Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4, 1457-1465
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i4.6320

Economic study on the feasibility of building a standard model to predict wheat yield production in Iraq using the bootstrapping and time series method

Alaa K. Frahan[1*], T.A. Hatim[2], Sarah A. Bandar[3]
[1]Department of Plant Protection, College of Agriculture, University of Misan, Misan, Iraq; alaa.k.f@uomisan.edu.iq (A.K.F.)
[2,3]Department of Mathematics, College of Education, University of Misan, Misan, Iraq; Mr.tagrid@uomisan.edu.iq (T.A.H.) Sara.ab@uomisan.edu.iq (S.A.B.)

Abstract: The wheat crop is of great importance in the agricultural economy and dietary pattern, as it represents the main source of bread in Iraq. To bridge the gap, production increased through the efficient use of economic resources. To minimize this problem, the Bootstrapping method is applied, while the time series analysis method ultimately resolves it, making the comparison between the two methods very important. Wheat yield data in Iraq were obtained from published and unpublished secondary sources, including the Ministry of Agriculture, the Ministry of Planning - Central Bureau of Statistics, and the Arab Organization for Agricultural Development in Iraq. Using SPSS (version 22), a model was developed to describe the relationship between production and costs, utilizing the Bootstrapping method and the time series method. Both equations explain 68% of the variation in wheat production, and both equations are statistically significant. However, the Bootstrapping method cannot determine the presence or absence of autocorrelation in residuals. Moreover, the time series method confirms that no autocorrelation exists between the residuals.
Keywords: *ARIMA, Autocorrelation, Autoregressive, Bootstrapping.*

1. Introduction

Food security is one of the primary considerations of any country, therefore, countries must provide the needs of their people with basic commodities that concern their daily livelihood, in which wheat is one of their most important priorities [1]. Wheat is one of the oldest field crops far-famed to be cultivated within the world [2]. In Iraq, wheat grows in large areas, particularly in the northern governorates. It occupies an important place in Iraq's agricultural economy. It is considered one of the main crops as it is used as food for the majority of the population directly or indirectly, as many food industries depend on it. Wheat production increased from about 1.1 million tons in 1997, to about 6.3 million tons in 2020 [3] due to the use of high-yielding breeds and the increase in the agricultural area of the crop, which increased from about 5,498 thousand dunums in 1997, to about 8,574 thousand dunums in 2020 [4]. But despite these increases, it is noted that Iraq still imports large quantities of wheat as a result of the population growth that amounted to about 18 million people during the study period (1997 - 2020) [5-7], and this necessitates the need to rationalize consumption and encourage the cultivation of the crop using modern methods to achieve the highest production per unit area rtant place in the Iraqi balance of payments [8]. Therefore, it is necessary to draw the features of a future development strategic policy according to indicative planning that chooses the best ways to predict the actual needs of this crop in light of the steady increasing population growth rates.

The problem of the study lies in testing a model with strong predictive ability for wheat production in Iraq. The existence of autocorrelation affects the model's ability to predict and to reduce this problem, the bootstrap method is relied on, while the time series analysis model finally saves us from this

History: Received: 3 February 2025; Revised: 10 March 2025; Accepted: 14 March 2025; Published: 17 April 2025
* Correspondence: alaa.k.f@uomisan.edu.iq

problem. Therefore, the comparison between the two methods is very important to reach a standard model that describes the relationship between variables and formulate this relationship in its mathematical form in preparation for estimation, forecasting and the development of appropriate future economic policies. One of the most common statistical methods for measuring the relationship between variables is regression analysis and correlation analysis, where the correlation coefficient is a statistical tool in regression analysis. The correlation coefficient explains the degree or strength of the relationship between two or more random variables, with one (independent) explaining changes in the other (dependent), thus reaching a predictive statistical model with strong estimates. The objectives of the research are as follows:

1. Description of a standard economic model of the relationship between wheat production and wheat production costs, based on the bootstrap method and the time series analysis method. A comparative study between the results of the time-series method and the bootstrap.
2. Measure the efficiency of each method in the forecasting process to reach an efficient standard model that enables decision-makers to predict future events, take corrective actions and develop appropriate economic policies.

2. Analytical Method

The analysis of any economic phenomenon requires finding the factors that affect and are related to it, by finding a formula (or model) that expresses it and incorporates it into its main component. This is one of the most important goals of regression models, which is defined as a statistical method used to explore the relationship between a variable known as a dependent variable and one or several independent variables. The bootstrap method, which is one of the suggested methods, is one of the methods of sampling with replacement (i.e., returning the sample to the population). Time series analysis (ARIMA) is based on the idea of finding an appropriate mathematical model for the nature of the data so that the residuals or errors (which are the difference between the real values of this phenomenon and the values estimated in this model) are as few as possible and there is no autocorrelation between them. Autocorrelation is one of the problems that leads to inaccurate measurement of economic relations coefficients when using the least squares method in regression models, where autocorrelation indicates a correlation between consecutive values of the random limit. One of the most commonly used tests in this field is the Durbin-Watson (DW) test [9, 10].

$$y = \beta x + e \qquad (1)$$

Where:

- y : dependent variable vector
- x: matrix of the independent variables
- e: error vector

The autocorrelation between the residuals may be of the first order, the second order, or a higher order. In the case of autocorrelation of the first order, we find that each value of the random limit is related to the values that precede it only. This case (first order autoregressive model) can be represented as follows:

$$e_t = \rho e_{t-1} + u_t \qquad (2)$$

- e_t: the value of the random error in the current period
- e_{t-1}: the value of the random error in the previous period
- ρ: autocorrelation coefficient $|\rho| \leq 1$
- u_t: random error in the previous equation, in which $u_t \sim N(0, \sigma^2)$

The DW test is performed as follows: Null hypothesis: the autocorrelation coefficient of the random variable is not different from zero; $H_0: \rho = 0$. Alternative hypothesis: the autocorrelation coefficient of the random variable equals zero; $H_1: \rho \neq 0$. To obtain the value of DW statistic, the following formula is used [11, 12].

$$d^* = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \quad (3)$$

The DW statistic depends on $\hat{\rho}$:

- If $\hat{\rho} = 0$ then $d = 2$ (no autocorrelation).
- If $\hat{\rho} = 1$ then $d = 0$ (positive autocorrelation).
- If $0 < d < 2$ suggests possible positive autocorrelation.
- If $2 < d < d$ suggests possible negative autocorrelation.

Comparing d with critical values d_L and d_u from the DW table helps determine the presence of autocorrelation:

- If $d < d_L$ reject the null hypothesis (positive autocorrelation).
- If $4 - d_L < d < 4$ reject the null hypothesis (negative autocorrelation).
- If $d_u < d < 4 - d_u$ accept the alternative hypothesis (no autocorrelation).
- If $d_u > d > d_l$ or $4 - d_u > d > 4 - d_L$ the test is inconclusive.

2.1. Bootstrapped Durbin- Watson test

The bootstrap method enhances the DW test's robustness. The procedure, as suggested by Jinook and Seoung, includes the following steps [13]:

- Estimate $\hat{\beta}$ from the equation (1) using ordinary least squares and calculate $\hat{e}$.
- Estimate $\hat{\rho}$ from the equation (2) using ordinary least squares and calculate $\hat{u}$.
- Generate the bootstrap residual vector u^*.
- Form the residual vector e^* using an equation (2)
- Compute y^* using x , u^*.
- Recalculate DW statistic d^* using x, y^*
- Repeat steps 3-6 to build the empirical distribution of d^* , f_{d^*} and compare f_{d^*} with d from the original data.

2.2. Bootstrapped P(B-P) test

An alternative bootstrap test, P(B-P), involves:

1. Follow steps 1-3 as in the bootstrap DW test.
2. Select a residual from u_t^*, denoted , and form $e_1^*$by:

$$e_1^* = \frac{u_1^*}{\sqrt{1-\rho^2}} \quad (4)$$

3. Configure p by:

$$e_t^* = \hat{\rho} e_{t-1}^* + u_t^* \quad (5)$$

4. Estimate ρ_1^* is calculated use least squares:

$$\rho_1^* = \frac{\sum e_t^* e_{t-1}^*}{\sum (e_{t-1}^8)^2} \quad (6)$$

3. Result and Discussion

3.1. Linear Regression with Bootstrap

The model's performance is summarized by the R-squared value, which indicates that **78.3%** of the variance in wheat production can be explained by the model. The Durbin-Watson statistic (DW) is 1.619, and the corresponding lower (d_l = 0.507) and upper (d_u = 2.097) limits suggest that the test is inconclusive with respect to autocorrelation (d_l < DW < d_u). This indicates we cannot definitively confirm autocorrelation between the residuals.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1457-1465, 2025
DOI: 10.55214/25768484.v9i4.6320

Table 1.
Model Summary for Linear Regression Models.

Parameter	Model 1	Model 2
R	0.885	0.825
R Square	0.783	0.68
Adjusted R Square	0.667	0.666
Std. Error of the Estimate	0.7372	0.7385
Durbin-Watson	1.619	1.088

3.2. ANOVA Results

The significance of the model can be confirmed with the F-statistic. Since the significance value (Sig.) is 0.001 for Model 1, which is less than the threshold value of 0.05, we conclude that the model is statistically significant.

Table 2.
ANOVA for Linear Regression Models.

Parameter	Model 1			Model 2		
Sum of Squares	**Regression**	**Residual**	**Total**	**Regression**	**Residual**	**Total**
df	29.382	8.151	37.53	25.534	11.999	37.533
Mean Square	8	15	23	1	22	23
F	3.673	0.543		25.534	0.545	
Sig.	6.759			46.817		

3.3. Regression Coefficients

The coefficients table shows that no individual explanatory variables are significant in Model 1, as their significance values (Sig.) exceed the threshold of **0.05**. However, the overall model is still significant, which indicates that the combination of variables explains a substantial portion of wheat production variability.

Table 3.
Coefficients for Linear Regression Models.

Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.
	B	**Std. Error**	**Beta**	
Model 1	(Constant)	0.604	0.83	
	Work	-9.13E-06	0	-0.174
	Automated Work	1.40E-05	0	0.479
	Fuel	0	0	-0.53
	Seeds	-3.15E-05	0	-0.218
	Fertilizer	4.50E-05	0	0.51
	Pesticides	0	0.001	-0.476
	Other	2.96E-05	0	0.674
	Rent	0.001	0.001	0.6
Model 2	(Constant)	0.727	0.289	
	Total	8.41E-06	0	0.825

3.4. Model with Total Production Costs

It is clear from the previous table that all the explanatory variables are not significant. where all values of sig > 0.05. This is despite the fact that the linear model as a whole is significant as we explained earlier. This is what made us design a linear model between wheat production and total production costs. It is clear from the previous table that R^2=0.68 and this means that the model explains 68% of the change in wheat production in million tons. It is also clear that the Durban-Watson statistic (DW) =1.088 and the value of the lower limit of the statistic (d_l) = 1.037 and the upper limit (d_u) = 1.199, (n=24, k=1). ∴d_u>d>d_l then the test still is not final in the sense that we cannot ascertain whether there is an autocorrelation between the values of the random variable. It is clear from the previous table that the model is significant, since sig= 0.00 < 0.05. It is clear from the previous table that all the explanatory variables (constant, total costs) are significant. where all values of sig < 0.05. The regression equation can be written as:

$$y = 0.727 + 8.41 * 10^{-6}x \tag{7}$$

Where:
y: Production of wheat in million tons,
X: total costs of the production of wheat.

3.5. *ARIMA Time Series Model*

The ARIMA model selected was ARIMA (2,1,0), which captures two autoregressive terms (AR), one differencing term (d), and no moving average term (MA). The coefficients for AR are significant (p-value = 0.02), which suggests that this model effectively captures the temporal dynamics in wheat production data. The ARIMA model can be expressed as:

$$z_t = \phi_1 z_{t-1} - \theta_1 a_{t-1} + a_t \; ; \tag{8}$$

$$z_t = y_t - y_{t-1} \; ; \tag{9}$$

$$z_t = -.480 z_{t-1} + a_t \; ; \tag{10}$$

Where:
z_t: The first differences series of the series quantity of wheat production in million tons
ϕ: Estimated autoregressive coefficient
θ: Estimated moving average coefficient
a_t: Random error.

Table 4.
ARIMA Model Parameters and Model Description for Production in Million Tons.

Model ID	**Model_1**
Model Type	ARIMA(2,1,0)
Production in million tons	Natural logarithm
ARIMA Model	ARIMA(2,1,0)
AR Lag 2 Estimate	-0.48
AR Lag 2 SE	0.191
AR Lag 2 t-stat	-2.519
AR Lag 2 p-value	0.02

3.6. *Model Statistics*

The ARIMA model fits the data with an R-squared value of 0.683, explaining 68.3% of the variance in wheat production. Additionally, the Ljung-Box test for autocorrelation of residuals shows that the model's residuals are well-behaved, as the significance value (0.399) is greater than 0.05, confirming that no autocorrelation is present in the residuals.

Table 5.
Model Statistics for Production in Million Tons (Model_1).

Model		**Production in million tons-Model_1**
Number of Predictors		1
Model Fit statistics	Stationary R-squared	0.213
	R-squared	0.683
Ljung-Box Q(18)	Statistics	17.843
	DF	17
	Sig	0.399

It is clear from the previous table that the previous model explains the percentage 68.3% of the change in the dependent variable (wheat production in million tons), where $R^2 = 0.683$. The table also shows the results of the Box-Liung test - test the fit of the model at certain time gaps. The model is suitable if $Q_{(k)}$ is less than the tabular value of χ^2at degrees of freedom (k-p-q) , where $Q_{(k)}$: statistic value for the Box-Liung test, k: a set of autocorrelation coefficients for the residuals. The value of the Box-Liung test statistic was calculated at a time gap of (18) and it was = 17.843 less than the tabular value $\chi^2(\alpha, k-p-q)$, meaning that the degrees of freedom (18-1-0), $\left(\chi^2{}_{(.05,17)} = 27.59\right)$to confirm the good matching of the observed values to the predicted values.

3.7. Comparison Between Bootstrap and ARIMA Models

The Bootstrap method outperforms the ARIMA model in both Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), indicating that the Bootstrap method provides more accurate forecasts. To evaluate and compare the efficiency of the Bootstrap method and the ARIMA time series method in forecasting, several metrics can be used. The most common metrics include Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Pearson Correlation, Mean Squared Error (MSE). For this study, we focus on MAE and MAPE due to their widespread use in forecasting accuracy assessments.

3.8. Mean Absolute Error (MAE)

MAE measures the average magnitude of errors in a set of forecasts, without considering their direction. It is calculated using the formula:

$$MAE = \frac{\sum_{i=1}^{n} |x_i|}{n} \tag{12}$$

where:

- x_i: Error value which is the difference between the actual value and the predicted value, $(y_i - \widehat{y_i})$
- n: Number of data.

3.9. Mean Absolute Percentage Error (MAPE)

MAPE measures the average absolute percentage difference between the actual and predicted values. It is calculated using the formula

$$MAPE = \frac{\sum_{i=1}^{n} \left(\frac{|x_i|}{y_i}\right)}{n} \tag{13}$$

where:

- x_i: Error value which is the difference between the actual value and the predicted value, $(y_i - \widehat{y_i})$
- y_i: The actual value
- n: Number of data

Figure 1 compares the predictive performance of the Bootstrap and ARIMA models by plotting predicted values against actual observed values. The Bootstrap method enhances prediction and it suggests that the model effectively captures data patterns. Further validation can be conducted using error metrics like RMSE, MAE, and R^2 to quantify predictive performance.

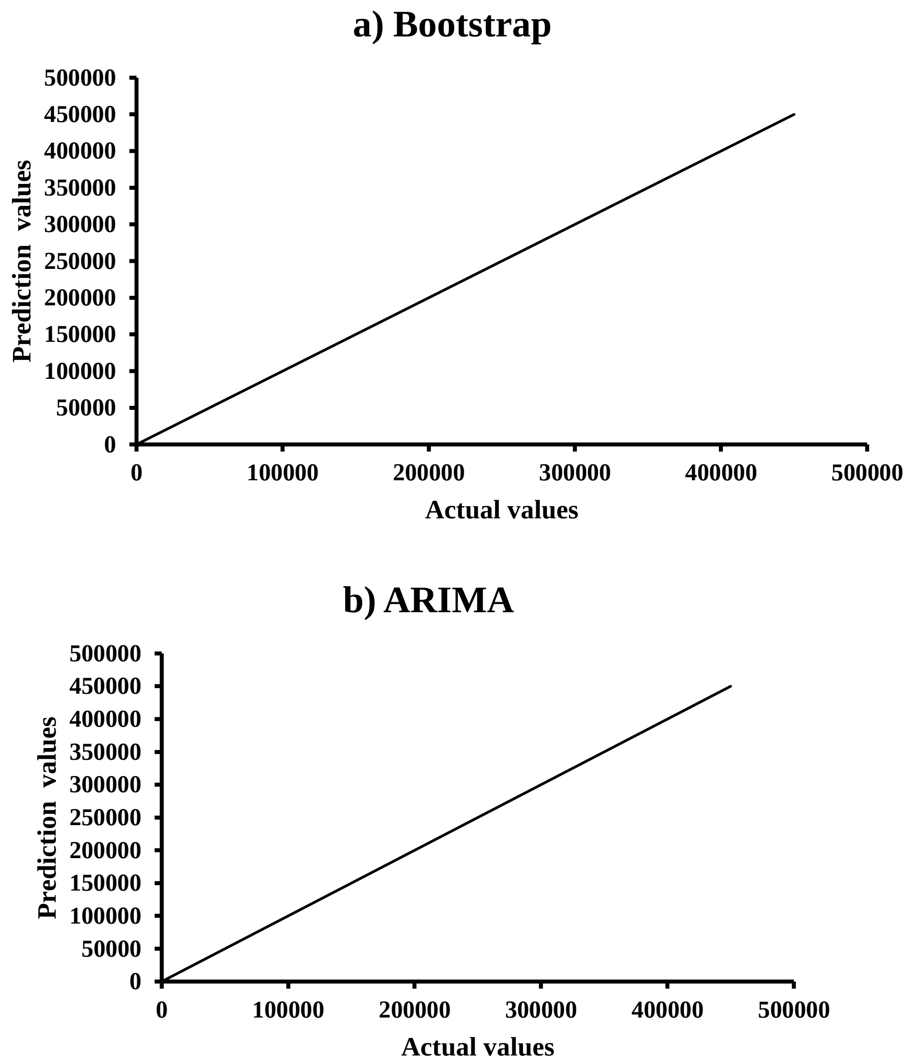

Figure 1.
Prediction vs. Actual Values for a) Bootstrap and b) ARIMA.

From the Table 6, it is evident that the Bootstrap method has lower values for both MAE and MAPE compared to the ARIMA (2,1,0) model. This indicates that the Bootstrap method yields predictions that are closer to the actual values than those made by the ARIMA model.

Table 6.
Comparison between bootstrap and ARIMA (2,1,0).

	Bootstrap	ARIMA(2,1,0)
MAE	0.5098	0.542
MAPE	0.2723	27.558

The Table 7 provides the forecast for wheat production in Iraq from 2021 to 2026, along with the Upper Control Limit (UCL) and Lower Control Limit (LCL) for each year.

Table 7.
Forecast for production of wheat in Iraq.

Model		25	26	27	28	29	30
Production in million tons-Model_1	Forecast	5.9	5.2	5.6	6.2	6.3	6.2
	UCL	11.5	12.8	14.6	17.1	18.7	20.1
	LCL	2.7	1.6	1.6	1.7	1.4	1.2

The Figure 2 displays observed wheat production in Iraq from 1997 to 2020, along with the forecasted wheat production for the period from 2021 to 2026, measured in million tons.

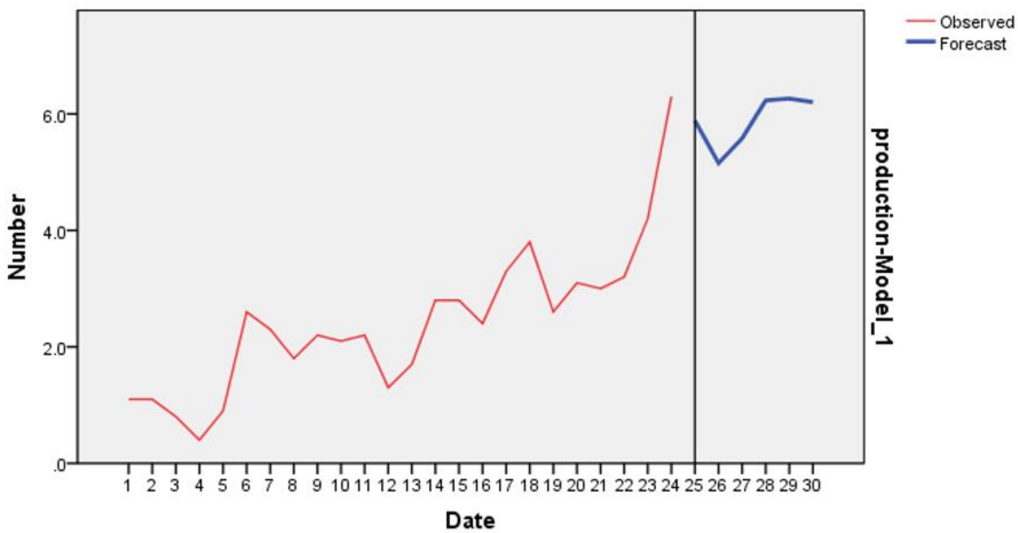

Figure 2.
Observed and forecast for production of wheat in Iraq.

4. Conclusion

A model describing the relationship between the amount of wheat production in Iraq and the production costs of this strategic crop, using the bootstrap method and the ARIMA time series method. Comparing the estimated equation of wheat production in Iraq using the bootstrap method and the estimated equation using the time series method. Both equations are important, and explain the 68% change in the amount of wheat production. We find that the bootstrap method is not able to resolve the presence or absence of a autocorrelation between the residuals (using the Darbin Watson test), while the time-series method ensures that there is no autocorrelation between the residuals. A forecast has been made for the quantity of wheat production in Iraq until 2026.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgments:

We would like to express our sincere gratitude to the Ministry of Agriculture - Agricultural Statistics Department for providing us with the agricultural data on wheat production, which greatly contributed to the completion of this research.

References

[1] P. Shetty, "Incorporating nutritional considerations when addressing food insecurity," *Food Security*, vol. 1, pp. 431-440, 2009. https://doi.org/10.1007/s12571-009-0039-6
[2] Y. Kusunose *et al.*, "Sustaining productivity gains in the face of climate change: A research agenda for US wheat," *Global Change Biology*, vol. 29, no. 4, pp. 926-934, 2023. https://doi.org/10.1111/gcb.16538
[3] R. G. Nwry, S. H. Abdulqader, and S. A. Hussain, "Effect of plant population and cultivars on growth, yield and its component of bread wheat (Triticum aestivum L.) under the rain-fed condition in Kurdistan-Iraq," *Tikrit Journal for Agricultural Sciences*, vol. 21, no. 3, pp. 41-51, 2021. https://doi.org/10.25130/tjas.21.3.5
[4] T. F. Abdul Rahman, K. J. Osamah, and N. J. Bilal, "Impact of improved seeds and some modern technologies on increasing the supply of wheat crop in Iraq for the agricultural season 2019-2020," *Iraqi Journal of Agricultural Sciences*, vol. 54, no. 1, pp. 176-188, 2023. https://doi.org/10.36103/ijas.v54i1.1689
[5] O. K. Abbas, F. I. Rahim, and M. K. Mohammed, "Economic analysis of the most important variables affecting the food security factors for the wheat crop in Iraq for the period 1995-2022," in *IOP Conference Series: Earth and Environmental Science*, 2024, vol. 1371, no. 6: IOP Publishing, p. 062013.
[6] L. A. Naji, A. A. Faisal, H. M. Rashid, M. Naushad, and T. Ahamad, "Environmental remediation of synthetic leachate produced from sanitary landfills using low-cost composite sorbent," *Environmental Technology & Innovation*, vol. 18, p. 100680, 2020. https://doi.org/10.1016/j.eti.2020.100680**:contentReference[oaicite:2]{index=2}
[7] L. A. Naji, S. H. Jassam, M. J. Yaseen, A. A. Faisal, and N. Al-Ansari, "Modification of Langmuir model for simulating initial pH and temperature effects on sorption process," *Separation Science and Technology*, vol. 55, no. 15, pp. 2729-2736, 2020. https://doi.org/10.1080/01496395.2019.1705067
[8] H. D. Alhassany, S. M. Abbas, M. Tostado-Véliz, D. Vera, S. Kamel, and F. Jurado, "Review of bioenergy potential from the agriculture sector in Iraq," *Energies*, vol. 15, no. 7, p. 2678, 2022. https://doi.org/10.3390/en15072678**:contentReference[oaicite:2]{index=2}
[9] J. Bai, S. H. Choi, and Y. Liao, "Feasible generalized least squares for panel data with cross-sectional and serial correlations," *Empirical Economics*, vol. 60, pp. 309-326, 2021. https://doi.org/10.1007/s00181-020-01977-2
[10] K. ArunKumar, D. V. Kalaga, C. M. S. Kumar, G. Chilkoor, M. Kawaji, and T. M. Brenza, "Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-regressive integrated moving average (ARIMA) and seasonal Auto-regressive integrated moving average (SARIMA)," *Applied Soft Computing*, vol. 103, p. 107161, 2021. https://doi.org/10.1016/j.asoc.2021.107161
[11] S. Nestler and S. Humberg, "A lasso and a regression tree mixed-effect model with random effects for the level, the residual variance, and the autocorrelation," *Psychometrika*, vol. 87, no. 2, pp. 506-532, 2022. https://doi.org/10.1007/s11336-021-09787-w
[12] A. Ares El Dean and M. Gad, *Statistical methods for economic and financial modeling*. Cairo: Cairo University Press, 2021.
[13] Satyanarayana, I. Beary, and K. A. Rao, "Empirical likelihood ratio test for autocorrelation in least squares regression," *Communications in Statistics-Simulation and Computation*, pp. 1-17, 2024. https://doi.org/10.1080/03610918.2024.2221234