ECONOMETRIC MODELING USING "GOODNESS OF FIT OF BEHAVIOR"

S. I. Noskov

A wide range of statistics are commonly used in statistical applications and in econometrics. These include the coefficient of multiple determination, Fisher, Durbin-Watson, and Theil statistics, bias, t-test, average relative approximation and prediction errors, etc. [1-4]. Each statistic corresponds to a particular characteristic of the model description of the system or the process and is formally expressed in terms of random approximation errors. The formal and substantive value of these statistics is unquestionable because of their deep theoretical foundations and their universal use in various combinations in all standard econometric models.

Yet the adequacy of a model is a multifaceted notion which includes a wide variety of particular characteristics, whose number is steadily increasing with the development of research in this field. Thus, an important aspect in evaluating the quality of econometric models is not related directly with approximation accuracy: it actually reflects the goodness of fit of the variation (behavior) of the observed and the model variables on various observations in the sample. Below we propose some techniques for formalizing such a "goodness of fit of behavior" test and adjusting the regression estimators on the basis of this test. The qualitative properties of the adjusted estimators are not analyzed in this paper.

Formalization of the "Behavior Goodness of Fit" Statistic

A necessary element of any econometric model is a general regression equation of the form

\[ y_t = f(\alpha; x_{t1}, x_{t2}, ..., x_{tk}) + \varepsilon_t, \quad k = 1, n, \]  

where \( y \) is the dependent variable, \( x_i, i = 1, ..., m, \) are the independent variables, \( \alpha \) is the vector of estimated parameters, \( f \) is the approximating real function, \( \varepsilon_t \) are random approximation errors, \( n \) is the sample size.

In regression analysis, the parameter \( \alpha \) of regression (1) is estimated by minimizing a particular loss function:

\[ I(\alpha) = \sum_{k=1}^{n} \varphi(\varepsilon_t). \]

The real function \( \varphi \) is monotone nondecreasing (usually convex) with nonnegative values. Classical loss functions [1, 2] include, in particular, the functions of Huber, Andrews, Meshalkin, and also functions of the form

\[ I_\nu(\alpha) = \sum_{k=1}^{n} |\varepsilon_t|^\nu, \quad \nu \geq 1, \]

where \( \nu = 1 \) corresponds to the least absolute values method and \( \nu = 2 \) to the least squares method.

Denote by \( \hat{y}_t, k = 1, ..., n \), the estimated values of the dependent variable calculated with the estimator \( \hat{\alpha} \), which has been obtained using the given loss function \( I(\alpha) \):

\[ \hat{y}_t = f(\hat{\alpha}; x_{t1}, x_{t2}, ..., x_{tk}), \quad k = 1, n. \]
In econometric modeling we may obtain a very poor fit between the calculated and the actual trajectories characterizing the variation of the dependent variables even when the regression model represents an almost functional dependence and the loss functions are very small. This may be reflected in different signs of the differences \( y_{k+1} - y_k \) and \( \dot{y}_{k+1} - \dot{y}_k \) for some pairs of observations \( k \) and \( k + 1 \), which lowers the quality of the model equation and reduces its usefulness for prediction, as it provides an unsatisfactory explanation of the observed process. The reason for poor fit of behavior is either omission of some essential factors from the list of dependent variables in the regression or an incorrect choice of the approximating function \( f \) or the loss function \( I(\alpha) \).

The "behavior goodness of fit" statistic (BGF) intended to detect such cases can be represented as an integer-valued function \( \Phi_1 \):

\[
\Phi_1(\alpha) = \sum_{k=1}^{n-1} \text{sign} (y_{k+1} - y_k) \text{sign} (\dot{y}_{k+1} - \dot{y}_k) 
\]

The value \( \Phi_1(\alpha) = n - 1 \) corresponds to perfect fit of the vectors \( y \) and \( \dot{y} \) in the sense of the statistic (2). If the sum (2) consists only of components 0 and 1, the two vectors are assumed to have an almost perfect fit.

Depending on the research objectives, the BGF statistic can be calculated on arbitrary pairs, triples, etc., of observation indices. In particular, it sometimes makes sense to include in the BGF statistic second differences, which reflect the rates of increase (decrease) of the dependent variable:

\[
\Phi_2(\alpha) = \sum_{k=1}^{n-2} \text{sign}(2y_{k+1} + y_k) \text{sign}(\dot{2y}_{k+1} + \dot{y}_k) 
\]

Discontinuity of the functions \( \Phi_1 \) and \( \Phi_2 \) is an obstacle to the application of the BGF statistic for adjusting the parameters in regression (1) with the object of improving the fit of the vectors \( y \) and \( \dot{y} \). An alternative definition of the BGF statistic (2) is therefore provided by

\[
\Phi_3(\alpha) = \sum_{k=1}^{n-1} l_k 
\]

where

\[
l_k = \begin{cases} 
1 & \text{sign} (y_{k+1} - y_k) \text{sign} (\dot{y}_{k+1} - \dot{y}_k) = -1, \\
0 & \text{otherwise.}
\end{cases}
\]

It is easy to see that with the BGF statistic (3) zero values of the function \( \Phi_3 \) correspond to cases of perfect and almost perfect fit.

Adjustment of Regression Estimators Using the BGF Statistic

The BGF statistic is obviously not an alternative to the loss function and other performance statistics for regression models, because approximation accuracy is after all the most important integral characteristic of the adequacy of the model to the observed system or process. Yet the BGF statistic may be used as a supplementary tool for adjusting the estimator \( \alpha \) obtained by minimization of a given loss function \( I(\alpha) \). This estimator adjustment can be carried out in the following way.

Let \( I^* \) be the given minimum value of the loss function for regression (1) and \( \alpha^* \) the corresponding parameter estimator. Assume that the researcher (the modeler) is able to assign a certain increment \( \Delta I^* \) by which the minimum value \( I^* \) may be increased without a substantial deterioration of approximation quality. Then the problem of improving the behavior goodness of fit can be stated in the form

\[
\phi_1(\alpha) = \max_{\alpha \in A}
\]

\[A = \{ \alpha \mid I(\alpha) \leq I^* + \Delta I^* \}.
\]