Memento on EViews Output

by Jonathan Benchimol

1 Generalized Method of Moments

The starting point of 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. The theoretical relation is replaced by its sample counterpart 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 of 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 $f(\theta)$ and a set of instrumental variables z_t:

$$E\left[f(\theta)' Z\right] = 0$$

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:

$$J(\theta) = (m(\theta))' A m(\theta) \tag{1}$$

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

1.1 J-statistic

The J-statistic is the minimized value of the objective function, where we report (1) 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.

1.2 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 of 1.0 indicates that the regression line perfectly fits the data.**

1.3 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.

1.4 Mean Dependent Variable

The value of the Mean dependent variable is the mean of the observations of the dependent variable.

1.5 S.D. Dependent Variable

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

1.6 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.**

1.7 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.**

1.8 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).**

1.9 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 negative serial correlation. . As a rough rule of thumb, if 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.**

1.10 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).

2 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.

2.1 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 and 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.**

2.2 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.

2.3 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.

2.4 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.

2.5 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 be also as small as possible.** Therefore the model to be chosen should be the one with the lowest value of information criteria test.

2.6 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.**

3 Summary table

R^2 and Adjusted R^2	$\to 1$	$> 0,8$
J-statistic	$\to 0$	$< 0,1$
Mean dependant variable	$\to +\infty$	> 100
S.E. of Regression	$\to 0$	Choose the lower value (comparison)
Residual sum of squares	$\to 0$	Choose the lower value (comparison)
Prob(F-statistic)	$\to 0$	$< 0,05$
Durbin-Watson statistic	$\to 2$	$1 < \text{DW} < 3$ (Under conditions)
Determinant residual covariance	$\to 0$	Choose the lower value (comparison)
Log-Likelihood	$\to +\infty$	$> 10^3$
Average Log-Likelihood	$\to +\infty$	> 10
AIC	$\to -\infty$	Choose the lower value (comparison)
SIC	$\to -\infty$	Choose the lower value (comparison)
HQIC	$\to -\infty$	Choose the lower value (comparison)