Memento on EViews Output*

Jonathan Benchimol†

This version: December 27, 2020
First version: February 10, 2008

Abstract

Running a simple least square regression requires to satisfy several hypotheses. This technical guide explains outputs and interpretations from standard econometric procedures in Eviews. Simple examples and estimations are detailed to avoid spurious econometric statements, unfortunately, frequent in economic research.

Keywords: stationarity, spurious regression, robustness, identification.
JEL Classification: C13, C22, C87.

Please cite this paper as:
Benchimol, J., 2020. Memento on EViews output. Technical Guide, Mimeo.

*This paper does not necessarily reflect the views of the Bank of Israel. I thank Michael Bell, Itamar Caspi, Guillaume Chevillon, André Fourçans, and Ben Schreiber for their constructive comments.

†Bank of Israel, Jerusalem, Israel. Email: jonathan.benchimol@boi.org.il.

1 Introduction

This paper is beneficial in understanding fundamental econometric concepts and, more specifically, avoiding spurious regressions and misinterpretations of Eviews outputs. It intends to become a useful guide for economists desiring to conduct proper estimations through Eviews. The data and replication files are available online.

2 Ordinary Least Squares

The Ordinary Least Squares (OLS) method is one of the most used estimation techniques, both in research and industry. This linear least-squares method estimates the unknown parameters in a linear regression model: it chooses the parameters of a linear function of a set of explanatory variables by minimizing the sum of the squares of the differences between the observed dependent variable[1] in the given dataset and those predicted by the linear function.

Before starting coding or writing, always ask this question: which research question do I want to answer? If the objective is to understand the connection and causalities between x_t and y_t, which are two economic variables, the corresponding data (time series) have to be available for your study.

For instance, how energy and consumer prices are related? To answer this, we have to select the relevant data corresponding to the research question. We choose the Domestic Producer Prices Index (Manufacturing) for Israel (x_t) and the Consumer Price Index Energy for Israel (y_t) to analyze this question.[2] Our analyses span from August 1997 to May 2017, at a monthly frequency.

2.1 Stationarity

In order to use stationary time series without affecting our results by seasonal effects, we compute the percentage growth of these two seasonally-adjusted[3] time series, dx_t and dy_t. Table 1 presents the stationarity tests based on Dickey and Fuller (1979).

Table 1 shows that our time series, dx_t and dy_t, are stationary. This property is *essential*[4] for OLS estimation, as we will see below.

[1] Values of the variable being predicted.

[2] Energy includes electricity, gas and other fuels & fuels and lubricants for personal transport equipment. It excludes water. Energy is 7.309 % of the CPI all items in 2008.

[3] We adjust for seasonality by using X12-ARIMA(0,1,1).

[4] Stationarity is even a *necessary* condition for a non cointegration analysis.

Table 1. Unit Root Tests

Null Hypothesis: dx_t has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-10.87532	0.0000
Test critical values:	1% level	-3.457984	
	5% level	-2.873596	
	10% level	-2.573270	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: $\Delta\,(dx_t)$
Method: Least Squares
Sample (adjusted): 1997M10 2017M05
Included observations: 236 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
dx_{t-1}	-0.672878	0.061872	-10.87532	0.0000
C	0.143256	0.063116	2.269741	0.0241

R-squared	0.335742	Mean dependent var	-0.001983
Adjusted R-squared	0.332903	S.D. dependent var	1.160250
S.E. of regression	0.947646	Akaike info criterion	2.738766
Sum squared resid	210.1396	Schwarz criterion	2.768121
Log likelihood	-321.1744	Hannan-Quinn criter.	2.750599
F-statistic	118.2725	Durbin-Watson stat	2.002750
Prob(F-statistic)	0.000000		

Null Hypothesis: dy_t has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.75033	0.0000
Test critical values:	1% level	-3.457984	
	5% level	-2.873596	
	10% level	-2.573270	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: $\Delta\,(dy_t)$
Method: Least Squares
Sample (adjusted): 1997M10 2017M05
Included observations: 236 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
dy_{t-1}	-0.893900	0.065009	-13.75033	0.0000
C	0.274881	0.133557	2.058150	0.0407

R-squared	0.446902	Mean dependent var	-0.001479
Adjusted R-squared	0.444538	S.D. dependent var	2.721591
S.E. of regression	2.028383	Akaike info criterion	4.260793
Sum squared resid	962.7545	Schwarz criterion	4.290147
Log likelihood	-500.7736	Hannan-Quinn criter.	4.272626
F-statistic	189.0716	Durbin-Watson stat	1.997194
Prob(F-statistic)	0.000000		

Note: Augmented Dickey-Fuller unit root tests (e.g., stationarity tests) for dx_t (left panel) and dy_t (right panel).

Additional stationarity tests exist, namely the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and the Phillips-Perron (PP) tests. All three tests should be conducted as a robustness check for stationarity.

Like an Augmented Dickey-Fuller (ADF) test, the null hypothesis for the PP test is that the series possesses a unit root and is not stationary. One distinct advantage of the PP over the ADF test is that it is a non-parametric test. Consequently, it applies to a broad set of problems. The ADF test uses a parametric autoregression to approximate the series's error process, whereas the PP test does not assume such a functional form of the errors. The PP test also adjusts for serial correlation and heteroscedasticity in the errors by modifying the test statistic. A disadvantage of the non-parametric PP test is that it requires a large sample size as it relies upon asymptotic theory, and large datasets are not always readily available.

For the KPSS test, the null hypothesis is that the series is stationary, i.e., does not possess a unit root. The KPSS test treats errors similar to the PP test, considers the case of a general error process (assuming no functional form), and suitably modifies the test statistic like in the PP test.

2.2 Causality

A pairwise Granger (1969) causality test is presented in Table 2 and shows that we cannot reject the hypothesis that dy_t does not Granger cause dx_t but we do reject the hypothesis that dx_t does not Granger cause dy_t. Therefore it appears that Granger causality runs one-way from dx_t to dy_t and not the other way.

Table 2. Granger Causality

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
dy_t does not Granger Cause dx_t	235	1.64321	0.1956
dx_t does not Granger Cause dy_t		7.50906	0.0007

Note: Pairwise Granger causality tests between dx_t and dy_t.

In other but more precise words, Table 2 shows that dx_t statistically causes dy_t.

2.3 Correlogram

Table 3 presents the correlograms of dx_t and dy_t. The autocorrelation of the series dx_t is not very big at lag one, and quasi inexistent in the next lags. The partial autocorrelation of the series dx_t is quasi inexistent. However, the Ljung and Box (1978) Q-statistics and their p-values show that the series contains some autocorrelation at several orders. This correlogram could motivate the use of an AR(1) component to the next estimations, including dx_t as the variable to explain.

Table 3 shows there is no autocorrelation nor partial autocorrelation for the series dy_t.

2.4 Linear Estimation

Assuming that related assumptions concerning the OLS regression are verified, the results presented in Table 4 show a significant relationship between dx_t and dy_t, with a good coefficient of determination[5] (Adjusted R2 around 0.63) and without autocorrelation of order one.[6]

2.5 Validation

Our OLS regression satisfies all the linear regression assumptions presented below and is significant according to statistics examined about the regression (Adjusted R2, Durbin-Watson, t-stat/p-values) as well as about the residuals (cf. above).

2.5.1 Strict Exogeneity and Normality of the Residuals

Fig. 1 shows that residuals are normally distributed with a quasi-zero average. The Jarque-Bera test confirms residuals' skewness and kurtosis match a normal distribution.

2.5.2 Linear Dependence

According to a simple cross-correlation between the two series (Table 5), there is no collinearity between our variables.

---

[5] The coefficient of determination is explained in Section 3.3.

[6] The Durbin and Watson (1950, 1951, 1971) test is close to 2.

Table 3. Autocorrelation and partial correlation

Included observations: 237

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|** |	.|** |	1	0.326	0.326	25.495	0.000
.|* |	.|. |	2	0.111	0.006	28.488	0.000
.|. |	.|. |	3	0.060	0.025	29.362	0.000
.|. |	.|. |	4	0.055	0.031	30.101	0.000
.|. |	.|. |	5	-0.009	-0.043	30.121	0.000
.|. |	.|. |	6	0.017	0.031	30.188	0.000
.|. |	.|. |	7	0.026	0.015	30.357	0.000
*|. |	*|. |	8	-0.085	-0.113	32.142	0.000
.|. |	.|* |	9	0.038	0.112	32.493	0.000
.|. |	.|. |	10	0.068	0.034	33.653	0.000
.|. |	.|. |	11	0.008	-0.035	33.670	0.000
*|. |	*|. |	12	-0.106	-0.109	36.500	0.000
*|. |	.|. |	13	-0.108	-0.064	39.457	0.000
.|. |	.|. |	14	-0.016	0.060	39.525	0.000
.|. |	.|. |	15	-0.002	0.012	39.526	0.001
.|. |	.|. |	16	0.018	0.006	39.613	0.001
.|. |	.|. |	17	-0.038	-0.046	39.988	0.001
.|. |	.|* |	18	0.073	0.120	41.371	0.001
.|. |	.|. |	19	0.066	0.019	42.493	0.002
.|* |	.|. |	20	0.086	0.031	44.402	0.001
.|* |	.|. |	21	0.093	0.050	46.668	0.001
.|. |	.|. |	22	0.057	0.024	47.516	0.001
.|. |	*|. |	23	-0.040	-0.070	47.940	0.002
*|. |	*|. |	24	-0.071	-0.066	49.268	0.002

Included observations: 237

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|* |	.|* |	1	0.106	0.106	2.7008	0.100
.|. |	.|. |	2	-0.001	-0.013	2.7012	0.259
.|. |	.|. |	3	-0.014	-0.013	2.7489	0.432
.|. |	.|. |	4	0.006	0.008	2.7563	0.599
.|. |	.|. |	5	0.061	0.060	3.6569	0.600
.|. |	.|. |	6	0.026	0.013	3.8215	0.701
.|. |	.|. |	7	0.004	0.000	3.8250	0.800
.|. |	.|. |	8	-0.051	-0.050	4.4588	0.814
.|. |	.|. |	9	0.007	0.018	4.4720	0.878
.|. |	.|. |	10	0.054	0.048	5.1923	0.878
.|* |	.|. |	11	0.083	0.071	6.9292	0.805
*|. |	*|. |	12	-0.171	-0.192	14.306	0.282
*|. |	*|. |	13	-0.189	-0.152	23.379	0.037
.|. |	.|. |	14	-0.033	0.004	23.653	0.050
.|. |	.|. |	15	0.029	0.030	23.861	0.067
.|. |	.|. |	16	0.054	0.036	24.597	0.077
.|. |	.|. |	17	0.004	0.011	24.601	0.104
.|. |	.|. |	18	0.042	0.072	25.068	0.123
.|. |	.|. |	19	0.002	0.012	25.069	0.158
.|. |	.|. |	20	0.057	0.039	25.910	0.169
.|. |	.|. |	21	0.040	0.005	26.321	0.194
.|. |	.|* |	22	0.066	0.074	27.460	0.194
*|. |	.|. |	23	-0.085	-0.063	29.361	0.169
*|. |	*|. |	24	-0.104	-0.104	32.244	0.121

Note: autocorrelation and partial correlation for dx_t (left panel) and dy_t (right panel).

Table 4. CPI Energy: Estimation

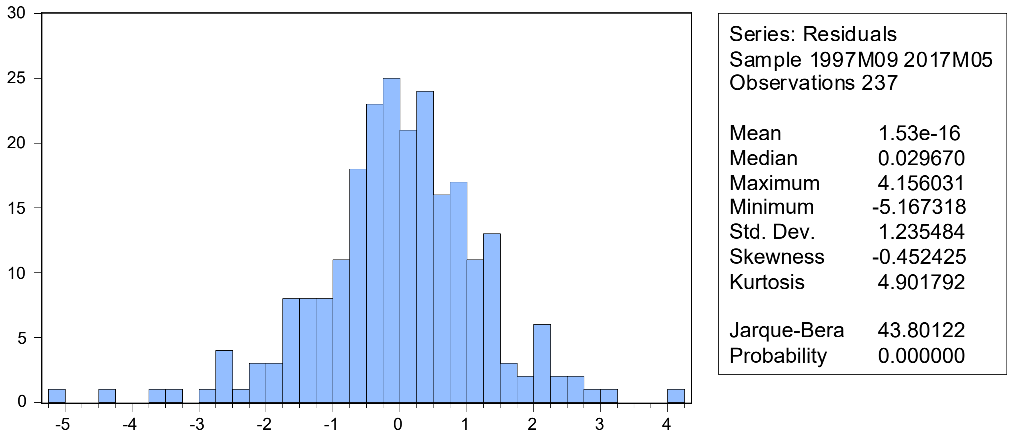
Dependent Variable: dy_t
Method: Least Squares
Sample (adjusted): 1997M09 2017M05
Included observations: 237 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.034746	0.082221	-0.422586	0.6730
dx_t	1.614157	0.080687	20.00522	0.0000

R-squared	0.630043	Mean dependent var	0.307295
Adjusted R-squared	0.628469	S.D. dependent var	2.031241
S.E. of regression	1.238110	Akaike info criterion	3.273452
Sum squared resid	360.2353	Schwarz criterion	3.302718
Log likelihood	-385.9041	Hannan-Quinn criter.	3.285248
F-statistic	400.2090	Durbin-Watson stat	2.175794
Prob(F-statistic)	0.000000		

Note: estimation of dy_t.

Figure 1. Histogram of Residuals



Note: the skewness measures the asymmetry of the distribution relative to the average. While it differentiates extreme values in one versus the other tail, kurtosis measures extreme values in either tail.

Table 5. Simple Cross-correlations

dx_t,dy_{t-i}	dx_t,dy_{t+i}	i	lag	lead
.|******** |	.|******** |	0	0.7938	0.7938
.|** |	.|** |	1	0.2105	0.2238
.|. |	.|* |	2	0.0291	0.0694
.|. |	.|. |	3	0.0006	0.0148
.|. |	.|. |	4	0.0417	0.0016
.|* |	.|. |	5	0.1123	-0.0334

Note: simple cross-correlation between dx_t and dy_t. Correlations are asymptotically consistent approximations.

2.5.3 Homoscedasticity

There is no heteroscedasticity according to several heteroscedasticity tests presented in Table 6.

2.5.4 Autocorrelation

According to a correlogram of the residuals, there is no autocorrelation for all lags considered. This is also the case when testing the square of the residuals (not displayed).

3 Generalized Method of Moments

The starting point of the Generalized Method of Moments (GMM) estimation is a theoretical relation that the parameters should satisfy. The idea is to choose the parameter estimates so that the theoretical relation is satisfied as "closely" as possible. Its sample counterpart replaces the theoretical relation, and the estimates are chosen to minimize the weighted distance between the theoretical and actual values. GMM is a robust estimator in that, unlike maximum likelihood estimation, it does not require information about the exact distribution of the disturbances. In fact, many common estimators in econometrics can be considered as special cases of GMM.

The theoretical relation that the parameters should satisfy are usually *orthogonality conditions* between some (possibly nonlinear) function of the parameters

Table 6. Heteroskedasticity Tests

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	1.276037	Prob. F(1,235)	0.2598
Obs*R-squared	1.279947	Prob. Chi-Square(1)	0.2579
Scaled explained SS	2.455077	Prob. Chi-Square(1)	0.1171

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Sample: 1997M09 2017M05
Included observations: 237

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.473073	0.199691	7.376753	0.0000
dx_t	0.221365	0.195965	1.129618	0.2598

R-squared	0.005401	Mean dependent var	1.519980
Adjusted R-squared	0.001168	S.D. dependent var	3.008764
S.E. of regression	3.007006	Akaike info criterion	5.048170
Sum squared resid	2124.891	Schwarz criterion	5.077436
Log likelihood	-596.2082	Hannan-Quinn criter.	5.059966
F-statistic	1.276037	Durbin-Watson stat	1.914559
Prob(F-statistic)	0.259789		

Heteroskedasticity Test: Harvey

F-statistic	1.281931	Prob. F(1,235)	0.2587
Obs*R-squared	1.285827	Prob. Chi-Square(1)	0.2568
Scaled explained SS	1.364489	Prob. Chi-Square(1)	0.2428

Test Equation:
Dependent Variable: LRESID2
Method: Least Squares
Sample: 1997M09 2017M05
Included observations: 237

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.163867	0.152199	-7.647003	0.0000
dx_t	0.169108	0.149359	1.132224	0.2587

R-squared	0.005425	Mean dependent var	-1.128033
Adjusted R-squared	0.001193	S.D. dependent var	2.293226
S.E. of regression	2.291857	Akaike info criterion	4.505005
Sum squared resid	1234.363	Schwarz criterion	4.534272
Log likelihood	-531.8431	Hannan-Quinn criter.	4.516801
F-statistic	1.281931	Durbin-Watson stat	1.685322
Prob(F-statistic)	0.258694		

Heteroskedasticity Test: ARCH

F-statistic	0.401994	Prob. F(1,234)	0.5267
Obs*R-squared	0.404734	Prob. Chi-Square(1)	0.5247

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Sample (adjusted): 1997M10 2017M05
Included observations: 236 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.461595	0.220163	6.638699	0.0000
RESID(-1)^2	0.041402	0.065300	0.634030	0.5267

R-squared	0.001715	Mean dependent var	1.524594
Adjusted R-squared	-0.002551	S.D. dependent var	3.014319
S.E. of regression	3.018162	Akaike info criterion	5.055611
Sum squared resid	2131.576	Schwarz criterion	5.084966
Log likelihood	-594.5621	Hannan-Quinn criter.	5.067444
F-statistic	0.401994	Durbin-Watson stat	2.000932
Prob(F-statistic)	0.526681		

Note: heteroskedasticity tests following the estimation presented in Table 4.

$f(\theta)$ and a set of instrumental variables z_t:

$$E\left[f(\theta)' Z\right] = 0 \tag{1}$$

where θ are the parameters to be estimated. The GMM estimator selects parameter estimates so that the sample correlations between the instruments and the function f are as close to zero as possible, as defined by the criterion function:

$$J(\theta) = (m(\theta))' A m(\theta) \tag{2}$$

where $m(\theta) = f(\theta)' Z$ and A is a weighting matrix.

3.1 Instrumental Variables

The need for an instrument variable arises due to the endogeneity of the explanatory variable in a regression. An explanatory variable is said to be endogenous if it is correlated with the model's error term.

An instrument variable is one of the ways by which we can overcome the endogeneity issue. An instrument variable needs to specify two critical properties:

- **Instrument Exogeneity**: The instrument should be uncorrelated with the error term. It is not possible to test this assumption. One could argue in favor of Instrument Exogeneity by appealing to established economic theory and behavior.

- **Instrument Relevance**: The instrument must be a valid proxy for the endogenous explanatory variable. This can be tested by running a regression of the endogenous variable on the instrument and assessing if the regression coefficient is significant.

When the number of Instrument Variables is as much as the number of Endogenous Variables, one of the methods used is the two-stage least squares (2SLS) method. For most statistical packages, including EViews, we have a single command to run a 2SLS regression. One needs to specify the dependent variable, the independent variable, and select a set of instruments. The interpretation of the results is similar to that of an OLS regression.

It is generally good practice to conduct tests for endogeneity and compare OLS and IV estimates. Provided that the choice of the instrument used is sound, it will generally be the case that IV estimates will be unbiased whereas OLS estimates will only be unbiased under certain conditions (Cov(X,U) = 0). The null hypothesis is that the variable is exogenous.

The three indicators and their threshold values for rejecting the null hypothesis to look out for when doing these tests are as follows:

- **Durbin-Waston** (DW): Reject the null if value is greater than 10.
- **Wu-Hausman** (F-stat): Reject the null if value is greater than 10.
- **Significance** (p-value): A low p-value depending on the significance level.

Rejecting the Null Hypothesis suggest that the instrument variable is endogenous.

3.2 J-statistic

The J-statistic is the minimized value of the objective function, where we report Eq. 2 divided by the number of observations. This J-statistic may be used to carry out hypothesis tests from GMM estimation. A simple application of the J-statistic is to test the validity of overidentifying restrictions. Under the null hypothesis that the overidentifying restrictions are satisfied, the J-statistic times the number of regression observations is asymptotically χ^2 with degrees of freedom equal to the number of overidentifying restrictions.

If the equation excluding suspect instruments is exactly identified, the J-statistic will be zero.

3.3 Coefficient of Determination

The *Coefficient of Determination* (R^2) is a statistic that will give some information about the goodness of fit of a model. In regression, the coefficient of determination is a statistical measure of how well the regression line approximates the real data points. **An R^2 value of 1.0 indicates that the regression line perfectly fits the data**. It's often a suspicious result. As presented in Table 7, an acceptable value for R^2 is superior to 0.5.

3.4 Adjusted Coefficient of Determination

The Adjusted Coefficient of Determination (Adjusted R^2) is a modification of R^2 that adjusts for the number of explanatory terms in a model. Unlike R^2, the Adjusted R^2 increases only if the new term improves the model more than would be expected by chance. The Adjusted R^2 can be negative (in very poorly specified regression equations.), and will always be less than or equal to R^2. Adjusted R^2 does not have the same interpretation as R^2. As such, care must be taken in interpreting and reporting this statistic. Adjusted R^2 is particularly useful in the feature selection stage of model building. Adjusted R^2 is not always better than R^2: adjusted R^2 will be more useful only if the R^2 is calculated based on a sample, not the entire population. For example, if our unit of analysis is a state, and we have data for all counties, then Adjusted R^2 will not yield any more useful information than R^2.

3.5 Mean Dependent Variable

The value of the *Mean Dependent Variable* is the mean of the observations of the dependent variable.

3.6 S.D. Dependent Variable

The value of the S.D. Dependent Variable is the estimated standard deviation of the dependent variable.

3.7 S.E. of Regression

The *S.E. of Regression* is a summary measure of the size of the equation's errors. The unbiased estimate of it is calculated as the square root of the sum of squared residuals divided by the number of usable observations minus the number of regressors (including the constant). **This measure should be closer to zero**.

3.8 Sum of Squared Residual

The residual sum of squares (RSS) is the sum of squares of residuals. It is the discrepancy between the data and our estimation model. **As smaller this discrepancy is, better our estimation will be.**

3.9 Prob(F-statistic)

To test the success of the regression model, a test can be performed on R^2. **Usually, we accept that the regression model is useful when the Prob(F-statistic) is smaller than the desired significance level, for example, 0.05 (for 5% significance level).**

3.10 Durbin-Watson Statistic

The Durbin-Watson statistic is a test statistic used to detect the presence of autocorrelation in the residuals from a regression analysis. Its value always lies between 0 and 4.

A value of 2 indicates there appears to be no autocorrelation. If the Durbin-Watson statistic is substantially less than 2, there is evidence of positive serial correlation and values much above 2 are indicative of the negative serial correlation. As a rough rule of thumb, if the Durbin-Watson statistic is less than 1.0, there may be cause for alarm. Small values of Durbin-Watson statistic indicate successive error terms are, on average, close in value to one another, or positively correlated. Large values of Durbin-Watson statistic indicate successive error terms are, on average, much different in value to one another, or negatively correlated. How much below or above 2 is required for significance depends on the number of usable observations and the number of independent variables (excluding the constant).

The Durbin-Watson test is a test for first-order serial correlation in the residuals of a time series regression. **A value of 2.0 for the Durbin-Watson statistic indicates that there is no serial correlation, but this result is biased toward the finding that there is no serial correlation if lagged values of the regressors are in the regression.**

3.11 Determinant Residual Covariance

The Determinant residual covariance is the determinant of the residual covariance matrix. **If the determinant of the residual covariance matrix is zero, the estimates are efficient.** But, if a comparison of two determinants of each's residual covariance matrix shows a value, for example, >100 for the original VAR and a value near to zero for the log-VAR, then a linearly dependent covariance matrix seems unlikely, the zero-value must be due to very small covariances (but these are caused by the transformation into log-units, and must not be due to a real improvement of the model).

4 Maximum-Likelihood

Maximum Likelihood Estimation (MLE) is a popular statistical method used to calculate the best way of fitting a mathematical model to some data. Modeling real-world data by estimating maximum-likelihood offers a way of tuning the free parameters of the model to provide an optimum fit.

The likelihood and log-likelihood functions are the basis for deriving estimators for parameters, given data. While the shapes of these two functions are different, they have their maximum point at the same value. In fact, the value of p that corresponds to this maximum point is defined as the Maximum Likelihood Estimate (MLE). This is the value that is “mostly likely" relative to the other values. This is a simple, compelling concept, and it has a host of good statistical properties.

4.1 Log-Likelihood

The shape of the log-likelihood function is important in a conceptual way. If the log-likelihood function is relatively flat, one can make the interpretation that several (perhaps many) values of p are nearly equally likely. They are relatively alike. This is quantified as the sampling variance or standard error. If the log-likelihood function is fairly flat, this implies considerable uncertainty. This is reflected in large sampling variances and standard errors, and wide confidence intervals.

On the other hand, if the log-likelihood function is fairly peaked near its maximum point, this indicates some values of p are relatively very likely compared to others. There is some considerable degree of certainty implied and this is reflected in small sampling variances and standard errors, and narrow confidence intervals. **So, the log-likelihood function at its maximum point is important as well as the shape of the function near this maximum point.**

4.2 Avg. Log-Likelihood

Average log-likelihood is the log-likelihood (i.e. the maximized value of the log likelihood function) divided by the number of observations. **The maximization of the log-likelihood is the same as the maximization of the average log-likelihood.** This statistic is useful in order to compare models.

4.3 Akaike Information Criterion

Akaike’s Information Criterion (AIC) is a measure of the goodness of fit of an estimated statistical model. It is grounded in the concept of entropy. The AIC is an operational way of trading off the complexity of an estimated model against how well the model fits the data.

The preferred model is the one with the lowest AIC value. The AIC methodology attempts to find the model that best explains the data with a minimum of free parameters. By contrast, more traditional approaches to modeling start from a null hypothesis. The AIC penalizes free parameters less strongly than does the Schwarz criterion.

4.4 Schwarz Information Criterion

The Bayesian information criterion (BIC) is a statistical criterion for model selection. The BIC is sometimes also named the Schwarz criterion, or Schwarz information criterion (SIC). It is so named because Gideon E. Schwarz (1978) gave a Bayesian argument for adopting it.

Given any two estimated models, the model with the lower value of BIC is the one to be preferred. The BIC is an increasing function of residual sum of squares and an increasing function of the number of free parameters to be estimated (for example, if the estimated model is a linear regression, it is the number of regressors, including the constant). That is, unexplained variation in the dependent variable, and the number of explanatory variables increase the value of BIC. Hence, lower BIC implies either fewer explanatory variables, better fit, or both. The BIC penalizes free parameters more strongly than does the Akaike information criterion.

4.5 Hannan-Quinn Information Criterion

Ideally, AIC and SBIC should be as small as possible (note that all can be negative). **Similarly, the Hannan-Quinn Information Criterion (HQIC) should also be as small as possible.** Therefore the model to be chosen should be the one with the lowest value of information criteria test.

4.6 Determinant Residual Covariance

Maximizing the likelihood value is equivalent to minimizing the determinant of the residual covariance matrix. Thus, the determinant of the residual covariance matrix and not the residuals itself are minimized. **As smaller this determinant is, better our estimation will be.**

5 Summary table

Summarizing all the statistical output generated following an estimation or a statistical test is impossible. Table 7 intends to provide a clue about some test results often used in the regular practice of econometrics and statistics.

Table 7. Summary Table

Type	Optimal	Acceptable
R^2 and Adjusted R^2	$\rightarrow 1$	> 0.5
J-statistic	$\rightarrow 0$	< 0.1
Mean Dependent Variable	$\rightarrow +\infty$	> 100
S.E. of Regression	$\rightarrow 0$	Choose the lower value (comparison)
Residual Sum of Squares	$\rightarrow 0$	Choose the lower value (comparison)
Prob(F-statistic)	$\rightarrow 0$	< 0.05
Durbin-Watson Statistic	$\rightarrow 2$	$1.8 < DW < 2.2$ (Under conditions)
Determinant Residual Covariance	$\rightarrow 0$	Choose the lower value (comparison)
Log-Likelihood	$\rightarrow +\infty$	$> 10^3$
Average Log-Likelihood	$\rightarrow +\infty$	> 10
AIC	$\rightarrow -\infty$	Choose the lower value (comparison)
SIC	$\rightarrow -\infty$	Choose the lower value (comparison)
HQIC	$\rightarrow -\infty$	Choose the lower value (comparison)

Note: the values provided in the right column are only indicative. They can change with respect to the type of econometric exercise.

References

Dickey, D. A., Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74 (366a), 427–431.

Durbin, J., Watson, G. S., 1950. Testing for serial correlation in least squares regression. I. Biometrika 37 (3/4), 409–428.

Durbin, J., Watson, G. S., 1951. Testing for serial correlation in least squares regression. II. Biometrika 38 (1/2), 159–177.

Durbin, J., Watson, G. S., 1971. Testing for serial correlation in least squares regression. III. Biometrika 58 (1), 1–19.

Granger, C. W. J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438.

Ljung, G. M., Box, G. E. P., 1978. On a measure of lack of fit in time series models. Biometrika 65 (2), 297–303.