SELECTING REGRESSION MODEL

J. Á. Víšek

Abstract: A new tool for the identification of regression model is proposed and its properties are established. The key importance of the new tool is that it is able to solve still not very well-known problem of *diversity of estimates*, as described in Víšek [22] and [25]. Main idea of the proposal is as follows. Having evaluated an estimate of regression coefficients for given data, the data are partitioned into two disjoint subsets (e.g. by a geometric rule applied in the factor space). Then for each subset of corresponding residuals we evaluate the estimate of their density, e.g. the kernel one. If the estimate of regression model is "near to the true model", the density of disturbances is the same in the both subsets, and hence also the estimates of density of residuals are approximately equal each to other. Therefore, finally, the estimates of density are compared by means of the weighted Hellinger distance. It implies that the significant difference between the estimates of density indicates that the given estimate of the regression model is not near to the "true" model or, in other words, that it is not "adequate" for the data. In the case when we have at our disposal more estimates of the regression model, and especially when the estimates are considerably different (each from other), the test statistic may be also used for selecting the estimate of the regression model. We just accept the estimate with the smallest weighted Hellinger distance. The result of the paper is illustrated by two simple numerical examples demonstrating especially the sensitivity of the test statistic to the difference between the estimates of density.

1991 AMS Mathematics Subject Classification: 62J05, 62J20.

Key words and phrases: Weighted Hellinger distance, diagnostics and choice of model, diversity of (robust) estimates.