



Using Multivariate Time Series Model VAR(P) to forecast Water Supply of Tigris And Euphrates Rivers in Iraq

(PP 220 - 228)

ID No. 2211

<https://doi.org/10.21271/zjhs.22.6.14>

Mohammed AbdulMajeed Badal

Dep. of Statistics/ Salahaddin University-Erbil

sinjary2013@gmail.com

Received: 17/04/2018

Accepted: 08/07/2018

Published: 27/12/2018

Abstract

In this paper, we will discuss the water crisis in Iraq that is one of the most important problems facing the Iraqi economy for water supply for two rivers (Tigris and Euphrates) in Iraq since (1980-2015) by using vector autoregressive VAR(P) model in time series for model that sufficient representation of the dynamic interactions in a system of variables (bivariate) applied for forecast, the model gives the minimum of ACC, BIC, FPE, SBIC, HQIC values and the maximum of R^2 value to forecast for the periods of ten years to forecasting the yearly rate of water imports of the Tigris and Euphrates rivers in Iraq by using VAR(P) and test Diagnostic checking by using STATA v.13 Program, We concluded that the model VAR(P), and the expected values of the yearly rate of water imports of the Tigris and Euphrates rivers are increasing after year 2015 for forecast for ten years (2016-2025).

Keywords: bivariate, VAR(P), VECM, FPE, SBIC, HQIC, R^2 , lag(P), Stationary

1. Introduction:

The first step for constructing a model for a specific purpose or for a particular sector of an economy is to decide on the variables to be included in the analysis, At this stage it is usually important to take into account what economic theory has to say about the relations between the variables of interest. Suppose we want to analyze the transmission mechanism of monetary policy. An important relation in that context is the money demand function, which describes the link between the real and the monetary sector of the economy, increasing the number of variables and equations does not generally lead to a better model because doing so makes it more difficult to capture the dynamic,

Inter temporal relations between them. In fact, in some forecast comparisons univariate time series models were found to be superior to large-scale econometric models. One explanation for the failure of the larger models is their insufficient representation of the dynamic interactions in a system of variables, vector autoregressive (VAR) processes are a suitable model class for describing the data generation process (DGP) of a small or moderate set of time series variables. In these models all variables are often treated as being a priori endogenous, and allowance is made for rich dynamics. Variables are called cointegrated if they have a common stochastic trend. If cointegrating relations are present in a system of variables, the VAR form is not the most convenient model setup. In that case it is useful to consider specific parameterizations that support the analysis of the cointegration structure. The resulting models are known as vector error correction models (VECMs) or vector equilibrium correction models.

The water crisis in Iraq is one of the most important problems facing the Iraqi economy, especially the agricultural sector, because Iraq was and is still an agricultural country, the magnitude of the water problem is large and if it does not appear today it will be stormy and dangerous in the future, even that some of the researchers and political analysts call the



millennium The third of the water cycle, instead of the era of oil that characterized the previous years, due to the fact that the water resources and the resulting problems are among the most prominent issues facing Iraq now and in the future, Iraq is fed by two major rivers, the [Tigris](#) and the [Euphrates](#), both of which originate outside of Iraq. These two rivers account for 98% of Iraq's surface water supply, their flow is therefore very vulnerable to dams and water diversions in Turkey, Syria and Iran. The Euphrates does not receive water from permanent tributaries within Iraq territory and is fed only by seasonal run off from [wad's](#). The average annual flow of the Euphrates at the border to Turkey is estimated at 30 km³, with a fluctual annual value ranging from 10 to 40 km³. The Tigris has an average annual runoff of 21.2 km³. Within Iraq, the Tigris River receives water from five main tributaries, namely the [Little Khabur](#), [Great Zab](#), [Little Zab](#), [Diyala](#) and Al Authaim. Yet, onlyyer lies entirely within Iraq, All together, 50% of the Tigris water comes from outside the country.

2. Materail and method:

2.1. The Models (VARs and VECMs)

we first introduce the basic vector autoregressive and error correction models, neglecting deterministic terms and exogenous variables.

For a set of K time series variables: $y_t = (y_{1t}, \dots, y_{kt})'$, a VAR model captures their dynamic interactions, the basic model of order p (VAR(p)) has the form

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad \dots(1)$$

where the A_i 's are $(K \times K)$ coefficient matrices and $u_t = (u_{1t}, \dots, u_{kt})'$,

is an unobservable error term. It is usually assumed to be a zero-mean independent white noise process with time-invariant, positive definite covariance matrix

$E(u_t u_t') = \Sigma u$, In other words, the u_t 's are independent stochastic vectors with $u_t \sim (0, \Sigma u)$.

Multivariate time series analysis is used when one wants to model and explain the interactions and commovements among a group of time series variables.

The process of VAR(p) is stable from the roots of

$$\det(I_k - A_{1z} - \dots - A_p z^p) \neq 0 \quad \text{for } |z| \leq 1 \quad \dots(2)$$

that is, the polynomial defined by the determinant of the autoregressive operator has no roots in and on the complex unit circle. On the assumption that the process has been initiated in the infinite past

($t=0, \pm 1, \pm 2, \dots$), it generates stationary time series that have time-invariant means, variances, and covariance structure. If the polynomial in (2) has a unit root (i.e., the determinant is zero for $z=1$), then some or all of the variables are integrated. For convenience we assume for the moment that they are at most I(1). If the variables have a common stochastic trend,

Vector autoregressive (VAR) processes are a suitable model class for describing the data generation process (DGP) of a small or moderate set of time series variables. In these models all variables are often treated as being a priori endogenous, and allowance is made for rich dynamics.

It is possible there are linear combination so of them that are I(0). In that case they are cointegrated. In other words, a set of I(1) variables is called cointegrated if a linear combination exists that is I(0). Occasionally it is convenient to consider systems with both I(1) and I(0) variables, there by the concept of cointegration is extended by calling any linear combination that is I(0) a cointegration relation, although this terminology is not in the spirit of the original definition because it can happen that a linear combination of I(0) variables is called a cointegration relation.



Although the model (1) is general enough to accommodate variables with stochastic trends, it is not the most suitable type of model if interest centers on the cointegration relations because they do not appear explicitly.

2.2. Vector Autoregressive and Vector Error Correction Models

Forecasting Two types of forecasts are available after you fit a VAR(p):

a one-step-ahead forecast and a dynamic h-step-ahead forecast. the one-step-ahead forecast produces a prediction of the value of an endogenous variable in the current period by using the estimated coefficients, the past values of the endogenous variables, and any exogenous variables. If you include contemporaneous values of exogenous variables in your model, you must have observations on the exogenous variables that are contemporaneous with the period in which the prediction is being made to compute the prediction. In Stata Program, these one-step-ahead predictions are just the standard linear predictions available after any estimation command, thus predict, produces one-step-ahead forecasts for the specified equation, produces the standard error of the linear prediction for the specified equation. The standard error of the forecast includes an estimate of the variability due to innovations, whereas the standard error of the linear prediction does not. The dynamic h-step-ahead forecast begins by using the estimated coefficients, the lagged values of the endogenous variables, and any exogenous variables to predict one step ahead for each endogenous variable.

2.3 Fitting the model:

The selection of the model is important, as under-fitting a model may not capture the true nature of the variability in the outcome variable, while an over-fitted model loses generality, Akaike Information Criteria (*AIC*) is then a way to select the model that best balances these drawbacks. Once a best model is selected, traditional null-hypothesis testing can then be used on the best model to determine the relationship between specific variables and the outcome of interest or *AIC* criterion asymptotically overestimates the order with

positive probability :

$$AIC = 2K - 2\log(L(\hat{\theta}/y)) \quad \dots(3)$$

where K is the number of estimable parameters (degrees of freedom) and $\log(L(\hat{\theta}/y))$ is the log-likelihood at its maximum point of the model estimated. Further refined this estimate to correct for small data samples:

$$AICc = AIC + \frac{2K(K+1)}{n-K-1} \quad \dots(4)$$

where n is the sample size and K and *AIC* are defined above. If n is large with respect to K , this correction is negligible and *AIC* is sufficient. *AICc* is more general, however, and is generally used in place of *AIC*. The best model is then the model with the lowest *AICc* (or *AIC*) score. It is important to note that the *AIC* and *AICc* scores are ordinal and mean nothing on their own.

Bayesian Information Criteria (*BIC*) is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower *BIC* means that a model is more likely to be the true model:

$$BIC = 2\log n - 2\log(L(\hat{\theta}/y)) \quad \dots(5)$$

The Box-Ljung test is a diagnostic tool used to test the lack of fit of a time series model. varsoc reports the final prediction error (FPE), Akaike's information criterion (*AIC*), Schwarz's Bayesian information criterion (SBIC), and the Hannan and Quinn information criterion (HQIC) lag order selection statistics for a series of vector auto-regressions of order 1, ..., max lag(p), BIC and HQ criteria estimate the order consistently under fairly general conditions if the true order p is less than or equal to P_{max} . A sequence of likelihood-ratio test statistics for all the full VARs of order less than or equal to the highest lag order is also

reported. In the post estimation version, the maximum lag and estimation options are based on the model just fit or the model specified in estimates (estname), the pre estimation version of varsoc can also be used to select the lag order for a vector error correction model (VECM). As shown by Nielsen (2001), the lag-order selection statistics discussed here can be used in the presence of I(1) variables. Many selection-order statistics have been developed to assist researchers in fitting a VAR of the correct order. Several of these selection-order statistics appear in the trend stationary [TS] var output. The varsoc command computes these statistics over a range of lags p while maintaining a common sample and option specification, varsoc computes four information criteria as well as a sequence of likelihood ratio (LR) tests. The information criteria include the FPE, AIC, the HQIC, and SBIC. For a given lag p , the LR test compares a VAR with p lags with one with $p-1$ lags. The null hypothesis is that all the coefficients on the p -th lags of the endogenous variables are zero. To use this sequence of LR tests to select a lag order,

As shown by Hamilton (1994, 295–296), the log likelihood for a VAR(p) is:

$$LL = \frac{T}{2} \left[\ln \left(\frac{1}{\sum \hat{u}_t u_t'} \right) - K \ln(2\pi) - K \right] \quad \dots(6)$$

where T is the number of observations, K is the number of equations, and $\sum$ is the

maximum likelihood estimate of $E(u_t u_t')$, where u_t is the

$K \times 1$ vector of disturbances, the log likelihood can be rewritten as:

$$LL(j) = 2\{LL(j) - LL(j-1)\} \quad \dots(7)$$

2.4 Model-order statistics

The formula for the FPE (final prediction error) given in Lutkepohl (2005, 147) is:

$$FPE = \sum u / \left(\frac{T + KP + 1}{T - KP - 1} \right)^K \quad \text{where written by}$$

$$FPE = \sum u / \left(\frac{T + \bar{m}}{T - \bar{m}} \right)^K \quad \dots(8)$$

where $\bar{m}$ is the average number of parameters over the K equations. this implementation accounts for variables dropped because of collinearity.

Lutkepohl (2005) advocates dropping the constant term from the log likelihood because it does not affect inference. The Lutkepohl versions of the information criteria are:

$$AIC = \ln \left(\sum u \right) + \frac{2PK^2}{T} \quad \dots(9)$$

$$SBIC = \ln \left(\sum u \right) + \frac{\ln(T)}{T} . PK^2 \quad \dots(10)$$

$$HQIC = \ln \left(\sum u \right) + \frac{2 \ln \{ \ln(T) \}}{T} . PK^2 \quad \dots(11)$$

where PK^2 is the total number of parameters in the model and LL is the log likelihood.

5.2. The Stationarity:

A **stationarity** time series is one whose properties do not depend on the time at which the series is observed, Inference after var and svar requires that variables be covariance stationary, the variables in y_t are covariance stationary if their first two moments exist and are independent of time. More explicitly, a variable y_t is covariance stationary if :

$a - E[y_t]$ is finite and independent of t , for all t .



$b\text{-Cov}[y_t, y_s]$ is a finite function of $|t-s|$ but not of t or s alone, for all t .

$c\text{-Var}[y_t]$ is finite and independent of t , for all t .

A VAR(P) process is stationary only if all the Eigen values of A_p are less than 1 in absolute value.

3. Results and discussion:

The water crisis in Iraq resulted from external and internal causes. The external causes were (climate change, global warming caused the rains to fall, the Tigris and Euphrates rivers are not from Iraq but from countries neighboring Iraq, political tension and the absence of international law) (Lack of water dams and lakes on the surface rivers in Iraq, non-creeping vegetation from river banks, lack of sedimentation, lack or weakness of water supply, lack of modern irrigation methods, low water cost, population increase).

We take the data of imports in Iraq governorate for the period sense 1980 to 2015 form the table (1), to forecasting the yearly rate of water imports of the Tigris and Euphrates rivers periods for the future years, using the VAR(P) model by equation (1):

Table(1): Water imports in Iraq of the Tigris and Euphrates rivers for years (1980-2015) billion cubic meters

Year	Euphrates	Tigris
1980	36.36	36.6
1981	29.8	9.99
1982	30.56	52.93
1983	31.43	54.4
1984	27.18	41.27
1985	37.22	34
1986	23.65	54.96
1987	17.22	32.6
1988	19.58	58.54
1989	46.73	86.66
1990	9.05	38.8
1991	12.4	30.87
1992	12.15	62.72
1993	12.37	66.36
1994	15.29	44.85
1995	23.9	66.4
1996	30.12	40.6
1997	27.64	42.75
1998	28.95	49.87
1999	18.61	18.6
2000	17.23	20.1
2001	9.59	21.28
2002	10.67	42.98
2003	15.71	51.13
2004	20.54	45.51
2005	17.57	38.07
2006	20.64	43.17
2007	19.33	37.76
2008	14.7	18.27
2009	13.57	27.97
2010	12.45	37.68
2011	14.62	32.94



2012	20.5	28.7
2013	17.85	27.46
2014	14.85	22.45
2015	13.54	21.8

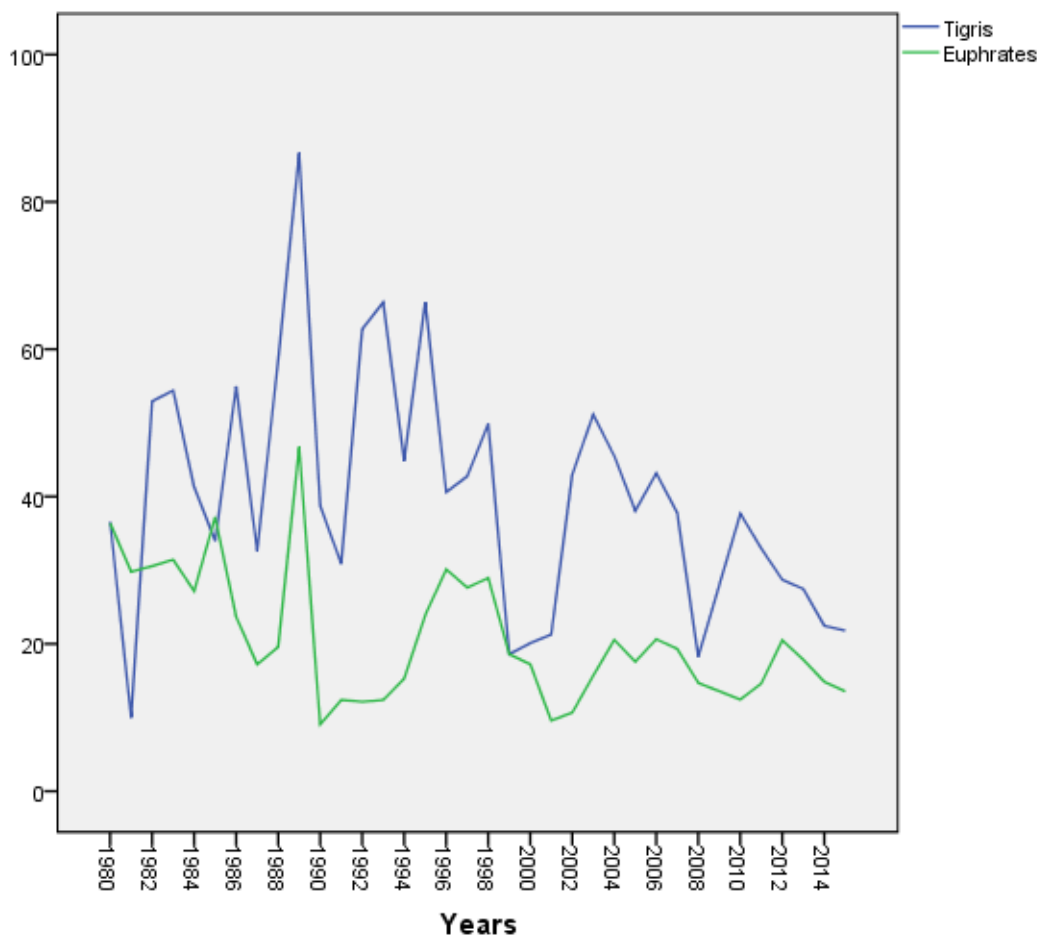
Source: Ministry of Water Resources in Iraq

We need to draw the original data in table (1) to know the distribution of water imports in Iraq for 1980-2015, and check the correlation between two rivers (Tigris and Euphrates) before multivariate in time Series analyses such as in table (2):

Table(2): Correlation between Tigris and Euphrates variables

		Tigris	Euphrates
Tigris	Pearson Correlation	1	.352*
	Sig. (2-tailed)		.035
	N	36	36
Euphrates	Pearson Correlation	.352*	1
	Sig. (2-tailed)	.035	
	N	36	36

*. Correlation is significant at the 0.05 level (2-tailed).



Figure(1):Time series of water imports of Tigris& Euphrates rivers in



Iraq(1980-2015)

Table(3): Vector Auto regression(VAR) for variables Tigris and Euphrates

Sample:	1982 - 2015	Number of obs.	= 34
Log likelihood =	-255.9988	AIC	= 15.52934
FPE	= 19068.13	HQIC	= 15.65182
Det (Sigma_ml) =	11884.57	SBIC	= 15.88848

From the table(3) the VAR model is the best because AIC (15.52934) value is less than HQIC ,SBIC values.

Table (4): Test of regression model (VAR) for variables Tigris and Euphrates

Equation	No. Parms.	RMSE	R-sq.	chi2	P>chi2
Tigris	4	18.275	0.8470	188.1752	0.0000
Euphrates	4	8.49077	0.8636	215.2803	0.0000

By using analysis for each equation has the same regresses — lagged values of y_{1t} and y_{2t} in Lag1 and Lag2, and give scientific in Lag(1) for Tigris with Tigris.

Table (5): Test of Auto regression model (VAR) for variables Tigris and Euphrates

		Coef.	Std. Err.	z	P> z	[95%Conf.Interval]	
Tigris	Tigris L1.	0.6095139	0.1951013	3.08	0.002	0.2212425	0.9977853
	L2.	0.1510763	0.2021691	0.75	0.455	-0.2451679	0.5473204
	Euphrates L1.	0.985785	0.4294622	0.23	0.818	-0.7431519	0.9403089
	L2.	0.2732492	0.3845697	0.71	0.477	-0.4804936	1.026992
Euphrates	Tigris L1.	0.1372861	0.0920399	1.49	0.136	-0.0431087	0.3176809
	L2.	0.0239557	0.0939298	0.26	0.799	-0.1601434	0.2080547
	Euphrates L1.	0.3198645	0.1995325	1.60	0.109	-0.0712119	0.7109409
	L2.	0.2962967	0.178675	1.66	0.097	-0.5224141	0.6464932

Before forecasting we must check weather our VAR model has satisfied all assumption or not (Diagnostic checking for residually):

Null: there is no serial correlation (no autocorrelation)

Alt: there is serial correlation (autocorrelation)

Table (6): Lagrange-multiplier test at lag order

lag	chi2	df	Prob. > chi2
1	9.7136	4	0.04554
2	15.8804	4	0.00318

Null: Residuals are normally distribution

Alt: Residuals are not normally distribution

We accepted assumption for Null because the residuals are normal in Lag(1) and Lag(2).

Table (7): Jarque-Bera test of VAR for normality

Equation	chi2	df	Prob. > chi2
Tigris	2.849	2	0.24058
Euphrates	1.545	2	0.46175
ALL	4.395	4	0.35520



From the table(7) the single-equation and overall Jarque–Bera statistics do not reject the null of normality, the null hypothesis is that the disturbance term in that equation has a univariate normal distribution, for all equations jointly, the null hypothesis is that the K disturbances come from a K-dimensional normal distribution.

Table(8) : analysis the matrix var(p) and test of stationary time series

rank	Parameters	LL	Eigen value	Trace statistic	Critical value 5%
0	6	-262.14105	.	23.2977	15.41
1	9	-254.52086	0.36125	8.0573	3.76
2	10	-250.49221	0.21099		

From the table (8) we see the series in model var (p) for Lags(2) it stationary because Eigen value less than one.

Therefore, the forecasting values of Water imports for ten years in Iraq of the Tigris and Euphrates rivers for years (2016-2025) by table(9):

Table (9): Forecasting Water imports in Iraq of the Tigris and Euphrates rivers for years (2016-2025) billion cubic meters

years	Forecasting Tigris	Forecasting Euphrates
2016	37.641691	16.381791
2017	42.678363	17.886183
2018	43.009483	18.677574
2019	42.555416	19.038232
2020	42.19404	19.231416
2021	41.996806	19.33271
2022	41.893724	19.388986
2023	41.839752	19.41956
2024	41.81066	19.436396
2025	41.794847	19.445569

4.Conclusions

1-We conclude that the expected values of Water imports in Iraq of the Tigris and Euphrates rivers increasing after the year 2015 by using the VAR(P), the vector autoregression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series.

2- The model VAR(p) give the best statistic by AIC (15.52934) because less than HQIC ,SBIC.

3-From test all assumption by Diagnostic to checking for residually to give best model of var(p) .

4-Based on the above, we find that the solution starts from inside Iraq through the construction of dams and lake reservoirs of water and not leave the water goes to the sea,

5- in addition to the Cree rivers and continuously to prevent the growth of plants that consume water by the water and lining the drains in order to reduce the waste water resulting from the project aims to encourage farmers to use modern irrigation methods in Iraqi agriculture and to disseminate experiments that proved the role of modern irrigation methods in reducing the consumption of water.



5. References:

- 1-Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle. In Second International Symposium on Information Theory, ed. B. N. Petrov and F. Csaki, 267–281. Budapest: Akailseoniai–Kiudo.
- 2-Amemiya, T. (1985), Advanced Econometrics. Cambridge, MA: Harvard University Press.
- 3-Hamilton, J. D. (1994), Time Series Analysis. Princeton: Princeton University Press.
- 4-Luo, C.S., Zhou, L., Qingfeng, W. (2013), "Application of SARIMA Model in Cucumber Price Forecast", Applied Mechanics and Materials, Vols. 373-375, pp. 1686-1690.
- 5-Lutkepohl, H. (1993), Introduction to Multiple Time Series Analysis. 2nd ed. New York: Springer.
- 6-Lutkepohl, H. (2005), New Introduction to Multiple Time Series Analysis. New York: Springer.
- 7-Paulsen, J. (1984), Order determination of multivariate autoregressive time series with unit roots. Journal of Time Series Analysis 5: 115–127.
- 8-Tsay, R. S. (1984), Order selection in nonstationary autoregressive models. Annals of Statistics 12: 1425–1433.
- 9- Franses, Ph. H. and Dijk, D.V. (2000), "Nonlinear Time Series Models in Empirical Finance" Cambridge University Press, ISBN 0 511 01100 8 virtual.
- 10-Gikungu, S.W., Waititu, A.G. (2015), "Forecasting Inflation Rate in Kenya Using SARIMA Model American Journal of Theoretical and Applied Statistics 4, 15-18.
- 11-Nielsen, B. (2001), Order determination in general vector autoregressions. Working paper, Department of Economics, University of Oxford and Nuffield College.
- 12-Michal, S. (2001), "Time Series Analysis" log linear publishing, Canada
- 13- Mira, S.K., Ahmad M.R. (2015), "Time Series Models for Average monthly Solar radiation in Malaysia", Research and Education in Mathematics, International Conference Kuala Lumpur, Malaysia.

پوخته

ئهم توێژینه وهیه بریتیه له شروقه کردنێک بۆ قهرانی ئاو له عیراق که یهکیکه له کیشه ههره گهره کانه که رووبه رووی ئابووری عیراق دهیتتهوه. توێژینه وهیه که مان شیکردنه وهیه که بۆ ئاستی داینکردنی ئاوی ههر دوو رووباره کانی دیجله وفورات له عیراقدا له سالی 1980 تا سالی 2015 ئه ویش له ریکای بهکارهینانی مۆدیلی (Vector autoregressive) له زنجیره ی کاتدا. ئهم مۆدیل به بریتیه له چالاکیه کانی دینامیکی (dynamic interaction) پتویست له سیسته می گورانکاری (bivariate) که بهرکارده هینریت بۆ پشینه کانه. مۆدیلی (VAR) شیکردنه وهیه کی وورد دهرده خات بۆ که مترین ریزه ی (ACC) (BIC, FPE, SBIC, HQIC) وه برزترین ریزه ی (R2) بۆ پشینه کاردنی ریزه ی سالانه ی ئاوی دیجله وفورات به درزیایی 10 سال له عیراقدا. ههره ها به رنامه ی (STATA v.13) مان به کارهیناوه بۆ پشکنینی ئاستی ئاوی هاورد هه کراوه بۆ عیراق. له دهرئه نجامی توێژینه وهیه که مان پشینه ده کړیت له بهر زبوونه وه ی ئاستی ئاوی هاورد هه کراوی ههر دوو رووباری دیجله وفورات له ده سالی ئاینده (2016-2025).

ملخص

تم في هذا البحث دراسة ازمة المياه في العراق وهي من الازمات الكبيرة والتي لها تأثيرها الاقتصادي المهم على البلد من خلال دراسة وتحليل موارد نهري دجلة والفرات من المياه في العراق لعام 2015-1980 خلال تطبيق نموذج الارتباط الذاتي AR(P) لمتعدد المتغيرات للسلسلة الزمنية وهو نموذج التقاطع بين المتغيرات ويعطي النموذج تحليل مفصل و جيد من خلال اختبار الفروض الواجب توفرها في السلسلة الزمنية باستخدام برنامج STATA 13. و لاكثر من متغير وعن طريق المقاييس الاحصائية التي تعطي اقل قيمة FPE, SBIC, HQIC, و اعلى قيمة ل R2 ووجد ان من المتوقع ان تزداد كمية الواردات لمياه دجلة والفرات خلال العشر السنوات القادمة من 2016-2025 بنسب متقاربة لكل من نهري دجلة والفرات.