LEAST-SQUARES THEORY BASED ON GENERAL DISTRIBUTIONAL ASSUMPTIONS WITH AN APPLICATION TO THE INCOMPLETE OBSERVATIONS PROBLEM

B. M. S. VAN PRAAG
ECONOMETRIC INSTITUTE
ERASMUS UNIVERSITY, ROTTERDAM

T. K. DIJKSTRA
UNIVERSITY OF GRONINGEN

J. VAN VELZEN
IBM, THE NETHERLANDS

The linear regression model \( y = \beta'x + \varepsilon \) is reanalyzed. Taking the modest position that \( \beta'x \) is an approximation of the “best” predictor of \( y \) we derive the asymptotic distribution of \( b \) and \( R^2 \), under mild assumptions.

The method of derivation yields an easy answer to the estimation of \( \beta \) from a data set which contains incomplete observations, where the incompleteness is random.

1. Introduction

Consider a \((k + 1) \times 1\) random vector \( z \) with mean \( \mu \) and non-singular covariance matrix \( \Sigma \), whose fourth-order moments exist.

Partition \( z \) as \([x', y']\), \( x \) has \( k \) elements and partition \( \mu \) and \( \Sigma \) conformably:

\[
\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}
\]

where \( \mu_x \equiv E(x), \Sigma_{xy} \equiv E(x - \mu_x)(y - \mu_y) \), etc.

Assume that a set of \( T \) independent observations on \( z \) is available:

\[ \{z_1, \ldots, z_T\}. \]

Define the sample mean \( \bar{z} \) and covariance matrix \( S \) as:

\[
\bar{z} \equiv \frac{1}{T} \sum_{t=1}^{T} z_t \quad \text{and} \quad S \equiv \frac{1}{T} \sum_{t=1}^{T} (z_t - \bar{z})(z_t - \bar{z})'
\]

and partition them as follows:

\[
\bar{z} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \quad S = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix}
\]

Let it be desired to find a function \( p(x - \mu_x) \) which predicts \( y - \mu_y \) in the best possible way, i.e., \( E(y - \mu_y - p(x - \mu_x))^2 \) is minimal. It is well-known, (see e.g., Rao, 1973, page 264) that \( p(x - \mu_x) \equiv E(y - \mu_y | x - \mu_x) \) is the optimal solution. Consequently, if the distribution of \( z \) were known, one could determine the form of \( p(x - \mu_x) \), and an estimator of \( p \), based on a set of independent observations, could at least in principle be obtained. In practice, the distribution of \( z \) is virtually always unknown and in general one does not...
take the trouble to estimate it. Instead, researchers frequently adopt the assumption of normality of \( z \) as a working hypothesis. This assumption simplifies the determination of \( p(x - \mu_x) \) and its estimation considerably: 
\[
p(x - \mu_x) = \beta(x - \mu_x);
\]
\( \beta \) is a \( k \times 1 \) vector of constants defined by \( \beta = \Sigma_x^{-1} \Sigma_{xy} \); its least-squares and maximum likelihood estimator 
\[
b = S_{xx}^{-1} S_{xy}
\]
is easily obtained and it possesses a number of desirable optimal properties.

So, if \( z \) is normal we know the form of the best predictor for \( y \) and we can estimate its parameters in an optimal way. It can be argued that the assumption of normality is not to be taken for granted because its violation may affect estimators of the sampling distribution of \( b \) and related statistics considerably (see below).

This paper is based on the following attitude and considerations:

1. We do not know the fourth-order moments of \( z \). In addition, we assume that we have a set of \( T \) independent observations on \( z \).
2. We want to predict \( y - \mu_y \) with the best possible linear approximation of 
\[
E[(y - \mu_y | x - \mu_x) - L'(x - \mu_x)]^2
\]
with the solution of:
\[
\minimize E[(L'(x - \mu_x) - L'(y - \mu_y | x - \mu_x))^2] \text{ w.r.t. all } (k \times 1) - \text{vectors } L.
\]
The solution is easily verified (see e.g., Cramér, 1946, p. 274) \( \beta = \Sigma_x^{-1} \Sigma_{xy} \). Moreover, the prediction error \( y - \mu_y - \beta(x - \mu_x) \) is uncorrelated with \( x \).

3. We estimate \( \beta \) by its sample counterpart \( b = S_{xx}^{-1} S_{xy} \).
4. We want to assess \( b \)'s accuracy or stability in a simple way without relying on a presumed validity of the normality assumption. And similarly for \( R^2 \), the sample counterpart of \( \rho^2 \) which is defined as the ratio between the variances of \( \beta'x \) and \( y \), i.e.:
\[
\rho^2 \equiv \frac{\beta' \Sigma_x \beta}{\Sigma_{yy}} \quad \text{and} \quad R^2 \equiv \frac{b' S_{xx} b}{S_{yy}}.
\]
The tools to be used are a few basic theorems of asymptotic theory and matrix differentiation. An excellent review of the former is given by Rao (1973), chap. 2.c and 6a.2) and Bishop, Fienberg and Holland, 1975, chap. 14). For matrix differentiation the reader is referred to Balestra (1976), Dhrymes (1978) or Graham (1981).

The structure of this paper is the following. In section 2 we derive the asymptotic distribution of \( b \). Some of the results are already known but the derivation is different from that in the literature. See also Van Praag (1978, 1980). We consider \( b \) as a function of a random sample covariance matrix \( S \). Some comparisons are made to the cases where we know more of the parent distribution, e.g., that \( z \) has an elliptical distribution or is multivariate normal. In section 3 we give the asymptotic distribution of the multiple correlation coefficient \( R^2 \). In section 4 we show how the derivation method can be adapted in a very simple way to deal with situations where some observations are incomplete, but it may be assumed that the incompleteness is generated by a random response process. In section 5 we present an application. These two sections are based on Van Praag and Van Velzen (1982). The final section lists a few references to extensions and related work. The Appendix assembles two lemmas, used in this paper.

2. The asymptotic distribution of \( b \)

The derivation of the asymptotic distribution of \( b \) proceeds in two steps: first, find the asymptotic distribution of \( S \) and second, apply an appropriate version of the well-known delta-method (see Lemma 1 of the Appendix).