Statistical Analysis of the Variables Affecting the Determine the Total Volume of Deposits

Maie M. Kamel

Abstract: The research aims to study the effect of influencing variables determine the size and direction of deposits in the banking system. On the basis that the deposit is the main source to give a strong batch of the National Investment and filling the financing gap in the national economy. The search period thirty-one year since 1980 to 2010, a period filled with lots of economic changes. This research is based on the standard statistical method and quantitative economic analysis.

Keywords:, Autocorrelation, Homo-scedasticity, Methodology of Econometrics, Multicollinearity

---

1- INTRODUCTION:

Search stretched over thirty-one year since 1980 to 2010, a period filled with many of the economic changes, Such as the appearance of the actual impact of economic openness, which began in 1974, and the period of economic reform in the beginning of the nineties and the beginning of a privatization since 1996, As well as the occurrence of some economic variables such as domestic and foreign second Gulf War in 1991. The decline in global economic activity and local levels and the emergence of a recession in the Egyptian economy.
In this study was to identify a number of factors or variables of importance to the national economy and directly affect the banking sector and specifically on the size of deposits.
Those variables are Gross Domestic Product (GDP), Money Supply, and the volume of Foreign Trade and population, which can be defined as follows:
- GDP is the sum of the values for each of the cash commodity production and service of the economy during a certain period of time. (Variable has a positive impact on the volume of deposits).
- Money Supply is the amount and effective means or domestic liquidity include all of the cash in circulation outside the banking system and deposits, current and non-current local currency and foreign currency deposits.(Variable has a negative impact on the volume of deposits).
- Foreign trade volume of exports and imports require open letters of credit banks. (Variable has a positive effect on the size of deposits).
- Population are based on the generation of GNP and all aspects of economic activity (Variable has a positive effect on the size of deposits).

2- METHODOLOGY

The economic theory test means verification of compatibility with the practical reality, and then be accepted, amended or rejected or to reach a new theory**.**
The econometric methodology is determined as follows:

2-1 Builds The Standard Model

this is known as the expression of economic theory in the form of an equation or set of equations, the equation is a relationship between the dependent variable and the independent variable and one or several independent variables.

Dependent Variable = f (Independent Variables)

The type of mathematical formula to offset what is suggested by economic theory, or that which is inspired by the shape of the spread of points dependent variable with the independent variables one at a time, or as evidenced Applied Studies earlier., And you must know the theoretical predictions of the relations between the variables under study, as there are some assumptions basic to build econometric models, and that the consequent drop one of them or all of them some of the problems which may affect the ability of the predictive model and interpret estimates.

2-1-1 Basic Assumptions To Build An Economic Model:
1) the dependent variable linear function in the coefficients of independent variables plus term of error. The model takes the following form:

$$Y_i = B_0 + B_1 X_{1i} + \ldots\ldots + B_k X_{ki} + e_i \qquad (1)$$

Where: i = 1, 2, n

Y_i refers to the dependent variable,
B_0 refers to the constant term.
$X_{1i}, \ldots, X_{ki}$ refers to the independent variables,
$B_1, \ldots\ldots, B_k$ refers to Partial regression coefficients of linear regression model.
e_i refers to the extent of the error,

And the drop that hypothesis occur an error the identification and formulation of the relationship of economic, as a result ignored or eliminated some of the important independent variables in determining the form, or the inclusion of some unimportant independent variables in the model, or that the true relationship between the dependent variable and the independent variables relationship is non linear or change of coefficient regression during the time period in which data were collected through them.
2) The expected value for error be equal to zero

$$\mathbf{E}(e_i) = \mathbf{O} \qquad (2)$$

And the consequent drop that hypothesis of a problem the constant Bias.
The bias is the difference between the mathematical expectation parameter estimates and estimated its real value,

(Bias = $E(\hat{\theta}) - \theta$)

3) **Homo-Scedasticity**

$$\mathrm{var}(e_i) = \mathrm{E}(e_i^2) = \sigma_e{}^2 \qquad (3)$$

This means that the variance error limit for not dependent on the values of the independent variables in the case of multiple regression, and the consequent drop that hypothesis a Hetero-scedasticity problem that limits the errors do not have the same variance.

4) Independence Errors From Each Other.
In the sense that the estimated value reduce error in a certain period of time independent of the appraised value reduce error in an unprecedented period of time

$$Cov\ (e_i e_j) = \mathrm{E}(e_i e_j) = 0 \qquad i \neq j \qquad (4)$$

And the consequent drop the hypothesis that the problem of Autocorrelation

5) Errors Follow Normal Distribution Allowing Testing Hypotheses.
Based on these assumptions (2-4) past can be incorporated as follows

$$e_i \sim i.i.d.N(0, \sigma_e^2) \qquad (5)$$

6) **Independent variables non-random variables** and the consequent drop the hypothesis of a record problems such as There are errors in the measurement of the independent variables.

7) Complete lack of linear relationship between the independent variables. And the consequent drop the hypothesis that the problem of Multicollinearity.

8) Error limit independence from the independent variable for each watch.

$$\mathrm{cov}(X_i u_i) = \mathrm{E}(X_i u_i) = 0 \qquad (6)$$

2-2 Estimate Standard Model

To reach accurate estimates of the values of the coefficients of the proposed models, and considers the ordinary least squares (OLS) method of the most used methods to obtain these estimates with the need for the availability of the previous assumptions. Since the method of least squares capabilities Best Linear Unbiased Estimators (BLUE)

2-3 Evaluation of the standard model is estimated
Estimates are evaluated based on three basic criteria:

2-3-1 First criterion: economic criterion (logical)
Represents the agreement of the results with the text of economic theory. Where they are sure of the extent of agreement signals the estimated regression coefficients with the expected signs according to economic theory .Also be sure of the extent of agreement estimated values for the regression coefficients with counterparts contained in economic theory.

2-3-2 Second criterion: statistical criterion
According to this criterion is the evaluation process in accordance with the following stages:

2-3-2-1 Evaluate the extent of the moral faculty for the estimated standard model.
Using the coefficient of determination R^2 to express the explanatory power of the models (the responsibility of the explanatory variables in explaining the changes in the dependent variable of the economic relationship). The coefficient of determination represents the ratio between the sum of the squares regression to the total sum of squares can be formulated as follows

$$R^2 = \frac{\text{SS regression}}{\text{SS total}} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (7)$$

The structure of the test of the relationship linear regression as follows:
Null hypothesis: all regression coefficients equal to zero.

H_0: all of $B_i = 0$ i=1,2,….,k

Against the alternative hypothesis:. No regression coefficient at least one not equal to zero

H_1: at least one of $B_i \neq 0$ i=1,2,….,k

the statistics used to test the Null hypothesis is the F statistic

$$F = \frac{R^2/k}{(1 - R^2)/n - k - 1} \qquad (8)$$

And the decision is made according to the following rule: accepted that the regression coefficients are not all equal to zero if the p-value (sig.) of the statistic is less than the used level of significant (α).This indicates that the estimated regression equation as a whole statistically significant in explaining the behavior of the dependent variable of the economic relationship.

2-3-2-2 Significant test for the estimated parameters of the regression model.
Parameters are evaluated regression models estimated separately, after ascertaining the extent of acceptance in accordance with the criteria of economic and make sure the model as a whole is significant and the structure of the test as follows

$H_0 : B_i = 0$ i=0,1,2,.....,k

$H_1 : B_i \neq 0$ i=0,1,2,.....,k

the statistics used to test the hypothesis is the t statistic

$$t = \frac{\hat{B}_i}{S.E(\hat{B}_i)} \qquad (9)$$

This test follows the probability distribution of student T. The decision is in accordance with the following rule: the alternative hypothesis is accepted the if the p-value (sig.) of the statistic is less than the used level of significant (α)

2-3-2-3 Test the overall performance of the estimated standard models.
It reflects the ability of the model to predict the estimated values of past or present during the time period of the estimate. This test is done according to one of the following methods

2-3-2-3-1 Calculate the coefficient of determination.
That whenever the value of the high coefficient of determination whenever explained on the quality of the estimated model to predict the current values during the time period of the estimate.

2-3-2-3-2 Test significant difference between the predicted value and the actual value
The structure of the test as follows

$$H_0 : \hat{y}_i - y_i = 0$$
$$H_1 : \hat{y}_i - y_i \neq 0$$

IJSER

Where: y_i is the watch for value of the variable, $\hat{y}_i$ is the predicted value.
The statistics used is the t-test and be worded as follows:

$$t = \frac{y_i - \hat{y}_i}{\hat{\sigma}_u \sqrt{1 + \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}}} \qquad (10)$$

This test follows the distribution of T. the decision is made according to the following rule:, the null hypothesis is accepted (no significant difference between the predicted value and the actual value),if the p-value (sig) is exceed the level of significance used. This decision which indicates the quality of the predictive ability of the model estimated with respect to No. i

2-3-2-3-3 Theil coefficient of inequality (U)

and knows Theil coefficient of inequality (U) as follows

$$U = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} Y_i^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i^2}} \qquad (11)$$

The value of the parameter U lies between zero and one

$0 \leq U \leq 1$

- If the parameter value is equal to zero U = 0, this indicates that fully reconcile the estimated model to predict during the time period of the estimate.
- If the parameter value is equal to the one true U = 1, it indicates poor ability Predictive models to predict estimated during the time period of the estimate.
- If the parameter value ranged between zero and one $0 < U < 1$, In this case must describes the various sources of error in forecasting.

2-3-3 Third criterion: The Econometric Standard

This criterion measures the extent Reliability criterion in assessing the statistical model Proposal. Among these tests DW (Durbin Watson) test for autocorrelation, and test Multicollinearity, and test of Herteroscedasticity

2-4 Using the model to Forecast

The main objective of the standard models is to predict future values of the variable values on the basis of the known or anticipated future of the independent variables .
It can predict the values of the dependent variable in two ways:

First : Predict point prediction point

It is intended to predict the future value of one variable 's , under the assumption of a certain value Independent variables separately.

Second : prediction for the Interval Prediction

It is intended to predict the period which can be the future of the true value of the significant value equal to (1- α)% Where the limits of confidence , which takes the following form:

$$\Pr\{\hat{Y}_t - t_{\alpha/2,(n-k)}\, se(\hat{Y}_t) \leq Y_t \leq \hat{Y}_t + t_{\alpha/2,(n-k)}\, se(\hat{Y}_t)\} = (1-\alpha) \qquad (12)$$

Where :

$$se(\hat{Y}_t) = \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{(X_t - \bar{x})^2}{\sum_{i=1}^{n} (X_i - \bar{x})^2}\right)} \qquad (13)$$

3-Applied study:

Data were obtained from the Egyptian Central Agency for Public Mobilization and Statistics (CAMPS) and Egyptian Cabinet Information and Decision Support Center (IDSC).Data covers the period from 1980 to 2010. Computer Program SPSS17, Minitab 16 and E-VEIWS7 are used in analysis and the same results had.
According to the concept of economic theory, the volume of deposits is estimated as a function of the independent variables affecting it, as shown next model

$$y_i = u + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_i$$

Where is :

The dependent variable is :

y The volume of deposits in the Egyptian banking system

The Independent Variables are:

X_1 Gross Domestic Product (GDP),
X_2 Money Supply ,
X_3 the volume of Foreign Trade
X_4 population.

e_i the error term where $e_{i \sim} IN(0, \sigma_e^2)$

Applying the method of ordinary least squares (OLS) on the data available. Obtaining the following results

Descriptive Statistics			
	Mean	Std. Deviation	N
Deposits	1.8691E5	1.65062E5	31
G.D.P	2.8370E5	3.02663E5	31
Money Supply	2.4815E5	2.61882E5	31
Foreign Trade	2.7058E4	22533.96567	31
Population	60.9742	12.78384	31

Table (1)

Table (1) shows the values of means and standard deviations and numbers of the dependent variable and the independent variables.

Correlations						
		Deposits	G.D.P	Money Supply	Foreign Trade	Population
Deposits	Pearson Correlation	1	.985**	.978**	.865**	.965**

	Sig. (2-tailed)		.000	.000	.000	.000
	N	31	31	31	31	31
G.D.P	Pearson Correlation	.985**	1	.993**	.830**	.950**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	31	31	31	31	31
Money Supply	Pearson Correlation	.978**	.993**	1	.863**	.963**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	31	31	31	31	31
Foreign Trade	Pearson Correlation	.865**	.830**	.863**	1	.884**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	31	31	31	31	31
Population	Pearson Correlation	.965**	.950**	.963**	.884**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	31	31	31	31	31
**. Correlation is significant at the 0.01 level (2-tailed).						

Table (2)

Table (2) Shows the correlation coefficients matrix and the level of significance between all the independent variables and the dependent variable. Through the matrix, we find that GDP has the greatest impact on the dependent variable, followed by the money supply and population.

We also find that there are internal correlations between the independent variables, suggesting that we are in the process of Multicollinearity problem.

To test the hypothesis H_0: $\rho = 0$ VS H_1 : $\rho \neq 0$

The p value is less than 0.0001 indicating that the correlation is significant. Which describes the strength of the linear relationship between variables.

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.994[a]	.989	.987	18779.64897	2.070

a. Predictors: (Constant), Population, Foreign Trade, G.D.P, Money Supply

b. Dependent Variable: Deposits

Table(3)

According to the Table (3)We note that the independent variables explain 98.9% of the change in the dependent variable and differences remaining 1.1% as a result of random variations and this is illustrated by the value of the coefficient of determination R^2.also find that the value of test D.W refers to the absence of the problem of autocorrelation.

ANOVA[b]					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	8.082E11	4	2.020E11	572.906	.000[a]
Residual	9.170E9	26	3.527E8		
Total	8.174E11	30			

a. Predictors: (Constant), Population, Foreign Trade, G.D.P, Money Supply

b. Dependent Variable: Deposits

Table (4)

According to Table (4) reports the ANOVA result for the model, we find that the significance of the model is less than 0.001 Therefore we can say that the statistically significant regression.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	-188210.177	52585.58		-3.579	0.001		
G.D.P	0.848	0.111	1.555	7.628	0.000	0.010	96.378
Money Supply	-0.657	0.149	-1.043	-4.408	0.000	0.008	129.63
Foreign Trade	1.314	0.361	0.179	3.639	0.001	0.178	5.629

IJSER

Population	4296.737	1082.106	0.333	3.971	0.001	0.061	16.278

a. Dependent Variable: Deposits

Table (5)

Table (5) shows the results of the regression analysis and estimating constant and regression coefficients where they are statistically significant.
The model is

Deposits = -188210.177 + 0.848 G.D.P - 0.657 Money Supply +1.314 Foreign Trade + 4296.737 Population

From standardized coefficient column, we find that GDP and money supply are more variables with the largest impact on the volume of deposits.
To detect the presence of Multicollinearity problem between the independent variables

We find that the value of the variance inflation factor ($VIF_i = \frac{1}{r_{ii}}$) if it is greater than 100 was evidence of a high linear-correlation between the independent variable associated with this parameter with the rest of the other independent variables.
We found that money supply has a higher VIF then G.D.P then Population.
To find out any independent variables requires entry into the model is the use of so-called Tolerance scale which is the inverted to VIF .Tolerance scale must be less than or equal to 0.01
According to this scale the variables must be in equation are money supply and G.D.P.

Evaluate the predictive ability of the model proposed:

Theil coefficient of inequality U = 0.0003, and that value refers to the quality of the estimated model to predict during the time period of the estimate.
In this search we cannot compare the actual values and the predicted values due to the political and economic conditions and the absence of stability experienced by the country since 2011 and so far.

4- Conclusion:

Statistical analysis showed a strong relationship between the total volume of deposits in the Egypt banking system and independent factors affecting it, such as GDP ,Remittances from Egyptians working abroad , Money Supply, Foreign Trade and Population.
From the study, we find a structural relationship between variables (the dependent variable and the independent variables), which corresponds to the economic theory.
One of the challenges that directs the Egyptian banking system is working to increase the volume of deposits because of their key role in financing economic development programs.
It should be noted that the presence of Multicollinearity is not considered a problem in itself, but the problem is in the degree of Multicollinearity. If the degree of Multicollinearity low can accept such couplings, where it is expected to continue the case of Multicollinearity between the variables in the future to the same degree as they were in appreciation of the period. If, however, a high degree of coupling will lead to the difficulty of determining the estimated values for the regression coefficients.

REFERENCES

[1] Doane, D. & L. Seward. 2007, Applied Statistics, Mc Graw..Hill Irwin,2nd Edition.

[2] Greene, W.H. 2000. Econometric Analysis ,4th Edition. New Jersey: Prentice Hall

[3] Maddala ,G.S. 1977, Econometrics. McGRAW HILL USA.

[4] Mirer, T. 1990, Economic Statistics and Econometrics: 2nd Edition Maxwell Macmillan International Editions.

[5] Stock, J. & M. Watson. 2005. Introduction to Econometrics. Pearson Education Inc.

Details about Author:

Maie M. Kamel
statistic Department
Faculty of commerce, Tanta University
Egypt
maie.m.kamel@gmail.com

IJSER