Springer Optimization and Its Applications 54

Pavel S. Knopov
Arnold S. Korkhin

Regression Analysis Under A Priori Parameter Restrictions

Springer

Springer Optimization and Its Applications

VOLUME 54

Aims and Scope
Optimization has been expanding in all directions at an astonishing rate during the last few decades. New algorithmic and theoretical techniques have been developed, the diffusion into other disciplines has proceeded at a rapid pace, and our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in all areas of applied mathematics, engineering, medicine, economics, and other sciences.

The series *Springer Optimization and Its Applications* publishes undergraduate and graduate textbooks, monographs and state-of-the-art expository work that focus on algorithms for solving optimization problems and also study applications involving such problems. Some of the topics covered include nonlinear optimization (convex and nonconvex), network flow problems, stochastic optimization, optimal control, discrete optimization, multi-objective programming, description of software packages, approximation techniques and heuristic approaches.

For further volumes:
http://www.springer.com/series/7393

Pavel S. Knopov • Arnold S. Korkhin

Regression Analysis Under A Priori Parameter Restrictions

Springer

Pavel S. Knopov
Department of Mathematical Methods
of Operation Research
V.M. Glushkov Institute of Cybernetics
National Academy of Science of Ukraine
03187 Kiev
Ukraine
knopov1@yahoo.com

Arnold S. Korkhin
Department of Economical Cybernetics
and Information Technology
National Mining University
49005 Dnepropetrovsk
Ukraine
korkhina@nmu.org.ua

ISSN 1931-6828
ISBN 978-1-4614-0573-3 e-ISBN 978-1-4614-0574-0
DOI 10.1007/978-1-4614-0574-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011935145

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Regression analysis has quite a long history. It is conventional to think that it goes back to the works of Gauss on approximation of experimental data. Nowadays, regression analysis represents a separate scientific branch, which is based on optimization theory and mathematical statistics. Formally, there exist two branches of regression analysis: theoretical and applied.

Up to recent time, developments in regression analysis were based on the hypothesis that the domain of regression parameters has no restrictions. Divergence from that approach came later on when equality constraints were taken into account, which allowed use of some a priori information about the regression model. Methods of constructing the regression with equality constraints were first investigated in Rao (1965) and Bard (1974).

Usage of inequality constraints in a regression model gives much more possibilities to utilize available a priori information. Moreover, the representation of the admissible domain of parameters in the form of inequality constraints naturally includes the cases when constraints are given as equalities.

Properties of the regression with inequality constraints are investigated in many papers, in particular, in Zellner (1971), Liew (1976), Nagaraj and Fuller (1991) and Thomson and Schmidt (1982), where some particular cases are considered. Detailed qualitative analysis of the properties of estimates in case of linear regression with linear constraints is given in the monograph (Malinvaud 1969, Section 9.8).

Asymptotic properties of the estimates of regression parameters in regression with finite number of parameters under some known a priori information are studied in Dupacova and Wets (1986), Knopov (1997a–c), Korkhin (1985), Wang (1996), etc. We note that the results obtained in Korkhin (1985) and Wang (1996) under different initial assumptions, almost coincide. There are many results concerning practical implementation of regression models with inequality constraints, for example, Liew (1976), Rezk (1996) and McDonald (1999), Thomson (1982), Thomson and Schmidt (1982). This problem was also studied in Gross (2003, Subsection 3.3.2).

In this monograph, we present in full detail the results on estimation of unknown parameters in regression models under a priori information, described in the form of inequality constraints. The book covers the problem of estimation of regression parameters as well as the problem of accuracy of such estimation. Both problems are studied is cases of linear and nonlinear regressions. Moreover, we investigate the applicability of regression with constraints to problems of point and interval prediction.

The book is organized as follows.

In Chapter 1, we consider methods of calculation of parameter estimates in linear and nonlinear regression with constraints. In this chapter we describe methods of solving optimization problems which take into account the specification of regression analysis.

Chapter 2 is devoted to asymptotic properties of regression parameters estimates in linear and nonlinear regression. Both cases of equality and inequality constraints are considered.

In Chapter 3, we consider various generalizations of the estimation problem by the least squares method in nonlinear regression with inequality constraints on parameters. In particular, we discuss the results concerning robust Huber estimates and regressors which are continuous functions of time.

Chapter 4 is devoted to the problem of accuracy estimation in (linear and nonlinear) regression, when parameters are estimated by means of the least squares method.

In Chapter 5, we discuss/consider statistical properties of estimates of parameters in nonlinear regression, which are obtained on each iteration of the solution to the estimation problem. Here we use algorithms described in Chap. 1. Obtained results might be useful in practical implementation of regression analysis.

Chapter 6 is devoted to problems of prediction by linear regression with linear constraints.

Kiev, Ukraine Pavel S. Knopov
Dnepropetrovsk, Ukraine Arnold S. Korkhin

Acknowledgments

We are very grateful to the scientific editor of this book, Professor Panos Pardalos, senior publishing editor, Elizabeth Loew, and to the associate editor in mathematics, Nathan Brothers, for their helpful support and collaboration in preparation of the manuscript.

We thank our colleagues from V.M. Glushkov Institute of Cybernetics of National Academy of Science of Ukraine for many helpful discussions on the problems and results described and presented in this book.

We thank our colleagues L. Belyavina, L. Vovk, V. Knopova, Yu. Kolesnik, E. Odinzova, for invaluable help during the preparation of our book for publication.

Contents

Notation

m.s.e.	Mean square error
ECLS estimate	Estimate of the regression parameter my means of the least squares method with equality constraints
$\lvert I \rvert$	Cardinality of the set I
ICLS estimate	Estimate of the regression parameter my means of the least squares method with inequality constraints
$\mathbf{J}_n$	Unit matrix of order n
LS	Least squares method
LS estimate	Estimate of the regression parameter my means of the least squares method without restrictions
$\mathbf{M}'$	Transposition of a matrix (vector) $\mathbf{M}$
$\mathbf{O}_{mn}$	Zero $(m \times n)$ matrix
$\mathbf{O}_n$	Zero n-dimensional vector
$p \lim$	Means convergence in probability
$\mathbf{1}_n$	$n-$ dimensional vector with entries equal to 1
$\lvert\lvert \cdot \rvert\rvert$	Euclidean norm of a vector (matrix)
$\stackrel{p}{\Rightarrow}$	Convergence in distribution
$\boldsymbol{\varepsilon} \sim N(\mathbf{M}_1, \mathbf{M}_2)$	$\boldsymbol{\varepsilon}$ has a normal distribution with mean $\mathbf{M}_1$ and covariance $\mathbf{M}_2$

Chapter 1
Estimation of Regression Model Parameters with Specific Constraints

Consider the regression

$$y_t = \tilde{f}(\mathbf{x}_t, \boldsymbol{\alpha}^0) + \varepsilon_t, \quad t = 1, 2, \ldots, \tag{1.1}$$

where $y_t \in \Re^1$ is the dependent variable, $\mathbf{x}_t \in \Re^q$ is an argument (regressor), $\boldsymbol{\alpha}^0 \in \Re^n$ is a true regression parameter (unknown), $\tilde{f}(\mathbf{x}_t, \boldsymbol{\alpha})$ is some (nonlinear) function of $\boldsymbol{\alpha}$, ε_t is a noise, and t is an observation number.

In what follows the symbol "$\prime$" denotes the transposition.

We will use the function $\tilde{f}(\mathbf{x}_t, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha} \in \Re^n$ is a dependent variable, for estimation of $\boldsymbol{\alpha}^0$ and for investigation of the obtained estimates.

For convenience we write

$$f_t(\boldsymbol{\alpha}) = \tilde{f}(\mathbf{x}_t, \boldsymbol{\alpha}), \quad t = 1, 2, \ldots \tag{1.2}$$

and call such a function the regression function.

Assume that a priori parameter constraints are known:

$$g_i(\boldsymbol{\alpha}^0) \leq 0, \quad i = \overline{1, m}. \tag{1.3}$$

System of inequalities (1.3) involves equalities as a particular case due to the fact that any equality can be represented in the form of two inequalities:

$$g_i(\boldsymbol{\alpha}^0) \leq 0 \quad \text{and} \quad -g_i(\boldsymbol{\alpha}^0) \leq 0.$$

Suppose that for $t \in [1, T]$ the values of y_t and $\mathbf{x}_t \in \Re^n$ are known. In the present chapter the estimation of the parameter $\boldsymbol{\alpha}^0$ will be done by means of the least squares method, i.e.

$$S(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{t=1}^{T}(y_t - f_t(\boldsymbol{\alpha}))^2 \rightarrow \min, \tag{1.4}$$

P.S. Knopov and A.S. Korkhin, *Regression Analysis Under A Priori Parameter Restrictions,* Springer Optimization and Its Applications 54,
DOI 10.1007/978-1-4614-0574-0_1, © Springer Science+Business Media, LLC 2012

under the constraints

$$g_i(\boldsymbol{\alpha}) \le 0, \quad i = \overline{1,m}, \tag{1.5}$$

where T is the length of the observed dynamic (time) series $\mathbf{x}_t$ and y_t.

Since the case of the linear regression and linear constraints on $\boldsymbol{\alpha}$ is extremely important and is used for nonlinear estimation algorithms, it will be discussed separately in Sect. 1.1.

Section 1.2 is dedicated to nonlinear estimation, i.e., to solving the problems (1.4) and (1.5) under rather general setting. Section 1.3 is dedicated to the perspective for economical applications in the case when the multivariate linear regression parameter with nonlinear equality constraints is analysed.

1.1 Estimation of the Parameters of a Linear Regression with Inequality Constraints

Assume that in (1.2) $f_t(\boldsymbol{\alpha}) = \tilde{f}_t(\mathbf{x}_t, \boldsymbol{\alpha}) = \mathbf{x}_t'\boldsymbol{\alpha}, t = 1, 2, \ldots$ and take in (1.5) $g_i(\boldsymbol{\alpha}) = \mathbf{g}_i'\boldsymbol{\alpha}, i = \overline{1,m}$, where $\mathbf{g}_i \in \Re^m$, $i = \overline{1,m}$ are known vectors. Then the estimation problems (1.4) and (1.5) can be written in the following form:

$$S(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\boldsymbol{\alpha})^2, \quad g_i(\boldsymbol{\alpha}) = \mathbf{g}_i'\boldsymbol{\alpha} - b_i \le 0, \quad i = \overline{1,m} \tag{1.6}$$

or

$$\frac{1}{2}||\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}||^2 \to \min, \quad \mathbf{G}\boldsymbol{\alpha} \le \mathbf{b}, \tag{1.7}$$

where $\mathbf{Y} = [y_1 \quad y_2 \ \ldots \ y_T]'$; $\mathbf{X}$ is some $(T \times n)$ matrix. The rows of this matrix are the vector rows $\mathbf{x}_t'$, $t = \overline{1,T}$; $\mathbf{G}$ is an $(m \times n)$ matrix with rows $\mathbf{g}_i'$, $i = \overline{1,m}$; $\mathbf{b} = [b_1 \quad b_2 \ \ldots \ b_m]'$.

We pose some additional assumptions on the regressor and the constraints, which will be used later on.

Assumption 1.1. Matrix $\mathbf{X}$ in (1.7) is of full rank.

Assumption 1.2. Matrix $\mathbf{G}$ in (1.7) is of full rank.

1.1.1 Method of Estimating the Solution to (1.7)

Taking into account the fact that the rank of $\mathbf{X}$ is equal to n (Assumption 1.1), we obtain its orthogonal expansion $\mathbf{X} = \mathbf{M}_1 \begin{bmatrix} \mathbf{M}_2 \\ \mathbf{O}_{T-n,n} \end{bmatrix} \mathbf{M}_3'$, $\mathbf{M}_1 = [\mathbf{M}_{11} \quad \mathbf{M}_{12}]$, where

$\mathbf{M}_1$ is an orthogonal $T \times T$ matrix, $T \times n$ is the dimension of the submatrix $\mathbf{M}_{11}$, $\mathbf{M}_2$ is a non-degenerate $(n \times n)$ matrix, $\mathbf{M}_3$ is an orthogonal $(n \times n)$ matrix.

Put $\mathbf{x} = \mathbf{M_2 M_3^{-1}} \boldsymbol{\alpha} - \mathbf{M}'_{11}\mathbf{Y}$. From the orthogonal decomposition of the matrix $\mathbf{X}$ and the properties of orthogonal matrixes mentioned above we obtain the following: for the cost function in (1.7),

$$\|\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}\|^2 = \left\| \mathbf{Y} - \mathbf{M}_1 \begin{bmatrix} \mathbf{J}_n \\ \mathbf{O}_{T-n,n} \end{bmatrix} (\mathbf{x} + \mathbf{M}'_{11}\mathbf{Y}) \right\|^2$$

$$= \mathbf{M}_1 \left\| \mathbf{M}'_1 \mathbf{Y} - \begin{bmatrix} \mathbf{x} \\ \mathbf{O}_{T-n} \end{bmatrix} - \begin{bmatrix} \mathbf{M}'_{11}\mathbf{Y} \\ \mathbf{O}_{T-n} \end{bmatrix} \right\|^2$$

$$= \left\| \begin{bmatrix} \mathbf{O}_n \\ \mathbf{M}'_{12}\mathbf{Y} \end{bmatrix} - \begin{bmatrix} \mathbf{x} \\ \mathbf{O}_{T-n} \end{bmatrix} \right\|^2 = ||\mathbf{M}'_{12}\mathbf{Y}||^2 + ||\mathbf{x}||^2,$$

while for the constraints in (1.7)

$$\mathbf{N}_1\mathbf{x} \leq \mathbf{N}_2$$

holds true, where $\mathbf{N}_1 = \mathbf{G}\mathbf{M}_3\mathbf{M}_2^{-1}$, $\mathbf{N}_2 = \mathbf{b} - \mathbf{G}\mathbf{M}_3\mathbf{M}_2^{-1}\mathbf{M}'_{11}\mathbf{y}$.

Getting rid of the term independent of $\mathbf{x}$, we obtain the transformed problem (1.7):

$$\frac{1}{2}||\mathbf{x}||^2 \to \min, \quad \mathbf{N}_1\mathbf{x} \leq \mathbf{N}_2. \tag{1.8}$$

This problem has a solution (as well as problem (1.7)) if the constraints are consistent.

Consider the following minimization problem (Lawson and Hanson 1974, Chapter 23 §5),

$$P(\mathbf{U}) = \frac{1}{2}||\mathbf{N}\mathbf{U} - \boldsymbol{\Phi}||^2 \to \min, \quad \mathbf{U} \geq \mathbf{O}_m, \tag{1.9}$$

where $\mathbf{U} \in \Re^m, \mathbf{N} = [\mathbf{N}_1 \ \vdots \ \mathbf{N}_2]', \boldsymbol{\Phi}' = [\mathbf{O}'_n \ \vdots \ 1]$.

Unlike (1.8), (1.9) always has a solution. In order to establish the connection between the problems (1.9) and (1.8) we introduce the following notation: $\hat{\mathbf{U}}$ is the solution to (1.9), $\mathbf{r} = \mathbf{N}\hat{\mathbf{U}} - \boldsymbol{\Phi}$.

The necessary and sufficient conditions for the existence of the minimum in (1.9) are:

$$\mathbf{N}'(\mathbf{N}\hat{\mathbf{U}} - \boldsymbol{\Phi}) + \boldsymbol{\Lambda} = \mathbf{O}_m, \quad \boldsymbol{\Lambda} \geq \mathbf{O}_m, \hat{\mathbf{U}}'\boldsymbol{\Lambda} = 0. \tag{1.10}$$

Hence, we obtain

$$\mathbf{N}'(\mathbf{N}\hat{\mathbf{U}} - \boldsymbol{\Phi}) \leq \mathbf{O}_m \tag{1.11}$$

and

$$\hat{\mathbf{U}}'\mathbf{N}'(\mathbf{N}\hat{\mathbf{U}} - \boldsymbol{\Phi}) = \mathbf{O}_m. \tag{1.12}$$

By arguments similar to those given in Lawson and Hanson (1974, Chapter 23 §4), we have $||\mathbf{r}||^2 = \mathbf{r}'\mathbf{r} = \hat{\mathbf{U}}'\mathbf{N}'(\mathbf{N}\hat{\mathbf{U}} - \boldsymbol{\Phi}) - r_{n+1}$, where r_{n+1} is $(n+1)$th component of $\mathbf{r}$. Using this equality and (1.12) we obtain $||\mathbf{r}||^2 = -r_{n+1} \geq 0$.

Suppose that $||\mathbf{r}|| > 0$, and assume that $\hat{\mathbf{x}} = -r_{n+1}\mathbf{N}_1'\hat{\mathbf{U}}$. Then

$$\mathbf{N}'(\mathbf{N}\hat{\mathbf{U}} - \boldsymbol{\Phi}) = \begin{bmatrix} \mathbf{N}_1 & \vdots & \mathbf{N}_2 \end{bmatrix} \begin{bmatrix} \mathbf{N}_1'\hat{\mathbf{U}} \\ r_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{N}_1 \vdots \mathbf{N}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}} \\ -1 \end{bmatrix} (-r_{n+1})$$
$$= (\mathbf{N}_1\hat{\mathbf{x}} - \mathbf{N}_2)||\mathbf{r}||^2 \leq \mathbf{O}_m, \tag{1.13}$$

which implies $\mathbf{N}_1\hat{\mathbf{x}} \leq \mathbf{N}_2$. We also would like to mention that if $||\mathbf{r}|| = 0$ the constraints in (1.8) are not consistent, see Lawson and Hanson (1974, Chapter 23 §4).

Now we can demonstrate that $\hat{\mathbf{x}}$ is the solution to (1.8).

Theorem 1.1. *If the constraints in (1.8) are consistent, then the solution is given by* $\mathbf{x} = \hat{\mathbf{x}} = ||\mathbf{r}||^{-2}\mathbf{N}_1'\hat{\mathbf{U}}$, *where* $\hat{\mathbf{U}}$ *is the solution to (1.9).*

Proof. The necessary and sufficient conditions for the existence of the minimum in (1.8) are:

$$\mathbf{x} + \mathbf{N}_1'\boldsymbol{\lambda} = \mathbf{O}_n, \quad \boldsymbol{\lambda} \geq O_m, \quad \boldsymbol{\lambda}_i'(\mathbf{N}_{1i}\mathbf{x} - \mathbf{N}_{2i}) = 0, \quad i = 1, \ldots, m, \tag{1.14}$$

where $\mathbf{N}_{1i}$ is the ith row of the matrix $\mathbf{N}_1$, $\mathbf{N}_{2i}$ is the ith component of the vector $\mathbf{N}_2$, $\boldsymbol{\lambda} \in \Re^m$ is the Lagrange multiplier, and $\boldsymbol{\lambda}_i$ denotes the ith component of $\boldsymbol{\lambda}$.

Substituting in (1.14) $\mathbf{x} = \hat{\mathbf{x}} = ||\mathbf{r}||^{-2}\mathbf{N}_1'\hat{\mathbf{U}}$, we obtain

$$\boldsymbol{\lambda} = ||\mathbf{r}||^{-2}\hat{\mathbf{U}} \geq \mathbf{O}_m. \tag{1.15}$$

Next we show that $\boldsymbol{\lambda}$ also satisfies the third condition in (1.14). From (1.10), (1.11), and (1.15) we derive

$$\hat{\mathbf{U}}'\boldsymbol{\Lambda} = 0 = \hat{\mathbf{U}}'\mathbf{N}'(\mathbf{N}\hat{\mathbf{U}} - \boldsymbol{\Phi}) = \hat{\mathbf{U}}'(\mathbf{N}_1\hat{\mathbf{x}} - \mathbf{N}_2)||\mathbf{r}||^2 = \boldsymbol{\lambda}'(\mathbf{N}_1\hat{\mathbf{x}} - \mathbf{N}_2).$$

Taking into account that $\boldsymbol{\lambda} \geq \mathbf{O}_m$, and according to (1.13) $\mathbf{N}_1\hat{\mathbf{x}} - \mathbf{N}_2 \leq \mathbf{O}_m$, we obtain from the latter equation the third condition in (1.14). Then the pair $(\hat{\mathbf{x}}, \boldsymbol{\lambda})$ satisfies the necessary and sufficient conditions for existence of the minimum in (1.8). Therefore, $\hat{\mathbf{x}}$ is the solution to (1.8). Theorem is proved. □

Thus, the solution to the problem (1.9) allows us to answer two questions: to determine the compatibility of the constraints in (1.8) (and, consequently, in (1.7)), and in case of compatibility to obtain the solution $\hat{\boldsymbol{\alpha}}$ by means of relatively easy transformation of the solution to (1.9). Namely,

$$\hat{\boldsymbol{\alpha}} = \mathbf{M}_3\mathbf{M}_2^{-1}(||\mathbf{N}\hat{\mathbf{U}} - \boldsymbol{\Phi}||^{-2}\mathbf{N}_1'\hat{\mathbf{U}} + \mathbf{M}_{11}'\mathbf{y}).$$

Corollary 1.1. *If Assumption 1.1 holds true and the problem (1.7) has a solution, then the related vector of Lagrange multipliers is given by* $\boldsymbol{\lambda} = \hat{\mathbf{U}}||\mathbf{N}\hat{\mathbf{U}} - \boldsymbol{\Phi}||^{-2}$.

Proof. The necessary and sufficient conditions for the existence of the minimum to (1.7) are:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\alpha} - \mathbf{X}'\mathbf{Y} + \mathbf{G}\overline{\boldsymbol{\lambda}} = \mathbf{O}_n, \quad \overline{\boldsymbol{\lambda}}'(\mathbf{G}\boldsymbol{\alpha} - \mathbf{b}) = 0, \overline{\boldsymbol{\lambda}} \geq \mathbf{O}_m.$$

From above, using the orthogonal transformation $\mathbf{X}$, we obtain

$$\mathbf{x} + \mathbf{N}_1'\overline{\boldsymbol{\lambda}} = \mathbf{O}_n, \quad \overline{\boldsymbol{\lambda}} \geq O_m, \qquad \overline{\boldsymbol{\lambda}}_i'(\mathbf{N}_{1i}\mathbf{x} - \mathbf{N}_{2i}) = 0, \quad i = 1, \ldots, m,$$

where $\overline{\boldsymbol{\lambda}}_i$ is the ith component of $\overline{\boldsymbol{\lambda}}$.

We see that these relations are satisfied when $\mathbf{x} = \hat{\mathbf{x}}$, $\overline{\boldsymbol{\lambda}} = \boldsymbol{\lambda}$, compared with (1.14). However, the pair of vectors $(\hat{\mathbf{x}}, \overline{\boldsymbol{\lambda}})$ is unique due to uniqueness implied by Assumption 1.1. On the other hand, as it was shown in the proof of Theorem 1.1, $\boldsymbol{\lambda}$ is given by (1.15). Hence the corollary follows. □

1.1.2 Algorithm of Finding the Solution to (1.9)

Assume that Assumptions 1.1 and 1.2 are satisfied. According to Lawson and Hanson (1974, Chapter 23 §3), we can proceed as follows.

Step 1 Let $\mathrm{P} = \varnothing$, $\Im = \{1, 2, \ldots, m\}$, $\mathbf{U} := \mathbf{O}_m$.

Step 2 Calculate the vector $\mathbf{w} = \mathbf{N}'(\boldsymbol{\Phi} - \mathbf{N}\mathbf{U}) \in \Re^m$.

Step 3 If the set $\Im$ is empty or $w_j \leq 0$ for all $j \in \Im$, go to Step 12. Here w_j is the jth component of $\mathbf{w}$.

Step 4 Find the index $i \in \Im$ such that $w_i = \max(w_j, j \in \Im)$.

Step 5 Move the index i from the set $\Im$ to the set P.

Step 6 Denote by $\mathbf{N}_\mathrm{P}$ the $((n+1) \times m_\mathbf{P})$-matrix, whose jth column is jth column of matrix $\mathbf{N}$, if $j \in \mathrm{P}$, $j = \overline{1, m}$.
Here m_P is the number of columns in the matrix $\mathbf{N}_\mathbf{P}$.
If $n + 1 \geq m$, then calculate the vector $\mathbf{z}_\mathbf{P} = (\mathbf{N}_\mathbf{P}'\mathbf{N}_\mathbf{P})^{-1}\mathbf{N}_\mathbf{P}'\boldsymbol{\Phi} \in \Re^{m_\mathbf{P}}$.
If $n + 1 < m$, then find $\mathbf{z}_\mathbf{P}$ with a minimal norm: $\mathbf{z}_\mathbf{P} = \mathbf{N}_\mathbf{P}'(\mathbf{N}_\mathbf{P}\mathbf{N}_\mathbf{P}')^{-1}\boldsymbol{\Phi} \in \Re^{m_\mathrm{P}}$.
Denote the components $\mathbf{z}_\mathbf{P}$ by z_j, $j \in \mathrm{P}$.
Put $z_j = 0$, $j \in \Im$.
Form a vector $\mathbf{z} = [z_j]$, $j = \overline{1, m}$.
Note that if $n+1 < m$, then the first $m_\mathbf{P} - 1$ components of $\mathbf{z}_\mathbf{P}$ are zero, the $m_\mathbf{P}$ of the component is equal to the element $(m_\mathbf{P}, (n + 1))$ of the matrix $\mathbf{N}_\mathrm{P}'(\mathbf{N}_\mathrm{P}\mathbf{N}_\mathrm{P}')^{-1}$ provided that $n + 1 < m$.

Step 7 If $z_j > 0$ for all $j \in \mathrm{P}$, then put $\mathbf{U} := \mathbf{z}$ and go to Step 2.

Step 8 Find the index $k \in \mathrm{P}$ such that

$$\frac{U_k}{U_k - z_k} = \min\left\{\frac{U_j}{U_j - z_j} : z_j \leq 0, j \in \mathrm{P}\right\}.$$

Step 9 Put $\gamma := U_k / U_k - z_k$.
Step 10 Put $\mathbf{U} := \mathbf{U} + \gamma(\mathbf{z} - \mathbf{U})$.
Step 11 Move all indices $j \in \mathrm{P}$ for which $U_j = 0$ from the set P to the set $\Im$. Then go to Step 6.
Step 12 Stop. The solution $\hat{\mathbf{U}} = \mathbf{U}$ is obtained.

1.1.3 Special Case of the Problem (1.7)

Usually a regression has a free term on which the constraints are often not imposed. Let us show that in the case when $\mathbf{x}_t = [1 \quad \tilde{\mathbf{x}}_t']'$, $\tilde{\mathbf{x}}_t \in \Re^{n-1}$, the solution of the estimation problem can be simplified by reducing the number of variables to one. The theorem presented below takes place.

Theorem 1.2. *If Assumption 1.1 holds true and no constraints are imposed on the free term α_1, then the solution to the problem (1.7) is of the form $\hat{\boldsymbol{\alpha}} = [\hat{\alpha}_1 \quad \hat{\tilde{\boldsymbol{\alpha}}}']'$, where $\hat{\alpha}_1 = \overline{y} - \hat{\tilde{\boldsymbol{\alpha}}}'\overline{\tilde{\mathbf{x}}}$, $\hat{\tilde{\boldsymbol{\alpha}}} \in \Re^{n-1}$ is the solution to*

$$\frac{1}{2}\tilde{\boldsymbol{\alpha}}'\mathbf{r}\tilde{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}'\mathbf{d} \to \min, \quad \mathbf{A}\tilde{\boldsymbol{\alpha}} \leq \mathbf{b}. \tag{1.16}$$

Here $\tilde{\boldsymbol{\alpha}} \in \Re^{n-1}$, $\mathbf{r} = \sum_{t=1}^{T}(\tilde{\mathbf{x}}_t - \overline{\tilde{\mathbf{x}}})(\tilde{\mathbf{x}}_t - \overline{\tilde{\mathbf{x}}})'$, $\mathbf{d} = \sum_{t=1}^{v}(\tilde{\mathbf{x}}_t - \overline{\tilde{\mathbf{x}}})(y_t - \overline{y})$, $\overline{\tilde{\mathbf{x}}} = \sum_{t=1}^{T}\tilde{\mathbf{x}}_t / T$, $\overline{y} = \sum_{t=1}^{T} y_t / T$, and $\mathbf{A}$ is the $m \times (n-1)$ matrix composed of $n-1$ last columns of the matrix $\mathbf{G}$.

Proof. We write the Lagrange function for the minimization problem (1.7) in the form $L(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{\alpha}'\mathbf{R}\boldsymbol{\alpha} - \boldsymbol{\alpha}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\lambda}'(\mathbf{G}\boldsymbol{\alpha} - \mathbf{b})$, where $\boldsymbol{\lambda}$ is the m-dimensional vector of Lagrange multipliers.

According to Assumption 1.1, the necessary and sufficient conditions for the existence of the minimum in (1.7) are of the form

$$\nabla_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \mathbf{R}\boldsymbol{\alpha} - \mathbf{X}'\mathbf{Y} + \mathbf{G}'\boldsymbol{\lambda} = \mathbf{O}_n, \tag{1.17}$$

$$\lambda_i(\mathbf{g}_i'\boldsymbol{\alpha} - b_i) = 0, \quad \lambda_i \geq 0, \ i = \overline{1, m}, \tag{1.18}$$

where $\nabla_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ is the gradient of the Lagrange function along the vector $\boldsymbol{\alpha}$, and λ_i is ith component of $\boldsymbol{\lambda}$.

Since no constraints are imposed on the free term α_1, the matrix $\mathbf{G}$ and its ith row $\mathbf{g}'_i$ are of the form

$$\mathbf{G} = \left[\mathbf{O}_m \ \vdots \ \mathbf{A}\right], \quad \mathbf{g}'_i = [0 \ \mathbf{A}_i], \tag{1.19}$$

where $\mathbf{A}_i$ is the ith row of the matrix $\mathbf{A}$. Then we have

$$\mathbf{G}'\boldsymbol{\lambda} = \begin{bmatrix} \mathbf{O}'_m \\ \mathbf{A}'\boldsymbol{\lambda} \end{bmatrix}. \tag{1.20}$$

Let us consider the condition (1.17). We consider the first of these equations, which by (1.20) can be rewritten as

$$\frac{\partial L(\boldsymbol{\alpha}, \boldsymbol{\lambda})}{\partial \alpha_1} = \alpha_1 T + \alpha_2 \sum_{t=1}^{T} x_{t1} + \alpha_2 \sum_{t=1}^{T} x_{t2} + \cdots + \alpha_n \sum_{t=1}^{T} x_{t,n-1} - \sum_{t=1}^{T} y_t = 0.$$

Dividing both sides of the above equation by the number of observations T, we obtain

$$\alpha_1 + \alpha_2 \overline{x}_1 + \alpha_3 \overline{x}_2 + \cdots + \alpha_n \overline{x}_{n-1} - \overline{y} = 0, \tag{1.21}$$

where $\overline{\mathbf{x}}_i = \sum_{t=1}^{T} x_{ti}/T$, $i = \overline{1, n-1}$, is the ith component of $\overline{\tilde{\mathbf{x}}}$.

Equation (1.21) must be satisfied for the estimates of the parameters, i.e.,

$$\hat{\alpha}_1 = \overline{y} - \hat{\alpha}_2 \overline{x}_1 - \hat{\alpha}_3 \overline{x}_2 - \cdots - \hat{\alpha}_n \overline{x}_{n-1} = \overline{y} - \hat{\tilde{\alpha}}' \overline{\tilde{\mathbf{x}}}, \tag{1.22}$$

which proves the first statement of the theorem.

Consider the ith equation ($i = \overline{2, n}$) in the system of equations (1.17),

$$\begin{aligned} \frac{\partial L(\boldsymbol{\alpha}, \boldsymbol{\lambda})}{\partial \alpha_i} &= \alpha_1 \sum_{t=1}^{T} x_{ti} + \alpha_2 \sum_{t=1}^{T} x_{ti} x_{t1} \\ &+ \alpha_3 \sum_{t=1}^{T} x_{ti} x_{t2} + \cdots + \alpha_n \sum_{t=1}^{T} x_{ti} x_{t,n-1} \\ &- \sum_{t=1}^{T} x_{ti} y_t + \mathbf{a}_i \boldsymbol{\lambda} = 0, \quad i = \overline{2, n}, \end{aligned}$$

where $\mathbf{a}_i$ is the ith row of the matrix $\mathbf{A}'$.

Substituting α_i from equality (1.21) in the above equation, we obtain

$$\sum_{j=2}^{n} \alpha_j \left[-\overline{x}_{j-1} \sum_{t=1}^{T} x_{ti} + \sum_{t=1}^{T} x_{ti} x_{t,j-1} \right] - \left[-\overline{y} \sum_{t=1}^{T} x_{ti} + \sum_{t=1}^{T} x_{ti} y_t \right] + a_i \lambda = 0. \tag{1.23}$$

After transformations, we get

$$\begin{aligned} -\overline{x}_{j-1} \sum_{t=1}^{T} x_{ti} + \sum_{t=1}^{T} x_{ti} x_{t,j-1} &= \sum_{t=1}^{T} (x_{ti} - \overline{x}_i)(x_{t,j-1} - \overline{x}_{j-1}) - \overline{y} \sum_{t=1}^{T} x_{ti} + \sum_{t=1}^{T} x_{ti} y_t \\ &= \sum_{t=1}^{T} (x_{ti} - \overline{x}_i)(y_t - \overline{y}). \end{aligned}$$

Substituting the last two expressions in equality (1.23), we find

$$\sum_{j=2}^{n}\alpha_j\sum_{t=1}^{T}(x_{ti}-\overline{x}_i)(x_{t,j-1}-\overline{x}_{j-1})-\sum_{t=1}^{T}(x_{ti}-\overline{x}_i)(y_t-\overline{y})+\mathbf{a}_i\boldsymbol{\lambda}=0,\quad i=\overline{2,n}.$$

We express the obtained system of equations in the vector form using the notation for problem (1.16):

$$\mathbf{r}\tilde{\boldsymbol{\alpha}}-\mathbf{d}+\mathbf{A}'\boldsymbol{\lambda}=\mathbf{O}_n. \tag{1.24}$$

Consider the condition (1.18). Taking into account the structure of the matrix $\mathbf{G}$ (see representation (1.19)), we obtain

$$\boldsymbol{\lambda}_i(\mathbf{A}_i\tilde{\boldsymbol{\alpha}}-b_i)=0,\quad \boldsymbol{\lambda}_i\geq 0,\ i=\overline{1,m}. \tag{1.25}$$

Since (1.17) and (1.18) have a unique solution, the (1.24) and (1.25) obtained from them possess the same property. Then these equations are the necessary and sufficient conditions of the existence of the minimum in problem (1.16), which holds true for the vector $\tilde{\boldsymbol{\alpha}}=\hat{\tilde{\boldsymbol{\alpha}}}$. Hence, the subvector $\hat{\tilde{\boldsymbol{\alpha}}}$ of the vector $\hat{\boldsymbol{\alpha}}$ is the solution to (1.16). Theorem is proved. □

Based on Theorem 1.2, one can reduce the estimation of the regression parameter to solving a quadratic programming problem in which the elements of the matrix of the objective function are by modules less than one. Put

$$\boldsymbol{\beta}=\mathbf{B}\tilde{\boldsymbol{\alpha}}, \tag{1.26}$$

where

$$\mathbf{B}=\sigma_y^{-1}\boldsymbol{\sigma}_{\mathbf{x}},\boldsymbol{\sigma}_{\mathbf{x}}=\mathrm{diag}(\sigma_{xi}),\quad i=\overline{1,n},$$

$$\sigma_y=\sqrt{T^{-1}\sum_{t=1}^{T}(y_t-\overline{y})^2},\quad \sigma_{xi}=\sqrt{T^{-1}\sum_{t=1}^{T}(x_{ti}-\overline{x}_i)^2},\quad i=\overline{1,n}.$$

Denoting $\mathbf{r}_{\boldsymbol{\beta}}=\boldsymbol{\sigma}_{\mathbf{x}}^{-1}\mathbf{r}\boldsymbol{\sigma}_{\mathbf{x}}^{-1}$, $\mathbf{d}_{\boldsymbol{\beta}}=(\sigma_y\boldsymbol{\sigma}_{\mathbf{x}})^{-1}\mathbf{d}$, and $\mathbf{A}_{\boldsymbol{\beta}}=\mathbf{A}\mathbf{B}^{-1}$, we obtain from problem (1.16) that

$$\frac{1}{2}\boldsymbol{\beta}'\mathbf{r}_{\boldsymbol{\beta}}\boldsymbol{\beta}-\boldsymbol{\beta}'\mathbf{d}_{\boldsymbol{\beta}}\to\min,\quad \mathbf{A}_{\boldsymbol{\beta}}\boldsymbol{\beta}\leq\mathbf{b}. \tag{1.27}$$

The advantage of the solution $\hat{\boldsymbol{\beta}}$ of such estimation problem is that $\hat{\boldsymbol{\beta}}$ does not depend on the scale of measurement of variables. The elements of the matrix $\mathbf{r}_{\boldsymbol{\beta}}$ and of the vector $\mathbf{d}_{\boldsymbol{\beta}}$ vary in the same range: from -1 to 1, which allows to reduce the round-up errors.

Therefore, as is well known (Draper and Smith 1998, Sect. 2.1.3; Maindonald 1984, Sects. 1.8 and 1.10), the numerical solution to the problem (1.27 is more exact than that obtained as a numerical solution to (1.7) (if no constraints are imposed on both problems). Under constraints, the accuracy of the numerical solution to (1.27) will also be higher than that of problem (1.7) since the main error is due to the inversion of matrices involved in the objective functions of the problems above.

The components of $\hat{\boldsymbol{\beta}}$ are standardized estimates of $(n-1)$ last component of the vector $\hat{\boldsymbol{\alpha}}$. We call them "beta weights" in analogy with the term used in regression analysis without constraints. Such weights can be conveniently used for estimation and comparison of the influence force of independent variables on the dependent variable.

For the problem described in (1.27) we use the calculation scheme described in Sect. 1.1.1. For this we reduce problem (1.27) to the least squares estimation problem with constraints, see (1.7).

Let $\mathbf{X}^-$ be a $T \times (n-1)$ matrix, with tth row $\tilde{\mathbf{x}}_t' - \bar{\tilde{\mathbf{x}}}'$, and let $\mathbf{Y}^-$ be a vector whose tth element is $y_t - \overline{y}$.

Let us transform the objective function in (1.16) by adding the element $\frac{1}{2}(\mathbf{Y}^-)'\mathbf{Y}^-$, which is a constant; thus, adding this element has no impact on the solution of our optimization problem. After transforming (1.26) we have

$$
\begin{aligned}
\frac{1}{2}\tilde{\boldsymbol{\alpha}}'\mathbf{r}\tilde{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}'\mathbf{d} + \frac{1}{2}(\mathbf{Y}^-)'\mathbf{Y}^- &= \frac{1}{2}\tilde{\boldsymbol{\alpha}}'(\mathbf{X}^-)'\mathbf{X}^-\tilde{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}'(\mathbf{X}^-)'\mathbf{Y}^- + \frac{1}{2}(\mathbf{Y}^-)'\mathbf{Y}^- \\
&= \frac{1}{2}||\mathbf{Y}^- - \mathbf{X}^-\tilde{\boldsymbol{\alpha}}||^2 = \frac{1}{2}||\mathbf{Y}^0 - \mathbf{X}^0\boldsymbol{\beta}||^2\sigma_y^2 \\
&= \left(\frac{1}{2}\boldsymbol{\beta}'\mathbf{r}_{\boldsymbol{\beta}}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{d}_{\boldsymbol{\beta}} + \frac{1}{2}(\mathbf{Y}^0)\mathbf{Y}^0\right)\sigma_y^2,
\end{aligned}
$$

where $\mathbf{X}^0$ is the $T \times (n-1)$ matrix with (t, i)th component given by $x_{ti} - \overline{x}_i/\sigma_{xi}$; and tth element of $\mathbf{Y}^0 \in \Re^T$ is $y_t - \overline{y}/\sigma_y$, $t = \overline{1, T}$.

Therefore we have

$$
\frac{1}{2}\boldsymbol{\beta}'\mathrm{r}_{\boldsymbol{\beta}}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{d}_{\boldsymbol{\beta}} = \frac{1}{2}||\mathbf{Y}^0 - \mathbf{X}^0\boldsymbol{\beta}||^2 - \frac{1}{2}(\mathbf{Y}^0)'(\mathbf{Y}^0). \tag{1.28}
$$

It follows from (1.28) that solution to the problem

$$
\frac{1}{2}||\mathbf{Y}^0 - \mathbf{X}^0\boldsymbol{\beta}||^2 \rightarrow \min, \quad \mathbf{A}_{\boldsymbol{\beta}}\boldsymbol{\beta} \leq \mathbf{b} \tag{1.29}
$$

coincides with solution to (1.27).

The problem (1.29) is similar to (1.7), and thus can be solved by orthogonal transformation of the matrix $\mathbf{X}^0$ described in Sect. 1.1.1. We only need to replace $\mathbf{Y}$ by $\mathbf{Y}^0$, $\mathbf{X}$ by $\mathbf{X}^0$, $\boldsymbol{\alpha}$ by $\boldsymbol{\beta}$, $\mathbf{G}$ by $\mathbf{A}_{\boldsymbol{\beta}}$ in every equation of Sect. 1.1.1.

Thus we obtained the solution $\hat{\boldsymbol{\beta}}$ to (1.29), and taking into account (1.26) we can find $\hat{\tilde{\boldsymbol{\alpha}}} = \mathbf{B}^{-1}\hat{\boldsymbol{\beta}}$, since the matrix $\mathbf{B}$ is non-degenerate and diagonal.

To find $\hat{\boldsymbol{\alpha}}_1$ we use (1.22). Finally, the solution to the original problem (1.7) is given by $\hat{\boldsymbol{\alpha}} = [\hat{\boldsymbol{\alpha}}_1 \quad \hat{\tilde{\boldsymbol{\alpha}}}']'$.

1.2 Estimation of Parameters of Nonlinear Regression with Nonlinear Inequality Constraints

1.2.1 Statement of the Problem and a Method of Its Solution

Consider a regression $y_t = f_t(\boldsymbol{\alpha}^0) + \varepsilon_t$, where $y_t \in \Re^1$ is the dependent variable, $\boldsymbol{\alpha}^0 \in \Re^n$ is the unknown parameter; ε_t is the noise; t is the index of the observation.

1.2.1.1 Estimation with Constraints

Let $y_t \in \Re^1$ and $\mathbf{x}_t \in \Re^n$ be known vectors, $t \in [1, T]$. The estimate $\hat{\boldsymbol{\alpha}}$ for the regression parameters can be found by solving the problem (1.4) and (1.5):

$$S(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{t=1}^{T}(y_t - f_t(\boldsymbol{\alpha}))^2 \to \min,$$

$$g_i(\boldsymbol{\alpha}) \le 0, \quad i \in I = \{1, 2, \ldots, m\}. \tag{1.30}$$

The solution to (1.30) can be obtained by iterations. Let us linearize at each iteration the components $\mathbf{f}(\boldsymbol{\alpha}) = [f_1(\boldsymbol{\alpha}) \ \ldots \ f_T(\boldsymbol{\alpha})]'$, and the functions $g_i(\boldsymbol{\alpha}), i = \overline{1, m}$, in neighborhood of the point determined at the previous iteration.

The auxiliary problem obtained after linearization has the following form at the current point $\boldsymbol{\alpha}$:

$$\begin{cases} \frac{1}{2}||\mathbf{Y} - \mathbf{f}(\boldsymbol{\alpha}) - \mathbf{D}(\boldsymbol{\alpha})\mathbf{X}||^2 + \frac{1}{2}v\mathbf{X}'A(\boldsymbol{\alpha})\mathbf{X} \to \min, \\ \mathbf{G}_\delta(\boldsymbol{\alpha})\mathbf{X} + \mathbf{g}_\delta(\boldsymbol{\alpha}) \le \mathbf{O}_{m_\delta}, \end{cases}. \tag{1.31}$$

where $\mathbf{Y}$ is the vector defined in (1.7), v is a positive number; $\mathbf{D}(\boldsymbol{\alpha})$ is a $T \times n$ matrix, $\mathbf{D}(\boldsymbol{\alpha}) = [\partial f_t(\boldsymbol{\alpha})/\partial \alpha_j]$, $t = \overline{1, T}$, $j = \overline{1, n}$, the matrix $\mathbf{G}_\delta(\boldsymbol{\alpha}) = [\partial g_i(\boldsymbol{\alpha})/\partial \alpha_j]$, $i \in I_\delta(\alpha) \subseteq I$, $j = \overline{1, n}$, is of dimension $m_\delta \times n$, where m_δ is the number of elements in $I_\delta(\boldsymbol{\alpha})$; $\mathbf{g}_\delta(\boldsymbol{\alpha}) = [g_i(\boldsymbol{\alpha})]$, $i \in I_\delta(\boldsymbol{\alpha})$, $\mathbf{A}(\boldsymbol{\alpha})$ is a positive definite matrix with elements independent of $\mathbf{X}$.