Using Markov Chains to Predict Productivity of Maize in Iraq for the Period (2019-2025)

L.A.F.Al-Ani[1], & A.D.K. Alhiyali[1]

[1] Professor, Department of Agricultural Economics, College of Agricultural Engineering Sciences, University of Baghdad, Iraq

* Research is a Part of M.Sc. Thesis for the 1st author

Correspondence: A.D.K. Alhiyali, Department of Agricultural Economics, College of Agricultural Engineering Sciences, University of Baghdad, Iraq.

Abstract

The research aims to predict the productivity of one of the most important major crops in Iraq, which is Maize, using Markov chains, which is one of the most important predictive methods that depend on relatively recent historical data and based mainly on previous data that is not far away. This is the advantage that Markov chains have, as relying on somewhat old historical data may negatively affect the predicted values. The results of the research showed the superiority of the third state to predict the productivity of Maize depending on the availability of Markov chains prediction conditions for this state. The results of the research also showed the continued decline in productivity for the coming years, as well as the impact of the predictive values on changes in the cultivated area more than changes in production, which confirms the existence of horizontal expansion at the expense of vertical expansion, that is, there is no intensification of production per unit area. The research also found that the actual values of productivity have approached the estimated values of the following years, and the matter applies to the convergence of these results for the subsequent years with the previous years, which confirms the accuracy of the method of Markov chains, in other words that what happened in the recent past had a clear impact in the future near.

Keywords: transition matrix, Kolmgrove-Simrnov test, stochastic process

1. Introduction

The agricultural sector faces many problems, including the problem of low productivity for most crops, including the yellow corn crop(Maize), which casts a shadow on other aspects, such as self-sufficiency of this crop and endangering food security, in addition to the fact that productivity is closely related to the efficiency of resource utilization related to its production conditions, which makes the prediction process the future values are of great importance, which makes choosing one method without the other very important, especially since the predicted values will be relied upon in planning for the future and preparing the requirements for achieving the goals. It takes additional importance when forecasting the main crops, including the Maize crop, which has multiple uses, including human, animal and industrial use. The Markov method of forecasting is one of the most important prediction methods, especially since it does not depend on old historical data, that is, it relies on data that is not old to predict the future.

The research assumed a low productivity of the Maize crop in Iraq compared to global productivity, despite the increase in its values in recent years, but it remains less than what is recorded in the world, especially in cereal crops. The research also assumed that Markov chains are the most appropriate in estimating this type of data because they do not need much past historical data. The Markov chain method requires less stringent assumptions and provides more information than other methods such as regression.

The research aims to predict the productivity of maize crop in Iraq using Markov chains during the period 2019-2025.

2. Literature Review

Many researchers dealt with this topic in predicting, including Boalsbet (2015), who predicted wheat productivity in Algeria, as well as the researcher Ahmed D. A (2009) tried to predict the direction of the consumer price index numbers for basic commodities (foodstuffs-fabrics, clothing and shoes-fuel and lighting-transportation-medical services and

medicines-rent) using Markov chains as one of the statistical methods in forecasting the data at the present time. Transitional Probabilities Matrix with the greatest possibility method that was applied to monthly data taken from the Central Agency for Statistics and Information Technology in 2009).   The researchers, Kivslev & Amaid (1968) showed that the Markov chain model has different uses in economic applications.. The researcher Hollberry (1969) suggests in his article a method, based on multiple regression techniques, of replacing the constant transition probabilities with probabilities which are a function of various factors including structural characteristics in the industry. The results of this procedure compared quite favorably with those of the assumption of constant probabilities.   Matis and Others (1985) studied the Markov chain approach to predict crop yields, and concluded that the Markov chains approach requires less stringent assumptions and provides more information than other methods such as regression. Singh and Ibrahim (1996), in there study, multi date spectral data during crop growth period was used in Markov Chain Model to predict wheat yield. The results indicate that the use of spectral data near the maximum vegetative growth of wheat crop improves the efficiency and reliability of yield forecast about a month before its actual harvest.

Jasinthan. P and other (2015), targeted their search is to construct the Markov chain model for daily vegetable price movement in Jaffna. Two models are highlighted, where the price movement is considered as being in a state of (1) gain or loss and (2) large gain, or small gain or loss, or large loss. Twelve different types of vegetables are considered and above two models are used to analyze the price movement of each vegetable. These models were used to obtain transitional probabilities, steady state probabilities and mean recurrence times. The results indicate that the pattern of price movement of (eggplant) is similar to price movements of other vegetables, both in models (1) and (2). The investor is encouraged to invest in the vegetable market at any time in a way which leads to a greater chance of getting more gain than loss. Al-Ani and Al-Hiyali (2021) dealt with the use of Markov chains in predicting the productivity of the wheat crop for the period 2019-2025, and they reached the efficiency of using this method in prediction . Bairagi and Sarat (2017)   attempted have been made to study and predict the future market price of potatoes with the data collected from Lanka Regulated Market in the Nagaon District, Assam. Using the transition probability matrix and initial state vector the prediction for short period were made. The results obtained from the analysis were indistinguishable from real situation. Furthermore, considering the three possible stages of a particular day's price status in comparison to the previous day's price as 'increase', 'decrease' and 'remain same', the three-state transition matrix has been calculated and attempted to find out the long term behavior of the market price of potatoes. De Almeida Ferreira et al.(2020) concluded that Markov models are a special kind of stochastic process with a short-term memory of events. Hidden Markov chains are a special kind of Markov models that separate visible variables (the information on the production system) from the hidden variables (the harvest outcome). There are no assumptions to build a model other than some statistical relationship between the observed elements and the result. Additionally, Markov chains capture the temporal   process evolution in a very straightforward manner . Khawaldah, H.A. et al. (2020) concluded that The CA-Markov model has proved to be a powerful tool for analyzing LULC dynamic change and predicting future scenarios . Jain & Rangana Agrawal (1992), Ramasubramanian, V and Lalmohan Bahar (2014) and many researchers were using Markov chains method in their researches, and they found the efficiency of this method in prediction compared to other prediction methods.

3. Conceptual Framework

Markov chain

A Markov process is a stochastic process that satisfies the Markov property (Gagniuc 2017). (sometimes characterized as "memory lessness"). In simpler terms, it is a process for which predictions can be made regarding future outcomes based solely on its present state and—most importantly—such predictions are just as good as the ones that could be made knowing the process's full history, In other words, conditional on the present state of the system, its future and past states are independent (Oksendal, B. K. (Bernt Karsten) (2003). Markov chains are an essential part of stochastic processes, and they are widely used in many different disciplines. It is a stochastic process that satisfies a Markov property, which means that the past and future are independent when the present is known. This means that if a person knows the current state of the process, no additional information about their past states will be needed to make a better prediction in the future. This property allows to significantly reduce the number of parameters when studying such a process (Myers et al n.d). There are different concepts regarding Markov chains.   When approaching Markov chains there are two different types; discrete-time Markov chains and continuous-time Markov chains. This means that we have one case where the changes happen at specific states and one where the changes are continuous. In our report we will mostly focus on discrete-time Markov chains. One example to explain the discrete-time Markov chain is the price of an asset where the value is registered only at the end of the day. The value of the Markov chain in discrete-time is called the state and in this case the state corresponds to the closing price. A continuous-time Markov chain changes at any time. This can be explained with any example where the measured events happens at a continuous time and lacks "steps" in its appearance. One well known example of continuous-time Markov chain is the poisson process, which is often practised in queuing theory. For a finite Markov chain the state space S is usually given by $S = \{1, \ldots, M\}$ and

the countably infinite state Markov chain state space usually is taken to be *S = {0, 1, 2, . . . }*. These different variances differ in some ways that will not be referred to in this paper. A Markov chain can be stationary and therefore be independent of the initial state in the process. This phenomenon is also called a steady-state Markov chain and we will see this outcome in the example of market trends later on, where the probabilities for different outcomes converge to a certain value. However, an infinite-state Markov chain does not have to be steady state, but a steady-state Markov chain must be time-homogenous. Which by definition means that the transition probabilities matrix *Pi, j (n, n + 1)* is independent of n (Mayer et al n.d)

As for Al-Atom, he defined Markov chains as a method by which past changes and fluctuations are analyzed in order to predict future changes and take appropriate decisions. Statistically speaking, the Markov chain is a stochastic process that is described by a number of states, and any random process is defined as a Markov chain if the following two conditions are met(Al-Atom, 2006): first: System states change from stage to stage, Second: The probability that a system will be in a given state at some point depends at most on the previous stage. The possibilities of movement from one state to another are called Transition Probabilities, and they are represented by a matrix called the Transition Matrix, and the outputs of the process are described by a matrix called the revenue or cost matrix, and the elements of this matrix are returns or costs resulting from the transition from one state to another.

Markov chain assumptions

What makes stochastic processes are called Markov chains $\{X_n : n \in T\}$ Without other processes, this is what

was mentioned in the following hypotheses, which characterized Markov chains (Tag and Sarhan 2007):

1- The state space of this process is discrete (discrete state)

2- The parameter space (T) for this process is discrete (time-discrete) i.e. $\{X_n : n = 1,2,....\}$

3- This process checks the Markov property:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1,....,} X_1 = i_1) = P(X_{n+1} = j | X_n = i\|$$

because:

$X_n :$ Transitional Probability Matrix.

$X_{n+1} :$ Represents the value of the phenomenon in the subsequent period.

$P_{ij} :$ It represents the probability of a phenomenon moving from state i to state j .

It should be noted that the Markov chains method involves the following calculation (Tohme   2010):

1- Matrix of Transition Probabilities

2- Prediction Future shares

3- Steady- State Probabilities

The most important part will be addressed, which is the matrix of transitional probabilities (MTP). The probabilities of moving from one state to another during a certain period of time are called transitional probabilities and can be represented by a matrix called the transition matrix or Markov matrix (Al-Atom 2006).

The transition probabilities $P_{ij}, i, j = 0,1,2,3,.....$ are summed up in the form of a matrix, as follows

(Greenwell and et.al 2003)

$$P = \begin{array}{c} \\ i{=}0 \\ i{=}1 \\ i{=}2 \\ \\ \\ \end{array} \begin{pmatrix} \overset{j=0}{P_{00}} & \overset{j=1}{P_{01}} & \overset{j=2}{P_{02}} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ P_{20} & P_{21} & P_{22} & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{pmatrix}$$

It is called the Transition Probability Matrix of the mentioned Markov chain, as the matrix is divided into double rows and columns, meaning that the number of rows and the number of columns are equal. The symbol i represents the rows, while the symbol j represents the columns, and the element that is arranged (i, j) represents the probability (P_{ij}) which is the probability of a random process moving from state i to state j in one step during a specified period of time. The following figure shows the possibility of moving from one state to another (Scheepers,H 2019
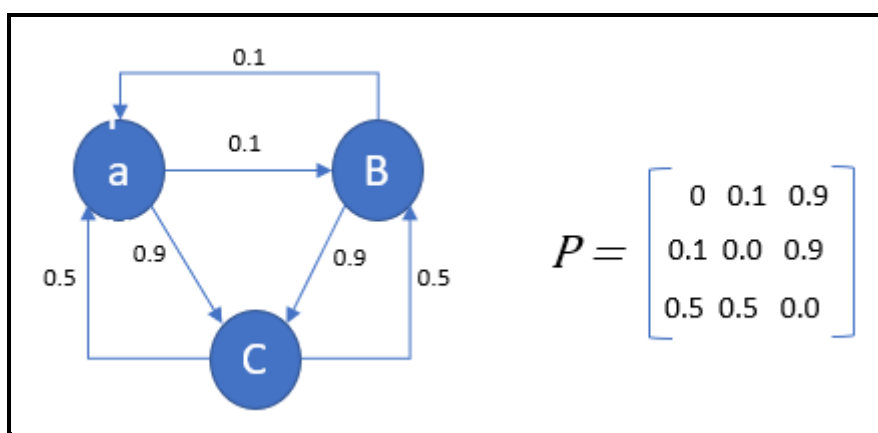

Figure 1. Possibility of moving from one state to another

Productivity

Productivity is an economic concept of very important scientific and practical value, as it is a general indicator of economic and social progress in general, and agricultural progress in particular. Productivity means the criterion by which it is possible to measure the good use of productive resources, and in light of this it is possible to determine and evaluate the degree of benefit from directing resources to reach the targeted results (Hammoud and Fakhoury 2008).

 Productivity means achieving the largest proportion of output from a given value of input. It can be defined (as an indicator that shows the ability of the various elements of production to achieve a certain level of output, compared to the inputs that were invested for the productive purpose) (Al-Hassan 2015). That the process of economic and social development and increasing productivity aims to achieve the optimal use of available economic resources, so the relationship between them is very close, as we can say that achieving economic development can be completed in multiple concurrent and organically interrelated ways, including the horizontal expansion (Horizontal Expansion) in production, i.e. Increasing the number of economic units by adding new capital components and expanding the available ones and increasing the number of workers in them, including the vertical expansion in production, by continuous improvement, in the level and degree of exploitation of economic resources, by raising the productivity of the resources used in the existing economic units (Al-Hasani   2018).

4. Data and Methodology

This research is based on the data of the time series of Maize productivity in Iraq during the period (2000-2018) taken from the Iraqi Ministry of Agriculture - Department of Agricultural Economics will be relied on and then predicted using Markov chains. The use of Markov chains in predicting the yield of maize goes through the following stages (Alani and Al-hiyali 2021):

1- After preparing the data for the phenomenon that we are going to predict its future path, we first divide it into certain levels, after we subtract the smallest value of the phenomenon R_{Min} from its largest value R_{Max} (range), then divide the result of the subtraction process by the number of levels specified previously:

$$Output = \frac{R_{Max} - R_{Min}}{number\ of\ specific\ states}$$

(Note: All possible cases are calculated and then the state is selected based on the passing of the approved statistical tests)

2- Then we configure the levels according to the number of specific cases, suppose we have 4 states:

States	First Limits	The second Limits
Stat1	R_{Min}	$Y1 = (R_{Min} + \dfrac{Range}{4})$
Stat2	$Y1$	$Y2 = (Y1 + \dfrac{Range}{4})$
Stat3	$Y2$	$Y3 = (Y2 + \dfrac{Range}{4})$
Stat4	$Y3$	$R_{Max} = (Y3 + \dfrac{Range}{4})$

3- The third stage: define the transition matrix, each element in this matrix expresses the probability that the phenomenon will move from one level to another provided that the sum of any row of the transition matrix is equal to the correct one. The following table indicates the transition matrix if the number of levels is 4

Table 1. Transition matrix if the number of levels is 4

	S1	S2	S3	S4	Total
S1	P_{11}	P_{12}	P_{13}	P_{14}	1
S2	P_{21}	P_{22}	P_{23}	P_{24}	1
S3	P_{31}	P_{32}	P_{33}	P_{34}	1
S4	P_{41}	P_{42}	P_{43}	P_{44}	1

4- Productivity values are averaged at each of the four levels, as in the following table:

Stat1	Stat2	Stat3	Stat4
$\sum Y1/N1$ Number of values for the first case	$\sum Y2/N2$ Number of values for the second case	$\sum Y3/N3$ Number of values for the third case	$\sum Y4/N4$ Number of values for the fourth case

5- A vector is formed whose elements are in the number of levels specified in the first step, provided that all of them are equal to zero except for one element that is equal to the right one and its position in the line is corresponding to the level in which the last value of the phenomenon is located. If the last value is in the fourth level, for example, the vector is written as follows:

Stat1	Stat2	Stat3	Stat4
0	0	0	1

6- Then this vector is multiplied by the transition matrix to get a new vector and the latter is multiplied by the transition matrix to get a new line vector. Then the new vector is multiplied by the averages computed in the fourth step, and we get the expected value of the phenomenon in the next year. By repeating the last two phases on the last vector, we get the values of the phenomenon in subsequent years.

Table 2 indicates the yield values of Maize in Iraq during the period 2000-2018

Table 2. Values of Maize Productivity

In Iraq during the period (2000-2018)

Year	Productivity Kg/Donum[*]
2000	558.9
2001	588.0
2002	786.2
2003	647.7
2004	562.2
2005	577.5
2006	606.7
2007	619.8
2008	587.2
2009	521.6
2010	570.1
2011	647.6
2012	830.9
2013	1041.7
2014	765.1
2015	796.1
2016	853.9
2017	831.7
2018	1134.4

Source: Iraqi ministry of agriculture/ department of agricultural economics

* Donum= 2500 m^2

5. Results and Discussion

Table (3) indicates that the predicted values of productivity for the third case during the mentioned period were relied upon after passing a set of aforementioned statistical tests, as it was found that the predicted values began to decline starting in 2019 and then stabilized at the level of 769.1 in 2025 and that the reason We relied on these values as they gave the lowest value for the MSE predictive accuracy measure compared to other cases. It should be noted that the predictive values of Maize yield for the fifth and sixth cases have been excluded because they violate one of the conditions of the transition matrix.

Table 3. The results of forecasting the productivity of maize crop in Iraq during the period (2019-2025) and according to the cases of Markov chains

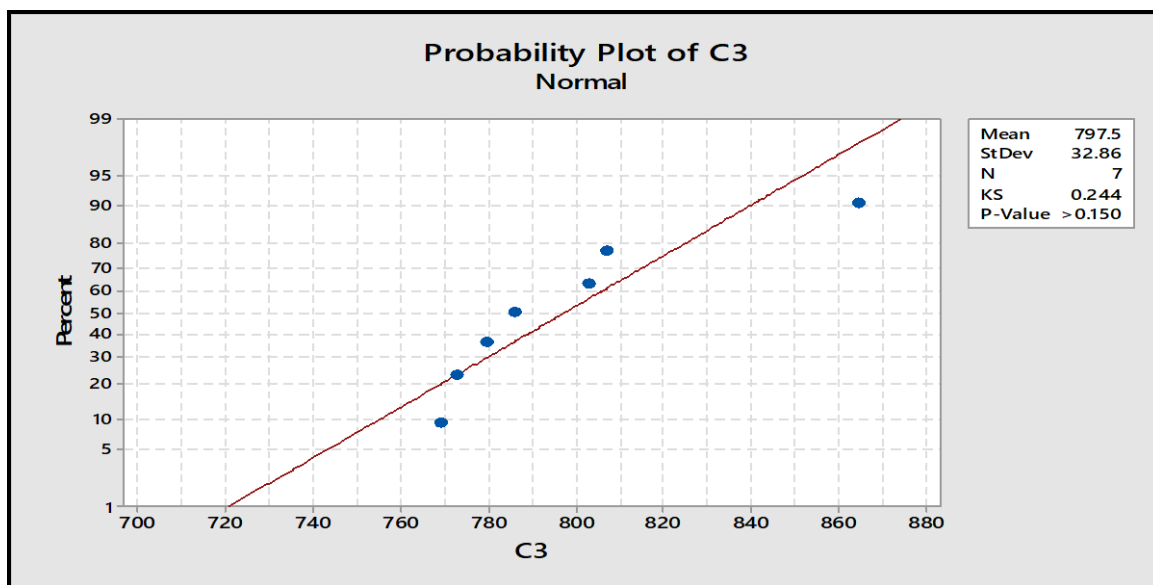| Year	Second state	Third state (selected)	Fourth state
2019	836.4	864.6	736.3
2020	809.7	806.9	793.7
2021	792.0	803.0	775.1
2022	780.3	786.1	751.3
2023	772.6	779.6	752.9
2024	767.5	773.0	753.8
2025	764.2	769.1	749.9
KS[*]	0.15	0.15	0.03

* K.S: Kolmogorov-Smirnov test

Figure 2. Normal distribution of maize yield data during the period (2019-2025) according to K-S test for the third case

It is clear from Figure (2) of the results of the Kolmgrove-Simrnov test that the P-value is greater than 0.05 level of significance, which supports the validity of the assumption that the data follow a normal distribution. It is also clear from the figure that most of the data are located on the straight line and very close to it, and this in turn agrees with the results of the KS test, which confirms the nature of the data.

Table (4) indicates that the predicted values of productivity have recorded a decrease, as the production value for the year 2019 recorded a remarkable decrease, amounting to about 864.6 kg / donum, and this is due to the fact that the changes in area were greater than changes in production, which clearly affected the predicted productivity rates and here we clearly conclude that the conditions surrounding the Maize crop, especially production, did not allow for a positive impact on productivity values, and the decline in productivity continued until it stabilized at the level of 769.1.

Table 4. Predicted values for each of the productivity, area and production of Maize in Iraq for the period (2019-2025)

year	Area 1000 Donum Fifth state	Δ%	Production 1000 Ton Third state	Δ%	Productivity Kg/Donum Third state	Δ%
2018	55.8		63.3		1134.4	
2019	252.8	353.05%	346.0	446.60%	864.6	-23.78%
2020	321.3	27.10%	356.9	3.15%	806.9	-6.67%
2021	346.9	7.97%	362.7	1.63%	803.0	-0.48%
2022	375.5	8.24%	366.8	1.13%	786.1	-2.10%
2023	394.1	4.95%	370.4	0.98%	779.6	-0.83%
2024	409.2	3.83%	373.7	0.89%	773.0	-0.85%
2025	420.3	2.71%	377.0	0.88%	769.1	-0.50%

6. Summery and Conclusion

The research hypothesis has been proven regarding the low productivity of the Maize crop in Iraq compared to the productivity at the level of the world and the Arab world. In addition to, The decrease in the areas planted with the Maize crop during certain years of the studied series, and this means that the crop under consideration is affected by the set of decisions and legislation issued from time to time, which is reflected negatively or positively according to the objectives set for the decisions and legislation. Also, The changes in the areas had the greatest impact on the productivity of the crops under consideration, which confirms the fact that there is a horizontal expansion at the expense

of the vertical expansion, that is, there is no intensification of production per unit area, and this has caused a clear decrease in the productivity of the crop under discussion.

The most important conclusions that came out of the research in the statistical aspect are, the research hypothesis of the low productivity of the yellow maize crop was also proven. The statistical analysis proved the research hypothesis that the predictive values of the selected cases confirmed a significant decrease in productivity. From the above it is shown that the economic conclusion agrees with the statistical one. Also, the results of the research confirmed that Markov chains do not need old historical data, which did not constitute a major obstacle in interpreting the prediction results, as it is common knowledge that future values, especially in agriculture, are affected by close previous data, which gives logical explanations for what will happen in the future depending on what happened in recent years, and this conclusion is supported by the existence of mathematical models used by researchers in economic analysis for the purpose of interpreting events in the future, and the most famous of these models is the dynamic Nerlove model, which is based on the idea that the variable to be estimated is affected by variables of previous years, such as $t_{-1}, t_{-2}, \ldots$. Finally, The results of the research proved the convergence of the actual values of the productivity of the four main crops with their estimated values for the following year, and the matter was applied to the convergence of these results for the subsequent years with the previous years, which confirms the high accuracy that characterizes Markov chains, in addition to the fact that the characteristics of these chains are affected by their influence in recent years. A positive result in obtaining these results, meaning that what happened in the recent past had the greatest impact in the near future.

References

Ahmed, D. A. (2009). Using Markov Chains to Predict Consumer Price Index Numbers in Iraq. *The first conference of the College of Science, University of Diyala, 6*(1), 1-18

ALani, L. A. F., & Alhiyali, A. D. K. (2021). Forecasting Wheat Productivity in Iraq for the period 2019-2025 Using Markov Chains. *Iraqi Journal of Agricultural Sciences, 52*(2), 411-421. https://doi.org/10.36103/ijas.v52i2.1302

Al-Atom, S. (2006). *Operations Research.* Amman, Jordan: Dar Al-Manahej for Publishing and Distribution.

Al-Hasani, F. A. T. (2018). *An Economic Analysis of the Determinants of Agricultural Growth in Iraq for the period 1980-2015.* Master's Thesis, coll., of Agriculture., Unvi. Of Baghdad.

Al-Hassan, S. A. N. (2015). *An Economic Analysis of the Productivity of Wheat Crop in Reclaimed and non-Reclaimed lands in light of Risk and Uncertainty in Iraq.* PhD Dissertation , coll., of Agriculture., Unvi. Of Baghdad.

Bairagi. A., & Sarat, C. K. (2017). Markov Chain Modeling for Prediction on Future Market Price of Potatoes with Special Reference to Nagaon District. *Journal of Business and Management, 19*(12), 25-31.

Boalsbet, A. Q. (2015). Using Markov Chains in Predicting Wheat Production in Algeria. *Journal of Human Sciences, 43,* 171-183.

De Almeida Ferreira, J. S., Ferreira, A. P. L., & Perez, N. B. (2020). A hidden Markov chain approach to crop yield forecasting. *IADIS International Journal on Computer Science & Information Systems, 15*(2), 148-160.

Gagniuc, P. A. (2017). *Markov Chains from Theory to Implementation and Experimentation.* NewDalhi, India: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119387596

Greenwell, R. N., Nathan, P. R., & Lial, M. (2014). *Calculus with Applications for the Life Sciences: eighth edition.* Canada : Pearson New International Edition.

Hammoud, K. K., & Fakhoury, H. Y. (2008). *Production and Operations Department.* Amman, Jordan: Dar Al-Safa Publishing.

Hollberry, M. C. (1969). Projecting the Size Distribution of Agricultural Firms—An Application of a Markov Process with Non-Stationary Transition Probabilities. *American Journal of Agricultural Economics, 51*(2), 289-302. https://doi.org/10.2307/1237580

Iraqi Ministry of Agriculture - Department of Agricultural Economics. Data for different years.

Jain, R. C., & Rangana, A. (1992). Crop yield probability model. *Biometric Journal, 34*(4), 501-511. https://doi.org/10.1002/bimj.4710340410

Jasinthan, P., Laheetharan, A., & Satkunanathan, N. (2015). A Markov Chain Model for Vegetable Price Movement in Jaffina. *Sri Lankan Journal of Applied Statistics, 16*(2), 93. https://doi.org/10.4038/sljastats.v16i2.7825

Khawaldah, H. A., Farhan, I., & Alzboun, N. M. (2020). Simulation and prediction of land use and land cover change using GIS, remote sensing and CA-Markov model. *Global Journal of Environmental Science and Management, 6*(2), 215-232.

Kivslev, Y., & Amaid, A. (1968). Linear and Dynamic Programming in Markov Chains. *American Journal of Agricultural Economics, 50*(1), 111-129. https://doi.org/10.2307/1237875

Matis, J. H., Satio, T., Grant, W. E., Jwig, W. C., & Ritcgie, J. T. (1985). A Markov Chain Approach to Crop Yield Forecasting. *Journal of Agricultural Systems, 18*(3),171-187. https://doi.org/10.1016/0308-521X(85)90030-7

Myers, D. S., Wallin, L., & Wikstrom, P. C. (n.d.) . An introduction to Markov Chains and their Applications Within Finance. [www page]. Retrieved from http://www.math.chalmers.se/.html

Oksendal, B. K. (Bernt Karsten), 1945-. (2003). *Stochastic differential equations : an introduction with applications.* Berlin ; New York :Springer,

Ramasubramanian, V., & Laimohan, B. (2014). Crop Yield forecasting by Markov Chain models and simulation. *Journal of Statistics and Applications, 12*(1&2),1-13.

Scheepers, H. (2019). *Markov Chain Analysis and Simulation using Python.* [www page]. Retrieved from https://towardsdatascience.com/.html

Singh, R., & Ibrahim, A. E. (1996). Use of spectral data in Markov Chain model for crop yield forecasting. *Journal of the Indian Society of Remote Sensing, 24,* 145-152. https://doi.org/10.1007/BF03007327

Tag, L., & Sarhan, A. M. (2007). *Introduction to Stochastic Processes.* Riyadh, Kingdom of Saudi Arabia: Scientific Publishing and Printing Press.

Tohma, H. (2010). *Decision-Making Theory.* Amman, Jordan: Dar Al-Safa Publishing and Distribution.