Using The Fully Modified M Method In Estimating The Effect Of Cultivated Area On Barley Production In Iraq For The Period 1970-2018*

A.S. Azhar Kadhim Jbarah**
azkdfr_2017@uomustansiriyah.edu.iq
Prof Dr. Ahmed Shaker Mohammed***
ahmutwali@uomustansiriyah.edu.iq

Abstract
The research is concerned with estimating the effect of the cultivated area of barley crop on the production of that crop by estimating the regression model representing the relationship of these two variables. The results of the tests indicated that the time series of the response variable values is stationary and the series of values of the explanatory variable were nonstationary and that they were integrated of order one (I(1)), these tests also indicate that the random error terms are auto correlated and can be modeled according to the mixed autoregressive-moving average models ARMA(p,q), for these results we cannot use the classical estimation method to estimate our regression model, therefore, a fully modified M method was adopted, which is a robust estimation methods, The estimated results indicate a positive significant relation between the production of barley crop and cultivated area.
Keywords: Robust Estimating Methods, Fully Modified M Method, Stationary And Nonstationary Time Series, Autoregressive-Moving Average Models.

1- Introduction:
Statistically the linear regression model based on some basic assumptions, Which facilitate the estimation process and the significance tests of the estimated model. However, these assumptions may not be realized all or some of them and in particular when the model variables data is in a time series format and the explanatory variable is unstable and the random error terms are auto correlated, which leads to the fact that the least squares estimators are inefficient, which calls for the search for alternative estimation methods, one of those methods is the fully modified M method, which is a robust estimation method.

2- The research goal:

---

*

** assistant teacher Statistics department / college of Administration & Economics Mustansiriyah University
*** Assistant Professor Statistics department / college of Administration & Economics Mustansiriyah University

Iraqi Journal For Economic Sciences/ ISSN:1812-8742/https://doi.org/10.31272/IJES2020.66.6

The research aims to adopt the fully modified M method in estimating the regression model with nonstationary explanatory variables (I(1)) ,and auto correlated random error terms. This model was used to estimate the effect of barley area on the production of that crop in Iraq.

3 - The theoretical side:

This side deals with the most important concepts related to the regression model, the subject of the research, and the robust estimation method, in addition to the tests that are used.

4 - Regression model with integrated regressors and ARMA error:

The regression model can be written as:

$$Y_t = X_t^T \beta + u_{0t} \quad (a)$$
$$X_t = X_{t-1} + u_{1t} \quad (b) \qquad \ldots\ldots\ldots 1$$
$$\emptyset(B) u_{0t} = \theta(B) e_t \quad (c)$$

Where X_t the explanatory variables matrix which are nonstationary, integrated, and correlated with the regression random error terms u_{0t} , the explanatory variable matrix was a full rank matrix that is mean they are endogenous variables and can be modeled according to the random walk model as in (b),[8]. The time series of regression random error terms are auto correlated and can be modeled according to the mixed autoregressive-moving average model ARMA(p,q) as in (c). The random error terms of the explanatory variables model (u_{1t}) are stationary time series with zero mean and may be correlated with the random error series (e_t) of the ARMA(p,q), which are also stationary time series with zero mean,[5]. The coefficients of the regression model in (1) can be detailed as follows:

$$\varphi = (\beta^T, \delta^T)^T \qquad \ldots\ldots\ldots 2$$

Where φ is the regression model coefficient vector of size (1*p+q+k+1), which consist of β the vector of the explanatory variables coefficient of size (1* k+1), and $\delta^T = (\emptyset^T, \theta^T)^T$ the vector of the ARMA(p,q) coefficient of size (1*p+q), where :

$$\emptyset = (\emptyset_1, \emptyset_2, \ldots\ldots\ldots, \emptyset_p) \qquad \ldots\ldots\ldots.. 3$$
$$\theta = (\theta_1, \theta_2, \ldots\ldots\ldots.., \theta_q) \qquad \ldots\ldots\ldots\ldots 4$$

5- The assumption of the random error terms:

Assume that $\underline{U}_t$ represent the vector of random errors which included (u_{0t}) the random errors of the regression model (1-a) and the random errors of the random walk model (u_{1t}), as shown below,[6] :

$$\underline{U}_t = (u_{0t}, u_{1t}) \qquad \ldots\ldots\ldots\ldots 5$$

This vector is assumed to satisfied the following assumptions:

a- It is completely stable and mixed sequence in mixed numbers α_m , that satisfied the following:

$$\sum_1^{\infty} \alpha_m^{(l-\beta)/l\beta} < \infty \quad , \quad l > \beta > 2 \qquad \ldots\ldots\ldots\ldots 6$$

b- The moment condition which states that:

$$\left\|\underline{U}_t\right\|_l < \infty \qquad \ldots\ldots\ldots\ldots 7$$

c- The probability density function h(0) of the random errors vector $\underline{U}_t$ is symmetric, positive, and continuous over the interval (-b,b) for b > 0 , [8].

Under the three assumptions outlined above, the long-term variance and covariance matrix of the vector U_t is exists and can be portions as follows:

$$\Omega_{uu} = \sum_{j=-\infty}^{\infty} E(U_t U_{t-j}^T) = \begin{bmatrix} \Omega_{00} & \Omega_{o1} \\ \Omega_{1o} & \Omega_{11} \end{bmatrix} \quad \ldots\ldots\ldots\ldots.. 8$$

Using the sine transformation for the random errors series u_{ot} as following:

$$v_t = sign(u_{ot}) \begin{Bmatrix} 1 & for & u_{ot} \geq 0 \\ -1 & for & u_{ot} < 0 \end{Bmatrix} \quad \ldots\ldots\ldots\ldots 9$$

Defining the error vector $Z_t = (v_t, u_{01}^T)$, then the long-term variance and covariance matrix of that vector is also exists under the three error assumptions, and because the error vector v_t is a finite function of the error vector u_{ot} . variance and covariance matrix of the vector Z_t can be portions as follows:

$$\Omega_{zz} = \sum_{j=-\infty}^{\infty} E(z_t z_{t-j}^T) = \begin{bmatrix} \Omega_{vv} & \Omega_{v1} \\ \Omega_{1v} & \Omega_{11} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots 10$$

In the same way we can defined the one direction long-term variance and covariance matrix for each of the two vectors U_t and Z_t respectively as follows,[9]:

$$\Delta_{uu} = \sum_{j=0}^{\infty} E(u_t u_{t-j}^T) = \begin{bmatrix} \Delta_{oo} & \Delta_{o1} \\ \Delta_{1o} & \Delta_{11} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots 11$$

$$\Delta_{ZZ} = \sum_{j=0}^{\infty} E(z_t z_{t-j}^T) = \begin{bmatrix} \Delta_{vv} & \Delta_{v1} \\ \Delta_{1v} & \Delta_{11} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots 12$$

6- Unit root:

Unit root is a characteristic of nonstationary linear stochastic process, it occurs when the one integer is the root of the characteristic equation of that stochastic process. It is worth noting that the stochastic process is nonstationary but does not necessarily always have a general trend,[10]. If the rest of the roots of the characteristic equation have an absolute value less than the one integer (in other words, lay outside the boundaries of the unit circle), then the first difference of that process is stationary, except that the stochastic process needs to make multiple differences to achieve stationarity. There are several tests used to test whether the time series has a unit root, the most important of these tests is KPSS test.

7- KPSS test:

This test was proposed by the researchers in 1992 to test the stationarity of the time series. This test is an expansion of the augmented Dicky Filler test. This test assumes that the observed time series y_t , t = 1,2,…,T , can be represented by a three-component aggregate model, the first represents the trend component (t) , the second

represent the random walk model (r_t), and the third is the stationary random errors ($\in_t$) as shown by the following equation:

$$y_t = Bt + rt + \in_t \quad .... 13$$

$$r_t = r_{t-1} + u_t \quad \ldots\ldots\ldots\ldots.. 14$$

Where $u_t \sim IID(0, \sigma_u^2)$

The KPSS test is concerned with testing the null hypothesis, which states that the random walk has variance equal to zero against the alternative hypothesis, which states that the series is stationary by the differences, i.e. the following hypothesis test,[1]:

$H_O : \sigma_u^2 = 0 \qquad V.S \qquad H_1 : \sigma_u^2 > 0$

The test statistics used to test the above hypothesis are based on the one side multiplied Lagrange test, and Locally best invariant (LBI) test. The formula for test statistics is as follows:

$$kpss = \frac{1}{T^2}\frac{\sum_{t=1}^{T} S_t^2}{S_\epsilon^2} \qquad \text{…………… 15}$$

where S_t^2 represents the subtotal squares of the estimated sequence of residuals S_t which computed as follows:

$$S_t = \sum_{i=1}^{t} e_i \quad ; \quad t = 1,2,\dots.T \qquad \text{……………. 16}$$

And S_ϵ^2 is the estimated variance of the random errors of the time series model.

If the calculated value of the test statistics is greater than the critical value of that test

the null hypothesis is rejected which indicates the nonstationary of the time series or it is integrated of order one (I(1)), that is mean that the time series follow the random walk model, [4].

8- Robust methods for estimating model parameters:

The adoption of traditional estimation methods for estimating the regression model with nonstationary explanatory variable and aout correlated random errors leads to inefficient estimates, this affects statistical inference which calls for the adoption of robust estimation methods leading to efficient estimates of model parameters. One of these methods is the fully modified M (FM-M).

9- Fully modified M Method (FM-M):

The M estimators are a general set of robust estimation methods used in estimating the vector parameters (φ) of the regression model shown in formula (1), which is the solution to the set of equations resulting from the minimization of the following objective function, [9] :

$$\varphi_M = \operatorname{argmin}[\sum_1^T \rho(Y_t - x_t^T\beta)] \qquad \text{……………… 17}$$

Where ρ is a weights function which may be take the form ($\rho(u) = |u|^\delta$, for $\delta \in [1,2]$), and it may be take the loss function for Huber according to the following formula:

$$\rho_c(u) = \begin{cases} \left(\frac{1}{2}\right)u^2 & for|u| \le c \\ c|u| - \left(\frac{1}{2}\right)c^2 & for\ |u| > c \end{cases} \qquad \text{……………….. 18}$$

The estimates $\widehat{\varphi}_M$ are the solution to the following set of equations after equal to zero:

$$\sum_{t=1}^{T} \psi\big(e_t(\varphi)\big)e_{\varphi t} = 0 \qquad \text{…………….. 19}$$

Where $e_t(\varphi)$ is the white noise vector of the mixed model autoregressive and moving average (ARMA) which represents the random error term u_{0t} , $e_{\varphi t}$ is the vector of the white noise derivatives with respect to all model parameters as follows:

$$e_{\varphi t} = \partial e_t(\varphi)/\partial\varphi \qquad \text{……………… 20}$$

That is:

$$e_{\varphi t}(\varphi) = (e_{\beta t}^T(\varphi), e_{\emptyset t}^T(\varphi), e_{\theta t}^T(\varphi))^T \qquad \text{……………….. 21}$$

Where:

$$e_{\beta t}(\varphi) = -\theta^{-1}(B)\emptyset(B)X_t \qquad \text{……………….. 22}$$

$$e_{\emptyset t}(\varphi) = -\theta^{-1}(B)\emptyset\big(B_p\big)(Y_t - X_t^T\beta) \qquad \text{……………….. 23}$$

$$e_{\theta t}(\varphi) = \theta^{-2}(B)\emptyset\big(B_q\big)(Y_t - X_t^T\beta) \qquad \text{……………….. 24}$$

When ρ is adifferentiable and concave function, and ($\psi = \acute{\rho}$) , then the two relations (17 and 19) are equivalent and in this case there is a unique solution to relation (19), [6]. The robust M estimators are consistent but biased of second order, because in spite of removing the autocorrelation between the random error terms u_{0t} using ARMA model , and so $\psi(e_t)$ will be uncorrelated, the white noise error e_t of the ARMA model still correlated with the weighted error $\psi(e_t)$ which lead to the bias in the estimation of the vector parameter $\underline{\beta}$. To treat this drop, Philips (1995), and Dong Wan Shin and Oesook Lee (2004), suggested to make a modified on the robust M estimators to get fully modified M estimators (FM-M), in this way we can treat the second order bias and lead to estimators with approximately normal distribution so we can use the standard tests of significant like t-test or wald test,[3].

The formula of the estimated regression parameters vector according to the FM-M method is as follows:

$$\beta_M^+ = \hat{\beta}_M - [(T^{-1} \textstyle\sum_1^T \acute{\psi}\,(\hat{u}_{0t}))X^TX)^T(X^T\Delta X\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{x\psi} + T\hat{\Delta}_{x\psi}^+)] \quad \ldots\ldots\ldots 25$$

Where:

$\hat{\Omega}_{xx}$ is the long run estimated variance-covariance matrix between u_{xt-j} and u_{xt} .

$\hat{\Omega}_{x\psi}$ is the long run consistent estimated variance-covariance matrix between u_{xt} and $\psi(u_{ot)}$, where :

$$\Omega_{x\psi} = \textstyle\sum_{j=-\infty}^{\infty} E\big(u_{xt}\psi(u_{ot+j})\big) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 26$$

$\hat{\Delta}_{x\psi}^+$ is the estimate of the one direction variance-covariance matrix between u_{xt} and $\psi(u_{ot})$, where :

$$\Delta_{x\psi}^+ = \Delta_{x\psi} - \Delta_{xx}\Omega_{xx}^{-1}\Omega_{x\psi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 27$$

And

$$\Delta_{x\psi} = \textstyle\sum_{j=0}^{\infty} E\big(u_{xt}\psi(u_{ot+j})\big) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 28$$

Variance-covariance matrices estimated using kernel function, that is mean that $\hat{\Omega}_{x\psi}$ and $\hat{\Delta}_{x\psi}^+$ are the consistent kernel estimator which is one of the nonparametric estimators,[2].

The approximated distribution of the estimated parameter vector according to FM-M method is the normal distribution with mean equal to β and variance-covariance matrix ($q(X^TX)^{-1}$) , the consistent estimator of the variance-covariance matrix q as follows,[7]:

$$\hat{q} = \hat{\Omega}_{\psi\psi} - \hat{\Omega}_{\psi x}\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{x\psi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 29$$

10- Significant test of the estimated parameter vector:

As maintained above, the robust fully modified M method gives an estimator with approximately normal distribution, so we can depend on the significant tests like t-test or wald test to test the following hypothesis:

$$H_0 = R\underline{\beta} - r = 0 \qquad V.S \qquad H_1 = R\underline{\beta} - r \neq 0$$

The formula of the t-statistics for testing the above hypothesis is as follows:

$$t_i = \frac{(\beta_{FM-OLS}^+ - \beta_i)}{s_i} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 30$$

Where s_i is calculated by using the following formula for all i= 1,2,… , k :

$$s_i = \sqrt{[\hat{q}(X^TX)^{-1}]_{ii}} \qquad i = 1,2,\ldots\ldots k \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 31$$

The null hypothesis can be rejected if the calculated value of the t-statistics is greater than the tabulated value of the t-distribution, [9].

We can also use wald statistics (W^{+}) which is calculated as follows:

$$W^{+} = (R\beta_{FM}^{+} - r)^{T}\{R\hat{q}(X^{T}X)^{-1}R^{T}\}^{-1}(R\beta_{FM}^{+} - r) / \hat{q} \quad \ldots\ldots\ldots\ldots\ldots 32$$

Under the null hypothesis wald statistics (W^{+}) distributed as qui square distribution (χ_{r}^{2}) with r degree of freedom, where r represents the number of leaner constraints on the model parameters,[6].

11- The application side:

This topic deals with the data which used in estimated the regression model under consideration. These data represent (49) pair of observation as two time series for the period 1970-2018 which represent the Production of barley crop in Iraq and cultivated area, these data were obtained from the records of the Ministry of Agriculture.

12- modeling the relation between production of barley crop and cultivated area:

Considering that the production of barley crop in Iraq is the response variable, which was represented by the symbol (Y) , and the cultivated area is as the explanatory variable (X), the relation between these two variables can be modeled according to the simple regression model after taken into consideration the natural logarithm for both variables , as follows:

$$\mathrm{Ln}Y_{t} = \beta_{o} + \beta_{1}\mathrm{Ln}X_{t} + u_{ot} \quad \ldots\ldots\ldots\ldots\ldots 33$$

13- Test of the normality assumption:

To demonstrate whether the response variable fulfills the normal distribution assumption, the Shapiro-Wilk W test is adopted using the GRETL statistical software, to test the null hypothesis which states that the response variable data follows the normal distribution against the alternative hypothesis that it does not follow the normal distribution. The results of the test are shown in Table (1), which indicated the acceptance of the null hypothesis, based on the probability value of that test, which was greater than the level of significance of 0.05.

Table (1) : The results of the Normality test

variable	Shapiro-Wilk W	p-value
lnY	0.971	0.281

*From the researcher's work

14- Checking the stability of the two regression model variables:

Before estimating the regression model shown in formula (33), the stability of the two time series of the response variable (Ln y) and the explanatory variable (Ln x) must first be verified. A kpss test was performed to test the null hypothesis (H_0: $\emptyset_1$= 1) against the alternative hypothesis(H_1: $\emptyset_1$= 1) using the GRETL statistical software, and table (2) showed the results of this test, which represented the critical values of the t statistics and their probability value, from these results, we conclude that the time series of the response variable (Ln Y) is stationary, and the time series of the explanatory variable (Ln X) is nonstationary, at a significant level of 0.05 and 0.10.

Table (2) : the results of unit root test (KPSS test)

		KPSS Test	Critical value	
			5%	10%
Ln Y	**Test with constant and trend**	**0.076**	**0.14**	**0.12**
Ln X	**Test with constant and trend**	**0.154**	**0.14**	**0.12**

*From the researcher's work

The test was repeated after taking the first difference of the time series values of the explanatory variable, the test results shown in Table (3) indicated the stationary of the time series of the explanatory variable.

Table (3) : the results of unit root test after taking the first difference

variables	Type of tese	KPSS Test	Critical value	
			5%	10%
Ln X	**Test with constant and trend**	**0.085**	**0.14**	**0.12**

15-Test the independence of random errors:

Based on the Durban Watson test, the independence of the random errors series values of the regression model under research was tested, the calculated value of the D.W statistics was equal to (2.83), by comparing this value with the tabulated values of that test which are (dl= 1.503 , du= 1.585), we conclude that the random error series is auto correlated at a significant level of 0.05 . By studying the behavior of the residual time series after taken the first difference, shown in Figure (1), and its auto correlation function and the partial auto correlation function shown in the two Figures (2) and (3), we got four models, one of which can be adopted to represent the random errors series of the regression model, these models shows in table (4) and the best one is AR(2).

Figure (1): residual time series after taking the first difference

Iraqi Journal For Economic Sciences/ ISSN:1812-8742/https://doi.org/10.31272/IJES2020.66.6

Figure (2): Autocorrelation Function for diff. residuals

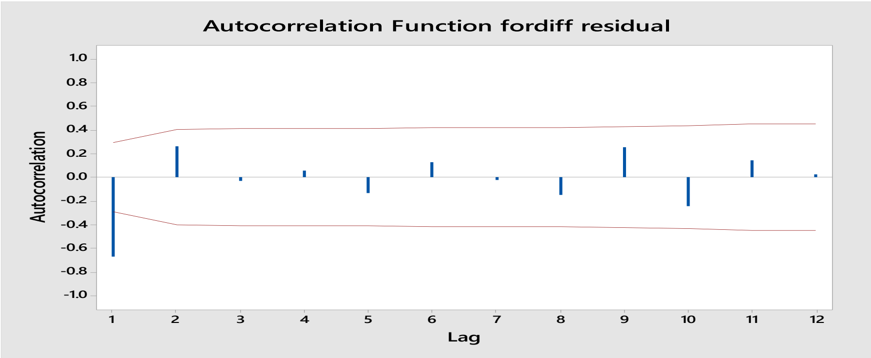
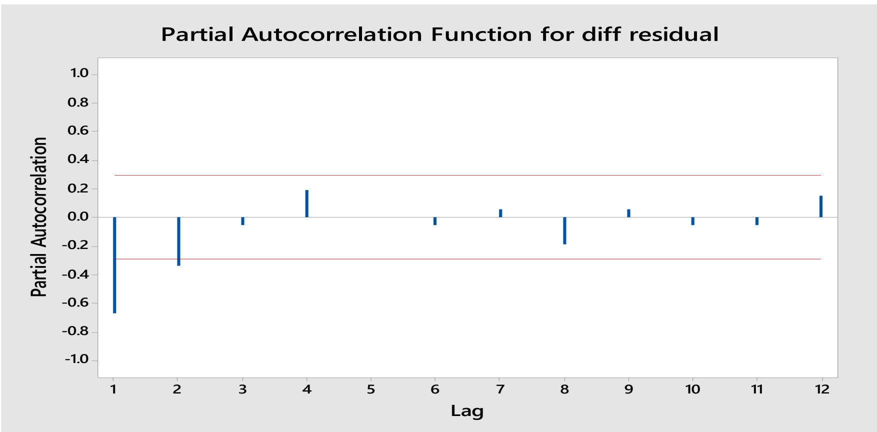


Figure (3): Partial Autocorrelation Function for diff. residuals

Model	Parameter estimate	MSE
AR(1)	ϕ_1 = - 0.671	5110.
MA(1)	θ_1 = 0.764	0.114
ARMA(1,1)	ϕ_1 = - 0.47 θ_1 = 0.496	0.107
AR(2)	ϕ_1 = -0.443	0.105
Model	Parameter estimate	MSE
AR(1)	ϕ_1 = - 0.671	5110.
MA(1)	θ_1 = 0.764	0.114
ARMA(1,1)	ϕ_1 = - 0.47 θ_1 = 0.496	0.107
AR(2)	ϕ_1 = -0.443	0.105

*From the researcher's work

Table (4): the suggested ARMA models for the residual series



16-

Estimating the regression model using FM-M:

Based on the results of previous tests which showed the stationary of the time series of the response variable (Ln y) (barley crop production), the time series of the explanatory variable (Ln x) (cultivated area) is integrated of first order I(1), Furthermore, the series of random errors are auto correlated and can be modeled by one of the mixed models ARMA. Therefore, the relationship between the response variable and the explanatory variable can be described according to the regression model shown in relation (1), and the traditional estimation methods cannot be used to estimate that model, on this basis, the robust fully modified M estimation method was used. The results of the estimation were obtained by writing a special program in Matlab language, these results shown in table (5), these results indicate the significance of the estimated parameters depending on the value of the t-statistics at the level of significance 0.05 and 0.01.In addition, the wald statistics indicated the significance of the model, as its calculated value was greater than the chi square. The value of the estimated slope parameter indicates a positive correlation between the cultivated area of the barley crop and the quantity of production for that crop. Based on the above, you can write an estimated regression model, as follows:

$$\hat{Y}_t = 3.38 + 0.36X_t \qquad \ldots\ldots\ldots\ldots\ldots\ldots 34$$

Table (5): estimated results of the regression model parameters

Estimated parameters		**Wald statistics**	**MSE**	**R^2**
$\hat{\beta}_0$	$\hat{\beta}_1$	**214.22****	**0.0123**	**0.837**
3.379 (0.4971)**	**0.358 (0.07521)****			

17- Conclusions:

a- The unit root test indicated that the time series of the production of barley crop (response variable) is stationary, while the series of the explanatory variable (cultivated area) is nonstationary and integrated of first order I(1).
b- Durban Watson test indicated that the random error series is auto correlated and can be modeled by one of the mixed ARMA models.
c- The statistical test indicated that the regression model, described in equation (33), does not fulfill the assumptions of analysis.
d- Based on the above conclusions, the relationship between the production of barley crop and cultivated area can be described according to the regression model shown in relation (1).
e- The estimated results according to fully modified M robust method indicate a positive significant relation between the production of barley crop and cultivated area.

18- Recommendations:

a- Other robust estimation methods such as : can be used to estimate the regression model parameters under consideration.
b- Adopt non-parametric estimation methods as an appropriate alternative to the robust estimation methods.

References:

1. شيخي ،محمد.(2012).طرق الاقتصاد القياسي، دار ومكتبة الحامد للنشر والتوزيع ،عمان.
2. Andrews,.d.w.k.(1991).Heteroskedasticity and autocorrelation consistent covariance ma-trix estimation.Econ.59(3).817-858.

3. Di Iorio Francesca & fachin Stefano.(2012).A Note on the Estimation of long-run relation ships in panel equations with cross-Section linkages. Econometrics vol.6, http://dx.doi.org/10.5018/economics-ejournal.ja
4. Kwiatkowski Denis& Philips peter C.B & Schmidt peter&Shin Yong cheol.(1992).Testing the null hypothesis of stationarity against the altrernative of aunit root .Journal of Econometric,159-178 north.
5. Maddala, G. S., & Kim, I. M. (1998). *Unit roots, cointegration, and structural change* (No. 4). Cambridge university press.
6. Philips peter, C.B. (1995) .Robust nonstationary regression.Cowies foundation for research in Economics yale university.
7. Phillips, P. C., & Loretan, M. (1991). Estimating long-run economic equilibria. *The Review of Economic Studies*, *58*(3), 407-436.
8. Serttas, F. O. (2010). Essays on infinite-variance stable errors and robust estimation procedures.
9. Shin, D. W., & Lee, O. (2004). M-Estimation for regressions with integrated regressors and arma errors. *Journal of Time Series* vol25,No,2,pp.283-299
10. Wei&Wiliam W.S. (2006).Time series analysis univariat and multivariate methods.2ned.