Non-linear regression and observational errors in the independent variable

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics

PROBABILITY THEORY, STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 76-12

Non-linear regression and observational errors in the independent variable

by

H.N. Linssen

Eindhoven, September 1976

The Netherlands

Non-linear regression and observational errors in the independent variable

by

H.N. Linssen

Paper contributed to the European Meeting of Statisticians 6-10 september, 1976.

1. Introduction

Sometimes a statistician is confronted with pairs of numbers:

$$(x_i, y_i) \qquad i = 1,\ldots,n \; .$$

He is told that x_i and y_i correspond to the independent respectively dependent variables of modelfunctions f_i with parametervector β. He is asked to produce an estimate for β.

To estimate β it is necessary to know in which experimental situation x_i and y_i were obtained, lest the wrong estimationmethod is used and inconsistent estimates result.

There is the situation that only the dependent variable is subject to measurement error. In formula:

$$y_i = f_i(x_i;\beta) + \text{error} \; .$$

Classical non-linear least squares yields consistent estimates for β.

Another well-known situation is called the Berkson-case; the deviation between the independent variable and a known fixed value follows a distribution, whose first two moments are known. In formula:

$$y_i = f_i(\xi_i,\beta) + \text{error}$$

$$\xi_i = x_i + \text{error} \; .$$

Berkson [1] showed classical least squares estimates to be consistant, provided f_i linear in ξ_i.

Fedorov [2] developed an algorithm to cope also with the non-linear case.

The third situation occurs when the independent variable is unknown but observed. In formula:

$$y_i = f_i(\xi_i,\beta) + \text{error}$$

$$x_i = \xi_i + \text{error}$$

There three situations are illustrated in figure 1.
In case III an estimate for β may follow from:

(1) $$\min_{\beta,\xi_i} \sum_{i=1}^{n} (x_i - \xi_i)^2 + (y_i - f_i(\xi_i,\beta))^2 ,$$

where for the sake of convenience the observational errors are assumed to be independent and homogeneously distributed.
We may reformulate (1) as follows:

(2) $$\min_{\gamma} \| z - g(\gamma) \|^2, \text{ where } \| a \|^2 \text{ equals } \sum_{i=1}^{n} a_i^2 \text{ and}$$

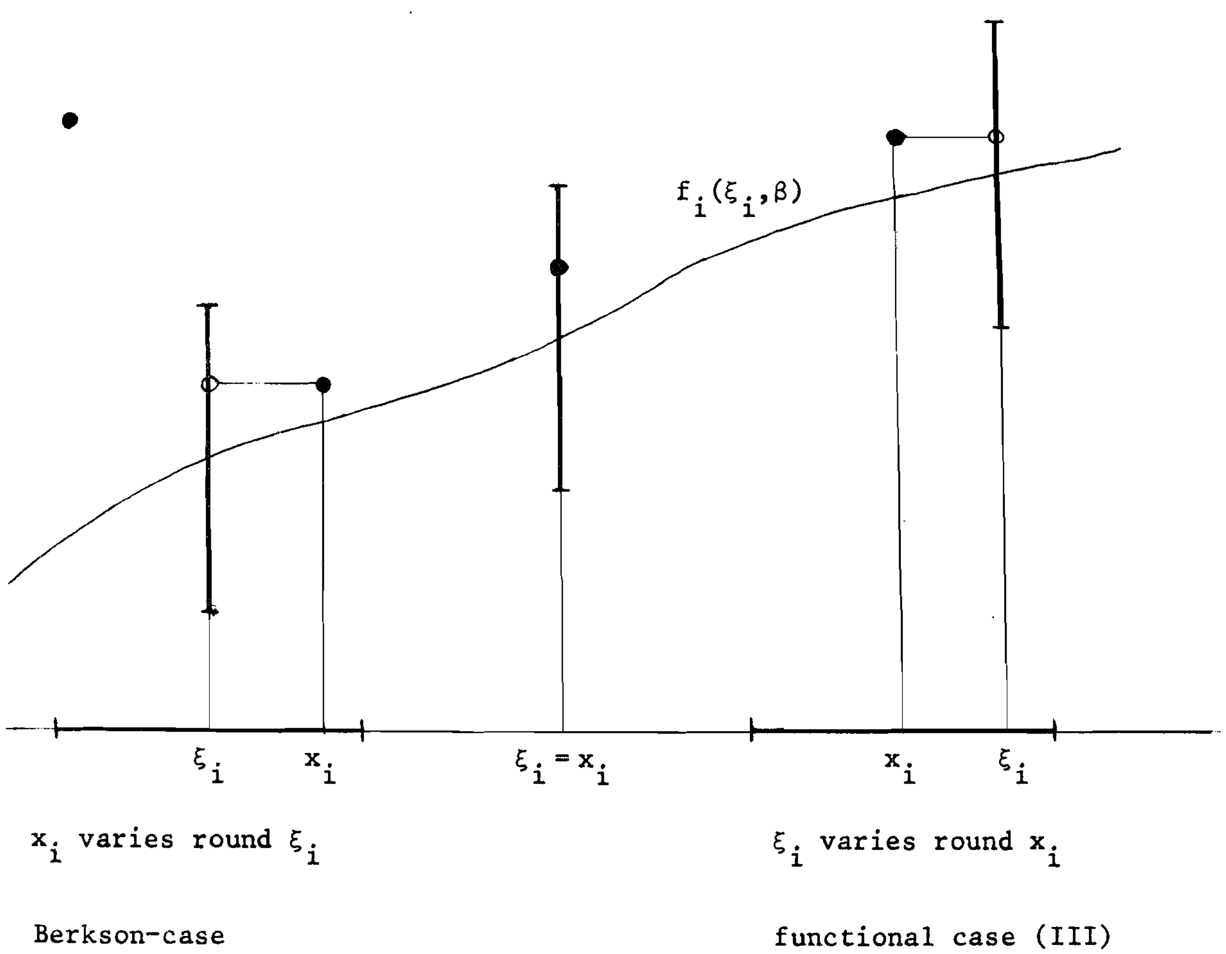

figure 1

(3) $z = \binom{y}{x}$; $y = (y_i)$; $x = (x_i)$; $\gamma = \binom{\xi}{\beta}$; $\xi = (\xi_i)$

and $g_i = f_i$ for $i \leq n$ and $g_i = \xi_i$ $i > n$.
The estimation problem includes now the determination of an estimate for the $(n + p)$-parametervector γ, given g and the observations z.
It is possible to determine an estimate for γ with the aid of the standard algorithms for nonlinear least squares as described for example in Chapter 3 of [3].
However, the number of unknowns is n + p and standardmethods will explode.
Knowledge about ξ_i is contained only in (x_i, y_i).
This feature of the observational model we employ in developing an efficient algorithm to solve (1).

2. The algorithm

2.1. General

A l. sq estimate $\hat{\gamma}$ for γ satisfies:

$$\min_{\gamma} \| z - g(\gamma) \|^2$$

This is solved iteratively. Denote the k^{th} approximation to γ by $\gamma^{(k)}$.
Then $\gamma^{(k+1)}$ satisfies:

$$\min_{\gamma} \| z - g(\gamma^{(k)}) - \frac{\partial g}{\partial \gamma^{(k)}} (\gamma - \gamma^{(k)}) \|^2$$

$\gamma^{(k+1)}$ is the solution to a linear regression problem and the iterative method turns out to be the solution of a sequence of linear regression-problems.

2.2. Linear regression and QR-decomposition

In the linear case

$g(\gamma) = G\gamma$ holds, where G is a known matrix of appropriate size.

$\hat{\gamma}$ follows from:

$$\min_{\gamma} \| z - G\gamma \|^2$$

and is equal to

$$(G^T G)^{-1} G^T z \ ,$$

provided G is column-regular.
An efficient and numerically stable way to actually compute $\hat{\gamma}$ employs QR-decomposition.

In case G is columnregular there exist an orthogonal matrix Q and an upper triangular matrix R, such that

$$G = Q\binom{R}{0} \ .$$

If G is $n * p$ then Q is $n * n$ and R is $p * p$.
Q and R result from orthogonalization f.e. by means of Householder transformations.
It is obvious that:

$$G^T G = R^T R \ . \qquad (4)$$

2.3. QR-decomposition and errors in both variables

In each step of the iterative solution to (2) QR-decomposition is applied. The Jacobian $\partial g/\partial \gamma^{(k)}$ is to be decomposed.
This matrix has the special structure (see (3)):

$$\frac{\partial g}{\partial \gamma^{(k)}} = \begin{pmatrix} \Lambda & X \\ I & 0 \end{pmatrix} \begin{matrix} n \\ n \end{matrix} \qquad \text{(columns: } n \ \ p\text{)}$$

Λ is diagonal because $\partial f_i/\partial \xi_j = 0 \quad (i \neq j)$.
X is a $n * p$-matrix with elements $\partial f_i/\partial \beta_j$.

Because the first n columns of $\partial g/\partial \gamma^{(k)}$ are orthogonal, the triangular factor of its decomposition can be written:

$$\begin{pmatrix} \Lambda_1 & A \\ 0 & R \end{pmatrix} \begin{matrix} n \\ p \end{matrix} \quad (\text{columns } n,\ p), \quad \Lambda_1 \text{ diagonal, } R \text{ upper triangular}$$

To determine Λ_1, A and R we employ identity (4) with the result:

$$\Lambda_1 = (\Lambda^2 + I)^{\frac{1}{2}}; \quad A = \Lambda_1^{-1}\Lambda X$$

and

(5) $$R^T R = X^T \Lambda_1^{-2} X .$$

From (5) we see that R is the triangular factor if $\Lambda_1^{-1}X$. The QR-decomposition of the 2n * (n + p)-matrix $\partial g/\partial\gamma^{(k)}$ is reduced the simple computation of Λ_1 and A and the QR-decomposition of the n * p-matrix $\Lambda_1^{-1}X$.

3. Consistency

In case the observational errors are normally distributed the least squares estimates for ξ and β are maximum likelihood. In the presence of many nuisance parameters the estimate for β need not be consistent (a well-known example is given by Neymand and Pearson [4]).

Now we shall prove the consistency of the least squares estimator of β in the very special case

$$f_i = \xi_i \cdot \beta \qquad (\beta \in \mathbb{R}^1)$$

The normal equations for ξ_i are given by:

$$(\xi_i - x_i) + (\beta\xi_i - y_i)\beta = 0 \qquad i = 1,\dots,n .$$

Solving this for ξ_i and substituting the result in (1) yields

$$\frac{1}{1+\beta^2} \sum_{i=1}^{n} (y_i - \beta x_i)^2 .$$

This quantity divided by n is called the mean residual sum of squares and denoted by MRSS.

Now assume

$$\frac{1}{n} \sum_{i=1}^{n} \xi_i^2 \to G \quad \text{for } n \to \infty .$$

Then for fixed β MRSS converges to:

$$1 + \frac{(\beta_0 - \beta)^2}{1 + \beta^2} G ,$$

where β^0 is the true value of β.
The behaviour of this function of β is sketched in figures 2 ($\beta_0 \neq 0$) and 3 ($\beta_0 = 0$).

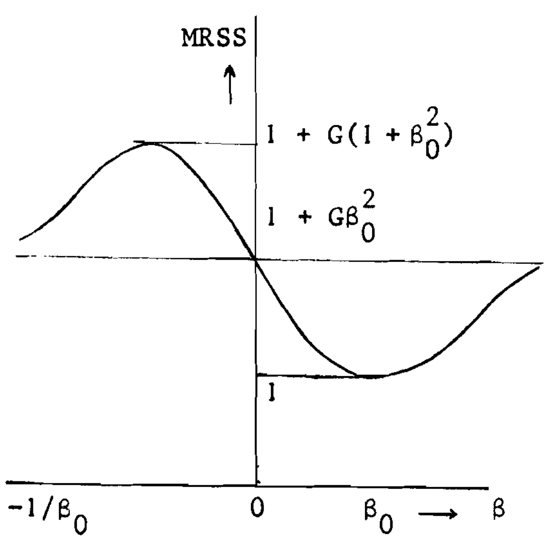

figure 2

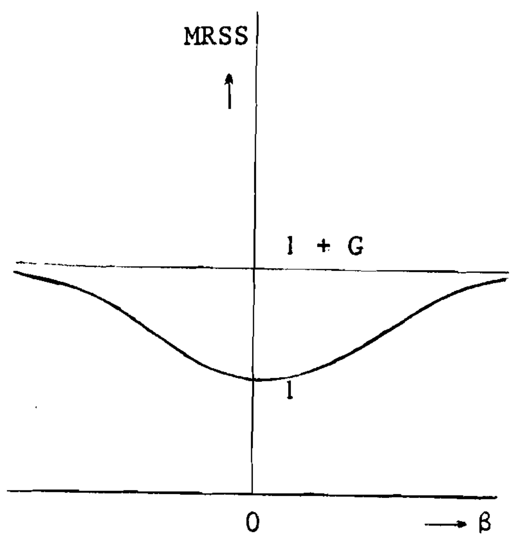

figure 3

From this we see that minimum sum of squares yields a consistent estimate of β.
The simple least squares estimate given by

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$$

is <u>inconsistent</u>, it converges to

$$\beta\cdot\frac{G}{G+1}\ .$$

I tried to show inconsistency by means of simulation for a number of model-functions which were nonlinear both in the independent variable and in the parameter. So far this has not been successful.
The approach sketched so far is different from the approach of Hodges and Moore [5], who tried to characterize the cases where error in the independent variable is not large enough to impair ordinary least squares estimation.

<u>References</u>

[1] Berkson, Are there two regressions? Journal of the American Statistical Association, 1950, pg. 164-180.

[2] Fedorov, V.V., Regression problems with controllable variables subject to error, Biometrika, 1974, pg. 49-56.

[3] Murray, ed., Numerical methods for unconstrained optimization, 1972, Academic Press.

[4] Neyman, J. and Scott, E.L., Consistent estimates based on partially consistent observations, Econometrica, 1948, pg. 1-32.

[5] Hodges, S.D. and Moore, P.G., Data Uncertainties and least squares regression, Applied Statistics, 1972, pg. 185-195.