α is a drift parameter. If a as α drift parameter, is positive (negative), Z_t drifts upward (downward). In this model, the average and variance, again violating the conditions of (weak) stationary, increase over time; then, the Random walk model with drift like the Random walk model without drift would be a non-stationary stochastic time series model.

2.1.3 Linear trend model

Time series data may show a linear trend which is determined as follows:

$$Z_t = c + bT_{t-1} \quad (3)$$

Suppose c and b are the constant and coefficient of the linear trend model, respectively. The linear trend model attempts to find the slope and intercept that give the best average fit for all the past data. The constant and coefficient might be estimated by the least square method:

$$b = \frac{\sum(Z_t - \bar{Z}_t)(T_t - \bar{T}_t)}{\sum(T_t - \bar{T}_t)^2} \tag{4}$$

$$c = \bar{Z}_t - b\bar{T}_t$$

The linear trend models are sensitive to outliers. These could significantly swing your model results.

2.1.4 Quadratic trend model

With a quadratic trend, the values of a time series tend to rise or fall at a rate that is not constant; it changes over time; As a result, the trend is not a straight line. The trend is defined as (Anderson, 2013):

$$Z_t = c + b_1 T_t + b_2 T_t^2 \tag{5}$$

Suppose c, b_1 and b_2 are coefficients. Like linear trend model, the quadratic trend model is sensitive to outliers, too.

2.1.5 The simple moving average model

One of the easiest time series models is the moving average, the forecast value of which for the next time would be the average of the previous values. Because of this matter, this model is called the moving average. In case the model is stationary, in many cases the moving average model might be used and of course, it would have a good accuracy; otherwise, with the help of the moving average, the trend of a time series might be recognized. In this model, the average of the previous period is used for the current period. In the same way, new information is used and constantly updated. Assume there are N observations, and t observations are used to estimate the average value as an MA (t), then the forecasting model would be as the equation 6:

$$\hat{Z}_{t+1} = \frac{z_1 + z_2 + \ldots + z_t}{t} \tag{6}$$

The main advantage of the simple moving average model is that it offers a smoothed line, less prone to whipsawing up and down in response to slight, temporary price swings back and forth. The simple moving average model's weakness is that it is slower to respond to rapid price changes that often occur at market reversal points. The simple moving average model is often favored by traders or analysts operating on longer time frames, such as daily or weekly charts.

2.1.6 Simple exponential smoothing model

Exponential smoothing follows the same moving average smoothing method. Thus, this method can be considered as a weighted average for time series data, which, in calculating the average, gives less weight to more distant data. The weight reduction of distant past values diminishes its importance in calculating and predicting future values, and present data will have a greater impact. The simple exponential smoothing model can be represented as follows:

$$\hat{Z}_t = \alpha Z_t + (1 - \alpha)\hat{Z} - t - 1 \tag{7}$$

Z_t is the real value of the series at time t and $\hat{Z}_t$ represents the predicted value of the series at time t by simple exponential smoothing model. So, the weighted average between the real value and the prediction at time t for time $t + 1$ is considered; thus, α is the smoothing constant, and the range of its value is between zero and one. On the other side, with the increase of this smoothing constant, the role of past observations in calculating and predicting future value decreases. The disadvantage of this model is the greater sensitivity of the simple exponential smoothing model, so it is more vulnerable to false signals and getting whipsawed back and forth. The use of the exponential smoothing model is suitable for time series without a trend.

2.1.7 Double Exponential Smoothing model

If time series have trends, using the double exponential smoothing method will give better results. It seems that double exponential smoothing might be considered as two uses of simple exponential smoothing. Exponential smoothing prediction can be obtained using two smoothing constants (α and β with values between zero to one) and the following equations:

$$Y_t = \alpha Z_t + (1 - \alpha)(Y_{t-1} + b_{t-1}) \tag{8}$$

$$b_t = \beta(Y_t - Y_{t-1}) + (1 - \beta)b_{t-1} \tag{9}$$

$$\hat{Z}_{t+j} = T_t + (b_t \times j) \tag{10}$$

Equation 8 is a smoothed-value Y_t, equation 9 calculates the trend value b_t, and finally, equation 10 computes the forecasted value for the amount of the next period predicted j. Y_0 and b_0 can be estimated by the least square method.

2.1.8 Exponential trend model

The exponential trend can be adjusted when a time series begins slowly, and then appears to be increasing at an increasing rate such that the percentage difference from observation to observation is constant. It is presented as follows:

$$\hat{Z} = b_0 b_1^t \tag{11}$$

The coefficient b_1 is relevant to the growth rate and b_0 is calculated the intercept (Zainal, 2010).

2.1.9 S-curve trend model

In many examples, the nonlinear trend is properly for time series data. The S-curve trend model (Pearl-Reed logistic trend model) is one of the nonlinear trend models and is determined as equation 12 (Weisberg, 2005):

$$\hat{Z} = \frac{1}{1 + e^{b_0 + b_1 t}} \tag{12}$$

Suppose b_0 and b_1 are constants. It should be noted that in the s-curve model, there is a risk of missing the turning point of the series.

2.1.10 ARIMA model

Box and Jenkins (1970) developed Autoregressive Integrated Moving Average (ARIMA) method, which was used by statisticians and economists to extract a model that would produce and forecast time series. This method includes four stages of identification, estimation, diagnosis, and forecast. Based on its past values and error sentences, ARIMA method models static time series. Then, it is a parametric method and no independent variables are used. ARIMA (p, q) model is determined as follows:

$$\begin{aligned} Z_t = {} & \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} \\ & + U_t + \theta_1 U_{t-1} + \theta_2 U_{t-2} + \ldots + \theta_q U_{t-p} \end{aligned} \tag{13}$$

$$U_t \sim iid(0, \sigma_u^2)$$

If Z_t is the original series, for every t, we suppose that is independent of $Z_{t-1} + Z_{t-2} + \ldots + Z_{t-p}$. The final aim of the proposed Box-Jenkins model is forecast. Therefore, the time series used must be static, because the instability of the time series makes the forecast of the future values of the series to be affected by a random or a definite trend in them and also to affect the results (Gujarati and Porter, 2004). So, if we use model ARIMA (p, q) for a non-static time series accumulated of order d, model ARIMA (p, d, q) will be obtained. At the identification stage, p and q values are defined using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) values. The limitation of the ARIMA model is that it's unable to detect and model the effects of cluster fluctuations for time series data related to financial data.

2.2 Diagnostic tests

The models fitted to the data will be adequate when the residuals are random. After fitting various models, the ACF and PACF of the residuals are estimated for every model. Three tests are used to examine the randomness of residual-based on ACF and PACF:

(i) Runs up and down test: computes the number of times the series runs up or down. This number is compared to the expected value of a random time series. When p-values are small, the time series is not purely random.

(ii) Runs above and below the median test: calculates the number of times the series moves above or below its median. This number is compared to the expected value of a random time series. When p-values are small (less than 0.05 if operating at the 5% significance level), the residuals are not purely random.

(iii) Ljung-Box test: measures a test statistic based on the first k residual autocorrelations. Like the above two tests, small p-values indicate that the residuals are not purely random (Forecasting Statgraphics, 2017).

As the p-values for all three tests are well above 0.05, there would be no reason to suspect that the residuals are white noise (Box et al., 2015).

2.3 Accuracy measurement

To measure the accuracy of the proper model, the following techniques are used:

2.3.1 Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m} e_{n+i}^2}{m}} \tag{14}$$

e_{n+i} is the residual term of $(n + i)$th the observation, and m is the number of observations. The best models have smaller RMSE values, which measure the variance of the forecasting errors; consequently, the minimum value of this measure recommends the best model with minimal forecasting error (Karim et al., 2010).

2.3.2 Akaike Information Criteria (AIC)

$$AIC = -2\, ln(\hat{L}) + 2k \tag{15}$$

Where $\hat{L}$ is the maximum value of the likelihood function for the model, k is the number of estimated parameters in the model (Burnham and Anderson, 2002; Akaike, 1974). The model represents the best model if its AIC value, compared to other fitted models, is minimal (Tsay, 2013).

2.3.3 Schwarz Bayesian Information Criteria (SBIC)

$$SBIC = -2\,ln(\hat{L}) + 2k\,ln(m) \quad (16)$$

The model with the minimal SBIC value is defined well as other fitted models (Tsay, 2013).

3 Results and Discussion

Descriptive statistics such as mean, standard deviation, coefficient of variation, minimum, maximum, range, standard skewness, and standard kurtosis of vegetable production data are given in Table 1. The average production of cucumber and onions are 1076010 and 1036090 tonnes, respectively; so cucumbers and gherkins, and also onions have the highest production in the country, which shows the policies adopted by the government as well as the higher consumption of these products in the country. Standard deviation is one of the scattering criteria that indicate, on average, how far the data is from the average. If the standard deviation of a set of data is close to zero, it shows that the data are close to the average and have little dispersed; while large standard deviation shows a significant scattering of data. The standard deviation value of tomato production is 2159770 tonnes that in comparison with other vegetables, shows a significant dispersion of the data; but in general, the standard deviation of all groups is high. The coefficient of variation represents the scattering rate per unit of the mean. The coefficient of variation is lower for garlic, i.e. 28 percent, and higher for tomatoes, i.e. 91 percent; this indicates that the production of these data is less and more scattered from the average values (Statgraphics output).

3.1 Selection of the model

Table 2 compares the results of fitting different models to the data of the vegetable group's production. The model with minimum RMSE, AIC, and SBIC values have been selected to generate the forecast data. The best model for eggplants production was model E, i.e. a quadratic trend model (quadratic trend = $141.399 + 41.6367t + -0.849445t^2$) to forecast the data. The best models for cucumbers and gherkins; garlic, onions, pumpkins, squash, and gourds, and tomatoes production were ARIMA (4,1,4) with constant, ARIMA (0,1,2) with constant, ARIMA (2,1,3) with constant, ARIMA (4,0,0) with constant, and ARIMA (5,0,4), respectively.

The output of the tests run on the residuals for determining the suitability of each model for the data is given in Table 3. An OK means that the model passes the test. '*', '**', and '***' mean that the model fails at the 95%, 99%, and 99.9% confidence level, respectively, which indicates that the residuals are not purely random. Note that the currently selected models for cucumbers and gherkins production, i.e. ARIMA (4,1,4) with constant, eggplants production, i.e. quadratic trend model, onions production, i.e. ARIMA (2,1,3) with constant, pumpkins, squash, and gourds production, i.e. ARIMA (4,0,0) with constant, tomatoes production, i.e. ARIMA (5,0,4), pass four tests, and for garlic production, ARIMA (0,1,2) with constant model passes five tests. Since no tests are statistically significant at the 95% or higher confidence level, the current model would be probably adequate for the data (Statgraphics output).

3.2 Diagnostic examination

The residual's normal probability plots of vegetable groups, The estimated ACF, and PACF between the residuals at different lags and 95% probability limits around 0 are shown in Fig. 2, Fig. 3, and Fig. 4, respectively. If the probability limits at a particular lag do not include the estimated coefficient, there is a statistically significant correlation at that lag at the 95% confidence level. Except for eggplants production, in which two of the partial autocorrelation coefficients are statistically significant at the 95% confidence level, in the rest of the groups, none of the autocorrelation coefficients and partial autocorrelation coefficients are statistically significant; that is the data could well be completely random. Also, three tests used to test for residual randomness based on ACF and PACF show that the residuals are white noise (Table 3).

3.3 Model summary

The summary of the forecasted models is shown in Table 4. The parameters of the models and their significance are given in the table. If the p-value is less than 0.05, the parameter is statistically different from zero, at the 95% confidence level in a significant way. For example in the cucumbers and gherkins, the selected model is an ARIMA (4,1,4) with constant. The p-value for the AR (4), MA (4), and the constant term is less than 0.05, so they are significantly different from zero (Statgraphics output).

3.4 Predicted model

Table 5 shows the forecasted values and real values in 1000 tonnes from the fitted models. The testing data sets are vegetable production group data from 2016-17 to 2019-20.

The forecasted data of the vegetable production groups, based on the fitted models for the next eight years (2020-21 to 2027-28), are given in Table 6. For these periods, 95% prediction intervals for the forecasted data is shown. Assuming that the fitted models are appropriate for each vegetable production group, these prediction intervals show that the actual data

Table 1. Summary Statistics for vegetables production

Variables	Cucumbers and gherkins	Eggplants	Garlic	Onions	Pumpkins, squash, and gourds	Tomatoes
Mean	1076010	519050	53259.1	1036090	219735	235523
Standard deviation	699962	195751	14950.5	694194	148384	2159770
Coefficient of variation	65.05%	37.71%	28.07%	67.00%	67.53%	91.70%
Minimum	200000	150000	16708	100000	53174	110000
Maximum	3026090	1162850	84073	2426050	807500	6362900
Range	2826090	1012850	67365	2326050	754326	6252900
Standard skewness	3.7826	1.95687	-1.62331	0.790582	6.49434	1.54639
Standard kurtosis	2.26593	3.36201	1.06272	-1.81694	7.5165	-2.07597

Table 2. Selecting the best model based on criteria

Model	Selection criteria	Cucumbers and gherkins	Eggplants	Garlic	Onions	Pumpkins, Squash, and gourds	Tomatoes
A	RMSE	0.327451	0.200827	0.150321	0.202489	0.112784	0.372038
	AIC	2.53982	1.06049	5.42037	1.06214	9.45094	1.1838
	SBIC	2.5598	1.06049	5.42037	1.06214	9.45094	1.1838
B	RMSE	0.330104	0.203565	0.152652	0.202158	0.113758	0.364489
	AIC	2.54482	1.06986	5.51782	1.0652	9.50204	1.18309
	SBIC	2.54834	1.07453	5.56453	1.06872	9.53725	1.18661
C	RMSE	0.699962	0.195751	0.149505	0.694194	0.148384	2.15977
	AIC	2.69515	1.06204	5.47615	1.31194	10.0335	1.53894
	SBIC	2.69867	1.06671	5.52286	1.31546	10.0687	1.54246
D	RMSE	0.440998	0.144526	0.141833	0.192988	0.146831	0.67653
	AIC	2.60614	1.00803	5.43746	1.0593	10.0464	1.31017
	SBIC	2.61318	1.01737	5.53087	1.06635	10.1168	1.31722
E	RMSE	0.442613	0.134438	0.140819	0.193683	0.138672	0.545767
	AIC	2.61026	1.00022	5.48979	1.06341	9.96592	1.27061
	SBIC	2.62082	1.01423	5.62991	1.07398	10.0716	1.28117
F	RMSE	0.490624	0.15671	0.146481	0.415435	0.151112	1.62975
	AIC	2.62747	1.02421	5.50196	1.21264	10.1038	1.48602
	SBIC	2.63451	1.03355	5.59537	1.21969	10.1743	1.49306
G	RMSE	0.63577	0.147144	0.14585	0.613457	0.151865	2.22668
	AIC	2.6793	1.01162	5.49332	1.2906	10.1138	1.54843
	SBIC	2.68634	1.02096	5.58673	1.29764	10.1842	1.55548
H	RMSE	0.364595	0.174721	0.171267	0.214914	0.110191	0.394705
	AIC	2.5647	1.0393	5.74795	1.07744	9.43833	1.19902
	SBIC	2.56822	1.04398	5.79466	1.08096	9.47354	1.20254
I	RMSE	0.32707	0.159819	0.148083	0.196818	0.104178	0.36688
	AIC	2.54298	1.02148	5.45704	1.05985	9.3261	1.1844
	SBIC	2.5465	1.02615	5.50374	1.06337	9.48483	1.18792
J	RMSE	0.361121	0.168742	0.153777	0.198718	0.11046	0.34823
	AIC	2.56278	1.03234	5.5325	1.06177	9.4432	1.17396
	SBIC	2.5663	1.03701	5.57921	1.06529	9.47842	1.17748
K	RMSE	0.334299	0.148689	0.149102	0.195713	0.105266	0.338279
	AIC	2.55074	1.0137	5.53742	1.06211	9.38077	1.17155
	SBIC	2.55778	1.02305	5.63084	1.06915	9.45119	1.1786
L	RMSE	0.386609	0.175027	0.157487	0.199876	0.116126	0.347222
	AIC	2.57642	1.03965	5.58018	1.06293	9.54325	1.17338
	SBIC	2.57995	1.04433	5.62689	1.06645	9.57847	1.18704
M	RMSE	0.265133	0.147982	0.101486	0.15814	0.0965093	0.27705
	AIC	2.52811	1.01275	5.10099	1.03834	9.30877	1.15535
	SBIC	2.53982	1.02209	5.24111	1.05947	9.36131	1.1769

Models: (A) Random walk, (B) Random walk with drift, (C) Constant mean, (D) Linear trend, (E) Quadratic trend, (F) Exponential trend, (G) S-curve trend, (H) Simple moving average of 2 terms, (I) Simple exponential smoothing with alpha, (J) Brown's linear exponential smoothing with alpha, (K) Holt's linear exponential smoothing with alpha and beta, (L) Brown's quadratic exponential smoothing with alpha, and (M) ARIMA

Table 3. Tests to adequate the best model for data

Model	Diagnostic test	Cucumbers and gherkins	Eggplants	Garlic	Onions	Pumpkins, squash, and gourds	Tomatoes
A	RUNS	OK	OK	OK	OK	OK	OK
	RUNM	OK	OK	OK	OK	OK	OK
	AUTO	OK	*	OK	*	OK	OK
	MEAN	OK	OK	OK	OK	OK	OK
	VAR	***	**	*	***	*	***
B	RUNS	OK	OK	OK	OK	OK	OK
	RUNM	OK	OK	OK	OK	OK	OK
	AUTO	OK	*	OK	*	OK	OK
	MEAN	OK	OK	OK	OK	OK	OK
	VAR	***	**	*	***	*	***
C	RUNS	***	OK	OK	***	***	***
	RUNM	***	OK	OK	***	***	***
	AUTO	***	OK	OK	***	***	***
	MEAN	***	***	*	***	***	***
	VAR	***	OK	*	*	***	***
D	RUNS	***	OK	**	*	***	***
	RUNM	**	OK	OK	***	***	***
	AUTO	***	OK	OK	OK	***	***
	MEAN	OK	OK	OK	OK	*	OK
	VAR	***	**	OK	**	***	OK
E	RUNS	***	OK	OK	*	***	***
	RUNM	***	OK	OK	**	***	***
	AUTO	***	OK	OK	OK	***	***
	MEAN	OK	OK	OK	OK	**	OK
	VAR	***	***	OK	***	OK	***
F	RUNS	***	OK	**	OK	***	**
	RUNM	***	OK	OK	***	***	***
	AUTO	***	OK	OK	***	***	***
	MEAN	OK	OK	OK	OK	**	OK
	VAR	***	**	OK	***	***	***
G	RUNS	***	OK	OK	**	***	***
	RUNM	***	OK	OK	***	***	***
	AUTO	***	OK	OK	***	***	***
	MEAN	***	*	OK	***	***	***
	VAR	***	***	*	***	***	***
H	RUNS	**	OK	OK	OK	*	OK
	RUNM	**	OK	OK	OK	OK	*
	AUTO	OK	OK	*	***	OK	OK
	MEAN	OK	OK	OK	OK	OK	OK
	VAR	***	**	**	***	OK	***
I	RUNS	OK	OK	OK	OK	OK	OK
	RUNM	*	OK	OK	OK	**	OK
	MEAN	OK	OK	OK	*	OK	OK
	AUTO	OK	OK	OK	OK	OK	OK
	VAR	***	**	OK	***	**	***
J	RUNS	***	OK	OK	OK	*	OK
	RUNM	*	OK	*	*	**	OK
	AUTO	OK	OK	OK	OK	OK	OK
	MEAN	OK	OK	OK	OK	OK	OK
	VAR	***	**	OK	***	**	***
K	RUNS	OK	OK	OK	OK	OK	OK
	RUNM	OK	OK	OK	OK	**	OK
	AUTO	OK	OK	OK	OK	OK	OK
	MEAN	OK	OK	OK	OK	OK	OK
	VAR	***	***	OK	***	**	***
L	RUNS	***	OK	OK	OK	*	OK
	RUNM	**	OK	OK	***	***	OK
	AUTO	OK	OK	OK	OK	OK	OK
	MEAN	OK	OK	OK	OK	OK	OK
	VAR	***	**	OK	***	*	***
M	RUNS	OK	OK	OK	OK	OK	OK
	RUNM	OK	OK	OK	OK	OK	OK
	AUTO	OK	OK	OK	OK	OK	OK
	MEAN	OK	OK	OK	OK	OK	OK
	VAR	***	**	OK	***	***	***

RUNS = Test for excessive runs up and down, RUNM = Test for excessive runs above and below median, AUTO = Ljung-Box test for excessive autocorrelation, MEAN = Test for difference in mean 1st half to 2nd half, VAR = Test for difference in variance 1st half to 2nd half, OK = not significant ($p \geq 0.05$), * = marginally significant ($0.01 < p \leq 0.05$), ** = significant ($0.01 < p \leq 0.01$), and *** = highly significant ($p \leq 0.001$)

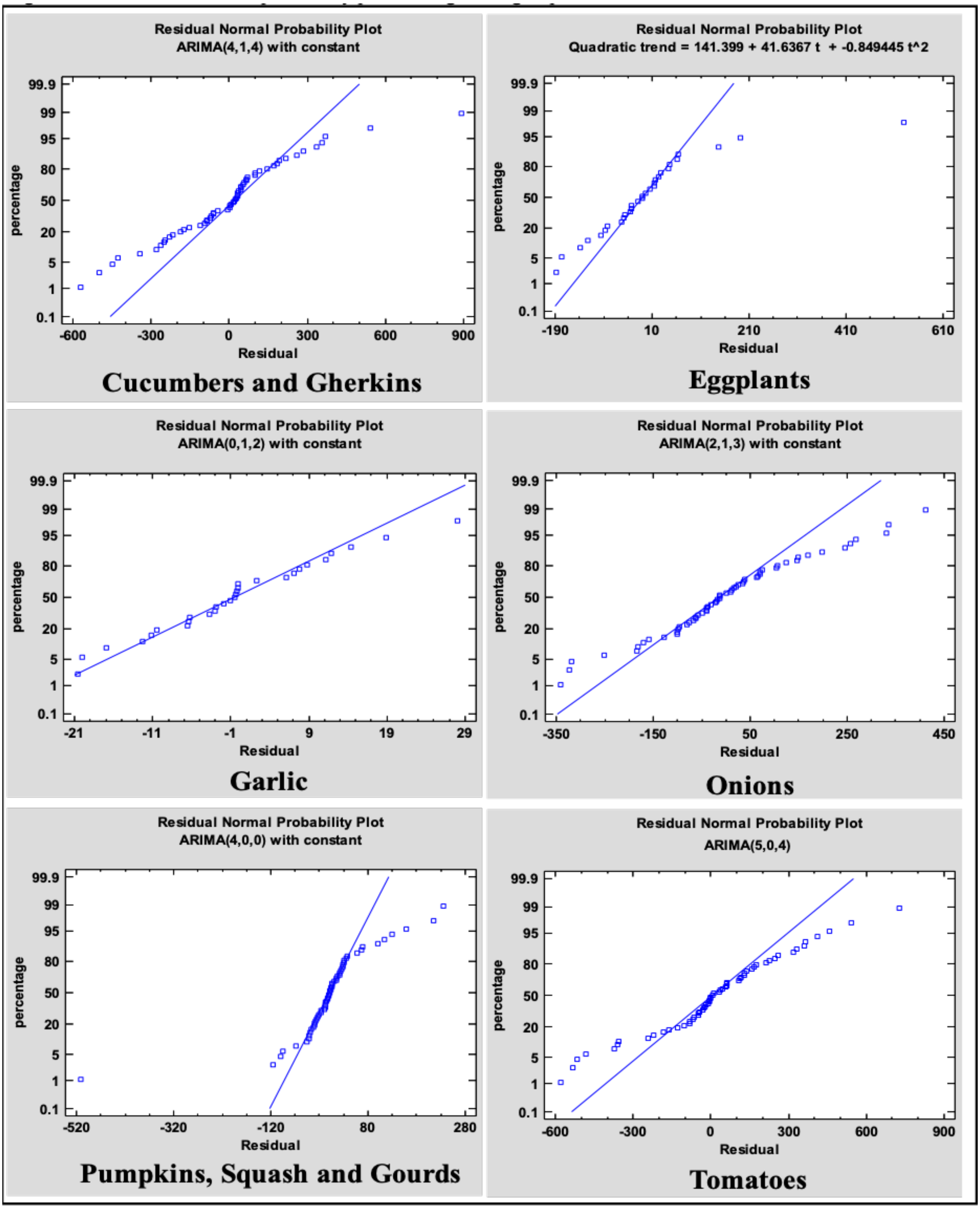

Figure 2. The residual normal probability plots of vegetables groups

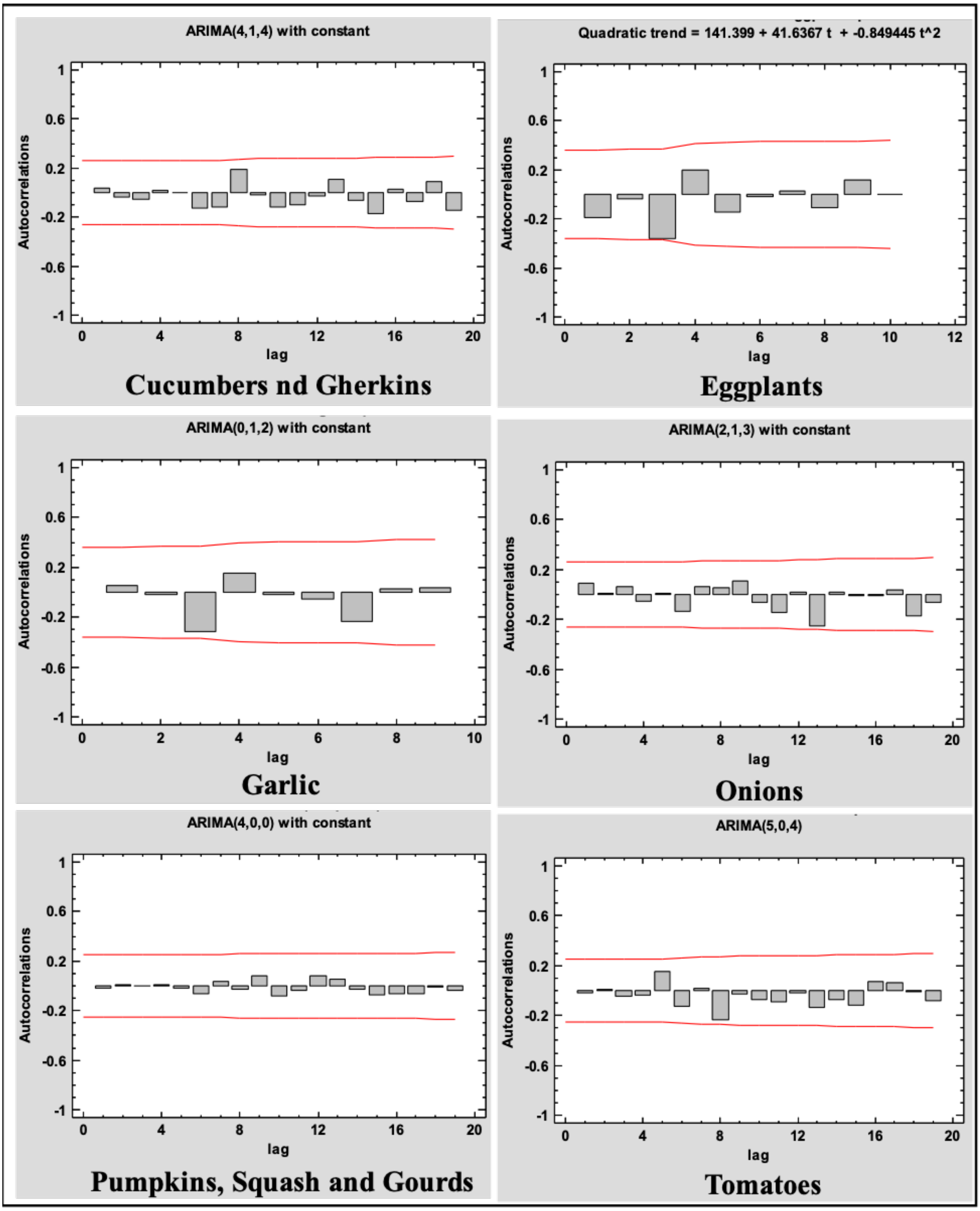

Figure 3. The Residual Autocorrelations Function plots of vegetables groups

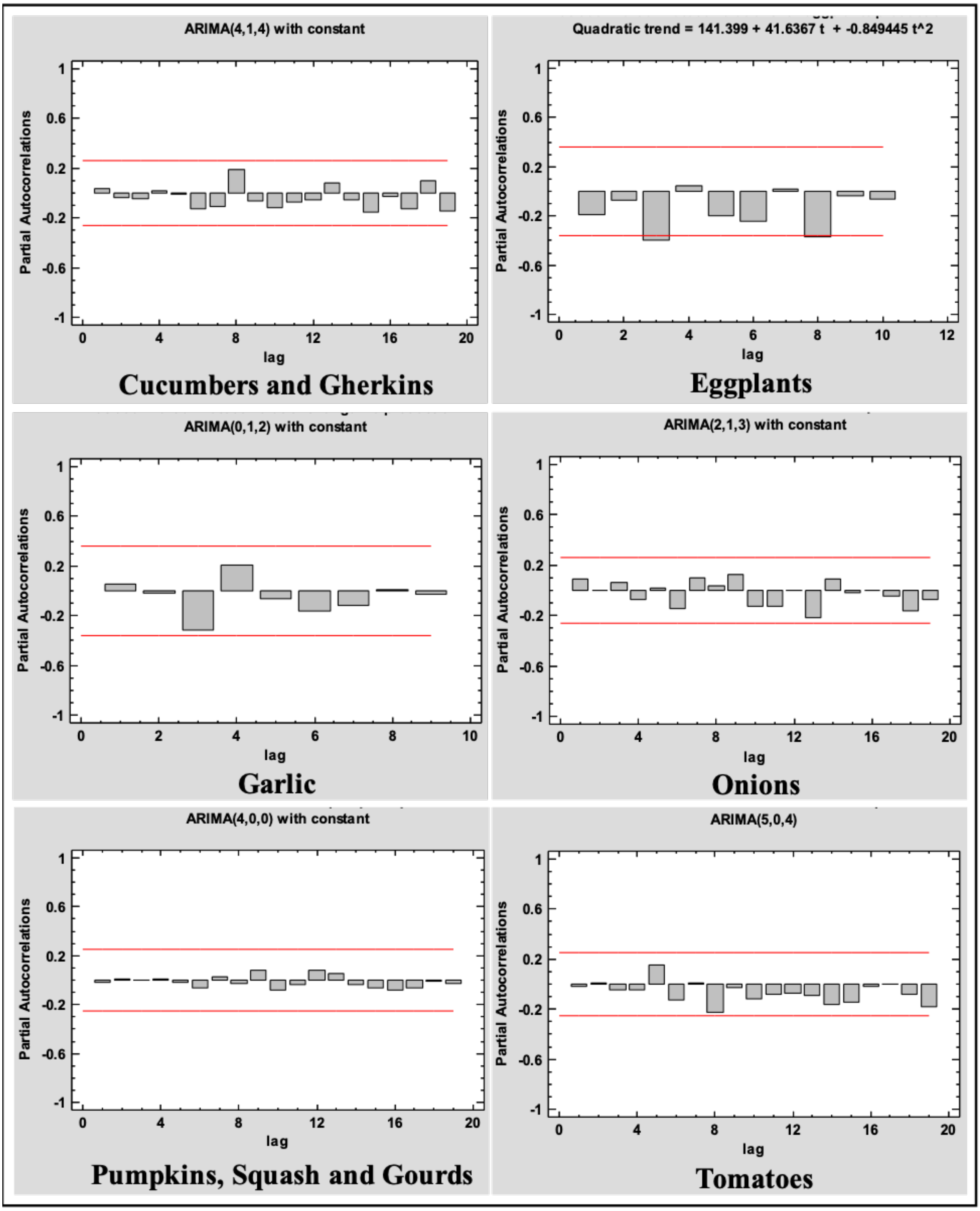

Figure 4. The Residual Partial Autocorrelations Function plots of vegetables groups

Table 4. Models summary

	Parameter	Estimate	Standard error	t	P-value
Cucumbers and gherkins Forecast model selected: ARIMA(4,1,4) with constant Estimated white noise standard deviation = 0.265512	AR(1)	-0.0066085	0.0970039	-0.0681262	0.945962
	AR(2)	1.1465	0.09604	11.9378	0
	AR(3)	0.245632	0.100718	2.4388	0.018406
	AR(4)	-1.06523	0.113816	-9.35926	0
	MA(1)	0.162863	0.0398128	4.09071	0.00016
	MA(2)	1.46066	0.0422342	34.5847	0
	MA(3)	0.169574	0.0521866	3.24938	0.002093
	MA(4)	-0.813736	0.0464328	-17.525	0
	Constant	23.8552	2.09761	16.7316	0
Eggplants Forecast model selected: Quadratic trend = 141.399 + 41.6367 t + -0.849445 t^2	Constant	141.399	78.8323	1.79367	0.084075
	Slope	41.6367	11.7226	3.55184	0.001429
	Quadratic	-0.849445	0.366915	-2.3151	0.028448
Garlic Forecast model selected: ARIMA (0,1,2) with constant Estimated white noise standard deviation = 0.119247	MA(1)	0.510106	0.1577	3.23466	0.003306
	MA(2)	0.648711	0.183469	3.5358	0.001549
	Constant	0.30751	0.0792795	3.87881	0.000641
Onions Forecast model selected: ARIMA (2,1,3) with constant Estimated white noise standard deviation = 0.163794	AR(1)	0.120149	0.115138	1.04353	0.301532
	AR(2)	-0.765763	0.118271	-6.47465	0
	MA(1)	0.638866	0.10494	6.08789	0
	MA(2)	-0.632573	0.107869	-5.86425	0
	MA(3)	0.889293	0.0570825	15.5791	0
	Constant	63.2596	2.27782	16.8764	0
Pumpkins, squash, and gourds Forecast model selected: ARIMA(4,0,0) with constant Estimated white noise standard deviation = 0.967709	AR(1)	0.588297	0.128708	4.57078	0.000029
	AR(2)	0.0945256	0.140407	0.673225	0.503675
	AR(3)	0.421584	0.140418	3.00235	0.004052
	AR(4)	-0.318356	0.128827	-2.47119	0.016652
	Constant	45.6083	56.1648	3.7955	0.000375
Tomatoes Forecast model selected: ARIMA(5,0,4) Estimated white noise standard deviation = 0.278162	AR(1)	1.06818	0.147419	7.24582	0
	AR(2)	0.15312	0.232498	0.658586	0.513183
	AR(3)	0.547105	0.212427	2.57549	0.013014
	AR(4)	-0.286516	0.215727	-1.32814	0.190162
	AR(5)	-0.518051	0.173101	-2.99276	0.004287
	MA(1)	0.630635	0.0620367	10.1655	0
	MA(2)	0.419127	0.0740804	5.65773	0.000001
	MA(3)	0.627165	0.0279604	22.4305	0
	MA(4)	-0.741904	0.0313519	-23.6638	0

Table 5. The forecasted and actual values of the vegetables production (‘000’ t) in 2016-17 to 2019-20

Variables	2016-17		2017-18		2018-19		2019-20	
	Actual	Forecasted	Actual	Forecasted	Actual	Forecasted	Actual	Forecasted
Cucumbers and gherkins	1681.78	1511.78	751.16	1001.76	697.426	579.911	871.692	1100
Eggplants	669.853	646.343	655.046	641.26	663.352	634.479	670.158	625.998
Garlic	58.835	59.3973	57.871	57.916	58.226	58.5662	58.582	58.7362
Onions	2400.59	2254.22	1700.94	2023.91	1564.44	1906.54	1779.46	2031.16
Pumpkins, squash, and gourds	178.004	199.2	182.762	188.665	186.919	180.798	191.077	191.402
Tomatoes	5828.56	5899.87	4894.96	5476.82	4661.13	5179.3	5248.9	4704.95

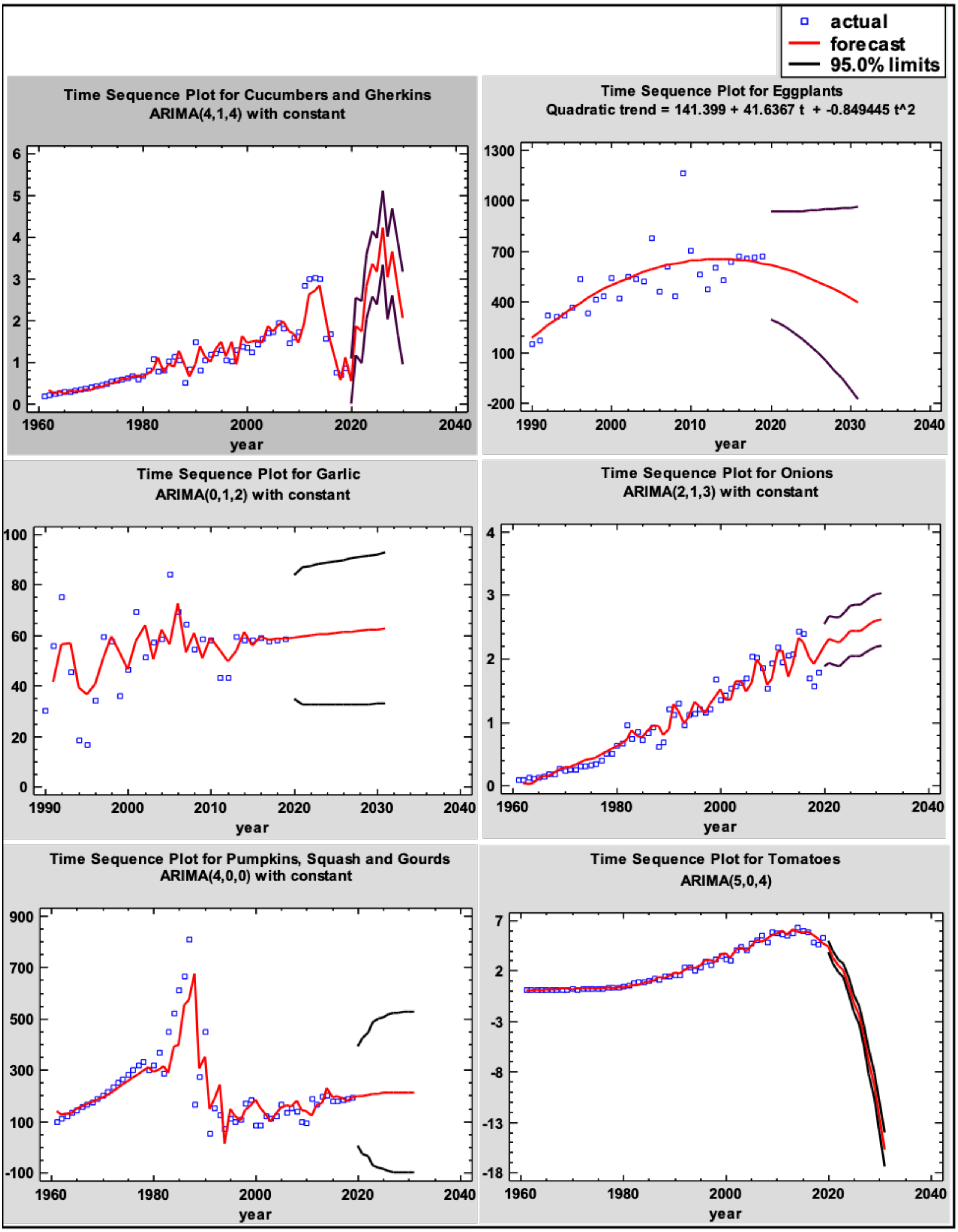

Figure 5. Time series forecast plots for vegetables production data ('000' t)

Table 6. The predicted values of the vegetables production groups ('000' t) based on the selected models for 2020-21 to 2027-28

Variable		2020-21	2021-22	2022-23	2023-24
Cucumbers and gherkins	Forecast	540.23	1853.83	1723.37	2801.3
	95% Limits	6.66, 1073.8	1160.24, 2547.42	976.83, 2469.9	2024.66, 3577.93
Eggplants	Forecast	615.819	603.94	590.363	575.087
	95% Limits	296.05, 935.59	272.41, 935.48	244.93, 935.8	213.53, 936.64
Garlic	Forecast	59.1889	59.5965	59.904	60.2115
	95% Limits	34.68, 83.70	32.30, 86.89	32.33, 87.48	32.37, 88.06
Onions	Forecast	2204.69	2299.4	2272.24	2259.72
	95% Limits	1876.01, 2533.37	1934.64, 2664.16	1895.58, 2648.9	1875.56, 2643.88
Pumpkins, squash, and gourds	Forecast	196.068	199.635	202.635	205.517
	95% Limits	2.05, 390.08	-25.46, 424.73	-38.15, 443.42	-74.47, 485.49
Tomatoes	Forecast	4399.59	3296.7	2469.91	2035.09
Variable		2025-26	2026-27	2027-28	2023-24
	95% Limits	3840.88, 4958.29	2686.85, 3906.54	1849.78, 3090.05	1404.76, 2665.41
Cucumbers and gherkins	Forecast	2024-25	2025-26	2026-27	2027-28
	95% Limits	3344.2	3168.98	4220.18	3021.31
Eggplants	Forecast	2556.2, 4132.2	2367.41, 3970.54	3332.06, 5108.3	2040.73, 4001.88
	95% Limits	558.112	539.438	519.065	496.993
Garlic	Forecast	178.19, 938.03	138.88, 939.99	95.62, 942.52	48.42, 945.56
	95% Limits	60.519	60.8265	61.134	61.4415
Onions	Forecast	32.40, 88.63	32.44, 89.21	32.49, 89.78	32.53, 90.35
	95% Limits	2342.27	2425.04	2435.03	2436.1
Pumpkins, squash, and gourds	Forecast	1955.28, 2729.25	2028.49, 2821.58	2030.4, 2839.65	2030.86, 2841.34
	95% Limits	207.411	208.927	210.258	211.065
Tomatoes	Forecast	-83.36, 498.19	-87.85, 505.70	-95.94, 516.46	-97.43, 519.56
	95% Limits	375.912	-1159.31	-2482.91	-4486.66
		-445.03, 1196.86	-2027.08, -291.55	-3395.89, -1569.93	-5549.07, -3424.25

at a selected future time are within this distance with 95% confidence. Based on Table 6, the forecast data in the cucumbers and gherkins show that production values do not increase uniformly in the next eight years. In the third to fifth groups, i.e. garlic, onions, pumpkins, squash, and gourds, an increase in the amount of production of these groups is predicted in the future. These findings are consistent with the results of Amin et al. (2014), Khayati (2015), and Maghrebi et al. (2020) studies. But they are despite the other studies such as Arivarasi and Ganesan (2015) and Fauziyanti et al. (2020). On the other hand, in the eggplants, and tomatoes, the decrease in production is predicted in the next eight years, especially in the tomatoes. These results are in line with other studies carried out by Arivarasi and Ganesan (2015), and Fauziyanti et al. (2020). So, decreasing the production of some vegetables could be challenged in the future. These forecasted data for vegetable production can be a good solution to ensure the food security of people in the country. On the other hand, if the country is not able to import, underproduction might lead to more production gaps in the country for a particular commodity, and therefore might lead to serious food insecurity, especially in emergency conditions such as floods, earthquakes, etc. The proper design and strategy lead to improved production decisions (Zinyengere et al., 2011; Goodwin et al., 2010). In most studies i.e. Amin et al. (2014), Arivarasi and Ganesan (2015), and Khayati (2015), the ARIMA model was recognized as the best model for the data. In Fig. 5, the actual and forecasted data are shown for each vegetable production group with a limit plot of 95% (Statgraphics output).

4 Conclusion

The forecast data in the cucumbers and gherkins record that production values do not increase regularly in the next eight years. In the garlic, onions, pumpkins, squash, and gourds, an increase in the production of these groups is forecasted in the future; besides, the predicted data show a relatively large decrease in production in the eggplants and tomatoes for the next eight years, particularly in the tomatoes. The predicted data show a worrying decline in the production of these two products. If proper planning is not done, the production reduction of these products would increase prices and imports of these products in the future. Therefore, by using the results of this research, it is suggested that the government and the Ministry of Agriculture, increase the amount of production and control its price with proper planning for the production of vegetables, so that the price increases and imports are prevented. Due to the need of using annual data in this research and the need to study many observations, the disability to access appropriate data of other effective variables on an annual at the large scale level was one of the limitations of the study. Also, the study was carried out on a large scale and over the whole country. So, the local climatic variations in the country could not be taken into account. Regional climatic variations could be studied in the future. The results of this research can be useful for the government, the Ministry of Agriculture, and researchers. They can be able properly plan in future vegetable production along with other corps production as well as further studies.

Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

Akaike H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723. doi: 10.1109/tac.1974.1100705.

Amin M, Amanullah M, Akbar A, et al. 2014. Time series modeling for forecasting wheat production of Pakistan. The Journal of Animal & Plant Sciences 24:1444–1451.

Anderson A. 2013. Business statistics for dummies: How to classify trends in a time series regression model. Wiley.

Arivarasi R, Ganesan M. 2015. Time series analysis of vegetable production and forecasting using ARIMA model. Asia. J. Sci. Technol 6:1844–1848.

Box GEP, Jenkins GM. 1970. Time Series Analysis. Holden-Day, San Francisco, USA.

Box GEP, Jenkins GM, Reinsel GC, Ljung GM. 2015. Time series analysis: Forecasting and Control. John Wiley & Sons.

Burnham KP, Anderson DR. 2002. A practical information-theoretic approach. Model selection and multimodel inference 2.

FAO. 2021. Food and Agriculture Organization of the United Nations. https://www.fao.org/faostat/en/#data/QC.

Fauziyanti NU, Alfana MAF, Putri RF. 2020. A projection production and consumption of food crops in Bali Province towards 2021-2025. IOP Conference Series: Earth and Environmental Science 451:1–12. doi: 10.1088/1755-1315/451/1/012037.

Forecasting Statgraphics. 2017. Statgraphics Technologies, Inc. https://www.statgraphics.com.

Goodwin P, Onkal D, Thomson M. 2010. Do forecasts expressed as prediction intervals improve production planning decisions? European Journal of Operational Research 205:195–201. doi: 10.1016/j.ejor.2009.12.020.

Gujarati DN, Porter DC. 2004. Basic econometrics, 4th edn, the macgraw hill. New York .

Hanke JE, Wichern DW. 2008. Business Forecasting. Pearson Education, New Delhi.

Karim MR, Awal MA, Akter M. 2010. Forecasting of wheat production in Bangladesh. Bangladesh Journal of Agricultural Research 35:17–28. doi: 10.3329/bjar.v35i1.5863.

Khayati A. 2015. Forecasting major vegetable crops productions in tunisia. International Journal of Research 2:15–19.

Killeen PR. 2018. Predict, control, and replicate to understand: How statistics can foster the fundamental goals of science. Perspectives on Behavior Science 42:109–132. doi: 10.1007/s40614-018-0171-8.

Latifi Z, Shabanali Fami H. 2021. Forecasting wheat production in iran using time series technique and artificial neural network. Journal of Agricultural Science and Technology 24.

Maghrebi M, Noori R, Bhattarai R, Yaseen ZM, Tang Q, Al-Ansari N, Mehr AD, Karbassi A, Omidvar J, Farnoush H, Haghighi AT, Kløve B, Madani K. 2020. Iran's Agriculture in the Anthropocene. Earth's Future 8:e2020EF001547. doi: 10.1029/2020ef001547.

MoAJ. 2021. Agricultural statistics-crops. Ministry of Agriculture Jihad. https://www.maj.ir/.

Pakravan M, Gilanpour O. 2013. Investigating export potential vision and competitiveness of iran's agricultural product in middle east and north africa. Journal of Agricultural Economics and Development 27:51–63.

Rahman MR, Ando K, Takeda S, Kobayashi S, Mahzabin IA. 2016. Coping strategies adopted by the rice farmers in the coastal area of Bangladesh. Fundamental and Applied Agriculture 1:161–166.

Tsay RS. 2013. Multivariate time series analysis: with R and financial applications. John Wiley & Sons. doi: 10.1002/0471746193.ch8.

Ulger TG, Songur AN, Çırak O, Çakıroğlu FP. 2018. Role of vegetables in human nutrition and disease prevention. In: Vegetables - Importance of Quality Vegetables to Human Health. InTech. doi: 10.5772/intechopen.77038.

Vågsholm I, Arzoomand NS, Boqvist S. 2020. Food security, safety, and sustainability—getting the trade-offs right. Frontiers in Sustainable Food Systems 4:1–14. doi: 10.3389/fsufs.2020.00016.

Weisberg S. 2005. Applied Linear Regression. John Wiley & Sons, Inc. doi: 10.1002/0471704091.

Zainal M. 2010. Time series and their components. QMIS 320:5.

Zinyengere N, Mhizha T, Mashonjowa E, Chipindu B, Geerts S, Raes D. 2011. Using seasonal climate forecasts to improve maize production decision support in Zimbabwe. Agricultural and Forest Meteorology 151:1792–1799. doi: 10.1016/j.agrformet.2011.07.015.

The Official Journal of the **Farm to Fork Foundation**
ISSN: 2518–2021 (print)
ISSN: 2415–4474 (electronic)
http://www.f2ffoundation.org/faa