International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 9 Issue 8 August - 2025, Pages: 120-128

Using the box-Jenkins methodology to predict the number of divorces

Maryam Mehdi and[1] Nibras Talib Mohammed[2]

1Karbala University / Faculty of administration and Economics / Department of Statistics, mariam.m@uokerbala.edu.iq

2Karbala University / Faculty of administration and Economics / Department of Statistics, nbrass.t@uokerbala.edu.iq

Abstract: *The box-Jenkins methodology was applied to analyze time series and predict the number of monthly divorces in Karbala governorate for the period (2025-2024) using Karbala court data (2020-2024). The analysis showed that the time series is unstable, and stability was achieved after applying the second-order divergence. By analyzing the functions of autocorrelation (ACF), partial autocorrelation (PACF) and information parameters (AIC, BIC, HQ), the ARIMA model(3,1,3) was identified as the best predictive model. The results showed the efficiency of the model through Leung-box tests and residue analysis, where the annual expected value of divorces in 2025 was about 410 cases with a confidence level of 95%. The study found an upward trend in divorce cases and recommended the establishment of family counseling centers and the adoption of social policies to reduce this phenomenon.*

Keywords: Box-Jenkins, ARIMA, forecasting, Divorce cases

1. Introduction

Time series analysis is one of the most important statistical methods for forecasting, widely used in statistical and economic applications. This method predicts future variable changes based solely on the variable's historical behavior. In other words, time series models consider past patterns of change in a variable and use this information to predict its future changes, making them advanced and effective forecasting tools.

The **Box-Jenkins methodology** is among the most important approaches used for time series forecasting. It differs from other forecasting methods as it doesn't assume any predetermined pattern in historical data. Instead, the appropriate model is selected by comparing the time series' autocorrelation coefficients with theoretical distributions of different models. A model is considered good if its residuals—the differences between predicted values and historical data—are small, normally distributed, and independent.

Building a forecasting model using the Box-Jenkins methodology consists of four main stages: (Box, G. E. P., Jenkins, G. M., & Reinsel, G. C.,2008)

1. **Model Identification:** Selecting a mathematical model based on statistical metrics and research expertise.
2. **Model Estimation:** After selecting the model, its parameters are estimated using available data and specialized statistical methods.

2. Model Diagnostics and Forecasting

2.1 Model Diagnostics:

The model is tested for compatibility with observed data. If it passes statistical tests (e.g., residual independence and normal distribution), it's ready for forecasting. If it fails, the process returns to the model identification stage to select a new model.

2.2 Forecasting:

The final model is used to generate future predictions. As new data emerges, forecasting errors are calculated and monitored to ensure model accuracy. Figure (1) illustrates the key stages of building time series models using the Box-Jenkins methodology (Hyndman, R. J., & Athanasopoulos, G.,2021).

Advantages of Box-Jenkins Methodology and ARIMA Models

Advantages of Box-Jenkins Methodology (Shawrawi, 2004)**:**

1. A comprehensive and reliable system for time series modeling, covering all stages from model selection to forecasting.

2. Doesn't assume data independence; instead, it leverages autocorrelation patterns through **ARMA** models, providing accurate forecasts.
3. Delivers more precise predictions than other methods when sufficient data is available.
4. Provides reliable confidence intervals for predictions, whether for seasonal or non-seasonal data.

3. ARIMA Models (Autoregressive Integrated Moving Average):

In practical applications, most time series are non-stationary. To convert them into stationary series, **differencing** of order **d** is applied. These models are known as **ARIMA(p, d, q)**, expressed by the equation: (Dickey, D. A., & Fuller, W. A. ,1979).

$$wt = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \delta + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}$$

Where:

- $w_t = \Delta^d yt$ (d-th order difference of the original series yt).
- $\Delta yt = yt - yt - 1$ (first-order difference).

Using the **backshift operator (B)**, the equation simplifies to:

$$\phi(B)\Delta^d yt = \delta + \theta(B)\epsilon t$$

Where:

- $\phi(B) = 1 - \phi 1B - \phi 2B^2 - \cdots - \phi pB^p$ $(AR\ component)$.
- $\theta(B) = 1 - \theta 1B - \theta 2B^2 - \cdots - \theta qB^q$ $(MA\ component)$.

The values of **p** and **q** are determined based on autocorrelation (ACF) and partial autocorrelation (PACF) coefficients.

3.1 Seasonal Models (SARIMA)

Time series (e.g., monthly or quarterly data) often contain **seasonal patterns**. To model these, **SARIMA(p, d, q)(P, D, Q)s** is used, where: (Box, G. E. P., & Pierce, D. A. ,1970), (Chatfield, C. ,2003)

- **s**: Seasonal period (e.g., 12 for monthly data).
- **P, D, Q**: Seasonal coefficients for the AR, differencing, and MA components.

The SARIMA equation:

$$\Phi(B^s)\phi(B)\nabla_S^D \nabla^d yt = \Theta(B^s)\theta(B)\epsilon t$$

Where:

- $\nabla^d yt$: Non-seasonal differencing.
- $\nabla_S^D yt$: Seasonal differencing.

Applied aspect

In this chapter, Time Series models will be based on box Jenkins models and its prediction of the number of divorces in the Holy province of Karbala.

The data for the research was obtained by the (Karbala court) from the period (2020-2024), where the time series data included the preparation of divorce cases by months for the period from January 2020 to December 2024 with (48) views as shown in the table(1).

Table (1) preparation of divorce cases by months in the Holy province of Karbala

202	137	164	202	202
202	45	236	202	202
202	372	305	202	13
202	264	107	202	72
236	246	109	202	432

33	192	269	202	244
74	208	251	401	450
91	159	252	302	146
202	344	233	72	
226	204	190	1	

1. Analysis of time series data

The time series data representing the number of y_t divorce cases in the Holy province of Karbala were analyzed using box Jenkins models and relying on the ready-made program (SPSS) to obtain the results .

2. Box-Jenkins models for divorce numbers Y_t

The box-Jenkins method will be applied to the data of divorce cases, where the stability of the series is examined, the rank of the model is determined, the parameters of the model are estimated, its morale is tested, as well as the morale of the model is tested, and then compared with the rest of the models in terms of preference in order to use the best model for forecasting.

2.1 Chain stability

After collecting the data, which is the first stage of the box and Jenkins methodology, we draw a data series representing the preparation of divorce cases in the Holy province of Karbala to identify the behavior of the series and its initial characteristics, and Figure (1) represents the drawing of the series.

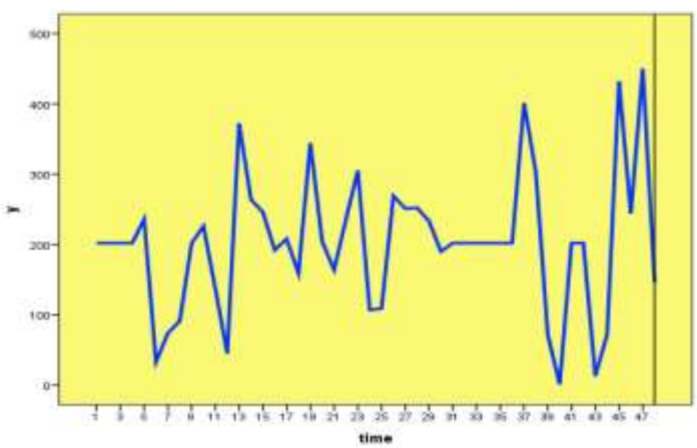

Figure (1) shows the series of divorces Y_t

Through Figure (1) we note that the vertical axis represents the values of the numbers of divorces and the horizontal axis represents the months and we note the stability of the time series as it expresses the general trend with time, and for more accuracy we draw both the autocorrelation function and the partial autocorrelation function, as in Figure (2) :

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 9 Issue 8 August - 2025, Pages: 120-128

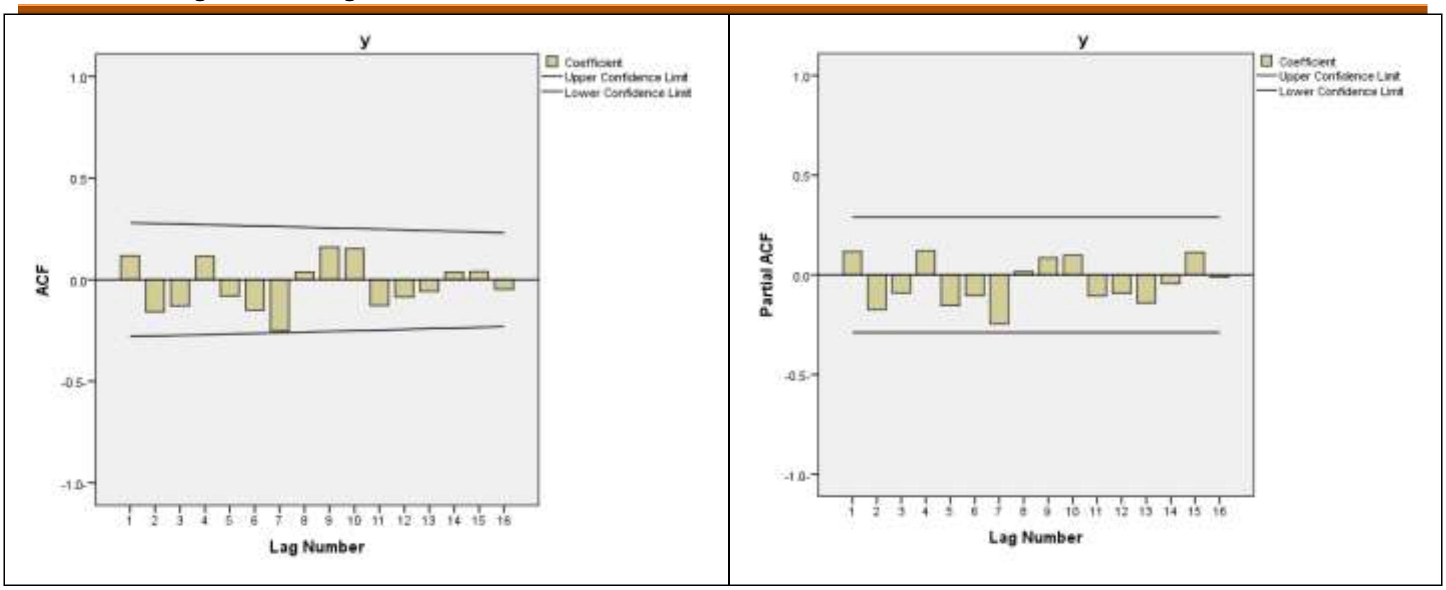

Figure (2) shows the graph of the autocorrelation function ACF and the partial autocorrelation PACF for divorces y_t

We note from **Figure (2)** that all the coefficients of the ACF autocorrelation function are within the confidence limits at the level of 95%, as well as the coefficients of partial autocorrelation, and this is an indicator of the stability of the series .**Table (2)** shows the values of the autocorrelation function and the partial autocorrelation function for the series of divorces.

Table (2) shows the ACF and PACF values for divorces

Lag	ACF	PACF
1	.149	.149
2	-.153	-.179
3	-.101	-.050
4	.099	.103
5	-.103	-.173
6	-.150	-.085
7	-.255	-.260
8	.012	.021
9	.144	.077
10	.164	.108
11	-.104	-.105
12	-.087	-.101
13	-.058	-.135
14	.036	-.038
15	.039	.111
16	-.049	-.017

We note from Table (2) that most of the values of the functions of autocorrelation and partial autocorrelation are accelerated after the first offset in the sense that $|ACF\,| > \left|{}^{1.96}/_{\sqrt{n}}\right|$ and $|PACF\,| > \left|{}^{1.96}/_{\sqrt{n}}\right|$ to be sure, we use the extended Dickey - Fuller unit root Test (Unit Root Test), which is considered the most accurate criterion for testing the stability of the chain, and the following table (3) shows the extended Dickey-Fuller test :

Table (3) shows the results of the extended Dickie Fuller test for divorces Y_t

Augmented Dickey-Fuller test for y(t)	with constant and trend	with constant	NO constant and NO trend
estimated value	-0.900910	-0.882007	-0.173964
test statistic	-5.964683	-5.937305	-2.120617
p-value	0.0001	0.0000	0.039400

We note from Tables (3) that the values of (p-value) are less than the moral level of 0.05, which calls us to reject the hypothesis of nothingness, that is, the absence of unit roots and the time series is stable .

2.2 Determining the rank of the model

After testing the values of the time series for the numbers of divorce cases and achieving stability in the series, we diagnose the appropriate model for representing the time series by studying and comparing the theoretical behavior of the functions (ACF), (PACF) and Table (2) shows the behavior of the functions of self-correlation and partial self-correlation in determining the model, noting that the appropriate model is the mixed model ARIMA and more precisely in determining the rank of the model, a number of moral models are reconciled and the best model is chosen based on statistical criteria Akiki Information Standard (AIC) and the Hanan-Quinn criterion (H-Q) and the Schwartz criterion (SBC), the proposed models are as shown in Table (4).

Table (4) shows the proposed Box Jenkins models with significant estimates

The model	AIC	H-Q	SBC
ARIMA(1,1,1)	1462.553	1464.688	1467.823
ARIMA(1,1,0)	1454.048	1457.25	1461.953
ARIMA(2,1 ,2)	1419.973	1426.376	1435.781
ARIMA(2,1 ,3)	1448.717	1452.986	1459.256
ARIMA(3,1,0)	1443.625	1448.961	1456.799
ARIMA(3,1,3)	1416.619	1424.089	1435.062

Through Table 4, we came to the appropriate model for the series of preparation of divorce cases y_t ,which is the mixed model after taking the second difference of the series $y_t$and the model HUARIMA(3,1,3), which is the model proposed according to the above criteria, where their values in the proposed model are lower compared to other proposed models with significant parameters .

2.3 Estimation of model parameters and testing their morale

After determining the rank of the model, the next step comes from the stages of building the time series model according to the box Jenkins method, which is to estimate the parameters of the model and test its validity using the program (Gretl 1.9.11), the method of the exact greatest possibility was applied to estimate the parameters of the proposed model, as in Table (5).

Table (5) shows the estimation of the parameters of the diagnosed model with its morale test ($p \le 0.05$)

parameters	coefficient	std. error	Z	p-value
phi_1	-0.567974	0.0983785	-5.773	7.77 E-09 ***
phi_2	-0.994264	0.0568991	-17.47	2.25E-068 ***
phi_3	-0.301098	0.0981223	-3.069	0.0022 ***

theta_1	-0.622946	0.0492217	-12.66	1.04E-036 ***
theta_2	0.622946	0.0533225	11.68	1.56E-031 ***
theta_3	-1.00000	0.0548669	-18.23	3.21E-074 ***

We note the values of (p-value) and for all the parameters of the model are less than 0.05, so the model is significant and acceptable, and the formula of the model ARIMA (3,1,3) is as follows :

$$Y_t = -0.567974\ Y_{t-1} - 0.994264\, Y_{t-2} - 0.301098\, Y_{t-3} + a_t + 0.622946\, a_{t-1} - 0.622946\, a_{t-2} + a_{t-3}$$

2.4 Checking the suitability of the model

After diagnosing the model, determining its rank and estimating its parameters, it is necessary to verify the correctness of the suitability of the model and its efficiency, and this is done through :

1. Ljung - Box Test Test

When applying the Ljung-Box statistic to examine the suitability of the model, where the value of (Q=22.8787) appears at the offset (k=34), which is less than the value of χ_((13, 0.05))^2 Tabular (22.6333), this means accepting the hypothesis of nothingness and that the errors are not related to each other, that is, the model is good, appropriate and efficient .

$H_\circ: \rho_1 = \rho_2 = \cdots = \rho_K = 0$

$H_1: \rho_i \neq \rho_L \quad ; \quad for\ i \neq L$

2. Testing the vesicles

The coefficients of the autocorrelation (ACF) and the partial autocorrelation (PACF) of the residuals of the estimated model were extracted and drawn, and we note in Figure (3) that all the values of the (ACF) and (PACF) coefficients fall within the confidence limits, which means that the sequence of residuals is random and the model used is good and appropriate .

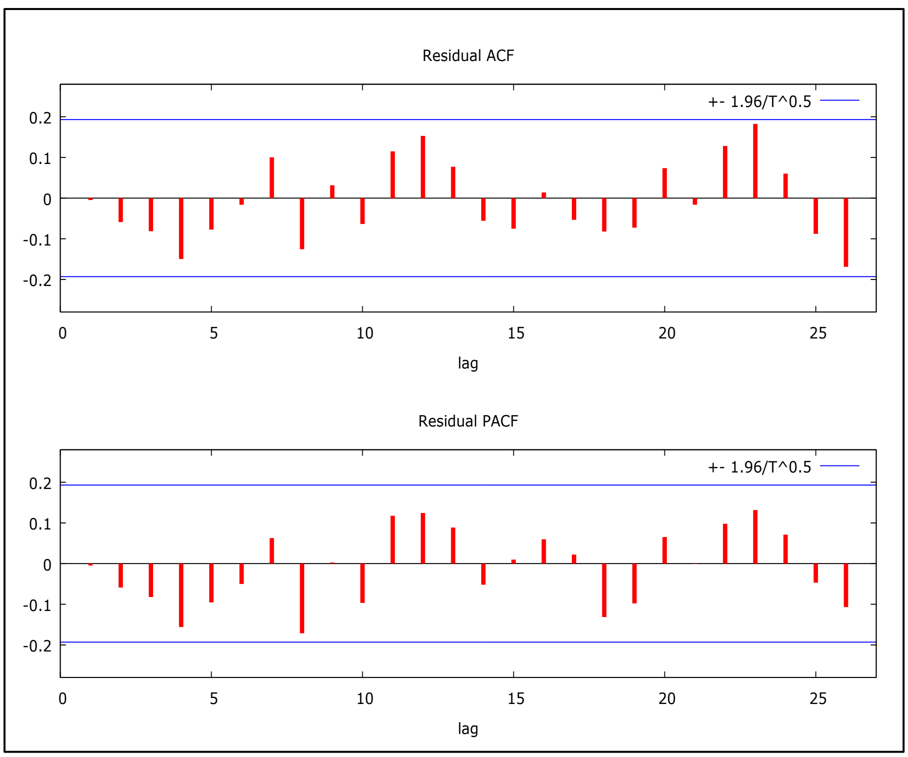

Figure (3) shows the diagram of the functions of autocorrelation ACF and partial autocorrelation PACF for the remainder

And Table (6) shows the values of the autocorrelation function and the partial autocorrelation function for the remainder, and we note that all the values of both functions fall within the confidence limit.

Table (6) shows the ACF and PACF values for the remainder

Lag	ACF	PACF
1	-0.0041	-0.0041
2	-0.0521	-0.0542
3	-0.0822	-0.0813
4	-0.1496	-0.1545
5	-0.0775	-0.0954
6	-0.0164	-0.0501
7	0.1004	0.0626
8	-0.1246	-0.1713
9	0.0311	0.0026
10	-0.0633	-0.0967
11	0.1123	0.1175
12	0.1511	0.1244
13	0.0777	0.0876
14	-0.0511	-0.0412
15	-0.0731	0.0091
16	0.0111	0.0511

3. Forecasting

After choosing the best model, Arima (3, 1, 3), it can be used to predict the number of divorces in the Holy province of Karbala for the period from January 2025 to December 2025, as shown in Table (7) below.

Table (7) represents the forecasting values for preparing divorce cases for the Holy province of Karbala

Months	forecasting value
January	344
February	435
March	455
April	542
May	378
June	213
July	447
August	434
September	333
October	543
November	456
December	345

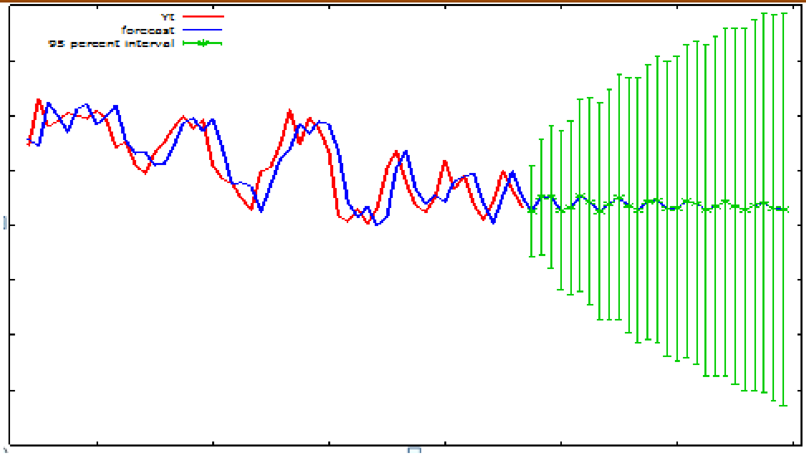

Figure (4) shows the drawing of forecasting values for the series of preparation of divorce cases according to the ARIMA model(3,2,3) with a confidence level of 95 %

In Figure (4), we note that the vertical axis represents the number of divorces and the horizontal axis represents the years, and we note that the predictive values within the sample have shown consistency with the real values . The number of divorce cases is expected to reach in 2025 at an annual rate of 410 divorces and a confidence level of 95%. An increase in the number of divorces.

Conclusions:

1. we note through the study of the series of divorce cases in Karbala governorate that it is unstable on average and that there is a clear general trend in the series, as the emergence of an impact in divorce cases .
2. the stability of the time series was achieved after taking the second difference of the data and after matching the coefficients of the autocorrelation and partial correlation of the time series with the theoretical behavior of the autocorrelation and partial functions, it turned out that the autocorrelation function gradually decreases with increasing displacement periods "
3. using the criteria for differentiation between several models, namely (AIC information criterion (AIC), Bayesian information criterion (BIC), it was found that the appropriate model for the data is the second-order self-regression model after taking the second difference of ARIMA (, 3,1, 3)
4. using this model to predict the number of divorce cases in Karbala governorate for the period (January-December 2025), the predictive values showed consistency with the original values of the series.

Recommendations

1. Work on the preparation of specialized analytical studies on the phenomenon of divorce in general and the divorce of citizens from female citizens in particular and finding appropriate solutions to it.
2. The establishment of marriage counseling offices, the establishment of a family counseling and Guidance Center affiliated to the Ministry of social affairs, as well as the establishment of Family Clinics in the neighborhoods through the rehabilitation of family and community doctors, including specialists in psychology, sociology and education who can deal with social problems with an enlightened vision and specialized thought.
3. The use of new systems to protect the family, especially children and care for divorced women.
4. Restoring the psychological compatibility of the divorced woman by integrating her into society and encouraging her to complete her studies, practice her hobbies and join social work and charitable organizations.

Reference

1. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008)."*Time Series Analysis: Forecasting and Control* (4th ed.). Wiley.
    - *Key reference for Box-Jenkins methodology foundations*.
2. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

3. Ljung, G. M., & Box, G. E. P. (1978)."On a measure of lack of fit in time series models." Biometrika, 65(2), 297-303.
4. Vandell, K. D. (1992). "Forecasting with time series models." Journal of Forecasting, 11(3), 215-234.
5. Shawrawi, A. (2004). Advanced Time Series Analysis. Cairo University Press.
6. Dickey, D. A., & Fuller, W. A. (1979). "Distribution of the estimators for autoregressive time series with a unit root." Journal of the American Statistical Association, 74(366), 427-431.
7. Box, G. E. P., & Pierce, D. A. (1970)."Distribution of residual autocorrelations in ARIMA models." Journal of the American Statistical Association, 65(332), 1509-1526.
8. Chatfield, C. (2003). The Analysis of Time Series: An Introduction (6th ed.). Chapman & Hall. SARIMA notation (Ch. 4).