RESEARCH ARTICLE

Comparison of Time Series Models before and after Using Wavelet Shrinkage Filtering to Forecast the Amount of Natural Gas in Iraq

Mohammad M. Faqe Hussein

Department of Statistics and Informatics, College of Administration and Economics, University of Sulaimani, Iraq

ABSTRACT

The procedure of reducing noise before analyzing the time series is important to get accurate outcomes when building models. Wavelet Shrinkage, consisting of thresholded wavelets, is a powerful mathematical technique for reducing the amount of noise that can be exposed. It has time series observations and selects the cutoff threshold appropriate for removing most of the noise. This paper used data on natural gas production in Iraq during (1981–2019) that contained non-stationary data, so the researcher must treat this problem before starting the analysis. Therefore, the researcher selected the model of time series analysis using Box–Jenkins models (ARIMA (p, d, q) models) and the method of wavelet shrinkage and the comparison between them based on some statistical criteria including the Akaike information criterion, Schwarz Bayesian information criterion, Bayesian information criterion, and Hannan–Quinn information criterion. Statistical programs such as Statgraphics XVII-X64 and MATLAB were used to analyze the data. The paper reached the efficiency of wavelet shrinkage filtering in treating the noise problem and obtaining efficient estimated models, in particular the wavelet shrinkage filter (Daubechies (db8)) with a soft threshold estimated by the fixed-form method and the possibility of linear models with lower one orders and higher efficiency for the obtained filtered observations compared to the corresponding models estimated from the original observations, that is, the suitable model for the data is a model ARIMA (0,1,1) for the data paved in a Daubechies method, it is preferable to do the waveform analysis of the data before analyzing the time series concerning the observations of natural gas production in Iraq. The preference was given to the model estimated by Box and Jenkins ARIMA (0,1,1) model using wavelet shrinkage Daubechies (db8) wavelet reduction.

Keywords: Time series, ARIMA, ACF, PACF, wavelets, akaike information criterion, schwarz bayesian information criterion, HQC

METHODOLOGY

Time Series

The time series is a random process of random variable X indexed by time (t). X(t) is a symbol used for discrete parameter processes (d.p.p.) and X(t) is a symbol for continuous parameter processes (c.p.p.). In this paper, we are discussing the discrete parameter process of time series.[1,2]

Box–Jenkins Model

The ARIMA (p,d,q) model is used to forecast time series; this model depends on previous data to supply predicting indexes. The ARIMA (p,d,q) model has autoregressive and moving average. With p is the # of AR terms, d is the # of differences non-seasonal, and q is the # of MA. The general equation of the model is: [3-6]

$$A = \alpha + \varnothing_1 Z_{t-1} + \varnothing_2 Z_{t-2} + \ldots + \varnothing_t Z_{t-q};$$
$$B = \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

$$Z_t = A - B + \varepsilon \qquad (1)$$

With $\varnothing$ is AR parameter, θ is MA parameter, $\alpha = \mu (1 - \varnothing_1 - \varnothing_2 - \ldots - \varnothing p)$, μ is the mean value of time series, e_t is error,

$\varepsilon_t = y_t - \widehat{y_t}$. The common equation of AR(p) is:

$$Z_t = \propto + \varnothing_1 Y_{t-1} + \varnothing_2 Y_{t-2} + \ldots + \varnothing_p Y_{t-p} + \varepsilon_t \qquad (2)$$

And the common equation of MA(q) is:

$$Z_t = \alpha - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q e_{t-2} + \varepsilon t \qquad (3)$$

The Procedures of the Box–Jenkins Ways

1- The first thing to note is that most of the time series is non-stop, and AR, MA members of the ARIMA model are

Corresponding Author:
Mohammad M. Faqe Hussein, College of Administration and Economics, Statistics and Informatics Department, University of Sulaimani, Iraq E-mail: mohammad.faqe@univsul.edu.iq

only related to the stop-time series. The random process of Zt is considered a stop if the μ and the σ^2 of the process do not change over time (t) means Zt is stationary. So to identify the ARIMA model, we have to do two steps: Based on the chart of time series, a chart of the PACF autocorrelation.[4-6]

2- Find the suitable values of p,d,q to identify the type of ARIMA (p,d,q) model. Where p is the # of AR terms, d is the # of differences (non-seasonal), and q is the #of MA. The PACF and ACF charts are used to determine p and q where ACF is the autocorrelation. PACF is the partial autocorrelation. Select the value of p if the PACF chart has a great value at lag 1,2,….p and descending after p. Choose the value of q if the ACF has a great value at lag 1,2,…,q and decreasing after q.

3- After calculating the regression coefficient of AR and the MA of the ARIMA model, then we will test the model.

Wavelets

The term short wave has flexible mathematical functions or a power concentrated in time to give a tool for analyzing transient, time varying phenomena. Hence, a short wave divides a function into different frequency components; each component can then be reviewed with a resolution that corresponds to its scale. Furthermore, a short wave is a function defined over a finite interval and has a mean of zero. The Fourier analysis is the original history of short wave analysis, and it includes both the Fourier transform and the Fourier series. The Fourier series is used to study systematic procedures, but the Fourier transforms in the frequency domain, and it has no information about the time environment. Hence, the Fourier transform indication of a signal or a function summarizes information of the data as a function of frequency, including the information of the original function.[7-9] as shown in Figure 1.

Wavelet Types

There are two types of wavelets that have gained popularity throughout the development of wavelet analysis (Discrete Wavelet and Continuous Wavelet) shown in Table 1:[10-12]

Haar wavelet

The Haar wavelet was created by Alfred Haar and was presented in 1909. The Haar function is the basis of the wavelet. The Haar is an extension from Fourier analysis to wavelet analysis.

The orthogonal basis for the space is Haar wavelet $\{L^2(R)\}$. Let $\phi(x)$, $\theta(x)$ be the functions defined on (R); consider the constraints of the (h_k) for (N=2):[11]

$$\theta(x) = \theta(2x) + \theta(2x-1) \qquad (4)$$

Now $\{h_0 = h_1 = 1\}$, and the stability condition implies that $\{h_0 + h_1 = 2\}$. Scaling function (θ) is satisfied as follow:

$$\theta(x) = \begin{cases} 1, 0 \le x < 1 \\ 0\, elsewhere \end{cases} \qquad (5)$$

If the wavelet function is defined as $\psi(x) = \psi(2x) + \psi(2x-1)$, then the function is below:

$$\psi(x) = \begin{cases} 1, 0 \le x < 0.5 \\ -1, 0.5 \le x < 1 \\ 0, otherwise \end{cases} \qquad (6)$$

The function (ψ) is the Haar wavelet, and the function $\theta(x)$ is the Haar scaling function.

The following properties of Haar wavelet:
- Orthogonal.
- Compact support.
- Symmetric the scaling procedure
- The anti-symmetric wavelet procedure.

Daubechies wavelet

A compactly supported wavelets. The family of smooth was created by Daubechies (1988), having the biggest number of die out moments for their support. Therefore, the structure of smooth orthonormal wavelets needs the wavelets $\psi(x)$ with many vanishing moments.

The Daubechies (db) wavelet of order (L) has (L) disappearing moments and is supported on the interval $[0,2L-1]$. The Daubechies wavelets also become smoother with increasing (L).

The Daubechies compactly supported wavelets are created using the usual condition. Consider (L=4) then the equations as follows: [13,14]

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 2$$
$$\pi_0 - \pi_1 + \pi_2 - \pi^3 = 0$$
$$-\pi_1 + 2\pi_2 - 3\pi_3 = 0$$
$$\pi_0\,\pi_2 + \pi_1\,\pi_3 = 0$$
$$\pi_0\,\pi_2 + \pi_1\,\pi_3 = 0 \qquad (7)$$
$$\pi_0^2 + \pi_1^2 + \pi_2^2 + \pi_3^2 = 2$$

Then, a unique solution for equation exists:

$$\pi_0 = \frac{1+(3)^{0.5}}{4} \quad \pi_1 \frac{3+(3)^{0.5}}{4}$$

$$\pi_2 = \frac{3+(3)^{0.5}}{4} \quad \pi_3 \frac{1-(3)^{0.5}}{4}$$

The corresponding Daubechies wavelet (db_2) is supported on interval [0, 3], as shown in Figure 2.

Daubechies wavelet has the following properties:
- Orthogonal.
- Compact support.
- There is no symmetry for N > 1.
- Has N vanishing moment.
- The filter length is 2N.

Coiflets wavelet

In 1989, Coifman proposed that orthonormal wavelet bases with disappearing moments be constructed not just for $\psi(x)$, but also for $\theta(x)$. Daubechies studied the structure of such short waves, called "Coiflet" since they were initially requested by Coifman, and the order became known as the order of Coiflet.[15]

The wavelet $\psi(x)$ associated with (CoifN) has (2N) vanishing moment but also the scaling function (has [2 −1]) vanishing moment. The two functions $\psi(x)$ and $\theta(x)$ have support with length (6N - 1).

The Coiflets wavelet satisfied the following properties:
- Orthogonal.
- Compact support.
- Almost symmetric.

Symlet wavelet

Symlet wavelets are created by altering the symmetry of Daubechies wavelets, and their properties are nearly identical to those of Daubechies wavelets. These are the Symlet wavelets, which are symmetrical wavelets. Wavelets corresponding to Symlet correspond to Sym N, where N is the order.[15]

Discrete Meyer wavelet

Meyer wavelets feature a symmetric scaling wavelet function and are orthogonal. It has an infinite number of supports and is band-limited, but it decays quicker than sync wavelet and is indefinitely differentiable.[16,17]

In the frequency domain, the Meyer wavelet is defined in terms of function z as an infinitely differentiable wavelet with infinite support.

$$\varphi(\omega) = \begin{cases} \frac{1}{(2\pi)^{0.5}} \sin\left(\frac{1}{2}\pi z\left(\frac{3|\omega|}{2\pi}-1\right)\right) e^{\frac{j\omega}{2}} & \text{if } \frac{2}{3}\pi < |\omega| < \frac{4}{3}\pi, \\ \frac{1}{(2\pi)^{0.5}} \cos\left(\frac{1}{2}\pi z\left(\frac{3|\omega|}{4\pi}-1\right)\right) e^{\frac{j\omega}{2}} & \text{if } \frac{4}{3}\pi < |\omega| < \frac{48}{3}\pi, \\ 0 & \text{otherwise} \end{cases} \quad (8)$$
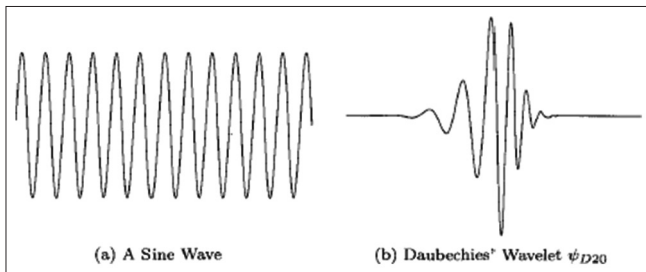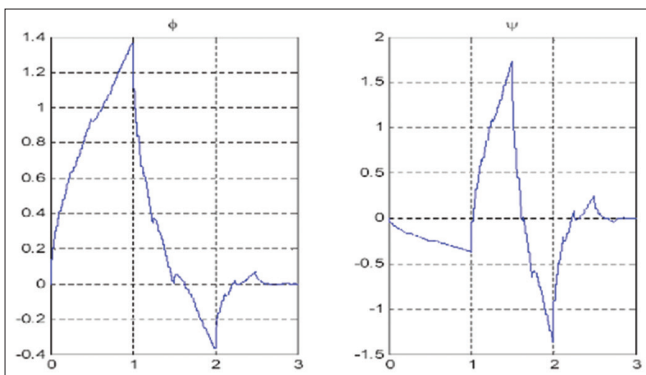

Figure 1: A wave and a wavelet

Figure 2: The scaling function and db$_2$ wavelet

Where

$z(x) = x^4 (35 - 84x + 70x^2 - 20x^3)$ if $0 < x < 1$

$$\varnothing(\omega) = \begin{cases} \frac{1}{(2\pi)^{0.5}} & \text{if } |\omega| < \frac{2}{3}\pi \\ \frac{1}{(2\pi)^{0.5}} \cos\left(\frac{1}{2}\pi z\left(\frac{3|\omega|}{2\pi}-1\right)\right) & \text{if } \frac{4}{3}\pi < |\omega| < \frac{48}{3}\pi, \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Thresholding

The threshold is the estimated noise class, it separates the wavelet coefficient into two parts, the values more significant than the threshold being considered signal and the smaller being considered noise; there are several rules to apply the wavelet coefficient thresholds and several ways to select a threshold as follows: [18]
1. Universal threshold.
2. Specific threshold.
3. Minimax threshold.
4. Cross-validation threshold
   - Two-fold cross-validation
   - Leave-one-out cross-validation

Thresholding Rules

Hard thresholding

Donoho and Johnstone proposed this technique. It is the most straightforward technique for implementing wavelet denoising, which interprets the "keep or kill" statement.

The wavelet coefficient is set to the vector ($\omega n^{(Ht)}$) with element.[18,19]

$$\omega_n^{(Ht)} = \begin{bmatrix} 0 & \text{if } |\omega_n| \leq \eta \\ \omega_n & \text{if } |\omega_n| > \eta \end{bmatrix} \quad (10)$$

If the coefficients are more significant than the threshold value (η), they remain without changing. However, if they are smaller than or equal to the threshold value (η), they are eliminated or set to zero.

Soft thresholding

The other standard procedure for wavelet denoising is soft thresholding of the wavelet coefficient, also suggested by Donoho and Johnstone, which is defined as follows: [18,19]

$$\omega_n^{(St)} = \text{sign } [\omega_n] \ (ABS(\omega_n) - \eta) + \ldots \quad (1)$$

Where

$$\text{Sign } [\omega_n] = \begin{bmatrix} +1 & \text{if } \omega_n > 0 \\ 0 & \text{if } \omega_n = 0 \\ -1 & \text{if } \omega_n < 0 \end{bmatrix} \quad (12)$$

$$(ABS(\omega_n) - \eta)_+ = \begin{bmatrix} (ABS(\omega_n) - \eta) & \text{if } (ABS(\omega_n) - \eta) \geq 0 \\ 0 & \text{if } (ABS(\omega_n) - \eta) < 0 \end{bmatrix} \quad (13)$$

Soft thresholding brings all coefficients to zero when the wavelet coefficient is less than the threshold and when the

coefficients are more significant than the threshold (η), they persist after reducing by the amount of the threshold, while the soft threshold is a continuous function is what the statement of "Shrink or Kill" interprets.

This provides the results of the wavelet thresholding estimator. For small samples, hard thresholding has lower mean square error and less sensitivity to small noises in the data. Soft thresholding has negligible deviation and an overall mean square error as shown in Figure 3.

Time Series Models Building Procedure

In the estimation phase, several models are tentatively selected and then the values of the corrected Akaike information criterion (AIC) (Hurvich and Tsai, 1989) and the Schwarz Bayesian information criterion (SBIC) and Hannan–Quinn information criterion (HQIC) proposed by Schwarz (1978) (Hannan and Quinn, 1979). The model system that has the minimum AIC, SBIC, and HQIC values among other model systems is selected as the most suitable model. Equation (14) describes the formula for calculating AIC, while Equation (15) describes the formula for SBIC and Equation (16) HQIC.[20]

$$AIC = \left(\frac{2n}{n-k-1} \right) \times k - 2 \times \ln |L_{max}|$$

$$\tag{14}]$$

$$SBIC = \ln [n] \times k - 2 \times \ln |L_{max}| \tag{15}$$

$$HQIC = 2\ln [\ln [n]] \times k - 2 \times \ln |L_{max}| \tag{16}$$

Where

- n = Sample size (e.g., data values, frequencies)
- k = parameter estimated (e.g., the normal distribution has 2: μ and σ)
- L_{max} = use maximized value of the log-likelihood to calculate the model by the (i.e., fitting the parameters by MLE and record the natural log of the likelihood)

RESULTS AND DISCUSSION

Data preparation concerns transformation and differentiation. Transformations of the data (such as square roots, logarithms, etc.) are useful to stabilize the variance in a time series; as the variation changes with class, the data are differentiated until there are no obvious patterns, such as trend or seasonality, in the data. Differentiating means taking the difference between consecutive observations or between observations in one part of the year. The differentiated data are often easier to model than the actual data. The data in Table 2 were obtained from the EIA U.S. Energy Information Administration during (1981–2019) obtained and used Statgraphics19 and MATLAB for data analysis.

Randomness test

First, the randomness test of gas production in Iraq should be done with the Box-Pierce test. The null hypothesis states that the time series is random, as opposed to the alternative hypothesis of non-randomness to learn that the data have a specific behavior rather than a random one. As described in Table 3 for time series; as $P < 0.05$ for each series, therefore, null hypotheses

H_0: The series is random versus H_1: The series is not random

The test depends on the sum of squares (S.S) of the first 24 autocorrelation coefficients. Because P-value for this test is 2.20016E-9, which is less than $\alpha = 0.05$, we can reject the hypothesis that the series is random with a 95% confidence level.

Data Transformation

After proving the non-randomness of the natural gas production time series, so to overcome this behavior, we will try to transform the data series. The data were transformed, for instance, the seasonal difference of order one. Then, the series was tested again for the existence of randomness. The hypothesis was accepted when $P > \alpha = 0.05$, as shown in Table 4.

The test is based on the S.S of the first 24 autocorrelation coefficients. Because P-value for this test is 0.275392, which is greater than or equal to $\alpha = 0.05$, we can accept the null hypothesis that the series is random with a 95% or higher confidence level.

Time Series Plots

The time series plots of the dataset (natural gas production) is illustrated. The plot shows that the data are clear of being non-stationary in the mean and in the variance. Seasonal and trend patterns with increasingly variation of variance because the variation in the magnitude of the fluctuation with time is referred to as non-stationary in the variance as shown in Figure 4.

Stationary

The time series plots of the data set (natural gas production) are shown. The plot shows that the data are clearly non-stationary in mean and variance. Seasonal and trend patterns with increasing variation in variance as variation in magnitude of variation over time are said to be non-stationary in variance, as shown in Figure 5.

We examine the autocorrelation function (ACF) and the partial ACF (PACF) for the original data to determine the original order of the series model. The graphs of ACF and PACF are shown in Figures 6 and 7.

After taking the differentiated series for natural gas production, which transforms into stationary series around the mean, and the ACF and PACF plots of the differentiated series, most of the lags are within the confidence limits, as in the mentioned Figures 8 and 9 are shown.

This Table 5 shows the estimated partial autocorrelations between values of adjusted natural gas production at various lags. The lag k partial autocorrelation coefficient measures the correlation between values of adjusted natural gas production at time (t, t+k), having accounted for the associations at all lower lags. It can be used to evaluator the order of the autoregressive model needed to fit the data. Also shown that 95% of the probability limit is around 0. If the probability limits at a specific lag do not contain the estimated coefficient, there is a statistically significant association at that lag at the 95% confidence level. In this case, one of the 24 partial autocorrelation coefficients is statistically significant at the 95% confidence level.

This Table 6 shows the estimated partial autocorrelations between values of adjusted natural gas production at different lags. The partial autocorrelation coefficient lag k measures the

Table 1: Types of wavelet (Discrete and Continuous)

Discrete wavelets		Haar	Daubechies	Coiflet	Symlet	Biorthogonal	Reverse biorthogonal	Discrete Meyer	Fejer-Korovkin filter
Continuous wavelet	Real valued	Beta	Hermitian		Hermitian hat	Mexican hat	Shannon		
	Complex valued	Mexican hat	Morley		Shannon	Modified Morlet			

Table 2: Natural gas production (billion cubic feet per year (bcfy))

Year	Gas production	Year	Gas production	Year	Gas production	Year	Gas production
1980	40.02	1990	140.59	2000	111.24	.	.
1981	23	1991	61.16	2001	.	.	.
1982	24	1992	79.78	2002	.	.	.
1983	18	1993	90.05	2003	.	.	.
1984	20	1994	111.94	2004	.	.	.
1985	29.19	1995	111.95	2005	.	.	.
1986	54.63	1996	114.42	2006	.	.	.
1987	132	1997	107.71	2007	.	.	.
1988	197	1998	104.18	2008	.	.	.
1989	227.41	1999	112.30	2009	.	.	.

Reference: EIA U.S. Energy Information Administration

Table 3: Natural gas production (bcf)

Series	Box-Pierce test	*P*-value
Natural gas production	67.6142	2.20016E-9

Table 4: Natural gas production (bcf)

Series	Box-Pierce test	*P*-value
Gas production natural	15.5309	0.275392

correlation between values of adjusted natural gas production at time (t, t+k), with the associations at all lower lags taken into account. It can be used to evaluate the autoregressive model order required to fit the data. Also shown that 95% of the probability limit is around 0. If the probability limit at a given lag does not include the estimated coefficient, then there is a statistically significant association at the 95% confidence level for that lag. In this case, one of the 24 partial autocorrelation coefficients is statistically significant at the 95% confidence level.

Choosing Appropriate Model

After studying the stationary time series through mean and variance, the model identification procedure determines the sufficient model, the visible ACF, and the PACF. The appropriate models for each time series are listed in Table 7 to identify the best models, lowest AIC, SBIC, and HQC.

Table 7 shows the comparative results of fitting different models to the data. The lowest value of AIC, HQC, and SBIC is ARIMA (0, 1, 1), the selected model used to generate the predicted value. Table 8 also summarizes the results of five tests performed on the residuals to determine whether each model
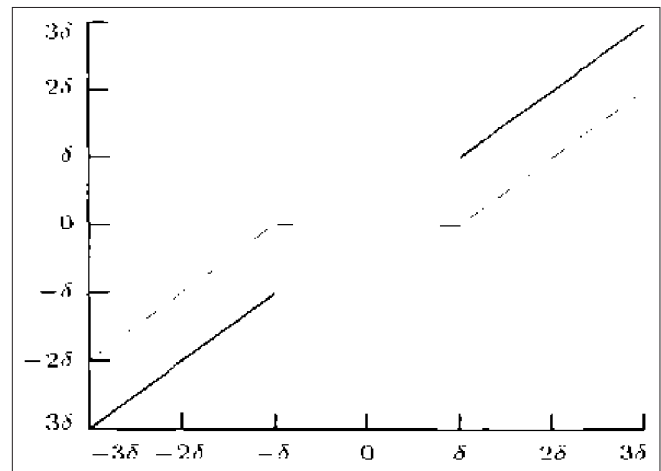

Figure 3: Hard and soft thresholding

fits the data. An OK means that the model passed the test. The currently selected model, model ARIMA (0,1,1), passes five tests. The present model is probably appropriate for the data.

Model Identification

In this point, we will examine the data ACF and PACF, after getting on the stationary for time series examine ACF and PACF. We get the model for the natural gas production ARIMA (0,1,1).

Estimation

It was found that the most suitable model representing the natural gas production series mentioned previous, the parameter was estimated, as shown in Table 9.
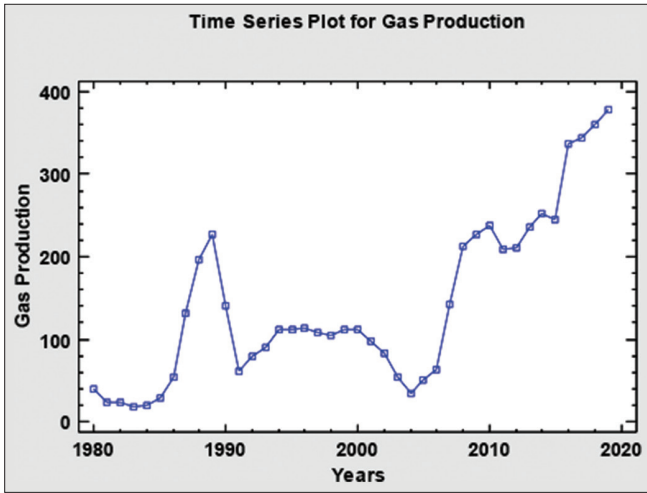
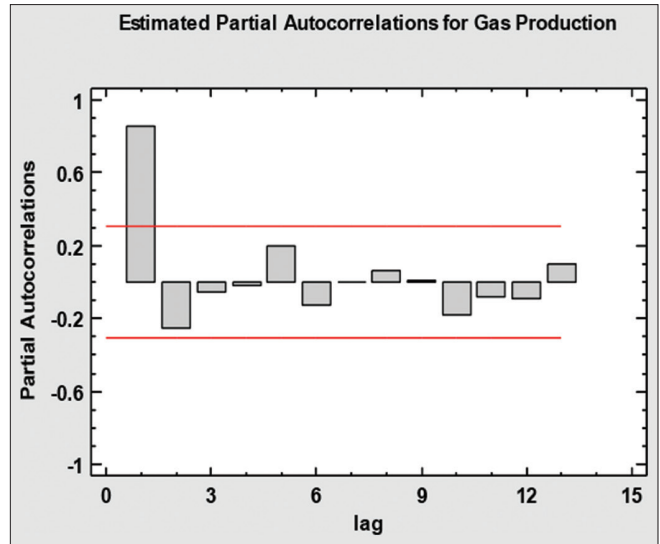**Figure 4:** The plot of nature gas production in Iraq from 1980 to 2019
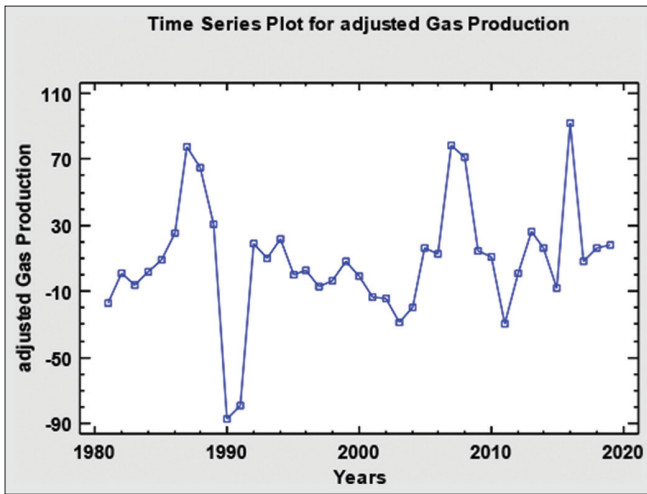

Figure 5: Natural gas production in Iraq after non-seasonal differencing of order one

Figure 6: ACF of natural gas production
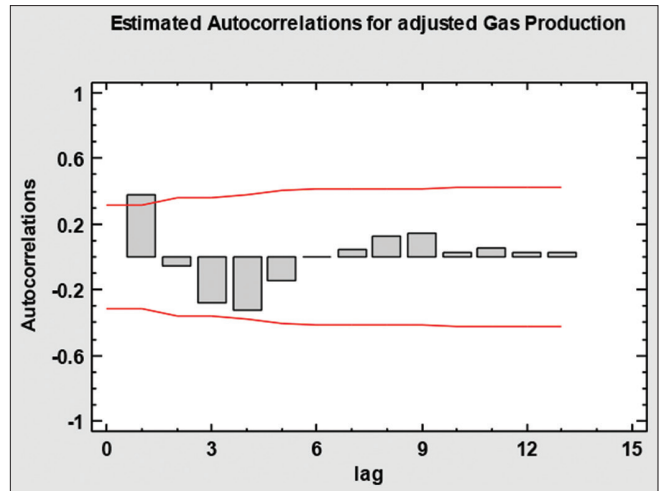

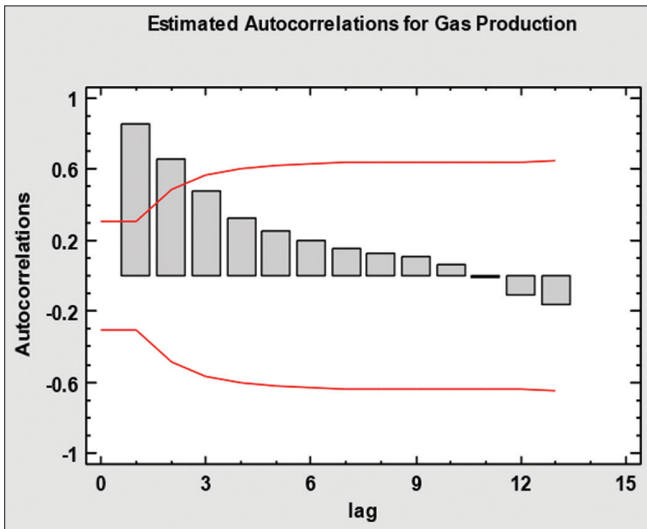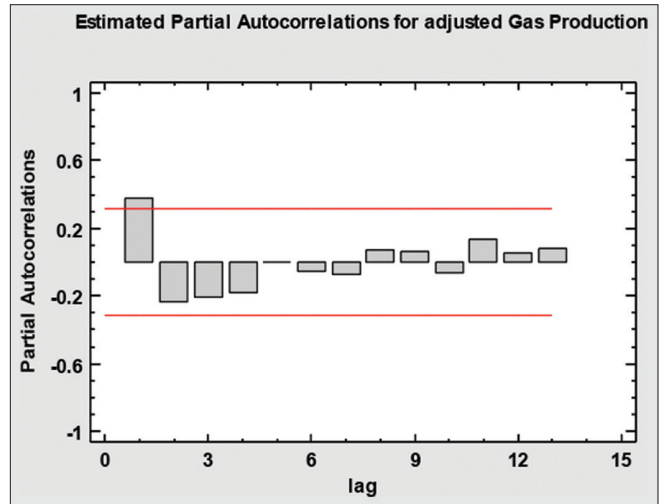**Figure 7:** PACF of natural gas production

Figure 8: ACF of natural gas production

Figure 9: PACF of natural gas production

Table 5: Estimated autocorrelations for adjusted gas production

Lag	Autocorrelation	Std. error	Lower 95% Prob. limit	Upper 95% Prob. limit
1	0.380673	0.160128	−0.313846	0.313846
2	−0.0531608	0.181858	−0.356436	0.356436
3	−0.280853	0.182256	−0.357216	0.357216
4	−0.328115	0.193035	−0.378342	0.378342
5	−0.141684	0.206841	−0.405403	0.405403
6	0.00220161	0.209315	−0.410251	0.410251
7	0.0434127	0.209316	−0.410252	0.410252
8	0.125133	0.209547	−0.410704	0.410704
9	0.14458	0.211454	−0.414443	0.414443
10	0.0274365	0.213974	−0.419381	0.419381
11	0.0547573	0.214064	−0.419558	0.419558
12	0.0278836	0.214423	−0.420262	0.420262
13	0.0299531	0.214516	−0.420444	0.420444

Model Checking

For a suitable predicting model, the residuals for the fitted model should go by simply white noise; in other words, the selected model has exhausted the whole behavior of the series. Therefore, we estimated the ACF and PACF of the residual and plotted it; we would hope to find no significant ACF and PACF. Furthermore, the fitted model should reflect adequate forecasting.

Autocorrelation and partial autocorrelation of residual test

His autocorrelation Figure 10 shows the calculated autocorrelations between the residuals at different lags. The lag-k autocorrelation coefficient measures the correlation between the residuals at time (t, t-k). Also shown that 95% of the probability limit is around 0. If the probability limit at a given lag does not include the estimated coefficient, then there is a statistically significant association at the 95% confidence level for that lag. In this case, none of the 24 autocorrelation coefficients is statistically significant, which means that the time series may well be random (white noise).

This PACF Figure 11 shows the estimated PACF between the residuals at different lags. In this case, none of the 24 ACF is statistically significant at the 95% confidence level.

Test of goodness of fit

It is probable to identify the suitable model ARIMA (0,1,1) for this model repeatedly used Box-Pierce test's ($P \geq \alpha = 0.05$), the null (H_o) hypothesis is not rejected, and the model above is suitable. Meaning that the residuals are also random, and the predicting will be suitable and reliable enough.

H_0: Adequate model versus H_1: Not adequate model

Table 10 above indicates that P-value for this test is greater than or equal to $\alpha = 0.05$; we cannot reject the hypothesis that the model is reasonable.

Table 6: Estimated partial autocorrelations for adjusted gas production

Lag	Partial autocorrelation	Std. error	Lower 95% Prob. limit	Upper 95% Prob. limit
1	0.380673	0.160128	−0.313846	0.313846
2	−0.231641	0.160128	−0.313846	0.313846
3	−0.207301	0.160128	−0.313846	0.313846
4	−0.178398	0.160128	−0.313846	0.313846
5	0.000865901	0.160128	−0.313846	0.313846
6	−0.0563702	0.160128	−0.313846	0.313846
7	−0.0764279	0.160128	−0.313846	0.313846
8	0.0706657	0.160128	−0.313846	0.313846
9	0.0646901	0.160128	−0.313846	0.313846
10	−0.0625562	0.160128	−0.313846	0.313846
11	0.132611	0.160128	−0.313846	0.313846
12	0.0507118	0.160128	−0.313846	0.313846
13	0.0820885	0.160128	−0.313846	0.313846

Table 7: Estimation period

Model	AIC	HQC	SBIC
ARIMA (0,1,1)	7.04975	7.06501	7.09197
ARIMA (1,1,0)	7.06818	7.08344	7.1104
ARIMA (1,0,1)	7.09645	7.12699	7.1809
ARIMA (2,1,0)	7.10137	7.1319	7.18581
ARIMA (0,1,2)	7.1152	7.14573	7.19965

Table 8: Summarizes the results

Model	RMSE	RUNS	RUNM	AUTO	MEAN	VAR
ARIMA (0,1,1)	33.1113	OK	OK	OK	OK	OK

Table 9: ARIMA model summary

Model	Model formula	Parameter	Estimate	Std. error	t	*P*-value
ARIMA (0,1,1)	$\widehat{Y}_t = \frac{1}{4} + Y_{t-1} - \,_1 e_{t-1}$	θ	−0.4491	0.148274	−3.02885	0.004397

Table 10: Hypotheses and probability value for residual series model

Series	Model	Box-Pierce test	Probability value (*P*-value)
Natural gas production	ARIMA (0,1,1)	5.39691	0.943393

Forecasting

In the next step, after a model has accepted the entire analytical test, it is suitable to be used for forecasting. To select a final model for forecasting, the model's precision can be checked to choose how the model performed in terms of forecasting as shown in Figure 12 and 13.

Forecasting use Wavelet Analysis

Five techniques have been used to decrease noise in the data of the production of natural production gas series in Iraq, as follows:

Haar wavelet

Figure 14 displays the actual data time series and the soft threshold of the Haar wave for the production of natural gas in Iraq. The red color is the original data, and the violet is a drawing of the Haar wave. It is clear from the figure that the difference is slight, which indicates the preservation of the data form and, as it is clear from the figure, the presence of blackness more and less noise.

Furthermore, by applying a model to the data smoothed by the Haar method, Table 11 shows the outcomes of calculating the model's parameters.

It is clear from the table the significance of the estimated parameters and the implementation of the condition of stationary, as the absolute values are less than 1. Figures 15 and 16 show the ACF of the model's residuals for the data smoothed by the Haar method. It is clear from the figure that the coefficients of autocorrelation are not significant within the confidence, meaning that the residuals show changes purely random.

Daubechies wavelet

Figure 17 explains the time series of actual data and soft threshold of the Daubechies wave for natural gas production in Iraq. The data smoothed with this wave revealed that their results are the same as the results in the Haar smoothed data (1,2) and different in (3,4,5,6,7,8,9,10).

It is clear from Table 12 the significance of the calculated parameters and the implementation of the condition of stationary, as the absolute values are less than 1. Figures 18 and 19 show the ACF of the model's residuals for the data smoothed by the Daubechies method. It is clear from the figure that the coefficients of autocorrelation are not
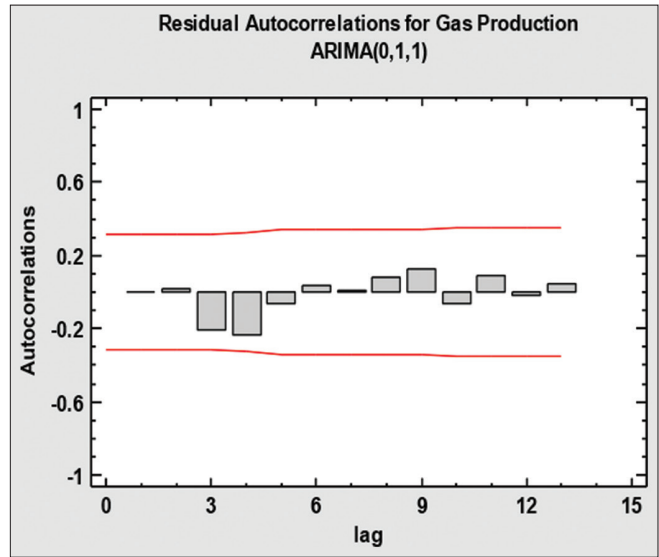

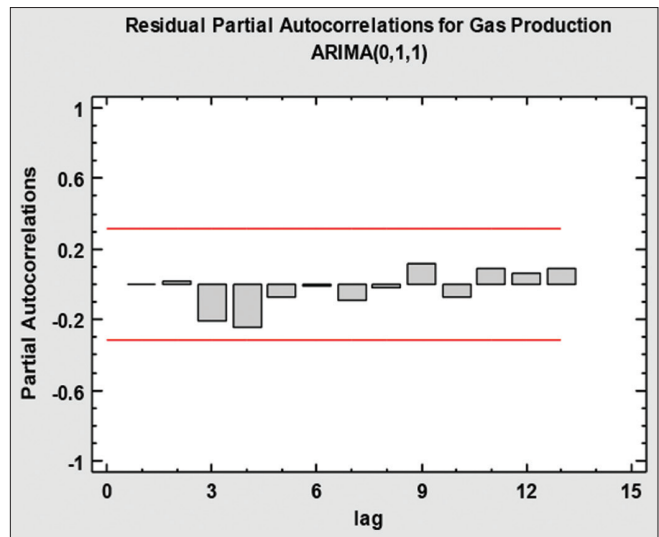**Figure 10: ACF of the residual natural gas production**

Figure 11: PACF of the residual natural gas production

significant within the confidence, meaning that the residuals show changes purely random.

Symlet wavelet

Figure 20 shows the time series number of the actual data and the smooth threshold of the Symlet wave for natural gas production in Iraq. From the figure, it can be seen that the difference between the actual data and the smoothed data are small in a Symlet way and this indicates that the data shape is preserved as it can be seen from the figure that there are more blackness and less noise.

It is clear from Table 13 the significance of the calculated parameters and the implementation of the condition of stationary, as the absolute values are <1. Figures 21 and 22
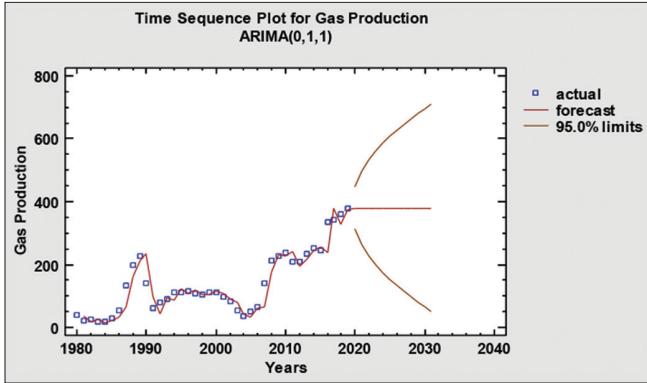

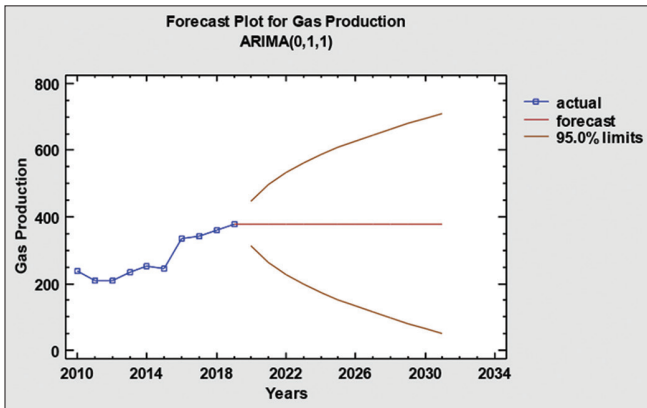**Figure 12:** The sequence plot for the gas production of ARMIA(0,1,1)

Figure 13: Forecasting plot for the gas production of ARMIA (0,1,1)
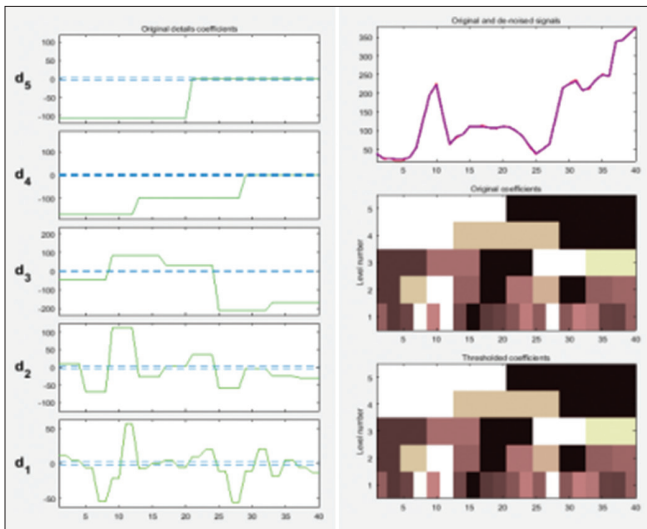

Figure 14: Time series of original data and Haar wavelet soft threshold for natural gas production

show the ACF of the model's residuals for the data smoothed by the Symlet method. It is clear from the figure that the coefficients of autocorrelation are not significant within the confidence, indicating that the residuals show changes purely random.

Coiflets wavelet

Figure 23 shows the time series number of the original data and the smooth threshold of the Coiflets wave for natural gas production in Iraq. It is clear from the figure that the difference between the original data and the smoothed data is small in a Coiflet fashion, and this indicates that the data shape is preserved, as it is clear from the figure that there are more blackness and less noise.

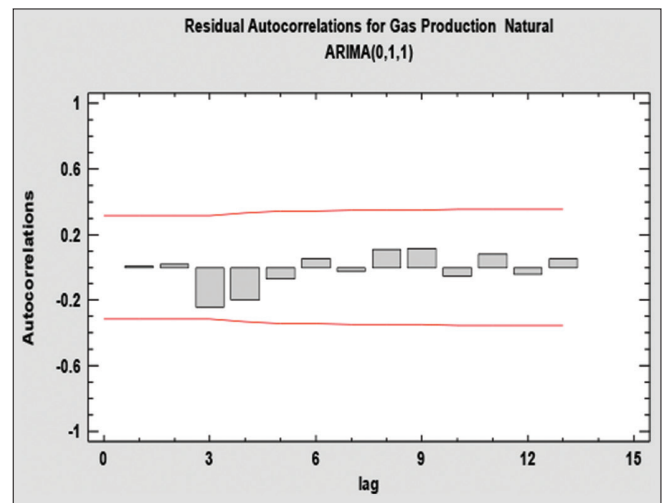From Table 14, the meaning of the calculated parameters and the implementation of the stationary condition becomes
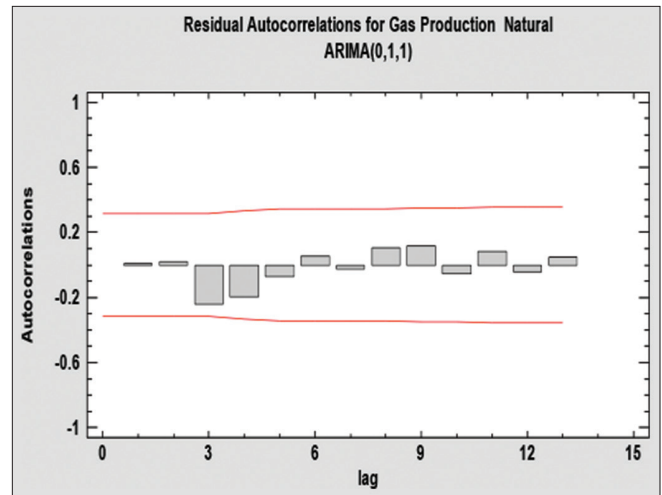

Figure 15: ACF for ARIMA (0,1,1) use Haar method

Figure 16: PACF for ARIMA (0,1,1) use Haar method

Table 11: Model estimation ARIMA (0,1,1) for natural gas production data using Haar method

	Model	Parameter	Estimate	Std. error	t	*P*-value	AIC	HQC	SBIC
Haar	ARIMA (0,1,1)	θ	−0.437605	0.149145	−2.9341	0.00565	7.00141	7.01667	7.04363

Figure 17: The original data time series and the soft threshold of Daubechies wavelet for natural gas production

clear, since the absolute values are <1. Figures 24 and 25 show the ACF of the model residuals for the data smoothed with the Coiflets method. It is clear from the figure that the autocorrelation coefficients within the confidence are not significant, indicating that the residuals show purely random changes.

Meyer D wavelet

Figure 26 shows the time series number of the original data and the smooth threshold of the Meyer D wave for natural gas production in Iraq. It is clear from the figure that the difference between the raw data and the smoothed data in the Meyer-D method is small, and this indicates that the data shape is preserved, since it is clear from the figure that there are more blackness and less noise there.

From Table 15, the meaning of the estimated parameters and the implementation of the steady-state condition

becomes clear, since the absolute values are <1 Figures 27 and 28 show the ACF of the model residuals for the data smoothed with the Meyer-D method. From the figure, it is clear that the within confidence autocorrelation coefficients are not significant, this means that the residuals show purely random changes.

Comparison between Model Estimation

In this part, a comparison is made between a model ARIMA (0,1,1) of the original data for the production of natural gas in Iraq and a model ARIMA (0,1,1) of the paved data in a methods (Haar, Daubechies, Symlet, Coiflets, and Meyer D)

From Table 16 that the values of AIC, HQC, and SBIC decrease between the models, since the values of AIC = 7.04975, HQC = 7.06501, and SBIC = 7.09197 before using the wavelet, while the values of AIC, HQC, and SBIC

Table 12: Model estimation ARIMA (0,1,1) for natural gas production data using Daubechies method

		Models	Parameters	Estimate value	Std. error	t-test	P-value	AIC	HQC	SBIC
Daubechies	1	ARIMA (0,1,1)	θ	−0.437605	0.149145	−2.9341	0.00565	7.00141	7.01667	7.04363
	2	ARIMA (0,1,1)	θ	−0.437605	0.149145	−2.9341	0.00565	7.00141	7.01667	7.04363
	3	ARIMA (0,1,1)	θ	−0.495006	0.146536	−3.378	0.0017	6.9028	6.91807	6.94502
	4	ARIMA (0,1,1)	θ	−0.495876	0.147017	−3.3729	0.00172	6.91675	6.93202	6.95897
	5	ARIMA (0,1,1)	θ	−0.465073	0.147599	−3.1509	0.00317	6.92173	6.937	6.96395
	6	ARIMA (0,1,1)	θ	−0.461273	0.147652	−3.1241	0.00341	6.91783	6.9331	6.96005
	7	ARIMA (0,1,1)	θ	−0.473796	0.146474	−3.2347	0.00252	6.90718	6.92244	6.9494
	8	ARIMA (0,1,1)	θ	−0.498346	0.144843	−3.4406	0.00143	6.86407	6.87933	6.90629
	9	ARIMA (1,2,1)	θ	0.501546	0.143699	3.49024	0.00129	6.86551	6.88078	6.90773
			θ	1.1013	0.00074181	1484.62	0.00000			
	10	ARIMA (0,1,1)	θ	−0.494748	0.145456	−3.4014	0.00159	6.87393	6.8892	6.91616

Table 13: Model estimation ARIMA (0,1,1) for natural gas production data using Symlet method

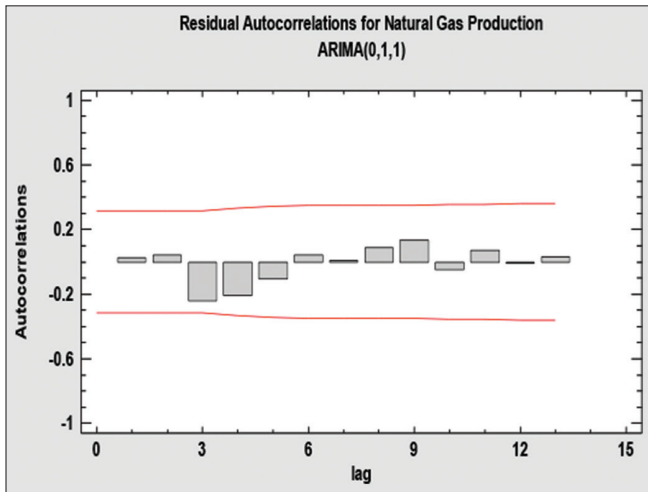|		Model	Parameter	Estimate	Std. error	t	P-value	AIC	HQC	SBIC
Symlet	2	ARIMA (0,1,1)	θ	−0.486617	0.146764	−3.3157	0.00202	6.91261	6.92788	6.95484
	3	ARIMA (0,1,1)	θ	−0.495006	0.146536	−3.378	0.0017	6.9028	6.91807	6.94502
	4	ARIMA (0,1,1)	θ	−0.4896	0.146117	−3.3507	0.00183	6.89975	6.91502	6.94198
	5	ARIMA (0,1,1)	θ	−0.499416	0.146321	−3.4132	0.00154	6.88965	6.90492	6.93187
	6	ARIMA (0,1,1)	θ	−0.485861	0.14574	−3.3338	0.00192	6.89689	6.91215	6.93911
	7	ARIMA (0,1,1)	θ	−0.487812	0.146686	−3.3255	0.00196	6.8857	6.90097	6.92792
	8	ARIMA (0,1,1)	θ	−0.488614	0.145876	−3.3495	0.00184	6.89811	6.91338	6.94033

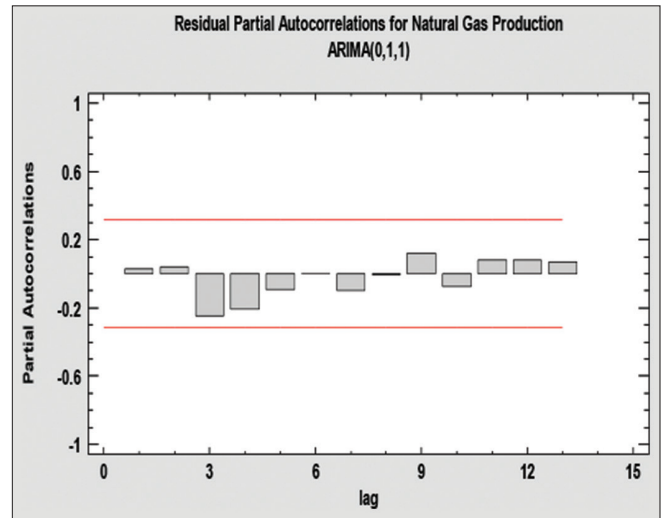**Figure 18: ACF for ARIMA (0,1,1) use Daubechies method**

Figure 19: PACF for ARIMA (0,1,1) use Daubechies method

drop after using the Daubechies (db8) method are the same (6.86407, 6.87933, and 90629), respectively. It is preferable to use wavelet analysis before time series analysis.

CONCLUSIONS

According to the results of the previous chapter, the following conclusions are obtained.

1. Gas is the second-largest source of energy in Iraq, and at the same time, Iraq has considered the fourth country in the world in the expulsion of natural gas without investment.

2. As a result of the use of gas and its advantages, it has made it the most widely used among the sources of non-renewable fossil energy.

3. The wavelet shrinkage method with the residual model and the estimated values proved preferable to the natural gas data in Iraq after the wavelet shrinkage process. The model for the smoothed data by Daubechies (db8)

Table 14: Model estimation ARIMA (0,1,1) for natural gas production data using Coiflets method

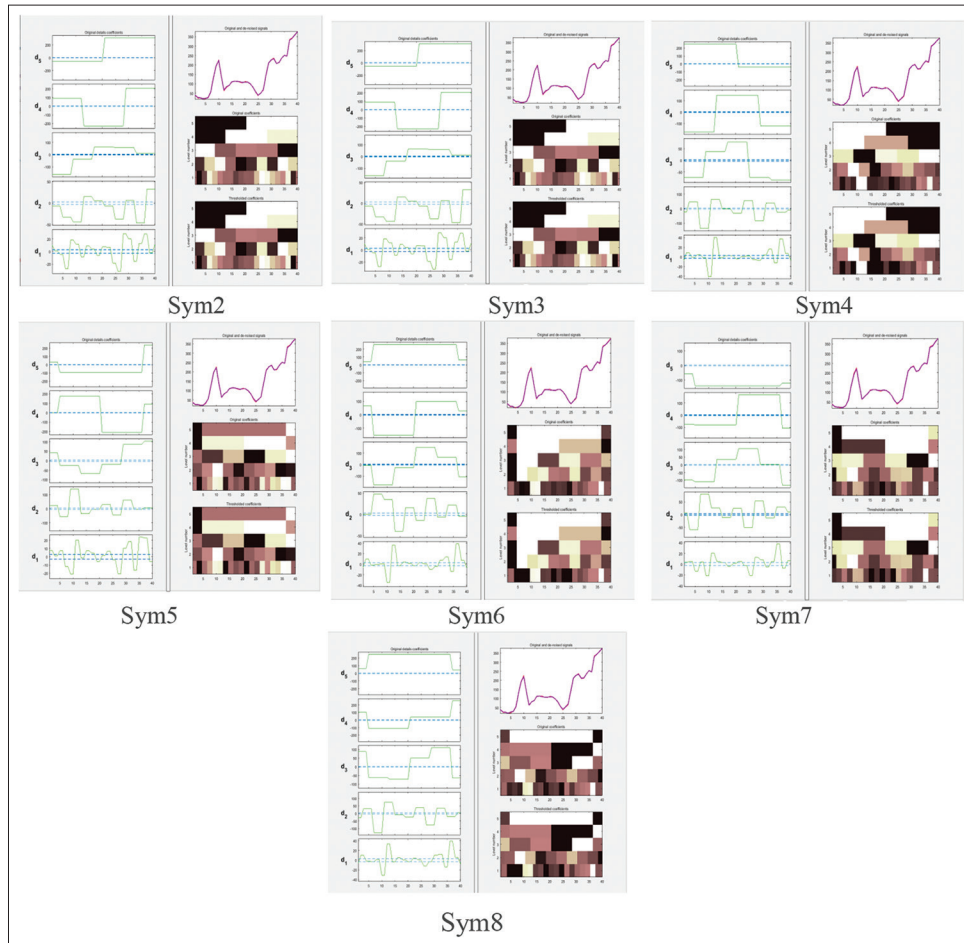|		Model	Parameter	Estimate	Std. error	t	*P*-value	AIC	HQC	SBIC
Coiflets	1	ARIMA (0,1,1)	θ	−0.438559	0.149165	−2.9401	0.00556	6.97715	6.99242	7.01937
	2	ARIMA (0,1,1)	θ	−0.477477	0.146572	−3.2576	0.00237	6.91079	6.92605	6.95301
	3	ARIMA (0,1,1)	θ	−0.481209	0.146402	−3.2869	0.00219	6.90506	6.92032	6.94728
	4	ARIMA (0,1,1)	θ	−0.482222	0.14615	−3.2995	0.00211	6.8993	6.91456	6.94152
	5	ARIMA (0,1,1)	θ	−0.481759	0.146099	−3.2975	0.00212	6.89982	6.91508	6.94204

Figure 20: Time series of the original data and the soft threshold of the Symlet wavelet for natural gas production
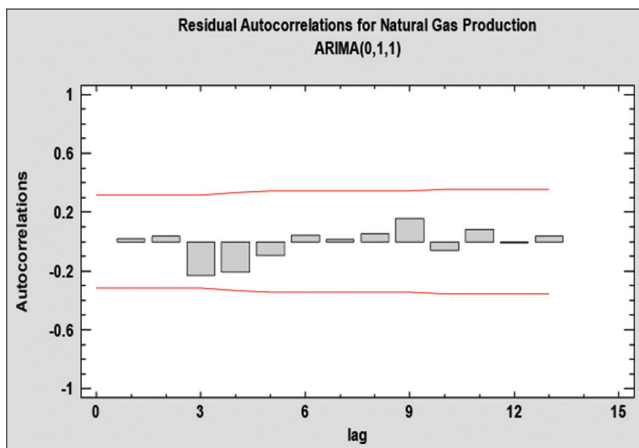

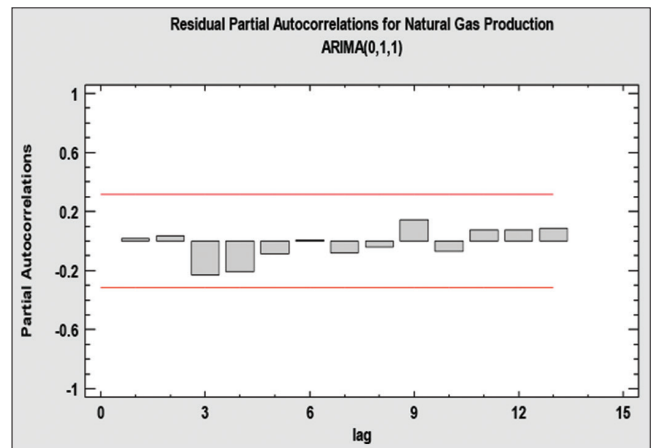**Figure 21: ACF for ARIMA (0,1,1) use Symlet method**

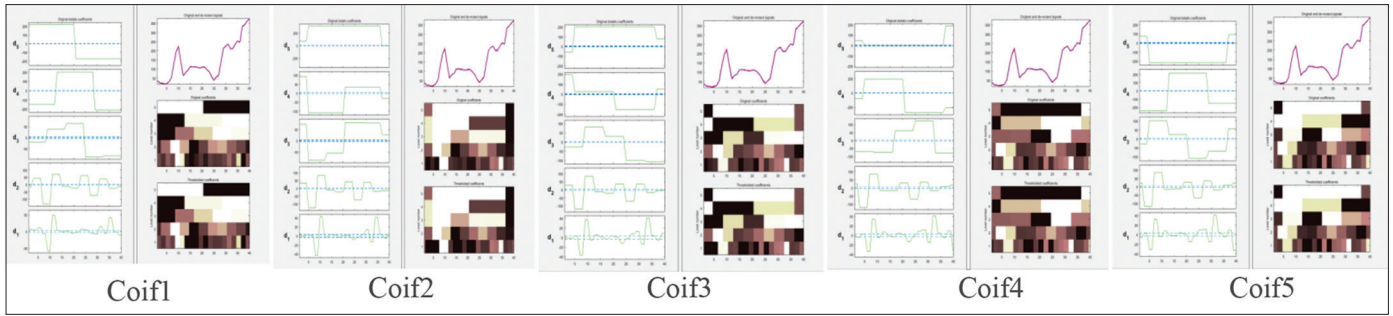**Figure 22: PACF for ARIMA (0,1,1) use Symlet method**

Figure 23: Time series of the original data and the soft threshold of the Coiflets wavelet for natural gas production
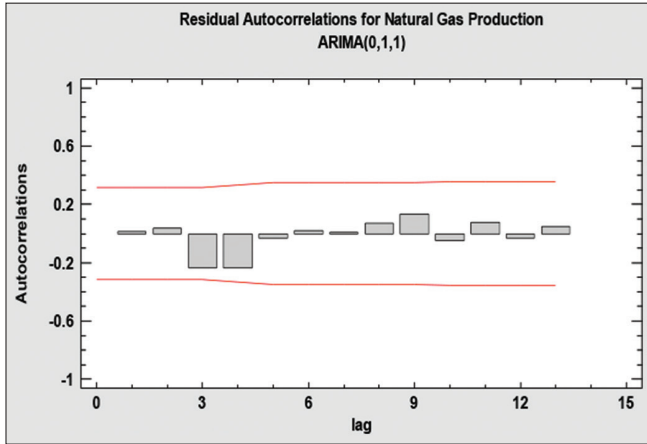

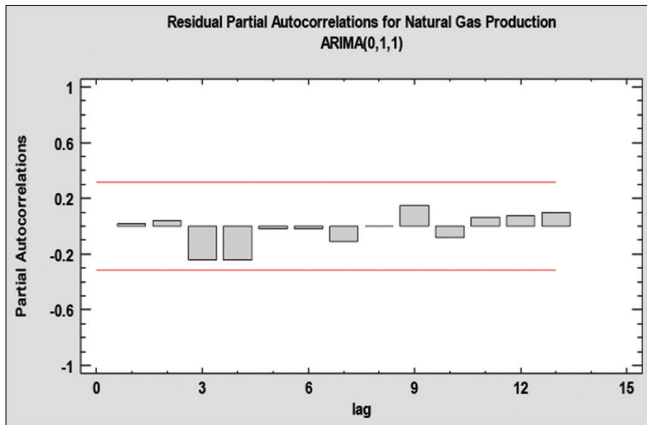**Figure 24:** ACF for ARIMA (0,1,1) use Coiflets method

Figure 25: PACF for ARIMA (0,1,1) use Coiflets method

Figure 26: Time series of the original data and the soft threshold of the Meyer D wavelet for natural gas production
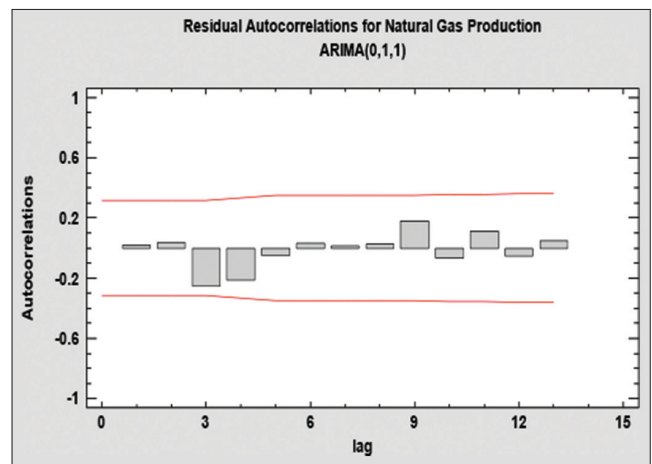
Figure 27: ACF for ARIMA (0,1,1) use Meyer D method

method with the soft threshold cut is the model ARIMA (0,1,1) is better using some statistical criteria than the original data for the production of natural gas in Iraq.

4. The wavelet Daubechies (db8) is better and more efficient than all other methods indicated in the analysis summary, from getting the best criteria comparison AIC, HQC, and SBIC. We note that natural gas production has increased over the years by forecasting the selected model – the forecasted values (2020=370.617 bcfy,….) for natural gas production.

5. If original data are available, the forecast values from the fitted model and the residuals are also displayed. For the time after the end of the series, prediction limits of 95%

are shown for the predictions. These limit show where the original data value is likely to be at a selected future point in time with 95% confidence, provided that the fitted model is appropriate for the data.

6. Wavelet shrinkage be used to treat the problem of pollution and heterogeneity in time series observations

Table 15: Model estimation ARIMA (0,1,1) for natural gas production data using Meyer D method

	Model	Parameter	Estimate	Std. error	t	P-value	AIC	HQC	SBIC
Meyer D	ARIMA (0,1,1)	θ	−0.491145	0.145599	−3.3733	0.00172	6.87297	6.88824	6.9152

Table 16: Comparison between model estimation

|		Model	Parameter	Estimate	Std. error	t	P-value	AIC	HQC	SBIC
Before using wavelet		ARIMA (0,1,1)	θ	−0.4491	0.148274	−3.0289	0.0044	7.04975	7.06501	7.09197
Haar		ARIMA (0,1,1)	θ	−0.437605	0.149145	−2.9341	0.00565	7.00141	7.01667	7.04363
Daubechies	1	ARIMA (0,1,1)	θ	−0.437605	0.149145	−2.9341	0.00565	7.00141	7.01667	7.04363
	2	ARIMA (0,1,1)	θ	−0.437605	0.149145	−2.9341	0.00565	7.00141	7.01667	7.04363
	3	ARIMA (0,1,1)	θ	−0.495006	0.146536	−3.378	0.0017	6.9028	6.91807	6.94502
	4	ARIMA (0,1,1)	θ	−0.495876	0.147017	−3.3729	0.00172	6.91675	6.93202	6.95897
	5	ARIMA (0,1,1)	θ	−0.465073	0.147599	−3.1509	0.00317	6.92173	6.937	6.96395
	6	ARIMA (0,1,1)	θ	−0.461273	0.147652	−3.1241	0.00341	6.91783	6.9331	6.96005
	7	ARIMA (0,1,1)	θ	−0.473796	0.146474	−3.2347	0.00252	6.90718	6.92244	6.9494
	8	ARIMA (0,1,1)	θ	−0.498346	0.144843	−3.4406	0.00143	6.86407	6.87933	6.90629
	9	ARIMA (1,2,1)	θ	0.501546	0.143699	3.49024	0.00129	6.86551	6.88078	6.90773
			θ	1.1013	0.00074181	1484.62	0.00000			
	10	ARIMA (0,1,1)	θ	−0.494748	0.145456	−3.4014	0.00159	6.87393	6.8892	6.91616
Symlet	2	ARIMA (0,1,1)	θ	−0.486617	0.146764	−3.3157	0.00202	6.91261	6.92788	6.95484
	3	ARIMA (0,1,1)	θ	−0.495006	0.146536	−3.378	0.0017	6.9028	6.91807	6.94502
	4	ARIMA (0,1,1)	θ	−0.4896	0.146117	−3.3507	0.00183	6.89975	6.91502	6.94198
	5	ARIMA (0,1,1)	θ	−0.499416	0.146321	−3.4132	0.00154	6.88965	6.90492	6.93187
	6	ARIMA (0,1,1)	θ	−0.485861	0.14574	−3.3338	0.00192	6.89689	6.91215	6.93911
	7	ARIMA (0,1,1)	θ	−0.487812	0.146686	−3.3255	0.00196	6.8857	6.90097	6.92792
	8	ARIMA (0,1,1)	θ	−0.488614	0.145876	−3.3495	0.00184	6.89811	6.91338	6.94033
Coiflets	1	ARIMA (0,1,1)	θ	−0.438559	0.149165	−2.9401	0.00556	6.97715	6.99242	7.01937
	2	ARIMA (0,1,1)	θ	−0.477477	0.146572	−3.2576	0.00237	6.91079	6.92605	6.95301
	3	ARIMA (0,1,1)	θ	−0.481209	0.146402	−3.2869	0.00219	6.90506	6.92032	6.94728
	4	ARIMA (0,1,1)	θ	−0.482222	0.14615	−3.2995	0.00211	6.8993	6.91456	6.94152
	5	ARIMA (0,1,1)	θ	−0.481759	0.146099	−3.2975	0.00212	6.89982	6.91508	6.94204
Meyer D		ARIMA (0,1,1)	θ	−0.491145	0.145599	−3.3733	0.00172	6.87297	6.88824	6.9152
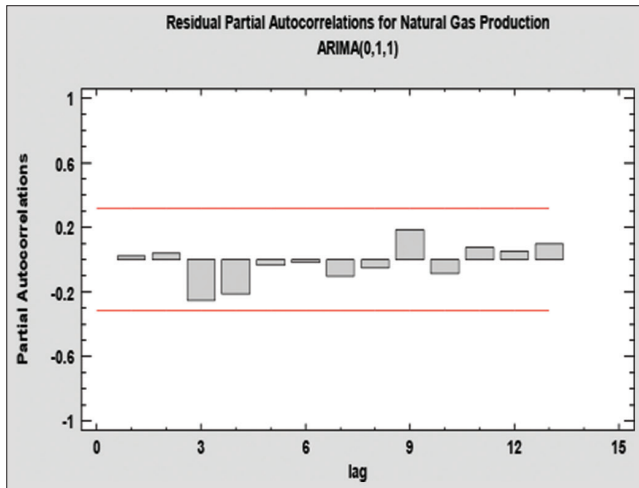
Figure 28: PACF for ARIMA (0,1,1) use Meyer D method

RECOMMENDATIONS

The researcher recommends the following points:

1. Due to the good results yielded from the shrinkage thresholding estimator in reducing noise, it is recommended for future studies to apply other wavelet filters and apply other threshold rules of the wavelet coefficients.

2. To confirm the increase in the activities of exploration and exploration for natural gas in Iraq, and using Modern techniques to increase Iraq's proven and potential reserves of natural gas.

3. Reconstruction, rehabilitation, and development of the infrastructure of the oil and gas industry, which was subjected to significant damage due to wars, economic blockade, and neglect, which led to the obsolescence of the existing ones, and which weakened its production and marketing capabilities.

REFERENCES

1. M. M. Ali, M. Z. Babai, J. Boylan and A. A. Syntetos. Supply chain forecasting when information is not shared. *European Journal of Operational Research*, vol. 260, no. 3, pp. 984-994, 2017.

2. S. E. Hikichi, E. G. Salgado and L. A. Beijo. Forecasting number of ISO 14001 certifications in the Americas using ARIMA models. *Journal of Cleaner Production*, vol. 147, pp. 242-253, 2017.

3. A. W. Omer, H. T. A. Blbas and D. H. Kadir. A comparison between brown's and holt's double exponential smoothing for forecasting applied generation electrical energies in Kurdistan region. *Cihan University-Erbil Scientific Journal (CUESJ)*, vol. 5, no. 2, pp. 56-63, 2021.

4. D. H. Kadir. Time series modeling to forecast on consuming electricity. *Journal of Al Rafidain University College*, vol. 2020, no. 46, pp. 473-485, 2020.

5. H. R. Wang, C. Wang, X. Lin and J. Kang. An improved ARIMA model for hydrological simulations. Nonlinear processes in geophysics discussions. *Nonlinear Processes in Geophysics Discussions*, vol. 1, no. 1, pp. 841-876, 2014.

6. K. W. Wang, C. Deng, J. P. Li, Y. Y. Zhang, X. Y. Li and M. C. Wu. Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiol Infect*, vol. 145, p. 1118-1129, 2017.

7. D. L. Fugal. *Conceptual Wavelets in Digital Signal Processing*. San Diego, California: Space and Signals Technical Publishing, 2009.

8. A. S. Hamad. *Using some Thresholding Rules in Waveletshrinkage to Denoise Signals for Simple Regression Withapplic Ation in Rezgary Hospital*. Sulaimaniyah, Iraq: University of Sulaimani-College of Administration and Economics, 2010.

9. L. Xu, D. Zhang and K. Wang. Wavelet-based cascaded adaptive filter for removing baseline drift in pulse waveforms. *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 11, pp. 1973-1975, 2005.

10. S. A. Broughton and K. Bryan. *Discrete Fourier Analysis and Wavelets*. United States: John Wiley & Sons, 2009.

11. C. S. Burrus, R. A. Gopinath and H. T. Guo. *Introduction to Wavelets and Wavelet Transforms*. United States: Prentice Hall, 1998.

12. P. Bloomfield. *Fourier Analysis of Time Series: An Introduction*. 2nd ed. United States: John Wiley & Sons, 2013.

13. I. L. Cascio. *Wavelet Analysis and Denoising: New Tools for Economists*. London, UK: Queen Mary College, 2007.

14. C. K. Chui. *Wavelets: A Tutorial in Theory and Applications*. San Diego,United State: Academic Press Professional, 1992.

15. I. Daubechies. *Ten Lectures on Wavelets*. United States: Society for Industrial and Applied Mathematics, 1992.

16. F. W. Elliott, D. J. Horntrop and A. J. Majda. A Fourier-Wavelet monte carlo method for fractal random fields. *Journal of Computational Physics*, vol. 132, no. 2, pp. 384-408, 1997.

17. S. Abbasion, A. Rafsanjani, A. Farshidianfar and N. Irani. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing*, vol. 21, no. 7, pp. 2933-2945, 2007.

18. H. Y. Gao and A. G. Bruce. Waveshrink with Firm Shrinkage. *Statistica Sinica,* vol. 7, pp. 855-874, 1997.

19. J. Kovacevic and W. Sweldens. Wavelet families of increasing order in arbitrary dimensions. *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 480-496, 2000.

20. *Comparing Fitted Models using the SIC, HQIC or AIC Information Criteria. Vose*. Available from: https://www.vosesoftware.com/riskwiki [Last accessed on 2022 Jan 01].