REMARKS ON THE ESTIMATION OF COEFFICIENTS OF A REGRESSION

IN THE PRESENCE OF UNKNOWN EXPLANATORY VARIABLES

Herman Chernoff*

Massachusetts Institute of Technology

and

Harvard University

In the linear regression model $Y = \beta_1 X_1 + \beta_2 X_2 + u$, the coefficients β_1 and β_2 may be estimated by least squares. If the explanatory variable X_2 is not observed, the regression of Y on X_1 will give an estimate of β_1 whose bias will depend on the correlation between X_1 and X_2. However qualitative knowledge about X_2 can be exploited. We treat the case where the known and unknown explanatory variables and the coefficients are nonnegative and where it is known that for some, but not which, data points, the unknown explanatory variables are relatively small.

1. Introduction.

  A source of difficulty in estimating the effect of one variable on

another, especially in observational studies, is that the explanatory model may

omit a causal variable. Under some circumstances, this difficulty may be

serious. If the omitted variable is unimportant, i.e. it has a relatively small

effect, it may be safe to ignore it. If it is uncorrelated with the other

explanatory variables, it may also be ignored in linear regression models. If

it is correlated with the explanatory variables, and one desires only to use

these for prediction, one may proceed without it, as long as that correlation is