NUMERICAL ASPECTS OF SOME NONSTANDARD REGRESSION PROBLEMS

V. Fedorov
A. Vereskov

May 1921
WP-85-32

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

ABSTRACT

Regression models are extremely popular in different areas conjunct with systems analysis. The variety of these models is immense and now, as a consequence, there exist many computerized versions of the corresponding statistical methods. In this paper, we attempt to unify (from the computational viewpoint) at least some statistical approaches. We understand that similar attempts have repeatedly been made in statistical practice (see, for instance, BMDP 1983), but none of them can be considered completely successful. Nevertheless, any new attempt (this one, we hope) gives a more profound and comprehensive understanding of the situation and the future directions of the investigations.

CONTENTS

NUMERICAL ASPECTS OF SOME NONSTANDARD REGRESSION PROBLEMS

V. Fedorov and A. Vereskov

I. INTRODUCTION

The paper gives a short survey of models and estimation methods which are closely connected with the least square method. The first two sections are devoted to traditional regression models. All other models and estimation methods are considered in the third chapter. The reason why the main attention was paid to traditional case is very simple – all other models and estimation procedures under consideration can be transformed in some way to this case. Therefore statistical ideas and basic properties are practically the same for all considered statistical problems, and more or less detailed analyses presented in the first two chapters allow, firstly, avoidance of repetitions and, secondly, understanding of some common features of these problems.

The final chapter deals with technical aspects of ls-program set (which were prepared in VNIISI, Moscow) and some numerical examples are presented in it.

II. TRADITIONAL REGRESSION MODEL AND LEAST SQUARE METHOD

When functional relation between some variable y (usually it is called "response") and variables $x = (x_1, \cdots, x_k)^T$ (control variables or predictors) must be analyzed, based on empirical data, the model

$$y_i = \eta(x_i\ ,\ \Theta_t) + \varepsilon_i\ ,\ i = \overline{1,n} \tag{1}$$

is traditionally used. In (1), y_i is a result of an observation, x_i are conditions of this observation, $\eta(x, \Theta)$ is a given function, $\Theta = (\Theta_1, \cdots, \Theta_m)^T$ are unknown parameters, subscript "t" stands for true value, ε_i is an "error" of an observation, "i" is a number of an observation. The errors ε_i, $i = \overline{1,n}$, can describe either the deviation of function $\eta(x, \Theta)$ from the real function or "noise" of an experiment. To justify the model (1) it is necessary to make some formal assumptions about their behavior as well. In traditional cases, it is assumed that these errors are random with zero means ($E[\varepsilon_i] = 0$), independently identically distributed with finite variance ($E[\varepsilon_i^2] = \delta^2\gamma^2(x_i)$, where δ^2 can be unknown and $\gamma(x)$ is a given function).

Under these assumptions it is reasonable from a statistical point of view to use the least square estimation (l.s.e.):

$$\hat{\Theta}_n = \mathrm{Arg} \min_{\Theta} v_n^2(\Theta)\ , \tag{2}$$

where $v_n^2(\Theta) = \sum_{i=1}^{n} \gamma^{-2}(x_i)[y_i - \eta(x_i, \Theta)]^2$.

When $\eta(x, \Theta)$ is a linear function of Θ (e.g., $\eta(x, \Theta) = \Theta^T f(x)$) the minimization problem can be solved explicitly and the solution is

$$\hat{\Theta}_n = M_n^{-1} Y_n \quad , \tag{3}$$

where $M_n = \sum_{i=1}^{n} \gamma^{-2}(x_i) f(x_i) f^T(x_i)$, $Y_n = \sum_{i=1}^{n} \gamma^{-2}(x_i) y_i f(x_i)$. The sum of residual squares gives the estimator of the variance:

$$\hat{\delta}^2 = (n - m)^{-1} v_n^2(\hat{\Theta}) \ .$$

This result is too well known to be mentioned but formula (3) will be occasionally used in what follows. The statistical properties of l.s.e. are subject to classical linear regression analysis and can be found practically in any serious statistical monograph (see, for instance, C.R. Rao 1965). For this paper it will be enough to mention that $D(\hat{\Theta}) = \hat{\delta}^2 M_n^{-1}$ is a dispersion (or variance–covariance) matrix of $\hat{\Theta}_n$ and $d[\eta(x, \hat{\Theta})] = \hat{\delta}^2 f^T(x) M_n^{-1} f(x)$ is a variance of the forecasting: $\eta(x, \hat{\Theta}_n) = \hat{\Theta}_n^T f(x)$, at a point x.

In what follows the main attention will be paid to nonlinear models and fortunately most of the properties of linear cases are fulfilled in general situations at least asymptotically (see Jennrich 1969, Wu 1981).

Let us spell out the principal ones:

- L.s.e. are (strongly) consistent, i.e., they converge almost surely to true value Θ_t when $n \to \infty$.
- The parameter δ^2 (when $\gamma(x_i) \equiv 1$ it is the variance of ε_i) consistently estimated by $\hat{\delta}_n^2 = (n - m)^{-1} v_n^2(\hat{\Theta}_n)$.
- The dispersion matrix of $\sqrt{n}(\hat{\Theta}_n - \Theta_t)$ is consistently estimated by

$$\bar{D}(\hat{\Theta}_n) = n^{-1}\hat{\delta}^2 [\sum_{i=1}^{n} f(x_i, \hat{\Theta}_n) f^T(x_i, \hat{\Theta}_n)]^{-1}$$

where $f(x, \Theta) = \partial\eta(x, \Theta)/\partial\Theta$.

- The asymptotical distribution ($n \to \infty$) of $\sqrt{n}(\hat{\Theta}_n - \Theta_t)$ is normal with zero mean and variance matrix

$$\bar{D}_0(\Theta_t) = \lim_{n \to \infty} n^{-1}\hat{\delta}^2 \sum_{i=1}^{n} f(x_i, \Theta_t) f^T(x_i, \Theta_t)$$

These properties are fulfilled under rather mild assumptions which can be roughly formulated in the following way:

- Functions $\eta(x_i, \Theta)$ and $f(x_i, \Theta) = \partial\eta(x_{il}, \Theta)/\partial\Theta$ are smooth enough for any x_i in the vicinity of Θ_t,
- The conditions of observations have to guarantee nondegeneracy of the estimation problem; for instance, the limit function

$$v^2 = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} [\eta(x_i, \Theta) - \eta(x_i, \Theta_t)]^2$$

  has to exist in the vicinity of Θ_t and the function $v^2(\Theta)$ has to have unique minimum for $\Theta = \Theta_t$. Additionally, the limit

$$\bar{M}(\Theta_t) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} f(x_i, \Theta_t) f^T(x_i, \Theta_t)$$

  has to exist and matrix $\bar{M}_t(\Theta_t)$ has to be regular (rank $\bar{M}(\Theta_t) = m$).

The above mentioned properties allow us to use in practice the following approximations:

$$D(\hat{\Theta}_n) \simeq \hat{\delta}_n^2 [\sum_{i=1}^{n} \gamma^{-2}(x_i) f(x_i, \hat{\Theta}_n) f^T(x_i, \hat{\Theta}_n)]^{-1} , \qquad (4)$$

$$d[\eta(x, \hat{\Theta}_n)] \simeq f^T(x, \hat{\Theta}_n) D(\hat{\Theta}_n) f(x, \hat{\Theta}_n) \qquad (5)$$

and the value

$$\frac{n - m}{m - q} \frac{v_n^2(\hat{\hat{\Theta}}_n) - v_n^2(\hat{\Theta}_n)}{v_n^2(\hat{\Theta}_n)}$$

has F-distribution with $m - q$ and $n - m$ degrees of freedom, where $v_n^2(\hat{\hat{\Theta}}_n)$ represents the residual sum of squares obtained by fitting the model with a reduced number $q < m$ of free parameters.

III. MAIN LEAST SQUARE (L.S.) ALGORITHMS

Minimization problem (2) can be considered as a specific case of a non-linear minimization problem and any available general algorithm can be used to get its solution $\hat{\Theta}$. (In what follows index n will usually be omitted).

However, as was repeatedly discussed in statistical literature (see, for instance, Chambers 1973, Fedorov and Uspensky 1975, Jennrich and Ralston 1978), due to the specific structure of $v^2(\Theta)$ some algorithms are much more effective than general ones. Moreover, one should keep in mind that besides estimates $\hat{\Theta}$ for any statistical analysis, it is necessary to calculate matrices $D(\hat{\Theta})$ and function $d[\eta(x, \hat{\Theta}_n)]$ for a prescribed set $x_1, \cdots, x_{n'}$.

Most of the algorithms used in statistical practice can be represented in the form

$$\Theta_s = \Theta_{s-1} + \rho_s H_s Y_s \quad , \tag{6}$$

where s is the number of an iteration, ρ_s is the length of a step, H_s is some matrix (positively semidefinite in most cases), $Y_s = \sum_{i=1}^{n} \gamma^{-2}(x_i)[y_i - \eta(x_j, \Theta_s)]f(x_i, \Theta_s)$.

If $H_s \equiv I$ then (6) is an algorithm of the gradient type.

When

$$H_s = (M_s + \gamma_s A_s)^{-1} , \tag{7}$$

where $\gamma_s > 0$, $A_s \geq 0$, and

$$M_s = \sum_{i=1}^{n} \gamma^{-2}(x_i) f(x_i, \Theta_s) f^T(x_i, \Theta_s) , \tag{8}$$

one deals with algorithms of the Gauss-Newton type. The regularizators γ_s and A_s are usually non-zero for ill-conditioned minimization problem (2).

Algorithms of Gauss-Newton type are probably most popular in statistical software. They converge much faster than algorithms of the gradient type and deliver some additional information (for instance matrix M_s^{-1}) which is necessary for statistical analysis. But they demand calculation of derivatives $f(x_i, \Theta_s)$ at every step (by the way, any algorithm (6) demands it). Unfortunately, in modern empirical research, the calculation of functions $\eta(x, \Theta)$ and $f(x, \Theta)$ happens to be a serious numerical problem (for instance, $\eta(x, \Theta)$ can be described by some system of differential equations) or very tedious for a programmer. Of course, a straightforward use of finite difference approximation

$$f_\alpha(x, \Theta_s) \simeq \frac{\eta(x, \Theta_s + \Delta_s e_\alpha) - \eta(x, \Theta_s)}{\Delta_s} ,$$

where e_α is a vector with zero components except the α-th one, can save from tedious programming but it is not reasonable from numerical point of view.

In general nonlinear programming, there is a variety of different methods for avoiding calculation of derivatives (see, for instance, Chambers

1973, Fedorov and Uspensky 1975, Himmelblau 1973). The most efficient among them are based on different quadratic approximations of $v^2(\Theta)$. But all of them were defeated in statistical applications by methods which are based on linear approximation of $\eta(x,\Theta)$ using the "history" of iterative procedure in conjunction with the idea of Gauss-Newton method.

In these methods (presumably, first suggested by Peckham (1970)), at every step of iterative procedure the following minimization problem must be solved.

$$f_{is} = \operatorname{Arg}\min_{f_i} \sum_{s'=0}^{s'=s-1} \omega_{ss'}[\eta(x_i,\Theta_{s'}) - \eta(x_i,\Theta_{s-1}) - f_i^T(\Theta_{s'} - \Theta_{s-1})]^2 ,$$

where $\omega_{ss'}$ are some weights. In the simplest case (see Ralston and Jennrich 1978) $\omega_{s,s-1} = \omega_{s,s-2} = \cdots = 1$ and all rest weights equal to zero. Similarly to (2) and (3), one can get that

$$f_{is} = Q_s^{-1} U_{is} ,$$

where $Q_s = \sum_{s'=0}^{s'=s-1} \omega_{ss'}\Delta_{ss'}\Delta_{ss'}^T$, $U_{is} = \sum_{s'=0}^{s'=s-1} \omega_{ss'}\Delta_{ss'}u_{iss'}$,

$\Delta_{ss'} = (\Theta_s - \Theta_{s-1})$, $u_{iss'} = \eta(x_i,\Theta_{s'}) - \eta(x_i,\Theta_{s-1})$.

It is reasonable to suggest that approximately

$$f(x_i,\Theta_{s-1}) = f_{is} .$$

Now one can use iterative procedure (6) and (7) with

$$M_s = \sum_{i=1}^{n} \gamma^{-2}(x_i) f_{is} f_{is}^T$$

After some elementary transformation, this procedure can be presented in a more convenient form:

$$\Theta_s = \Theta_{s-1} + \rho_s Q_s \bar{M}_s^{-1} \sum_{i=1}^{n} \gamma^{-2}(x_i) U_{is}(y_i - \eta(x_i,\Theta_{s-1})) , \tag{9}$$

where $\bar{M}_s = \sum_{i=1}^{n} \gamma^{-2}(x_i) U_{is} U_{is}^T$.

The properties of iterative procedure (9) were studied by Vereskov (1981) and it should be pointed out that they depend upon not only structure of $\eta(x, \Theta)$ and sequence $x_1, \cdots, x_n$ but also upon sample $\varepsilon_1, \cdots, \varepsilon_n$. In other words, all assertions about convergency, for instance have a probabilistic character and there exist with non-zero probability some samples when the procedure does not converge in spite of "good quality" of $\eta(x, \Theta)$ and $x_1, \cdots, x_n$.

The iterative procedure (9) is the basic part of all algorithms discussed below and realized in program *ls0*.

IV. MODELS AND ESTIMATORS BASED ON THE LEAST SQUARE METHOD

Variance of Errors of Observations Depending upon Unknown Parameters

In the traditional case, it is assumed (see Ch. 2) that the function $\gamma^2(x_i)$ is known. In the more general case, it is natural to assume that variance of random values ε_i depends upon unknown parameters:

$$E[\varepsilon_i^2] = \gamma^2(x_i, \Theta) \tag{10}$$

If all parameters in (10) coincide with some parameters of response function $\eta(x, \Theta)$ then the estimator defined by the iterative procedure:

$$\hat{\Theta} = \lim_{q \to \infty} \Theta_q \ , \tag{11}$$

$$\Theta_q = \operatorname{Arg} \min_{\Theta} \sum_{i=1}^{n} \gamma^{-2}(x_i, \Theta)[y_i - \eta(x_i, \Theta_{q-1})]^2 \tag{12}$$

is acceptable from a statistical point of view (see Fedorov 1974). To solve (12) one can use iterative procedure (9). The corresponding algorithm is realized in program *ls1*.

If function $\gamma^2(x,\Theta)$ contains parameters which are not involved in $\eta(x,\Theta)$, then a more complicated procedure is needed to get the estimator of unknown parameters (see Malyutov 1982):

$$\hat{\Theta} = \lim_{q \to \infty} \Theta_q \ , \tag{13}$$

$$\Theta_q = \text{Arg}\min_{\Theta} \sum_{i=1}^{n} \{\gamma^{-2}(x_i,\Theta_{q-1})[y_i - \eta(x_i,\Theta)]^2 + \tag{14}$$

$$\frac{1}{2}\gamma^{-4}(x_i,\Theta_{q-1})[[y_i - \eta(x_i,\Theta_{q-1})]^2 - \gamma^2(x_i,\Theta)]\} \ .$$

The properties of (13) for normally distributed ε_i are studied by Malyutov (1982). In other cases, estimator (13) is still consistent and asymptotically normally distributed under some reasonably mild conditions.

Naturally, one can use program *ls0* to solve (14). The computer realization of (13) and (14) is included in the system under the name *ls2*.

Estimation of Parameters of Distribution Density Function

Let x be a random value with density $p(x,\Theta)$, $x \in X$, $\Theta \in \Omega \subset R^m$. The i-th experiment consists of r_i observations of numbers of cases n_{ij} when $x \in X_{ij} \subset X$, $X_{ij} \cup X_{il} = 0$, $j \neq l$. The method of estimation of Θ closely connected with the l.s. method was suggested by Vereskov and Pshennikov (1983). They also discussed its links with traditional approaches. The corresponding numerical procedure has the following form:

$$\hat{\Theta} = \lim_{q \to \infty} \Theta_q \ ,$$

$$\Theta_q = \text{Arg}\min_{\Theta} \sum_{i=1}^{n} \sum_{j=1}^{r_i} \frac{N_i[u_{ij} - p_{ij}(\Theta)]^2}{p_{ij}(\Theta_{q-1})}$$

where $N_i = \sum_{i=1}^{r_i} n_{ij}, p_{ij}(\Theta) = \int_{X_{ij}} p(x,\Theta)dx / \sum_{j=1}^{r_i} \int_{X_{ij}} p(x,\Theta)dx$, $u_{ij} = n_{ij}/N_i$.

In the simplest case one can use approximation $p_{ij}(\Theta) \simeq p(x_{ij},\Theta)\delta_{ij}$, $\delta_{ij} = \int_{X_{ij}} dx$. In general case, integrals are calculated numerically. The algorithm is realized in program ls3.

Multiresponse Case

If in model (1) the response is a vector, $y_i \in R^l$, and $E[\varepsilon_i \varepsilon_i^T] = D_i$, where D_i are given for all $i = \overline{1,n}$ then the solution of the minimization problem

$$\hat{\Theta} = \operatorname{Arg} \min_{\Theta} \sum_{i=1}^{n} [y_i - \eta(x_i,\Theta)]^T D_i^{-1} [y_i - \eta(x_i,\Theta)] \tag{15}$$

can be used as the estimate of Θ. It is clear that (15) has the same structure as (2). The corresponding numerical procedure is contained in program *ls* 4.

If the dispersion matrices D_i are unknown but there is the prior information that

$$E[\varepsilon_i \varepsilon_i^T] \equiv D(\Theta_t) \ , \ \operatorname{rank} D(\Theta_t) = l \ ,$$

then the estimator of Θ can be defined by the iterative procedure (Fedorov 1977, Phillips 1976)

$$\hat{\Theta} = \lim_{q \to \infty} \Theta_q \ , \tag{16}$$

$$\Theta_q = \operatorname{Arg} \min_{\Theta} \sum_{i=1}^{n} [y_i - \eta(x_i,\Theta)]^T D_{q-1}^{-1} [y_i - \eta(x_i,\Theta)] \ ,$$

$$D_{q-1} = n^{-1} \sum_{i=1}^{n} [y_i - \eta(x_i,\Theta_{q-1})][y_i - \eta(x_i,\Theta_{q-1})]^T \ .$$

Under some mild condition $\hat{\Theta}$ is strongly consistent and asymptotically random values $\sqrt{n}\,(\hat{\Theta} - \Theta_t)$ are normally distributed. Moreover,

$$D(\hat{\Theta}) = n^{-1} \sum_{i=1}^{n} [y_i - (x_i, \hat{\Theta})][y_i - \eta(x_i, \hat{\Theta})]^T$$

is a consistent estimator of $D(\Theta_t)$.

Procedure (16) is realized in program $ls5$.

Regression-Autoregression Models

In this case

$$y_i = \eta(x_i, y_{i-1}, \cdots, y_{i-k}, \Theta) + \varepsilon_i \quad . \tag{17}$$

Formally, variables $y_{i-1}, \cdots, y_{i-k}$ can be joined to the set of independent variables x_i (see, for instance, Anderson 1971) and the values

$$\hat{\Theta} = \text{Arg} \min_{\Theta} \sum_{i=1}^{n} \gamma^{-2}(x_i)[y_i - \eta(x_i, y_{i-1}, \cdots, y_{i-k}, \Theta)]^2 \tag{18}$$

can be used as estimates for Θ.

The problem (18) practically coincides with (2) but for convenience only set $(y_i, x_i)_1^n$ is used as an input in $ls6$

Observation of Deterministic Dynamic System

In systems analysis very often a response function $\eta(x, \Theta)$ can be described by a system of ordinary differential equation

$$\frac{\partial \eta}{\partial x} = \Psi(x, \Theta) \tag{19}$$

This specific case of regression model (2) can be treated with the help of $ls7$.

M-Estimates

When a researcher suspects that between random values ε_i can be outlined the so-called robust estimation methods are recommended. One of the most popular methods is M-method and corresponding estimators are defined by the minimization problem

$$\hat{\Theta} = \text{Arg min} \sum_{i=1}^{n} \rho[\,|y_i - \eta(x_i, \Theta)|\,] \ , \tag{20}$$

where $\rho(|z|)$ is usually some monotonous nondecreasing function of $|z|$.

The solution of (20) under some mild conditions (see, for instance, Mudrov and Kushko 1976) can be found with the help of the iterative procedure:

$$\hat{\Theta} = \lim_{q \to \infty} \Theta_q \ , \tag{21}$$

$$\Theta_q = \text{Arg min} \sum_{i=1}^{n} \frac{\rho[\,|y_i - \eta(x_i, \Theta_{q-1}|\,]}{[y_i - \eta(x_i, \Theta_{q-1})]^2} [y_i - \eta(x_i, \Theta)]^2$$

where the auxiliary minimization problem is the l.s. problem. Usually the stabilization of Θ_q happens after 3–4 iterations. Procedure (21) is realized in *ls*8 comprehensive discussion of statistical properties of (20) can be found, for instance, in Ershov (1978), Huber (1972).

Predictors Subject to Error

If a researcher wished to observe his system under conditions x_i but because of some random impact they happened to be $u_i = x_i + h_i$, then, instead of (1) one deals with the model

$$y_i = \eta(x_i + h_i, \Theta) + \varepsilon_i \ , \ i = \overline{1,n} \ , \tag{22}$$

where all ε_i and h_i are independent and $E[\varepsilon_i^2] = \gamma^2(x_i), E[h_i h_i^T] = D(x_i)$.

For model (22) the following estimation can be used (see Fedorov 1974):

$$\hat{\Theta} = \lim_{q \to \infty} \Theta_q \ , \tag{23}$$

$$\Theta_q = \text{Arg} \min_{\Theta} \sum_{i=1}^{n} \lambda(x_i, \Theta_{q-1})[y_i - \bar{\eta}(x_i, \Theta)]^2 \ ,$$

where

$$\bar{\eta}(x, \Theta) = \eta(x, \Theta) + \frac{1}{2} tr \ D(x) \frac{\partial^2 \eta(x, \Theta)}{\partial x \, \partial x^T} \ ,$$

$$\lambda^{-1}(x, \Theta) = \gamma^2(x) + \frac{\partial \eta(x, \Theta)}{\partial x^T} D(x) \frac{\partial \eta(x, \Theta)}{\partial x} \ .$$

The procedure (23) is realized in *ls*9.

In conclusion we emphasize once more that all algorithms described in this section are based on ls0.

V. NUMERICAL EXAMPLES

To illuminate the possibilities of the considered set of programs three simple regression problems will be considered. The data are borrowed from the paper by C. Marchetti, (1983) (see printout 1), and are extended for a few additional years. They describe the car population in Italy. In the cited paper it was suggested to use a regression model with logistic response function and with additive uncorrelated random errors (the latter statement is not expressed explicitly but it follows from the context):

$$y_i = \eta(x_i, \Theta) + \varepsilon_i = \frac{\Theta_3 e^{\Theta_1 + \Theta_2 x_i}}{1 + e^{\Theta_1 + \Theta_2 x_i}} + \varepsilon_i \ , \tag{24}$$

where $\Theta = (\Theta_1, \Theta_2, \Theta_3)^T$ are unknown parameters, x_i stands for time.

Formula (24) does not describe completely the regression model. It is still necessary to clarify the assumed properties of errors ε_i.

Three possible variants will be considered here:

1) Variance $E[\varepsilon_i^2]$ is independent of x_i and constant.

2) Variance $E[\varepsilon_i^2]$ is equal to $\delta^2\eta^2(x_i, \Theta)$, where δ^2 has to be also estimated ($\sqrt{E[\varepsilon_i^2]}/\eta(x_i, \Theta)$ = const, "relative error" is constant).

3) $E[\varepsilon_i^2] = \delta^2\eta(x_i, \Theta)$.

In the first case, program ls0 was used. The results are evident from printout 1.

It is interesting to stress two facts. Firstly, the residuals (see column "Y-F" and "NO.RES" and comments in the printout) have a tendency to increase. Secondly, their signs are definitely not randomly distributed.

Therefore, one may suspect that the more complicated second case is closer to reality and moreover the errors are correlated or response function does not reflect reality. We are not concerned here with accurate statistical analysis of the problem but only with illumination of how the software is working and therefore we restrict ourselves only by struggle with nonhomogeneity of errors using the hypothesis of cases 2 and 3.

The results are on printout 2. As initial values for estimated parameters, the estimates from the previous case were taken. The relative discrepancy between estimates happened to be more than 5%. Of course, this is not too much but is several times more than their standard errors. Therefore one can assert that the correponding two models give significantly (in a statistical sense) different results.

Unfortunately for the second model, the residuals have an inverse tendency: they decrease in average. Recollecting that in growth processes the observed value very often distributed according to Poisson's law, the third version with $E[\varepsilon_i^2] = \eta(x_i, \Theta)$ was analyzed. The final results are in printout 3. It is clear that now residuals have no tendency to increase or decrease systematically but their signs appear in long series.

It seems that all technical details on programming are clear from the printouts and marginal comments. Usually the content of input information are defined by questions which appear on the screen after application to a program from the considered set. Subprograms for response functions and weight functions should be located in auxiliary files "resp" and "weight," correspondingly. For other programs it becomes necessary to use some additional auxiliary files. For instance, program ls3 uses file "DEN" for a density function $p(x, \Theta)$, program ls6 uses file "AVT" for autoregression function, program ls7 uses file "DIFUR" for $\Psi(x, \Theta)$ (see (19)) and so on. More detailed information can be obtained from IIASA's Computer Services.

P R O G R A M   LS0

see Note No.		DATA: Y	X	W
(1)	1	0.3420e+00	0. e+00	0.1000e+01
	2	0.6130e+00	0.3000e+01	0.1000e+01
	3	0.6910e+00	0.4000e+01	0.1000e+01
	4	0.8610e+00	0.5000e+01	0.1000e+01
	5	0.1031e+01	0.6000e+01	0.1000e+01
	6	0.1231e+01	0.7000e+01	0.1000e+01
	7	0.1393e+01	0.8000e+01	0.1000e+01
	8	0.1659e+01	0.9000e+01	0.1000e+01
	9	0.1976e+01	0.1000e+02	0.1000e+01
	10	0.2449e+01	0.1100e+02	0.1000e+01
	11	0.3030e+01	0.1200e+02	0.1000e+01
	12	0.3913e+01	0.1300e+02	0.1000e+01
	13	0.4675e+01	0.1400e+02	0.1000e+01
	14	0.5473e+01	0.1500e+02	0.1000e+01
	15	0.6357e+01	0.1600e+02	0.1000e+01
	16	0.7295e+01	0.1700e+02	0.1000e+01
	17	0.8266e+01	0.1800e+02	0.1000e+01
	18	0.9174e+01	0.1900e+02	0.1000e+01
	19	0.1019e+02	0.2000e+02	0.1000e+01
	20	0.1129e+02	0.2100e+02	0.1000e+01
	21	0.1248e+02	0.2200e+02	0.1000e+01
	22	0.1343e+02	0.2300e+02	0.1000e+01
	23	0.1430e+02	0.2400e+02	0.1000e+01
	24	0.1506e+02	0.2500e+02	0.1000e+01
	25	0.1593e+02	0.2600e+02	0.1000e+01
	26	0.1647e+02	0.2700e+02	0.1000e+01
	27	0.1624e+02	0.2800e+02	0.1000e+01
	28	0.1713e+02	0.2900e+02	0.1000e+01
	29	0.1702e+02	0.3000e+02	0.1000e+01
	30	0.1770e+02	0.3100e+02	0.1000e+01
	31	0.1845e+02	0.3200e+02	0.1000e+01

(2)

```
NUMBER OF PARAMETERS    3
NUMBER OF VARIABLES    1
NUMBER OF CASES   31
INITIAL PARAMETERS -0.1000e+02  0.5000e+00  0.2500e+02
```

Note (1)

Data should be saved in the auxiliary file "ent.data". The first column is observations (dependent variable or response). The next columns contain the values of predictors (independent variables, controls). The last column describes "weights" of observations: $W_t = \sigma_t^{-2}$ in the simplest case. When only ratios between variances are known, then $W_t = \sigma^2\gamma^{-2}(x_t)$, where σ^2 will be estimated any initially and reasonable positive number can be used in input statement (details see in the main text or in subsequent columns). If one wish to get forecasting at some prescribed points which do not belong to "data" these points may be introduced with "small" weights (i.e., 10^{-9}).

Note (2)

Number of estimated parameters, number of independent variables, number of observations, initial value of estimated parameters.

```
              INTERNAL CONSTANTS

(3.1)    PARAMETER ERROR = Ø
         MAXIMUM NUMBER OF ITERATIONS    5Ø
(3.2)    NUMBER OF FREE PARAMETERS     3
(3.2)    THEIR NUMBERS    1  2  3
(3.3)    DELTA INITIAL H= Ø.1Øe+ØØ,  DELTA LAST H1= Ø.1Øe-Ø1
         LIMIT FOR PIVOTING: TOL=  Ø.1Øe-Ø9
(3.4)    NUMBER OF DIVISIONS: K1= 2
(3.5)    NUMBER OF DIVISIONS FOR RANDOM VECTOR: K2= 2
(3.6)    CONSTANTS FOR CONVERGENCE CRITERION: L1= 2,   L2= 2
(3.7)      REDUCTION OF RES.SUM: REDS=    1.ØØØØØe+Ø8

        ITER. ADD.   RES.SUM        PARAMETERS

                   Ø.11721e+Ø4 -Ø.1ØØe+Ø2  Ø.5ØØe+ØØ  Ø.275e+Ø2
                   Ø.1Ø566e+Ø4 -Ø.1ØØe+Ø2  Ø.55Øe+ØØ  Ø.25Øe+Ø2
                   Ø.71478e+Ø3 -Ø.1ØØe+Ø2  Ø.5ØØe+ØØ  Ø.25Øe+Ø2
                   Ø.5859Øe+Ø3 -Ø.11Øe+Ø2  Ø.5ØØe+ØØ  Ø.25Øe+Ø2
          1    Ø   Ø.14496e+Ø3 -Ø.115e+Ø2  Ø.598e+ØØ  Ø.178e+Ø2
          2    2   Ø.14117e+Ø3 -Ø.114e+Ø2  Ø.589e+ØØ  Ø.178e+Ø2
          3    1   Ø.53154e+Ø2 -Ø.745e+Ø1  Ø.388e+ØØ  Ø.177e+Ø2
          4    Ø   Ø.14Ø38e+Ø2 -Ø.366e+Ø1  Ø.2Ø1e+ØØ  Ø.184e+Ø2
          5    Ø   Ø.283Ø4e+Ø1 -Ø.444e+Ø1  Ø.238e+ØØ  Ø.181e+Ø2
          6    2   Ø.27811e+Ø1 -Ø.446e+Ø1  Ø.239e+ØØ  Ø.181e+Ø2
          7    Ø   Ø.19678e+Ø1 -Ø.433e+Ø1  Ø.232e+ØØ  Ø.184e+Ø2
          8    Ø   Ø.16112e+Ø1 -Ø.426e+Ø1  Ø.226e+ØØ  Ø.187e+Ø2
          9    Ø   Ø.1Ø859e+Ø1 -Ø.431e+Ø1  Ø.224e+ØØ  Ø.19Øe+Ø2
         1Ø    Ø   Ø.1Ø828e+Ø1 -Ø.43Øe+Ø1  Ø.224e+ØØ  Ø.19Øe+Ø2
         11    Ø   Ø.1Ø828e+Ø1 -Ø.43Øe+Ø1  Ø.224e+ØØ  Ø.19Øe+Ø2
         12    2   Ø.1Ø828e+Ø1 -Ø.43Øe+Ø1  Ø.224e+ØØ  Ø.19Øe+Ø2
```

Note (3.1)

If only the ratios between variances of observations are known, then 0 should be used. If all variances (or weights) are known, then 1 should be put in.

Note (3.2)

Sometimes it is useful to fix some of the estimated parameters. Then their numbers should not appear here.

Note (3.3)

Define the size of operability region of linear approximation (see page 7).

Note (3.4)

Upper bound for a number of step-length divisions.

Note (3.5)

Upper bound for a number of halfings of the additional random vector length. Usually moving in some randomly chosen direction helps to avoid singularity of approximation (9).

Note (3.6)

L_1 is the number of iterations when the inequality $\frac{v_{s-1}^2 - v_s^2}{v_s^2} < 10^{L_2}$ must be fulfilled ($v_s^2 = \sum_{i=1}^{n} W_i[y_i - \eta(x_i, \theta_s)]^2$) and the program will be terminated.

Note (3.7)

Additional choice of the terminating of the program. The program will be terminated if $v_0^2 / v_s^2 \geq REDS$.

Printout 1 — continued

(4)

	Y	F	Y-F	W	W*(Y-F)**2	STA.DEV.F	NO.RES
1	0.3420	0.25443	0.08757	1.0000	0.00767	0.01706	0.20
2	0.6130	0.49229	0.12071	1.0000	0.01457	0.02685	0.38
3	0.6910	0.61207	0.07893	1.0000	0.00623	0.03083	0.17
4	0.8610	0.75980	0.10120	1.0000	0.01024	0.03513	0.27
5	1.0310	0.94136	0.08964	1.0000	0.00803	0.03963	0.22
6	1.2310	1.16355	0.06745	1.0000	0.00455	0.04424	0.12
7	1.3930	1.43402	-0.04102	1.0000	0.00168	0.04875	0.05
8	1.6590	1.76117	-0.10217	1.0000	0.01044	0.05298	0.29
9	1.9760	2.15381	-0.17781	1.0000	0.03162	0.05661	0.89
10	2.4490	2.62071	-0.17171	1.0000	0.02948	0.05945	0.84
11	3.0300	3.16979	-0.13979	1.0000	0.01954	0.06119	0.56
12	3.9130	3.80723	0.10577	1.0000	0.01119	0.06187	0.32
13	4.6750	4.53618	0.13882	1.0000	0.01927	0.06152	0.55
14	5.4730	5.35558	0.11742	1.0000	0.01379	0.06073	0.39
15	6.3570	6.25906	0.09794	1.0000	0.00959	0.06015	0.27
16	7.2950	7.23430	0.06070	1.0000	0.00369	0.06059	0.11
17	8.2660	8.26314	0.00286	1.0000	0.00001	0.06241	0.00
18	9.1740	9.32262	-0.14862	1.0000	0.02209	0.06504	0.64
19	10.1910	10.38686	-0.19586	1.0000	0.03836	0.06779	1.13
20	11.2940	11.42948	-0.13548	1.0000	0.01835	0.06939	0.54
21	12.4840	12.42622	0.05778	1.0000	0.00334	0.06913	0.10
22	13.4250	13.35706	0.06794	1.0000	0.00462	0.06703	0.14
23	14.3040	14.20752	0.09648	1.0000	0.00931	0.06382	0.27
24	15.0600	14.96914	0.09086	1.0000	0.00825	0.06046	0.24
25	15.9250	15.63909	0.28591	1.0000	0.08175	0.05868	2.32
26	16.4660	16.21915	0.24685	1.0000	0.06094	0.05989	1.74
27	16.2410	16.71455	-0.47355	1.0000	0.22425	0.06437	6.50
28	17.1250	17.13272	-0.00772	1.0000	0.00006	0.07153	0.00
29	17.0230	17.48224	-0.45924	1.0000	0.21090	0.08024	6.54
30	17.6960	17.77196	-0.07596	1.0000	0.00577	0.08952	0.19
31	18.4500	18.01047	0.43954	1.0000	0.19319	0.09873	6.68

Note (4)

Y = observations

F = response function

STA.DEV.F is the estimation $\hat{d}[\eta(x,\theta)]$ of the variance of $\eta(x,\hat{\theta})$ at point x.

$$\text{NO.RES} = \frac{[Y_i - \eta(x_i,\hat{\theta})]^2}{\hat{\sigma}^2 - \hat{d}[\eta(x,\theta)]}$$

It is useful to remark that the positive and negative deviations appear here systematically. Therefore the hypothesis on independency of errors should be rejected.

```
(5)    RESIDUAL SUM W*(Y-F)**2                RESIDUAL SUM /(N-M)
       0.108277e+01                           0.386702e-01

       ESTIMATE OF PARAMETERS
       -0.433109e+01  0.224266e+00  0.193261e+02

       STANDARD DEVIATIONS OF PARAMETER ESTIMATES
           0.626d-01      0.414d-02      0.163d+00

       ESTIMATE OF COVARIANCE MATRIX
      0.391d-02
      -.247d-03 0.172d-04
      0.625d-02 -.535d-03 0.265d-01

       ESTIMATE OF CORRELATION MATRIX
      0.100d+01
      -.953d+00 0.100d+01
      0.614d+00 -.793d+00 0.100d+01

(6.1)  HISTOGRAM OF RESIDUALS Y-F

             INTERVAL    DENSITY
            -0.47446     0.0645
            -0.37957     0.
            -0.28468     0.0323
            -0.18978     0.1935
            -0.09489     0.0968
             0.          0.2903
             0.09489     0.2258
             0.18978     0.0323
             0.28468     0.0323
             0.37957     0.0323
```

Note (5)

Estimates of unknown parameters and their covariance and correlation matrices. It is useful to remember that residual sum/(N-M) = σ^2, see also Note (1).

Note (6.1)

Additional outputs which can be useful in the model testing and forecasting.

TEST ON INDEPENDENCE OF RESIDUALS

MEAN VALUE OF VECTOR X 16.9355

(6.2)

	NUMBER OF POINT	DISTANSE	Y-F
1	16	0.064516	0.060705
2	15	0.935484	0.097943
3	17	1.064516	0.002859
4	14	1.935484	0.117423
5	18	2.064516	-0.148624
6	13	2.935484	0.138824
7	19	3.064516	-0.195855
8	12	3.935484	0.105772
9	20	4.064516	-0.135478
10	11	4.935484	-0.139794
11	21	5.064516	0.057776
12	10	5.935484	-0.171706
13	22	6.064516	0.067942
14	9	6.935484	-0.177814
15	23	7.064516	0.096482
16	8	7.935484	-0.102172
17	24	8.064516	0.090857
18	7	8.935484	-0.041025
19	25	9.064516	0.285913
20	6	9.935484	0.067448
21	26	10.064516	0.246853
22	5	10.935484	0.089637
23	27	11.064516	-0.473549
24	4	11.935484	0.101202
25	28	12.064516	-0.007725
26	3	12.935484	0.078933
27	29	13.064516	-0.459236
28	2	13.935484	0.120714
29	30	14.064516	-0.075956
30	31	15.064516	0.439535
31	1	16.935484	0.087573

TABLES OF RESPONSE F(X,P) AS FUNCTION OF X

(6.3) TABLE 1

	F(X,P)	X(1) , . . .
1	0.2544	0.
2	0.7598	5.0000
3	2.1538	10.0000
4	5.3556	15.0000
5	10.3869	20.0000
6	14.9691	25.0000
7	17.4822	30.0000
8	18.4939	35.0000
9	18.8494	40.0000
10	18.9682	45.0000

Note (6.2)

DISTANCE = $\sqrt{(x_i - \bar{x})^T(x_i - \bar{x})}$, $\bar{x} = N^{-1}\sum_{i=1}^{N} x_i$. This information is useful in the case of multidimensional x.

$E[\varepsilon_t^2] = \sigma^2 \eta(x_t, \theta)$

```
                         P R O G R A M  LS1

        DATA: Y,X

     1  0.3420e+00  0.     e+00
     2  0.6130e+00  0.3000e+01
     3  0.6910e+00  0.4000e+01
     4  0.8610e+00  0.5000e+01
                 .
                 .
                 .
    29  0.1702e+02  0.3000e+02
    30  0.1770e+02  0.3100e+02
    31  0.1845e+02  0.3200e+02

    NUMBER OF PARAMETERS    3
    NUMBER OF VARIABLES    1
    NUMBER OF CASES   31
    INITIAL PARAMETERS -0.4300e+01  0.2250e+00  0.2000e+02

          INTERNAL CONSTANTS

    PARAMETER ERROR = 0
    MAXIMUM NUMBERS OF ITERATIONS     3    50
    NUMBER OF FREE PARAMETERS    3
    THEIR NUMBERS   1  2  3
    DELTA INITIAL H= 0.30e-01,  DELTA LAST H1= 0.50e-02
    LIMIT FOR PIVOTING: TOL=  0.10e-09
    NUMBER OF DIVISIONS: K1= 2
    NUMBER OF DIVISIONS FOR RANDOM VECTOR: K2= 2
    CONSTANTS FOR CONVERGENCE CRITERION: L1= 2,  L2= 3

   ITER. ADD.   RES.SUM        PARAMETERS

             0.44965e+00 -0.430e+01  0.232e+00  0.200e+02
             0.38465e+00 -0.443e+01  0.225e+00  0.200e+02
             0.31867e+00 -0.430e+01  0.225e+00  0.206e+02
             0.23643e+00 -0.430e+01  0.225e+00  0.200e+02
     1    0  0.81424e-01 -0.412e+01  0.206e+00  0.200e+02
     2    0  0.78516e-01 -0.416e+01  0.206e+00  0.204e+02
     3    0  0.78314e-01 -0.415e+01  0.206e+00  0.203e+02
             0.27709e+00 -0.428e+01  0.206e+00  0.203e+02
                                .
                                .
                                .
             0.78449e-01 -0.416e+01  0.207e+00  0.202e+02
     6    4  0.78449e-01 -0.416e+01  0.207e+00  0.202e+02
             0.79161e-01 -0.416e+01  0.207e+00  0.203e+02
             0.84577e-01 -0.418e+01  0.207e+00  0.202e+02
             0.80401e-01 -0.416e+01  0.208e+00  0.202e+02
             0.78449e-01 -0.416e+01  0.207e+00  0.202e+02
     7    0  0.78448e-01 -0.416e+01  0.207e+00  0.202e+02
```

Number of iterations within basic ls0. Usually it should be equal to 3–5.

Number of iterations described by (11), (12), page 8.

	Y	F	Y-F	W	W*(Y-F)**2	STA.DEV.F	NO.RES
1	0.3420	0.30966	0.03234	10.4313	0.01091	0.00913	5.65
2	0.6130	0.56933	0.04367	3.0861	0.00588	0.01246	2.53
3	0.6910	0.69602	-0.00502	2.0649	0.00005	0.01369	0.02
4	0.8610	0.84967	0.01133	1.3857	0.00018	0.01504	0.07
5	1.0310	1.03544	-0.00444	0.9331	0.00002	0.01660	0.01
6	1.2310	1.25919	-0.02819	0.6309	0.00050	0.01855	0.19
7	1.3930	1.52744	-0.13444	0.4288	0.00775	0.02119	2.97
8	1.6590	1.84727	-0.18827	0.2932	0.01039	0.02486	3.97
9	1.9760	2.22610	-0.25010	0.2019	0.01263	0.02993	4.82
10	2.4490	2.67131	-0.22231	0.1402	0.00693	0.03668	2.65
11	3.0300	3.18977	-0.15977	0.0983	0.00251	0.04527	0.97
12	3.9130	3.78714	0.12586	0.0698	0.00111	0.05567	0.43
13	4.6750	4.46703	0.20797	0.0501	0.00217	0.06762	0.84
14	5.4730	5.23013	0.24287	0.0366	0.00216	0.08064	0.84
15	6.3570	6.07332	0.28368	0.0271	0.00218	0.09408	0.85
16	7.2950	6.98905	0.30595	0.0205	0.00192	0.10712	0.75
17	8.2660	7.96509	0.30091	0.0158	0.00143	0.11902	0.55
18	9.1740	8.98485	0.18915	0.0124	0.00044	0.12927	0.17
19	10.1910	10.02831	0.16269	0.0099	0.00026	0.13793	0.10
20	11.2940	11.07347	0.22053	0.0082	0.00040	0.14582	0.15
21	12.4840	12.09821	0.38579	0.0068	0.00102	0.15452	0.39
22	13.4250	13.08211	0.34289	0.0058	0.00069	0.16601	0.26
23	14.3040	14.00795	0.29605	0.0051	0.00045	0.18188	0.17
24	15.0600	14.86282	0.19718	0.0045	0.00018	0.20273	0.07
25	15.9250	15.63845	0.28655	0.0041	0.00034	0.22797	0.13
26	16.4660	16.33109	0.13491	0.0038	0.00007	0.25626	0.03
27	16.2410	16.94089	-0.69989	0.0035	0.00171	0.28606	0.68
28	17.1250	17.47107	-0.34607	0.0033	0.00039	0.31594	0.16
29	17.0230	17.92705	-0.90405	0.0031	0.00254	0.34476	1.05
30	17.6960	18.31553	-0.61954	0.0030	0.00114	0.37177	0.48
31	18.4500	18.64389	-0.19389	0.0029	0.00011	0.39648	0.05

THE RESIDUAL SUM W*(Y-F)**2 THE RESIDUAL SUM /(N-M)
0.784476e-01 0.280170e-02

ESTIMATE OF PARAMETERS
-0.416336e+01 0.207376e+00 0.202167e+02

STANDARD DEVIATIONS OF PARAMETER ESTIMATES
0.312d-01 0.312d-02 0.550d+00

ESTIMATE OF COVARIANCE MATRIX
0.976d-03
-.230d-04 0.975d-05
-.826d-02 -.115d-02 0.302d+00

ESTIMATE OF CORRELATION MATRIX
0.100d+01
-.235d+00 0.100d+01
-.481d+00 -.668d+00 0.100d+01

Printout 3

$E(\varepsilon_i^2) = \sigma^2\eta^2(x_i, \theta)$

	Y	F	Y-F	W	W*(Y-F)**2	STA.DEV.F	NO.RES
1	0.3420	0.27282	0.06918	3.6626	0.01753	0.01145	2.96
2	0.6130	0.52002	0.09298	1.9219	0.01662	0.01722	2.85
3	0.6910	0.64332	0.04768	1.5536	0.00353	0.01945	0.61
4	0.8610	0.79462	0.06638	1.2579	0.00554	0.02179	0.95
5	1.0310	0.97963	0.05137	1.0204	0.00269	0.02418	0.46
6	1.2310	1.20490	0.02610	0.8296	0.00057	0.02657	0.10
7	1.3930	1.47777	-0.08477	0.6765	0.00486	0.02890	0.83
8	1.6590	1.80625	-0.14725	0.5535	0.01200	0.03116	2.05
9	1.9760	2.19871	-0.22271	0.4547	0.02255	0.03338	3.83
10	2.4490	2.66340	-0.21440	0.3754	0.01726	0.03571	2.91
11	3.0300	3.20781	-0.17781	0.3117	0.00985	0.03842	1.66
12	3.9130	3.83772	0.07528	0.2606	0.00148	0.04190	0.25
13	4.6750	4.55616	0.11884	0.2195	0.00310	0.04649	0.52
14	5.4730	5.36225	0.11075	0.1865	0.00229	0.05235	0.39
15	6.3570	6.25021	0.10679	0.1600	0.00182	0.05926	0.31
16	7.2950	7.20879	0.08621	0.1387	0.00103	0.06661	0.18
17	8.2660	8.22129	0.04471	0.1216	0.00024	0.07355	0.04
18	9.1740	9.26639	-0.09239	0.1079	0.00092	0.07921	0.16
19	10.1910	10.31983	-0.12883	0.0969	0.00161	0.08290	0.28
20	11.2940	11.35656	-0.06256	0.0881	0.00034	0.08436	0.06
21	12.4840	12.35307	0.13093	0.0810	0.00139	0.08397	0.24
22	13.4250	13.28948	0.13552	0.0752	0.00138	0.08269	0.23
23	14.3040	14.15085	0.15315	0.0707	0.00166	0.08205	0.28
24	15.0600	14.92781	0.13219	0.0670	0.00117	0.08362	0.20
25	15.9250	15.61632	0.30868	0.0640	0.00610	0.08839	1.03
26	16.4660	16.21695	0.24905	0.0617	0.00382	0.09635	0.66
27	16.2410	16.73376	-0.49276	0.0598	0.01451	0.10670	2.54
28	17.1250	17.17324	-0.04824	0.0582	0.00014	0.11836	0.02
29	17.0230	17.54321	-0.52021	0.0570	0.01542	0.13040	2.84
30	17.6960	17.85205	-0.15605	0.0560	0.00136	0.14212	0.26
31	18.4500	18.10803	0.34197	0.0552	0.00646	0.15310	1.26

THE RESIDUAL SUM W*(Y-F)**2 THE RESIDUAL SUM /(N-M)
0.179258e+00 0.640206e-02

ESTIMATE OF PARAMETERS
-0.424125e+01 0.219394e+00 0.192321e+02

STANDARD DEVIATIONS OF PARAMETER ESTIMATES
0.378d-01 0.309d-02 0.222d+00

ESTIMATE OF COVARIANCE MATRIX
0.143d-02
-.989d-04 0.956d-05
0.262d-02 -.490d-03 0.494d-01

ESTIMATE OF CORRELATION MATRIX
0.100d+01
-.847d+00 0.100d+01
0.312d+00 -.714d+00 0.100d+01

REFERENCES

Anderson, T.W. 1971. *The Statistical Analysis of Time Series*. New York: Wiley.

BMDP Statistical Software. 1983. W.J. Dixon, editor. University of California Press.

Chambers, J.M. 1973. Fitting nonlinear models: numerical techniques. *Biometrika*, 60:3–18.

Ershov, A.A. 1978. Robust method of estimation. *Automatika and Telemechanics*, 8:25–50.

Fedorov, V.V. 1974. Regression problems with controllable variables subject to error. *Biometrika*, 61:49–56.

Fedorov, V.V. and A.B. Uspensky. 1975. Numerical aspects of least squares method in design and analysis of regression experiments. Moscow State University (in Russian).

Fedorov, V.V. 1977. Estimation of regression parameters in multiresponse case. In *Regression Experiments*. Moscow: Moscow University (in Russian).

Jennrich, R.I. 1969. Asymptotic properties of nonlinear least squares estimator. *Ann. Math. Statist.* 1, 40:633–643.

Jennrich, R.I. and M.L. Ralston. 1978. Fitting nonlinear models to data. Technical Report No. 46. Los Angeles: University of California.

Huber, P.J. 1972. Robust statistics: a review. *Ann. Math. Statist.*, 43:1041–1067.

Malyutov, M.B. 1982. Asymptotical properties and applications of IRDEIN estimators of parameters in general regression model. In *Stochastic Process and Applications*. pp. 144–158. Moscow (in Russian).

Marchetti, C. 1983. The automobile in a system context: the past 80 years and the next 20 years. RR–83–18. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Mudrov, V.I. and V.L. Kushko. 1976. *Methods of Analysis of Observations*. Moscow: Soviet Radio (in Russian).

Peckham, G. 1970. A new method for minimizing a sum of squares without calculating gradients. *Computer J.*, 13:418–420.

Phillips, P.C.B. 1976. The iterated minimum distance estimator and the quasi-maximum likelihood estimator. *Econometrica*, 44:449–460.

Ralston, M.L. and R.I. Jennrich. 1978. DUD, a derivative-free algorithm for nonlinear least squares. *Technometrics*, 20:7–14.

Rao, C.R. 1965. *Linear Statistical Methods and Their Applications*. New York: Wiley.

Vereskov, A.I. 1981. Convergence of some derivative free algorithms. In *Methods of Complex Systems Studying*. pp. 58–62. Moscow: VNIISI (in Russian).

Vereskov, A.I. and A.S. Pshennikov. 1983. Estimation of aggregated parameters of work duration. In *Dynamics of Non-Homogeneous Systems*. Moscow: VNIISI (in Russian).

Wu, C.F. 1981. Asymptotic theory of nonlinear least squares estimations. *Ann. Stat.*, 9:501–513.