J Bangladesh Agril Univ 21(4): ***542-551****, 2023* https://doi.org/10.5455/JBAU.157099

ISSN 1810-3030 (Print) 2408-8684 (Online)

Journal of Bangladesh Agricultural University

Journal home page: http://baures.bau.edu.bd/jbau

Research Article

Modeling of Lemon Production: A Stepwise Regression Approach

Mashrat Jahan[1], Keya Rani Das[2]✉, Linnet Riya Barman[3] and Preetilata Burman[3]

[1] Department of Agricultural Economics, Bangabandhu Sheikh Mujibur Rahman Agricultural University Gazipur-1706, Bangladesh

[2] Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Agricultural University Gazipur-1706, Bangladesh

[3] Faculty of Agricultural Economics and Rural Development, Bangabandhu Sheikh Mujibur Rahman Agricultural University Gazipur-1706, Bangladesh

ARTICLE INFO

Article history
Received: 11 June 2023
Accepted: 13 December 2023
Published: 31 December 2023

Keywords
Citrus,
Climatic variable,
Multicollinearity,
Stepwise Regression,
MLRM

Correspondence
Keya Rani Das
✉: keyadas57@bsmrau.edu.bd

ABSTRACT

Lemon, which is both tropical and subtropical fruit, was formerly used for only entertaining the guests, but today, as a result of its many different uses; it is highly popular with both consumers and producers. Bangladesh is home to extensive lemon groves due to the fruit's high nutritional value, favorable growing conditions, and widespread demand on both domestic and international markets. This study examined the combined impact of farm specific factor and climatic variable on lemon production in Bangladesh. Here the stepwise regression technique with various diagnostic plots in R studio was used to assess the suitable multiple linear regression model and an effective autoregressive integrated moving average (ARIMA) model-based time series analysis was produced for future projection of lemon. The study found a positive link between lemon output and price but a negative correlation with climate variables. Annually, one substantial drop in the lemon cultivars as a result of climatic stress is initiated due to the most significant aspects of which are temperatures and rainfall. Result found, with each millimeter of increased AMR, output diminished by 0.0006419 times on average. The production drops by 0.3564144 times on average for every unit increase in annual mean temperature (AMT). Price increases for lemons may be traced back to covid-19, when production was increased in response to increasing demand about the health benefits of lemons. Also, the investigation found ARIMA (1,1,1) model with drift to the best for future projections of lemon production.

Introduction

Traditional forms of subsistence farming are by far the most important contributor to Bangladesh's agricultural production. Demand for Bangladesh's high-quality agricultural products has increased and thanks to the country's expanding middle class (now estimated at over 30 million) (ITA, 2022). Nevertheless, despite the fact that agriculture accounts for a significant portion of Bangladesh's overall economy, it has largely remained a subsistence industry, with unpredictable crop yields and ineffective infrastructure restricting farmers' ability to fully commercialize their production. Due to its geographical location and physiographic characteristics, its climate varies throughout the year, contributing to its erratic harvests (Amin et al., 2015). Considering all these challenges horticultural farming is gaining popularity over crop farming. The possibility of growing in a small-scale area with potential output price this production is reassuring farmers with great prospects. Among horticultural crops lemon being a common sour fruit rich in vitamin C (Sfgate, 2017) wins farmers heart due to its huge economic projections.

There are a number of reasons why lemon farming is so popular in Bangladesh, such as high demand in both domestic and overseas markets, rich sources of vitamins, suitable soil and favorable weather. The tree has a thorny appearance and is simple to grow in one's own backyard or other kind of garden. Despite the fact that the citrus fruits grown in farmhouse gardens are often intended for personal consumption (Roy and Sarker, 2016) and there is a limited scope for commercial cultivation can be seen for lemon. Although sometimes the homestead production of lemon and its sales found significant contribution on women empowerment and their livelihoods especially for female headed households (Kabir et al., 2016).

Cite This Article
Jahan, M., Das, K.R., Barman, L.R, and Burman, P. 2023. Modeling of Lemon Production: A Stepwise Regression Approach. *Journal of Bangladesh Agricultural University,* 21(4): 542-551. https://doi.org/10.5455/JBAU.157099

Lemon juice includes enzymes that are regarded to be a natural aid for controlling obesity. It also contains a variety of proteins that are responsible for the burning of human fat (Abobatta, 2019). Lemon is not only popular for its juicy character but also has a versatile use. In addition to its consumption practices, its pulp is recycled as animal feed, and its peels are reused in the production of vinegar, lactic acid, and citric acid, all of which are essential in the development of yeast enzymes. Moreover, the lemon tree's leaves and bark are believed to have therapeutic properties, which is another advantage of beginning lemon farming. Because of the widespread belief that a high-Vitamin-C diet would defend against the COVID-19 pandemic and is essential for those already infected with the fatal virus, lemons have been in high demand ever since the outbreak began (Seraj, 2021). As a result, the demand for lemons increased instantly. Because Bangladesh is home to such a large population, the nation's annual lemon consumption must account for a significant quantity. If it produces with a good quality assurance along with quantity the acceptance also increases vastly as Sánchez-Bravo et al. (2022) said that consumers were more receptive to lemons with a higher volatile component concentration that had been cultivated under organic circumstances. So, for many aspiring young business leaders, this may be a fantastic and lucrative idea.

Even though the majority of lemons come from the region of Sylhet, Chittagong and the Chittagong Hill Tracts it is regarded as a widely grown horticultural product across the whole nation (Kabir et al., 2016). It is also good for intercropping systems. Since citrus growth and production are affected by a wide range of abiotic factors, any change in climatic factors will have all round effects on citrus production (Walthall et. al., 2012). Also, the significance of taking into account location-specific phenophase changes within certain areas, since different trends may occur within a geographic area; this has significant repercussions for the future of agricultural planning and the supply of fruit crops to local and worldwide markets (Fitchett et al., 2014). The effects of climate change, particularly drought and high temperatures taken together, are having a negative impact on the production and development of plants (Mittler et al., 2012). However, there should be appropriate management methods in citrus orchards so that the choice of better shoot may promote citrus trees to generate larger yields with good fruit quality. This will limit the negative effects that plant physiological stressors have on the plant (Shafqat et. al., 2021).

Considering the above-mentioned facts, various comprehensive studies have been carried out on different aspects of lemon cultivation in Bangladesh. Bony et al. (2021) have assessed the productivity and profitability of a lemon-based agroforestry system through analysis of variance (ANOVA) and adjusting mean differences through Least Significant Difference (LSD) method. Their economic analysis showed that the net return, benefit-cost ratio (BCR) and land equivalent ratio (LER) was higher in the lemon-based agroforestry system than sole cropping. Hasan et al. (2016) have observed the effect of management practices on the growth and yield of lemon and lime following the statistical procedures for agricultural research. Wang et al. (2022) have noticed the effects of climatic factors on quality yield of lemon and established projection models to elucidate the impact of future climate change. They found temperature and humidity have a prodigious influence on quality growth of citrus fruit with increasing yield prediction due to future suitable climate conditions for growth and ripening. Iftikhar et al. (2010) have noticed the severity of yellow-vein cleaning viral illness in Eureka lemons was significantly correlated with meteorological factors such as maximum/minimum temperature and wind speed. Since citrus growth and production are affected by a wide range of abiotic factors, any change in climatic factors will have all round effects on citrus production (Walthall et. al., 2012). Though there have been quite many agronomical researches on the production of lemon and climatic effect on the output there hasn't been any particular research on the factors concerning both climatic and farm-oriented issues affecting the production of lemon that has been studied. So this study makes an attempt in finding out the fitted regression model from these combined factors affecting the production of lemon all over Bangladesh by the multiple linear regression method.

Materials and Methods

This section entails the detailed data description and empirical strategy adopted for obtaining the best suitable regression model of the lemon production in Bangladesh.

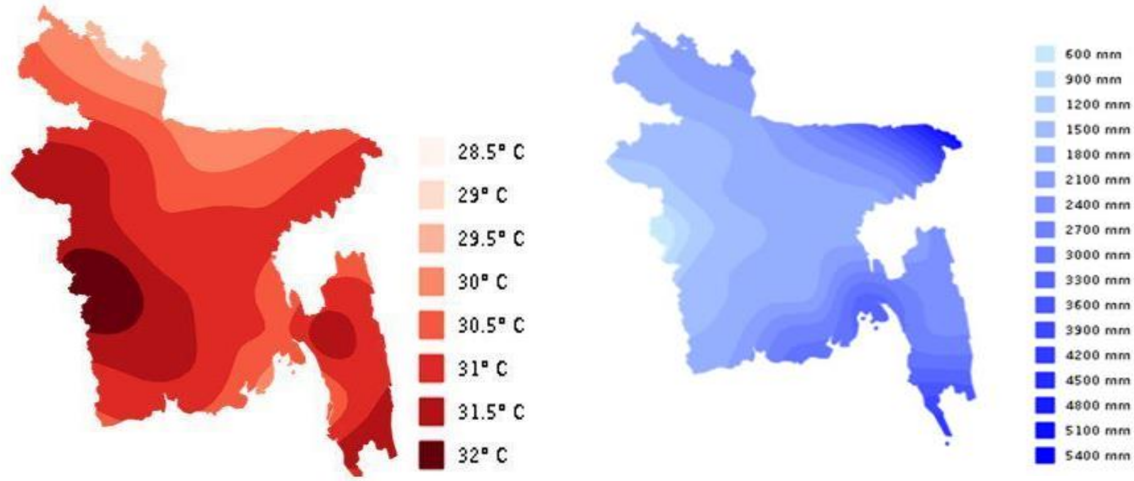

Figure 1. Annual Average Maximum Temperature and rainfall of Bangladesh

Data Description

This research exploits 60 years of annual time series data from 1961–2020 to investigate the influence of farm specific and climatic factor deployment on the production of lemon in Bangladesh. Here first two main factors (farm specific and climatic) were considered and then to find their effect altogether in production, these two categories were specified into several variables like HA in hector, yield in kg/ha, price in BDT and AMR in mm, AMT in celsius to find the best suitable regression model.

Empirical model

Farm specific variables along with climate variables have been chosen here to determine their effect on the production of lemon. This study is pursued to innovate the linear relation between lemon production and its commercial price, harvest area, and rainfall, temperature in Bangladesh. In this regard, multiple linear regression analysis is applied here to the time series data through R software.

MLRM

An expansion of the simple linear regression model for data with several predictor variables and a single result is the multiple linear regression models. Following the regression-built production model we observe the influence of farm specific variable and climatic stress variables on the production performance of lemon in Bangladesh over the last 60 years. It establishes a formal statistical association between the single incessant production output and the explanatory variables at the same time.

X_h (h = 1, 2, 3,4……, p − 1):

$Y_m = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{P-1} X_{P-1} + \varepsilon_m$

$\varepsilon_m \sim N(0, \sigma^2)$

$Y_i \sim N(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{P-1} X_{P-1}, \sigma^2)$. .... (1)

Correlation analyses

Correlation in the broadest sense is a measure of an association between variables (Schober and Schwarte, 2018). The most critical realization is that we can't leap from correlation to causation (Akoglu, 2018). A very well interpretation is displayed about the Pearson's and Spearman's correlation coefficients in Akoglu, H. 2018 and also with interpretations, assumptions of correlation coefficients is discussed in Janse at el., 2021. Pearson’s correlation coefficient calculation method was recommended to calculate the correlation degrees between variables for which the formula utilized is as follows:

$$r_{xy} = \frac{cov(xy)}{\sqrt{var(x)var(y)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Assumption checking

A sticky topic concerning the correct application of Multiple Regression Analyses is the substantial consideration of underlying assumption (Stade et. al., 2011). Normality, linearity, homoscedasticity, multicollinearity, and the absence of outliers are the assumptions that must be checked to conduct a linear regression analysis (Das and Imon, 2016). For a regression model to generalize we must be sure that underlying assumptions have been met, and to test whether the model does generalize we can look at cross-validating it (Field, 2009). So, the current study which focused on doing multiple linear regression there is a necessity of checking the assumptions.

Normality assumption

The violation of normality assumption resulted in bias model estimates (Schmidt and Finan, 2018). Although conventional normality tests work well with uncorrelated data, the existence of correlations among the residuals calls into doubt their usefulness when analyzing dependent data (Marange and Qin, 2019). In regression, it is the utmost expectation that variables on lemon production should be normally distributed. If there is any non-normality among the data, then the relationships and significance tests can be distorted by the variables (extremely skewed or kurtosis, or variables with significant outliers) (Osborne and Waters, 2002). So, a variety of data to test this assumption, including skew, kurtosis, normal P-P plots, normal Q-Q plots, which provide information on normality, and Kolmogorov-Smirnov tests, which provide inferential statistics on normality (Das and Imon, 2016).

Linearity assumption

Standard multiple regressions can only accurately estimate the relationship between variables if the relationships are linear in nature (Osborne and Waters, 2002). The checking is evaluated as an indispensable issue as the results of non-linearity in agricultural science like the prediction of production (i.e., lemon) will debase the original outcome.

Homoscedasticity Assumption

The violation of homoscedasticity indicates heteroscedasticity. Berry and Feldman (1985) [8] assert that if there is remarkable presence of heteroscedasticity in the test that means the interaction of an autonomous variable with some variables are not included in the model. The study inspects visual uneven scattered distribution of residuals as a presence of heteroscedasticity.

Outliers and high leverage points

In all areas of the social and natural sciences, outliers are regularly found during the data gathering phase of empirical or experimental studies. Due to the excessive or abnormal nature of the associated values, outliers are separated from the majority of the other data points. Outliers induce bias into statistical estimates, such as mean values, causing the final results to be either under or overestimated (Kwak and Kim, 2017). Dealing with outliers is necessary before continuing with the analysis of the data set that includes them. It's possible that observations brought about by high leverage points are the major cause of collinearity. When fitting a regression line using the ordinary least squares (OLS) method, the resultant residuals are the functions of leverages and high leverage points combined with large errors (outliers) may drag the fitted least squares line in such a manner that the fitted residuals related to such outliers are too tiny, resulting in concealment (false negative) of outliers (Habshah and Imon, 2009). This outliers and high leverage points of the current study can be identified by inspecting the Residuals vs Leverage plot.

Multicollinearity with the help of VIF

One potential remedy, in the event that multicollinearity is discovered in the data, is to centralize the data. In order to center the data, you will need to take each observation's score and then deduct the mean score for each independent variable. This is because multicollinearity renders some of the potentially important variables being studied to be non-significant (Shrestha, 2020). The most straightforward approach, on the other hand, is to use the VIF values to isolate the factors that are contributing to the multicollinearity problems and then exclude those variables from the regression. The VIF measures the degree to which one independent variable may be described by another independent variable. The higher the VIF, the greater the likelihood that another independent variable can explain the variable in question. Also, it can be applied stepwise regression if the dataset has a multicollinearity problem.

Stepwise regression

Stepwise regression is a subset of hierarchical regression technique that is an appropriate selection in case of more than one predictor. Using the least possible independent variables, the stepwise regression approach intends to optimize the estimating power. The stepwise regression method is a hybrid approach that combines forward and backward selection that involves an automatic process for selecting independent variables (Silhavy et. al., 2017). A variant of forward selection involves re-evaluating all of the model's aspirant variables after each iteration of variable addition to determine whether or not their significance has been decreased to an unacceptable level. In forward selection, the process of adding new model variables occurs in an iterative fashion, beginning with no inputs. An insignificant variable is dropped from the analysis when discovered.

ARIMA Model

ARIMA (Auto-Regressive Integrated Moving Average) model is widely used for time series forecasting and analysis. It is commonly denoted as ARIMA (p,d,q), where p, d and q are nonnegative integers. The three parameters have the following definitions: p is the number of autoregressive terms; d is the number of nonseasonal differences needed for stationarity; and q is the number of lags in the forecast errors in the prediction equation. Using the acquired difference steps, the ARIMA model enables the conversion of nonstationary series into stationary ones. The correct selection of the parameters p and q necessitates an analysis based on the residuals' autocorrelation function (ACF), partial autocorrelation function (PACF), Akaike's information criterion (AICC), and Schwarz's Bayesian information criterion (BIC). For this study

ARIMA (1,1,1) has been applied. ARIMA (1,1,1) is a model with one AR term and one MA term that are being applied to the variable-

$Z_t = X_t - X_{t-1}$

The AR component of ARIMA denotes that the developing variable of interest is regressed on its own lagged (i.e., previous) values. The MA component of the equation shows that the regression error is essentially a linear combination of error components, the values of which happened simultaneously and at distinct points in the past.

Results and Discussion

Although it is currently grown in all tropical and temperate regions of the world, the lemon originally has its origin from Southeast Asia.

The graph displays the average annual production of lemons in Bangladesh between 1960 and 2020 according to FAO data. As was previously mentioned, climate influences can cause significant changes in annual productivity.

As shown in Figure 2, the production growth of lemon has been widening since 2005, It is important to note that after the 20th century, lemon consumption consistently exceeded the expectations. The demand for lemons has risen in tandem with their cultivation as people have become more health conscious and have begun to consume lemonade in greater quantities than in the past.

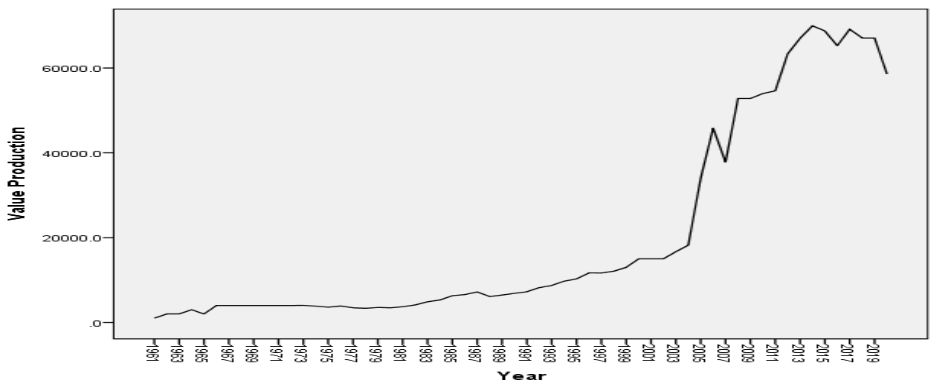

Figure 2. Production of citrus in Bangladesh from the year 1961 to 2020

Furthermore, Table 1 filed the data sources along with a thorough explanation. All the variables, with the exception of yield and AMR, are right-skewed, according to the descriptive measures of the variables, as displayed in Table 2. It appears that there is a lack of normal distribution to all the variables.

Table 1. Variables Specification

Variable	Definition	Source
HA	Harvest area under lemon cultivation in ha	
Yield	Per hector production	www.faostat.com
AMR	Average annual mean rainfall	www.climateknowledgeportal.wo
AMT	Average annual mean temperature	rldbank.org
Price	Consumer purchasing price at harvest	

Table 2. Descriptive statistics

	HA	Yield	Production	AMR	AMT	Price
Mean	8274.00	26803.73	20415.92	2245.0008	25.5795	12546.07
Median	2984.50	26667.00	7212.50	2228.7550	25.5550	4432.50
Max	400	9248	1000	1679.18	25.08	615
Min	64282	43179	69950	2844.49	26.58	42986
Skewness	3.092	-0.407	1.169	-0.103	0.860	1.169
Kurtosis	11.198	5.180	-.347	-.587	.525	-.347

Model approximation and exploration

Considering specific variables in relation to the equivalent production level of lemon, the fitted regression model for lemon production is as follows.

Production = $1.204e^{+01} - 2.301e^{-05}(HA) - 4.724e^{-05}(Yield) - 7.101e^{-04}(AMR) - 3.594e^{-01}(AMT) + 1.627e^{+00}(Price)$

According to the findings of the study, there is an inverse relationship between the production of lemons with the aggregate harvest area, yield under cultivation, mean annual rainfall and temperature but the sales price has a positive correlation with the production.

Table 3. Estimates and diagnostics of multiple linear regression model

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	$1.204e^{+01}$	$4.902e^{+00}$	2.456	0.01731 *
HA	$-2.301e^{-05}$	$1.723e^{-05}$	-1.336	0.18729
Yield	$-4.724e^{-05}$	$2.769e^{-05}$	-1.706	0.09374
AMR	$-7.101e^{-04}$	$2.402e^{-04}$	-2.956	0.00461 **
AMT	$-3.594e^{-01}$	$1.878e^{-01}$	-1.914	0.06098
Price	$1.627e^{+00}$	$1.492e^{-05}$	109048.479	$< 2e^{-16}$ ***
Residual standard error				0.4652
Multiple R-squared				Very near to 1
Adjusted R-squared				Very near to 1
F-statistic				$3.103e^{+10}$ (on 5 and 54 DF)
p-value				$< 2.2e^{-16}$

(Significance codes: 0 '***' 0.001 '**' 0.01 '*')

Result from Table 3 displays all the predictors except for price negatively affecting the production. For one unit incremental change in HA, AMR, AMT and yield, the production of lemon decreased by $2.301e^{-05}$, $7.101e^{-04}$ $3.594e^{-01}$and $4.724e^{(-05)}$ times on average while other independent variables remain fixed. Although the overall model found significance in table 3, most of the predictors except for price and annual rainfall are not significantly initiated. Price and AMR are significantly influenced by lemon production whereas it is not showing significant results for the rest of the predictor variables. However, the overall model shows significant results (F, p value < $2.2e^{-16}$), all of the individual predictor variables should be significant. So, this is an appropriate indication of violation of the assumption of MLRM of this dataset. Now it's the high time to check the assumption of MLRM.

Following the insignificant results that come out of the model, the next step is to construct a revised model that satisfies the criteria of having the highest degree of fitting with constant items and having t-test fallouts for each explanatory variable that have all elapsed.

Assumption checking of lemon

To diagnose how well the model is fitting with the data, R diagnostic plots were applied. These diagnostics include:

1. Residuals vs. fitted values
2. Q-Q plots
3. Scale location plots
4. Cook's distance plots

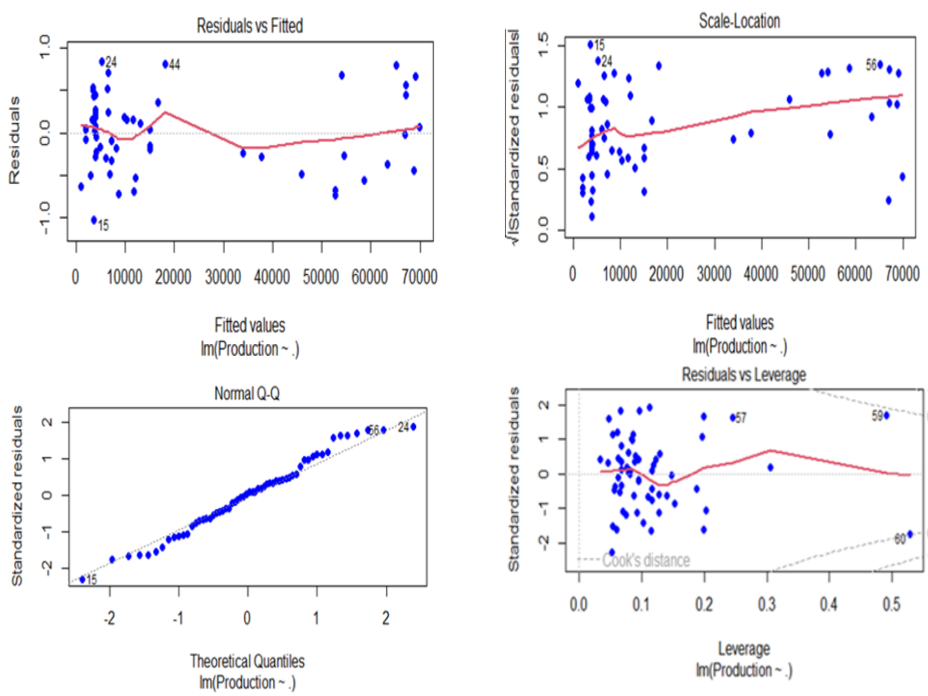

Figure 3. Assumption checking plots of MLRM

Linearity

The residuals vs fitted plot is to examine the assumption of linearity. The existence of a pattern can suggest that there is an issue with one of the components of the linear model. In our residual plot, there is no discernible pattern. So, suggesting from this plot the linearity link between the variables that were used as predictors and those that were used as outcomes does not match properly.

Normality

The normal Q-Q plot is used to investigate the assumption of the normal distribution. If everything goes according to plan, the normal probability plot of residuals will look straight line. In this particular instance, the points are located in close proximity to the reference line, though there are some deviations in the above and below part of the plot but this is very few and slight apart from the line. So that, this plot displays that the normality assumption has followed for the dataset.

Homogeneity

The scale–location plot, or spread–location plot, is another name for this kind of graph helps in identifying the assumption of homogeneity. This graphic demonstrates whether or not the residuals are evenly distributed over the ranges of the predictors. It is desirable to be able to make out a horizontal line with points distributed evenly throughout its length. It can be observed that the variety of the residual points decreases with the value of the fitted outcome variable in this example, which suggests that the residual error variances remain constant. This can be seen by looking at the graph. At first, the residuals appear randomly spread though there are some points (observation no. 22, 40 and 59) deviate extremely from the assumption. The red smooth line is horizontal because the residuals are becoming more dispersed to the x axis. Therefore, the homogeneity assumption does not appropriately match with this lemon production data.

Outliers and high leverage points

Examining the residuals versus leverage plot allows for the detection of outliers as well as high leverage areas. In the plot 57, 59 and 60 are the top three most extreme points, using something called a standardized residual above 1. Moreover, no outlier even exists here that exceeds 2 standard deviations, which is good. Every piece of data has a leverage statistic underneath 2(p+1)/n= 12/60=0.20. So, observation no. 57, 59, and 60 show patterns as leverage values. From the above assumption have been checked there is a lack of perfectly followed plot over the data to fit the MLRM for prediction. Also, there is an indication of a multicollinearity problem for this data. So, checking multicollinearity is now on demand.

Correlation analyses

Harvest area with Yield and annual mean rainfall has correlations of -0.29 and -0.18 respectively postulating a fairly weak negative linear relationship among them. This harvest area also has a positive weak linearity of 0.18 with annual mean temperature and strong correlation (0.80) with price. As the price increases the harvesting area of lemon also increases simultaneously. On the other hand, lemon yield establishes both positive and negative but weak relationships among annual mean temperature, annual mean rainfall and price by 0.15, 0.28 and -0.26 respectively. The mean rainfall and temperature have a correlation of -0.10 indicating that there is a weak negative or almost no relationship. Again, rainfall has a negative correlation of -0.33 but the temperature has weak positive linear (0.30) relationship with price.

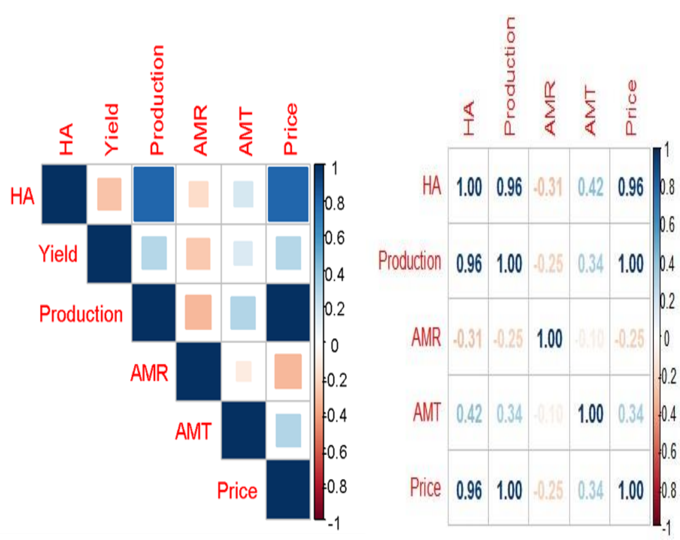

Figure 4. Correlation analysis of different variables for lemon production

Multicollinearity checking

Higher the VIF value encourages the higher multicollinearity. Here, the harvested area (HA) has a VIF of 12.873592 that indicates a potentially severe multicollinearity with HA as well as other predictors included in the model. In this particular instance, the coefficient estimates and p-values that are included in the result of the regression are probably not accurate. Yield has a VIF of 5.145907 that indicates a no multicollinearity problem with other predictor variables in the model. Annual mean rainfall (AMR) has a VIF of 1.161876 that indicates no multicollinearity between AMR and the model's other predictor variables, although the problem is usually not serious enough to warrant further investigation. Annual mean temperature (AMT) has a VIF of 1.109115 that indicates no multicollinearity between AMT and other predictors.

Table 4. Variance Inflation Factor

	HA	Yield	AMR	AMT	Price
Lemon	12.873592	5.145907	1.161876	1.109115	13.046727

Yet, the severity of this issue seldom calls for intervention. Price has a VIF of 13.046727 that indicates a potentially severe multicollinearity problem. We can also see from the output that the R-squared value for the model is 1. The global F-statistic is also visible, as it is $3.103e^{+10}$ because the relevant p-value is lower than $2.2e^{-16}$, it may be deduced that the entire regression model is statistically significant. Based on this actual situation, this result of high multi-collinearity between price and harvest area pushed the data now to follow a stepwise model for appropriateness. Therefore, stepwise regression analysis has found the best solution to test the conceivable multicollinearity.

Stepwise regression analysis

In the stepwise model the best lower value of Akaike information criterion (AIC) will construct the best model. The Table 5 shows the model 4 has the lowest AIC value for predicting variables.

Table 5. Best subset regression of lemon production

	Intercept	Price	HA	AMT	AMR	Yield	AIC
Model 1	√						1210.57
Model 2	√	√					-80.28
Model 3	√	√			√		-85.15
Model 4	√	√		√	√		**-86.86**

The result reveals that model 4 is the ultimate model in the middle of all the existing models that might be the pertinent according to the current collection of "possible predictors". This indicated that the model with the selected predictors was the best choice (i.e., Market price, AMR and AMT). In conclusion, the "suitable predictors" were therefore found to have been identified.

Table 6. List of appropriate predictors

Predictor	Lemon (estimate)
Intercept	10.6032
Price	1.6272
AMR	-0.0006
AMT	-0.3564

So, the final regression model for lemon is

$$Y = 10.6032768 + 1.6272727 \text{ Price} - 0.0006419 \text{ AMR} - 0.3564144 \text{ AMT}$$

From the final fitted regression model, price has a positive impact on lemon production. It signifies the possibility that if potential output price of lemon is encouraged in the market the farmer's concept of producing lemon in turn will be uplifted. Again, this study founds a 1 mm increase in AMR (annual mean rainfall) will decrease the production by 0.0006419 times on average if other predictors remain constant. Panigrahi (2014) also finds the almost same as his study found soil, climate, and cultivation also play important role in success under deficit irrigation while Arbona (2005) suggests just the opposite. On the other hand, an increase in AMT lemon production will decrease by 0.3564144 times on average while other independent variables remain fixed. Sandhu and Bal, (2013) observed lemons being a tremendously sensitive fruit to extreme temperature fluctuations. Sato, (2015), Balfagon et al., (2021) and Allen, (2000) also found the same outcome like this study that high temperature during maturation period of lemon cause fruit creasing and splitting, which lowers production with severe damage.

Forecasting with Time Series Data

It is possible to draw the conclusion that the time series under investigation is stationary if the p-value that was returned from the statistical test is significantly lower than the significance level that had been established beforehand, such as 0.05. For the dataset that contains information about lemon production, the result of the Augmented Dickey-Fuller (ADF) test statistic is -1.825, and the p-value for this test is 0.645. The outcome of the test suggests that the data being considered is non-stationary; this conclusion can be drawn from the contents of the dataset. In addition, the ADF test statistic produces a value of -1.861 when a logarithmic adjustment is applied to the production data, and this value is accompanied by a p-value of 0.630. This indicates that the null hypothesis is not supported. In addition, the result that was obtained suggests that there is some non-stationarity present in the data. We might be able to use the method of taking the initial difference of the data if we employ the ARIMA model. The Durbin-Watson test statistic values for the lemon production data and its logarithmic transformation are respectively 0.015 and 0.0004, with the former value being much higher.

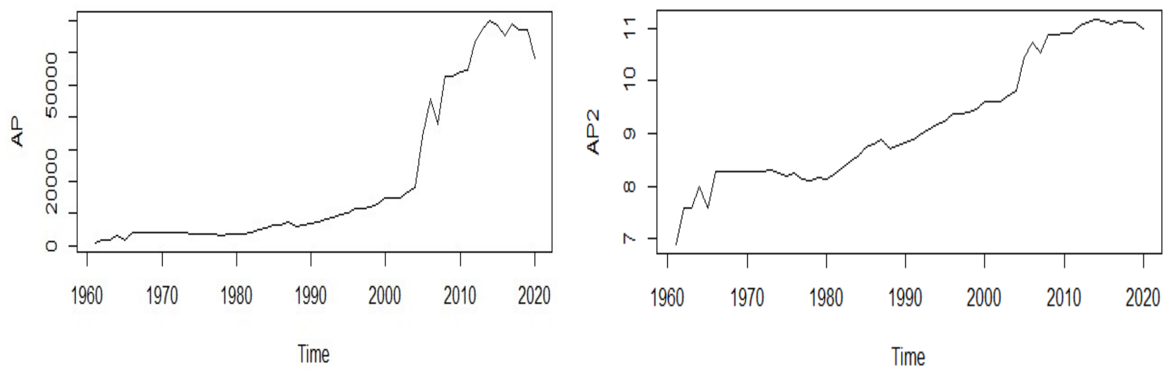

Figure 5. Time series plot of lemon production and logarithmic transformation of production data

Both of these findings point to the presence of a problem with positive autocorrelation within the dataset as a whole. In spite of the fact that applying a logarithmic transformation to the data on lemon production results in outcomes that are comparable in terms of stationarity and autocorrelation, we have decided to evaluate the original lemon production data as opposed to its logarithmic modification.

Table 7. Model performance checking

Parameter	Estimate	St. Error	AIC	AICc	BIC	RMSE	MAE	MAPE
AR1	-0.906	0.135						
MA1	0.658	0.187	-34.2	-33.5	-25.9	0.167	0.114	1.273
Drift	0.0671	0.019						

In this investigation, the data were analysed by fitting them to a number of different ARIMA models so that the effectiveness of each model could be evaluated. The ARIMA (1,1,1) model with drift was found to have the lowest values for Akaike's Information Criteria (AIC), Corrected Akaike's Information Criteria (AICc), Bayesian Information Criterion (BIC), root mean squared error (RMSE), mean absolute error (MAE), mean percentage error (MPE), and mean absolute percentage error (MAPE). This was discovered after comparing a number of different models. For the analysis of the lemon production data, it was determined that the ARIMA (1,1,1) model with drift was the most suitable option. With the help of this model, it will be feasible to make projections regarding the production of lemons that can be harvested in Bangladesh for any future years.

Conclusion

The target is to demonstrate multiple linear regression models for modelling lemon production in Bangladesh. For finding the best suitable variables to this production model is challengingly an intricate process. First, we identified severe correlation among the specific variables that are chosen for this study and ultimately the study set the stepwise regression as the best solution to the problem of multicollinearity and suggest the potential estimated model for lemon cultivars. This model may contribute to proper lemon production on the basis of climatic and farm specific predictors. Modelling from the potential regression may critically help lemon cultivars make informed decisions before harvest. Besides, a time series analysis added to forecast the lemon production in Bangladesh. For the time series analysis, ARIMA (1,1,1) model with drift was applied as per the model performance compared with others.

References

Abobatta W. F. 2019. Nutritional Benefits of Citrus Fruits. *American Journal of Biomedical Science & Research*. *3*(4), 303-306. https://doi.org/10.34297/AJBSR.2019.03.000681

Akoglu, H. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91-93. https://doi.org/10.1016/j.tjem.2018.08.001.

Allen, G. L. 2000. Principles and practices for communicating route knowledge. *Applied Cognitive Psychology*, *14*(4), 333–359. https://doi.org/10.1002/1099-0720(200007/08)14:4<333::AID-ACP655>3.0.CO;2-C

Amin M., Zhang J. and Yang M. 2015. Effects of climate change on the yield and cropping area of major food crops: a case of Bangladesh. *Sustainability, 7*(1), 898–915. https://doi.org/10.3390/su7010898

Arbona V, Iglesias DJ, Jacas J, Primo-Millo E, Talon M, and Gómez-Cadenas A. 2005. Hydrogel substrate amendment alleviates drought effects on young citrus plants. *Plant Soil*, *270*, 73-82. https://doi.org/10.1007/s11104-004-1160-0

Balfagon, D., Arbona, V., and Gomez-Cadenas, A. 2022. The future of citrus fruit. The impact of climate change on citriculture. *Metode Science Studies Journal*, *12*, 123-129. https://doi.org/10.7203/metode.12.20319

BBS. 2021. *Yearbook of agricultural statistics,* Bangladesh Bureau of Statistics. Statistics and Informatics Division, Ministry of Planning, Government of the People's Republic of Bangladesh. Pp.233. http://www.bbs.gov.bd/

Berry, W. D. and Feldman, S. 1985. *Multiple Regression in Practice.* Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-050. Newbury Park, CA: Sage Publications, Inc.

Bony, Z. F., Zakaria, M., Rahman, M., Riyadh, Z. and Saha, S. 2021. Productivity and Profitability Assessment of Lemon Based Agroforestry Systems in Bangladesh. *Asian Research Journal of Agriculture*, *14*, 42-49. https://doi.org/0.9734/ARJA/2021/v14i230123

Das, K. R., Adhikary, B. K. and Uddin, K. S. 2020. The Relationship between Students' Results and Spending Time on the Internet at Higher Education Level in Bangladesh. *ULAB Journal of Science and Engineering*, *11*(1): 1-8.

Das, K. R. and Imon, A. H. M. R. 2016. A Brief Review of Tests for Normality, *American Journal of Theoretical and Applied Statistic*, *5*(1): 5-12. https://doi.org/10.11648/j.atas.2016051.12

Field, A. 2009. *Discovering statistics using SPSS.* 3rd edition. pp 212-307. SAGE Publications Ltd. 1 Oliver's Yard, 55 City Road, London. ISBN 978-1-84787-906-6.

Fitchett, J.M., Grab, S.W., Thompson, D.I., and Roshan, G. 2014. Spatio-temporal variation in phenological response of citrus to climate change in Iran: 1960–2010. *Agricultural and Forest Meteorology*, *198*, 285–293. https://doi.org/10.1016/j.agrformet.2014.08.010

Geospatial Bangladesh: BMap viewer. 2022. Geospatial Bangladesh: bMap Viewer. https://geospatialbangladesh.info/

Habshah, M. N. and Imon A. H. M. R. 2009. The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics*, *36*(5): 507–520. https://doi.org/10.1080/02664760802553463

Hasan, M., Rahim, M. A., Islam, A. K. M. S., Rahman, M. M., Rahman, M. A., and Naher, N. 2016. Effect of management practices on the growth and yield of lime and lemon. *International Journal of Bioscience*, *8*(6):22-33.

Iftikhar, Y., Iqbal, Z., Ahmed, S., Awan, A., Saleem, U., and Sarwar, G., 2010. Effect of environmental factors on yellow vein clearing virus incidence in lemon. *Journal of Agricultural Research*, *48*, 87–92

ITA. 2022. *Country Commercial Guides Bangladesh, Agriculture Sectors.* https://www.trade.gov/country-commercial-guides/bangladesh-agriculture-sectors

Janse, R. J.,Hoekstra, T., Jager, K. J., Zoccali, C., Tripepi, G., Dekker, F. W., and Diepen, M. 2021. Conducting correlation analysis: important limitations and pitfalls. *Clinical Kidney Journal*, *14*(11), 2332–2337. https://doi.org/10.1093/ckj/sfab085

Kabir, M. S., Salehin, M. M., and Jovičić, E. 2016. Can minor fruit cultivation change the livelihood of the marginal peasants? A case study from Bangladesh. Faculty of Business Economics and Entrepreneurship. *International review*, (1-2), 91-100.

Kwak S. K. and Kim J. H. 2017. Statistical data preparation: management of missing values and outliers. *Korean journal of anesthesiology*. *70*(4):407-411.PMID: 28794835; PMCID: PMC5548942. https://doi.org/10.4097/kjae.2017.70.4.407

Marange C. S. and Qin Y. 2019. An empirical likelihood ratio based comparative study on tests for normality of residuals in linear models, *Metodoloski Zvezki. 16*(1), 1–16.

Mittler R, Finka A, and Goloubinoff P. How do plants feel the heat? *Trends in biochemical sciences*, *37*(3), 118-125. https://doi.org/10.1016/j.tibs.2011.11.007

Osborne, J. W., & Waters, E. 2002. Four Assumptions of Multiple Regression That Researchers Should Always Test. *Practical Assessment, research, and evaluation, 8*(1), 2.

Panigrahi P, Sharma RK, Hasan M, and Parihar SS. 2014. Deficit irrigation scheduling and yield prediction of 'Kinnow' mandarin (Citrus reticulate Blanco) in a semiarid region. *Agricultural Water Management, 140*, 48-60. https://doi.org/10.1016/j.agwat.2014.03.018

Roy, R and Sarker T. 2016. Farmers' Opinion towards Conservation and Causes of Depletion of Citrus at Barlekha Upazila of Moulvibazar District in Bangladesh. *Research & Reviews: A Journal of Life Sciences*, 6(2), 6-14.

Sánchez-Bravo P, Noguera-Artiaga L, Martínez-Tomé J, Hernández F, and Sendra E. 2022. Effect of Organic and Conventional Production on the Quality of Lemon "Fino 49". *Agronomy, 12*(5):980. https://doi.org/10.3390/agronomy12050980

Sandhu, S. and Bal, J. S. 2013. Effect of irrigationand soil cover on growth of lemon (Citruslimon (L.) Burm.) cv. Baramasi. *International Journal of Agriculture, Environment and Biotechnology*, *6*(3), 461-464.

Sato, K. 2015. Influence of drought and high temperature on citrus. In *Abiotic Stress Biology in Horticultural Plants; Springer, Tokyo*, 77-86. https://doi.org/10.1007/978-4-431-55251-2_6

Schmidt, A. F. and Finan, C. 2018. Linear regression and the normality assumption. *Journal of Clinical Epidemiology, 98*, 146-151. https://doi.org/10.1016/j.jclinepi.2017.12.006.

Schober, P., and Schwarte, L.A. 2018. Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia. 126*, 1763–1768

Seraj, S. (2021, April 16). Lemon cultivation changes fortunes for many. *The Daily Star.* https://www.thedailystar.net/country/news/lemon-cultivation-changes-fortunes-many-2078129

Sfgate. 2017. *Do Lemons Provide Vitamin C Like Oranges Do?.* sfgate.com. http://healthyeating.sfgate.com/lemons-provide-vitamin-c-like-oranges-do-3508.html.

Shafqat, W., A. Naqvi, S., Maqbool, R., Salman Haider, M., Jafar Jaskani, M., and A. Khan, I. 2021. Climate Change and Citrus. *Citrus - Research, Development and Biotechnology; Khan, MS, Khan, IA, Eds*, 147. doi: 10.5772/intechopen.95488

Shrestha, N. 2020. Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39-42. DOI: 10.12691/ajams-8-2-1.

Silhavy, R., Silhavy, P. and Prokopova, Z. 2017. Analysis and selection of a regression model for the Use Case Points method using a stepwise approach. *Journal of Systems and Software*, *125*, 1–14. https://doi.org/10.1016/j.jss.2016.11.029

Stade, M., Meyer, C., Niestroj, N. and Nachtwei, J. 2011. (Not) Everybody's Darling: Value and Prospects of Multiple Linear Regression Analysis and Assumption Checking. In B. Krause, R. Beyer und G. Kaul (Hrsg.), Empirische Evaluations method en Band 15 (S. 17-34). Berlin: ZeE.

Wang, S., Xie, W. and Yan, X. 2022. Effects of Future Climate Change on Citrus Quality and Yield in China. *Sustainability, 14*, 9366. https://doi.org/10.3390/su14159366

Walthall, C.L., J. Hatfield, P. Backlund, L. Lengnick, E. Marshall, M. Walsh, S. Adkins, M. Aillery, E.A. Ainsworth, C. Ammann, C.J. Anderson, I. Bartomeus, L.H. Baumgard, F. Booker, B. Bradley, D.M. Blumenthal, J. Bunce, K. Burkey, S.M. Dabney, J.A. Delgado, J. Dukes, A. Funk, K. Garrett, M. Glenn, D.A. Grantz, D. Goodrich, S. Hu, R.C. Izaurralde, R.A.C. Jones, S-H. Kim, A.D.B. Leaky, K. Lewers, T.L. Mader, A. McClung, J. Morgan, D.J. Muth, M. Nearing, D.M. Oosterhuis, D. Ort, C. Parmesan, W.T. Pettigrew, W. Polley, R. Rader, C. Rice, M. Rivington, E. Rosskopf, W.A. Salas, L.E. Sollenberger, R. Srygley, C. Stöckle, E.S. Takle, D. Timlin, J.W. White, R. Winfree, L. Wright-Morton, L.H. Ziska. 2012. Climate Change and Agriculture in the United States: Effects and Adaptation. USDA Technical Bulletin 1935. pp 23-146. Washington, DC.