Tatra Mt. Math. Publ. **22** (2001), 13–25

TATRA MOUNTAINS
Mathematical Publications

LINEARIZATION OF NONLINEAR REGRESSION MODELS BY SMOOTHING

ANDREJ PÁZMAN

Dedicated to Lubomír Kubáček on the occasion of his 70th birthday

ABSTRACT. We suggest linear estimators of parameters of a nonlinear regression model, which are derived from a linearization "by smoothing". We compare them with the estimators coming from a "standard" linearization (by the use of a Taylor formula). The proposed estimators has a better mean squared error (especially in the case that the regression is quadratic in the unknown parameters). The method can be applied also when the regression function has no derivatives, and in general, the necessary coefficients can be obtained by simulations. If the prior distribution of the parameters is much concentrated, the proposed and the standard methods coincide.

1. Introduction

In the 60ties and the beginning of 70ties prof. Kubáček introduced linear statistical methods in Slovakia and influenced by that and by his constant enthusiasm many people in science and applications. He supported much the study of excellent books of Linnik [10], and Rao [13], and wrote several papers on the subject (cf. [3], [4], [5] and a summarization in the book [6]). This interest came from the fact that besides being professor in mathematical statistics prof. Kubáček is also a highly qualified geodesist. In geodesy linear statistical methods are fundamental, although mathematical models which describe geodetic nets are nonlinear, but are linearized. Linearization has been important also to implement methods of experimental design in the optimization of geodetic nets, as presented in our common paper [9].

It is this and another linearization which are the subject of the present paper.

Any linearization brings some errors in the procedures of parameter estimations. These errors for the "standard" linearization have been studied in several

2000 Mathematics Subject Classification: 62J02, 2H12.
Keywords: linear estimators, mean squared error, bias, prior, distribution, approximate linear models.

Research supported by the VEGA-grant of the Slovak Grant Agency No. 1/7295/20.

papers. For example, by the influence of the mentioned "linear trends" in the 60ties, in [11] an expression for the bias of the estimator coming from the linearization has been presented. Problems of linearization are also behind the famous paper by Battes and Watts [1], which introduced curvatures as measures of nonlinearity of the model. Recently, Kubáček came back to the subject in a detailed study of different statistical consequences of the standard linearization (cf. [7]) and also in the thesis of his Ph.D. student (cf. [2]).

The linearization considered in these papers is based on the approximation of the regression function by the Taylor formula at a certain "prior point" in the parameter space. (Notice that a similar linearization appears in fact also when considering the asymptotic properties of the nonlinear least squares estimators.) The aim of the present paper is to show that one can advantageously consider another linearization, which takes into account the uncertainty of this prior point. In such case one does not use the Taylor formula any more, but choose the approximate linear model as the one which is the closest to the nonlinear model in a statistical sense.

2. Basic properties of a standard linearization

We consider a nonlinear regression model

$$\begin{aligned} y = \eta(\theta) + \varepsilon\,; \quad \theta \in \Theta\,, \\ E(\varepsilon) = 0\,, \quad \operatorname{Var}(\varepsilon) = \sigma^2 W\,, \end{aligned} \tag{2.1}$$

with $y \in \mathbb{R}^N$ the vector of observations, $\varepsilon \in \mathbb{R}^N$ the vector of random errors, $\theta \in \mathbb{R}^m$ the vector of unknown parameters localized in an open parameter space Θ. The component σ^2 of the variance matrix $\operatorname{Var}(\varepsilon)$ is usually unknown, but the positive definite component W is supposed to be known, and often we have $W = I$ (independent observations).

In this paper we shall suppose that $W = I$, since by the substitution

$$\begin{aligned} \eta(\theta) &\to W^{-1/2}\eta(\theta)\,, \\ y &\to W^{-1/2} y \end{aligned}$$

we obtain all results for the general case. Notice that

$$W^{-1/2} = U \operatorname{diag}\big(\lambda_1^{-1/2}, \dots, \lambda_N^{-1/2}\big) U^T,$$

where the columns of U are N orthonormal eigenvectors of W corresponding to the eigenvalues $\lambda_1, \dots, \lambda_N$.

When speaking about the standard linearization, the mapping $\eta\colon \Theta \to \mathbb{R}^N$ is supposed to be known, and to have continuous second order partial derivatives. Usual statistical tasks in the model are to construct estimators of θ and σ^2, or to test the model against some alternatives.

The standard linearization consists in taking a fixed "prior point" $\theta^o \in \Theta$, which is supposed to be "not too far" from the true value of θ. One approximates $\eta(\theta)$ by the linear part of its Taylor expansion at θ^o to obtain instead of (2.1) a linear model

$$\begin{gathered} y - \eta(\theta^o) = \frac{\partial\eta(\theta^o)}{\partial\theta^T}(\theta - \theta^o) + \varepsilon\,, \\ E(\varepsilon) = 0\,, \quad \mathrm{Var}(\varepsilon) = \sigma^2 I\,. \end{gathered}$$

In this (approximate) model we consider θ as totally unknown, and estimate it by the estimator $\hat{\theta}$, which by the Gauss-Markov theorem, resp. the Rao-Cramer inequality, is the minimum variance unbiased estimator when the errors are supposed normal, or when we consider only linear estimators and the errors are distributed arbitrary:

$$\hat{\theta} = \theta^o + \left(X^{oT}X^o\right)^{-1}X^{oT}\left[y - \eta(\theta^o)\right]. \tag{2.2}$$

Here

$$X^o = \frac{\partial\eta(\theta^o)}{\partial\theta^T}$$

is an $N \times m$ matrix with entries $\left.\dfrac{\partial\eta_i(\theta)}{\partial\theta_j}\right|_{\theta=\theta^o}$. The parameter σ^2 is estimated by

$$\hat{s}^2 = \frac{\left\|y - \eta(\theta^o) - X^o[\hat{\theta} - \theta^o]\right\|^2}{N - m}\,.$$

This approach is attractive by its simplicity, however, the prior knowledge about θ is used here quite differently as in a Bayesian approach. The explanation is that one uses θ^o just auxiliary to construct the estimators $\hat{\theta}$ and $\hat{s}^2$ (eventually to construct any other estimator or a test which are standard in linear models). The properties of these estimators have to be considered in the original model (2.1). The justification of this construction of estimators through linearization is in the fact that when the prior point θ^o is close to the true value of θ, then the properties of the estimators $\hat{\theta}$ and $\hat{s}^2$ are very good.

3. The linearization by smoothing with respect to the prior

3.1 The definition and motivation

We usually do not know the prior point θ^o exactly. Instead, we have rather more a prior probability distribution π on Θ, and we take θ^o equal either to the modus of the prior density or to the mean of π. The question is how to use the knowledge of π advantageously.

In [12] it is proposed to take the prior points θ^o several times at random according to the prior distribution, to apply a standard linearization each time, and to use a summary of the so obtained estimators. In fact there is a double linearization there, which can give supplementary errors.

Here we consider another approach. In the model (2.1) we need no more the differentiability of $\eta(\theta)$, but suppose that the mapping $\theta \in \Theta \to \eta(\theta)$ is measurable, and that the prior means $E_\pi(\theta)$ and $E_\pi[\eta(\theta)] = E_\pi(\eta)$ exist, the prior variance matrices $\mathrm{Var}_\pi(\theta)$ and $\mathrm{Var}_\pi[\eta(\theta)] = \mathrm{Var}_\pi(\eta)$ exist, and $\mathrm{Var}_\pi(\theta)$ is regular, and that the prior covariance matrix

$$\mathrm{Cov}_\pi[\eta, \theta^T] = E_\pi\Big[\big(\eta - E_\pi(\eta)\big)\big(\theta - E_\pi(\theta)\big)^T\Big]$$

exists and is of full rank. We denote its transpose by $\mathrm{Cov}_\pi[\theta, \eta^T]$.

Lemma 1. *The solution of the minimization problem*

$$\min_{\substack{X \in \mathbb{R}^{N\times m} \\ c \in \mathbb{R}^N}} E_\pi\Big[\big\|\eta(\theta) - [X\theta + c]\big\|^2\Big] \tag{3.1}$$

is equal to

$$X^* = \mathrm{Cov}_\pi[\eta, \theta^T]\ \mathrm{Var}_\pi^{-1}(\theta),$$
$$c^* = E_\pi(\eta) - \mathrm{Cov}_\pi[\eta, \theta^T]\,\mathrm{Var}_\pi^{-1}(\theta) E_\pi(\theta).$$

Proof. We can write

$$E_\pi\Big[\big\|\{[\eta(\theta) - E_\pi(\eta)] - X[\theta - E_\pi(\theta)]\} + \{E_\pi(\eta) - XE_\pi(\theta) - c\}\big\|^2\Big]$$
$$= E_\pi\Big[\big\|[\eta(\theta) - E_\pi(\eta)] - X[\theta - E_\pi(\theta)]\big\|^2\Big] + \big\|E_\pi(\eta) - XE_\pi(\theta) - c\big\|^2$$
$$\geq E_\pi\Big[\big\|[\eta(\theta) - E_\pi(\eta)] - X[\theta - E_\pi(\theta)]\big\|^2\Big]$$

with equality iff $c = E_\pi(\eta) - XE_\pi(\theta)$. Further we have

$$E_\pi\Big[\big\|[\eta(\theta) - E_\pi(\eta)] - X[\theta - E_\pi(\theta)]\big\|^2\Big]$$
$$= \mathrm{tr}\Big\{\big[X - \mathrm{Cov}_\pi[\eta, \theta^T]\,\mathrm{Var}_\pi^{-1}(\theta)\big]\,\mathrm{Var}_\pi(\theta)\big[X - \mathrm{Cov}_\pi[\eta, \theta^T]\,\mathrm{Var}_\pi^{-1}(\theta)\big]^T\Big\}$$
$$+ \mathrm{tr}\big\{\mathrm{Var}_\pi(\eta) - \mathrm{Cov}_\pi[\eta, \theta^T]\,\mathrm{Var}_\pi^{-1}(\theta)\,\mathrm{Cov}_\pi[\theta, \eta^T]\big\}$$

The second term does not depend on X, the first is minimized when $X = \mathrm{Cov}_\pi\left[\eta, \theta^T\right] \mathrm{Var}_\pi^{-1}(\theta)$. □

As a consequence of Lemma 1 one can consider a linear model given by

$$\begin{aligned} y - E_\pi(\eta) &= X^* \left[\theta - E_\pi(\theta)\right] + \varepsilon\,, \\ E(\varepsilon) = 0\,, &\quad \mathrm{Var}(\varepsilon) = \sigma^2 W \end{aligned} \tag{3.2}$$

which approximates the original nonlinear model in the sense of (3.1). Since the coefficients of the linear model (3.2) contain averaged values of θ and $\eta(\theta)$, we can say that it has been obtained by smoothing the original model (2.1).

Let us remark that in the case of ε distributed normally, $N\left(0, \sigma^2 I\right)$, (3.1) corresponds to minimizing the I-divergence between the nonlinear and a linear model. In the case that $\mathrm{Var}(\varepsilon) = \sigma^2 W$ one has to consider in (3.1) the norm $\|a\|_W^2 = a^T W^{-1} a$ instead of the Euclidean norm $\|a\|^2 = a^T a$.

In the linear model (3.2) the minimum variance unbiased estimators $\tilde\theta$ of θ and $\tilde s^2$ of σ^2, under the normality of the errors, are equal to

$$\begin{aligned} \tilde\theta &= E_\pi(\theta) + \left(X^{*T} X^*\right)^{-1} X^{*T} \left(y - E_\pi(\eta)\right) \qquad (3.3)\\ &= E_\pi(\theta) + \mathrm{Var}_\pi(\theta) \left\{\mathrm{Cov}_\pi\left[\theta, \eta^T\right] \mathrm{Cov}_\pi\left[\eta, \theta^T\right]\right\}^{-1} \mathrm{Cov}_\pi\left[\theta, \eta^T\right]\left(y - E_\pi(\eta)\right) \\ \tilde s^2 &= \frac{\left\|y - E_\pi(\eta) - X^*\left[\tilde\theta - E_\pi(\theta)\right]\right\|^2}{N - m}\,. \end{aligned}$$

3.2 Properties of $\tilde\theta$ and $\tilde s^2$ in model (2.1)

First notice that the terms in $\tilde\theta$ and $\tilde s^2$ can be computed numerically by simulations over the probability distribution of θ. For a large number n of random samples $\theta^{(i)}$ taken according to the prior distribution π, one obtains approximately

$$\mathrm{Cov}_\pi\left(\eta, \theta^T\right) \doteq \frac{1}{n}\sum_{i=1}^n \left[\eta\left(\theta^{(i)}\right) - \frac{1}{n}\sum_{k=1}^n \eta\left(\theta^{(k)}\right)\right]\left[\theta^{(i)} - \frac{1}{n}\sum_{k=1}^n \theta^{(k)}\right]^T$$

and similarly for the other terms.

The aim of the next statement is to show that when the prior distribution π approaches to a singular distribution concentrated at a point θ^o, then the linearization by smoothing approaches the standard linearization.

Theorem 2. *Let the assumptions for a standard linearization at θ^o be satisfied. Consider a sequence of priors π_n, $n = 1, 2, \dots$, which are all concentrated on a common compact subset of Θ. Let $F_n(\theta)$ be the distribution function on $\mathbb{R}^m$, which corresponds π_n and suppose that there are limits*

$$\lim_{n\to\infty} F_n(\theta) \begin{cases} 1 & \text{if } \theta_i > \theta_i^o \text{ for every } i = 1, \dots, m\,, \\ 0 & \text{if } \theta_i < \theta_i^o \text{ at least for one } i \in \{1, \dots, m\}\,, \end{cases}$$

i.e., that the sequence π_n ; $n = 1, 2, \dots$, *converges to a degenerate distribution concentrated at* θ^o. *Then*

$$\lim_{n\to\infty} E_n(\theta) = \theta^o, \qquad \lim_{n\to\infty} E_n(\eta) = \eta(\theta^o),$$
$$\lim_{n\to\infty} \mathrm{Cov}_n\big[\eta, \theta^T\big]\,\mathrm{Var}_n^{-1}(\theta) \;=\; \frac{\partial\eta(\theta^o)}{\partial\theta^T},$$

where by the subscript n *we denote the means and covariances with respect to* π_n.

P r o o f . The assumptions of the theorem mean that there is a sequence $\theta^{(n)}$, $n = 1, 2, \dots$, of random vectors converging to a constant vector θ^o in their distribution functions F_n. Consequently, the sequence converges also in probability, and the sequence $\eta\big[\theta^{(n)}\big]$, $n = 1, 2, \dots$, converges in the same way to $\eta[\theta^o]$, since $\eta(\cdot)$ is a continuous mapping (cf. [14], p. 19 and 24). Both sequences are uniformly bounded (since each π_n is concentrated on the same compact set), hence (cf. [14]) $\lim_{n\to\infty} E_n(\theta) = \lim_{n\to\infty} E\big(\theta^{(n)}\big) = \theta^o$, $\lim_{n\to\infty} E_n(\eta) = \lim_{n\to\infty} E\big(\eta(\theta^{(n)})\big) = \eta(\theta^o)$.

We can expand $\eta(\theta)$ according to the Taylor formula at θ^o, and obtain

$$\begin{aligned}\mathrm{Cov}_n\big(\eta, \theta^T\big) = {} & E_n\Big[\big(\eta(\theta^o) - E_n(\eta)\big)\big(\theta - E_n(\theta)\big)^T\Big] \\ & + \frac{\partial\eta(\theta^o)}{\partial\theta^T} E_n\Big[\big(\theta - \theta^o\big)\big(\theta - E_n(\theta)\big)^T\Big] \\ & + \frac{1}{2} E_n\Bigg[\sum_i (\theta - \theta^o)_i \frac{\partial^2\eta(\theta)}{\partial\theta_i\,\partial\theta^T}\Big|_{\xi^*(\theta)} \big(\theta - \theta^o\big)\big(\theta - E_n(\theta)\big)^T\Bigg],\end{aligned}$$

where $\xi^*(\theta)$ is a point between θ and θ^o which depends continuously on θ. Evidently

$$E_n\Big[\big(\theta - \theta^o\big)\big(\theta - E_n(\theta)\big)^T\Big] = \mathrm{Var}_n(\theta),$$

hence

$$\mathrm{Cov}_n\big(\eta, \theta^T\big)\,\mathrm{Var}_n^{-1}(\theta) = \frac{\partial\eta(\theta^o)}{\partial\theta^T} + \frac{1}{2} E_n\Bigg[\sum_i (\theta - \theta^o)_i \frac{\partial^2\eta(\theta)}{\partial\theta_i\partial\theta^T}\Big|_{\xi^*(\theta)} B(\theta)\Bigg],$$

where $B(\theta) = \big(\theta - \theta^o\big)\big(\theta - E_n(\theta)\big)^T\,\mathrm{Var}_n^{-1}(\theta)$ is a $m \times m$ matrix with $E_n\big[B(\theta)\big] = I$. Consequently, the sequence $B\big(\theta^{(n)}\big)$, $n = 1, 2, \dots$, converges in probability to I. Since $\big(\theta^{(n)} - \theta^o\big)$, $n = 1, 2, \dots$, converges in probability to zero, and

$\frac{\partial^2 \eta(\theta)}{\partial\theta_i \partial\theta^T}\Big|_{\xi^*(\theta^{(n)})}$ is bounded over the support of π_n, we obtain that

$$\lim_{n\to\infty} E_n\left[\sum_i (\theta-\theta^o)_i \frac{\partial^2\eta(\theta)}{\partial\theta_i\partial\theta^T}\Big|_{\xi^*(\theta)} B(\theta)\right]$$
$$= \lim_{n\to\infty} E\left[\sum_i \left(\theta^{(n)}-\theta^o\right)_i \frac{\partial^2\eta(\theta)}{\partial\theta_i\partial\theta^T}\Big|_{\xi^*\left(\theta^{(n)}\right)} B\left(\theta^{(n)}\right)\right] = 0\,.$$

Consequently

$$\lim_{n\to\infty} \mathrm{Cov}_n\left(\eta,\theta^T\right)\mathrm{Var}_n^{-1}(\theta) = \frac{\partial\eta(\theta^o)}{\partial\theta^T}\,.$$

□

Another interesting connection to the standard linearization is contained in the following statement.

Theorem 3. *Let $\eta(\theta)$ be a polynomial of degree 2 in the components of θ. Let $E_\pi(\theta) = \theta^o$, and let all third central moments of π be zero. Then*

$$\mathrm{Cov}_\pi\left(\eta,\theta^T\right)\mathrm{Var}_\pi^{-1}(\theta) = \frac{\partial\eta(\theta^o)}{\partial\theta^T}\,.$$

Remark. This statement does not mean that both estimators $\hat\theta$ and $\tilde\theta$ coincide, neither when $E_\pi(\theta) = \theta^o$, because $E_\pi(\eta) \neq \eta(\theta^o)$.

P r o o f. We proceed similarly as in the previous theorem to obtain

$$\mathrm{Cov}_\pi\left(\eta,\theta^T\right)\mathrm{Var}_\pi^{-1}(\theta) = \frac{\partial\eta(\theta^o)}{\partial\theta^T} + \frac{1}{2}E_\pi\left[\sum_i (\theta-\theta^o)_i \frac{\partial^2\eta(\theta^o)}{\partial\theta_i\partial\theta^T} B(\theta)\right]\,.$$

We have

$$\begin{aligned} E_\pi\left[\sum_i (\theta-\theta^o)_i \frac{\partial^2\eta(\theta^o)}{\partial\theta_i\partial\theta^T} B(\theta)\right] &= E_\pi\left[\sum_i \left(\theta - E_\pi(\theta)\right)_i \frac{\partial^2\eta(\theta^o)}{\partial\theta_i\partial\theta^T} B(\theta)\right] \\ &\quad + \sum_i \left(E_\pi(\theta)-\theta^o\right)_i \frac{\partial^2\eta(\theta^o)}{\partial\theta_i\partial\theta^T} E_\pi\left[B(\theta)\right] \\ &= 0\,. \end{aligned}$$

The first term on the right hand side is zero, because all third central moments of π are zero. □

4. The bias, the variance and the mean square error of the estimators

Denote by Z^o, resp. Z^* the g-inverses of X^o, resp. X^* defined by

$$Z^o = \left(X^{oT}X^o\right)^{-1}X^{oT}, \quad Z^* = \left(X^{*T}X^*\right)^{-1}X^{*T}.$$

From (2.2) and (3.3) we obtain that

$$\begin{aligned}\mathrm{bias}_\theta(\hat\theta) &= E_\theta(\hat\theta) - \theta \\ &= \theta^o - \theta + Z^o\left[\eta(\theta) - \eta(\theta^o)\right], \\ \mathrm{bias}_\theta(\tilde\theta) &= E_\pi(\theta) - \theta + Z^*\left[\eta(\theta) - E_\pi(\eta)\right] \end{aligned} \tag{4.1}$$

and

$$\begin{aligned}\mathrm{Var}_\theta(\hat\theta) &= \sigma^2\left(X^{oT}X^o\right)^{-1}, \\ \mathrm{Var}_\theta(\tilde\theta) &= \sigma^2\left(X^{*T}X^*\right)^{-1}.\end{aligned}$$

The variance matrices does not depend on θ. Evidently

$$E_\pi\left[\mathrm{bias}_\theta(\tilde\theta)\right] = 0$$

while $E_\pi\left[\mathrm{bias}_\theta(\hat\theta)\right] \neq 0$ in the general case, even if $\theta^o = E_\pi(\theta)$. However, it is more important to compare

$$\begin{aligned}E_\pi\left[\left\|\mathrm{bias}_\theta(\tilde\theta)\right\|^2\right] &= E_\pi\left[\mathrm{tr}\left\{\left(\mathrm{bias}_\theta(\tilde\theta)\right)\left(\mathrm{bias}_\theta(\tilde\theta)\right)^T\right\}\right] \\ &= \mathrm{tr}\ \mathrm{Var}_\pi[Z^*\eta - \theta]\end{aligned}$$

with

$$E_\pi\left[\left\|\mathrm{bias}_\theta(\hat\theta)\right\|^2\right] = \mathrm{tr}\ \mathrm{Var}_\pi[Z^o\eta - \theta] + \left\|Z^o\left[E_\pi(\eta) - \eta(\theta^o)\right] - \left[E_\pi(\theta) - \theta^o\right]\right\|^2,$$

which indicates a better performance of the estimator $\tilde\theta$. To have a global comparison of both estimators one has to consider the mean squared errors

$$\begin{aligned}\mathrm{MSE}(\tilde\theta) &= E_\pi\left[E_\theta\|\tilde\theta - \theta\|^2\right] \\ &= \sigma^2\left(X^{*T}X^*\right)^{-1} + \mathrm{tr}\ \mathrm{Var}_\pi[Z^*\eta - \theta]\end{aligned}$$

and

$$\begin{aligned}\mathrm{MSE}(\hat\theta) = E_\pi\left[E_\theta\|\hat\theta - \theta\|^2\right] &= \sigma^2\left(X^{oT}X^o\right)^{-1} + \mathrm{tr}\ \mathrm{Var}_\pi[Z^o\eta - \theta] \\ &+ \left\|Z^o\left[E_\pi(\eta) - \eta(\theta^o)\right] - \left[E_\pi(\theta) - \theta^o\right]\right\|^2.\end{aligned}$$

The comparison is especially favorable for the use of $\tilde\theta$ in the case that the nonlinear regression model is a polynomial of degree 2 (a model which is studied in [7])

Theorem 4. *Let the assumptions of Theorem 3 hold. Then*

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{MSE}(\tilde{\theta}) + \left\| Z^o \left[E_\pi(\eta) - \eta(\theta^o) \right] \right\|^2 .$$

P r o o f. It follows from $Z^o = Z^*$ and $E_\pi(\theta) = \theta^o$ in Theorem 3. □

4.1 Mean absolute bias of s^2

The estimators for σ^2 can be written in the form

$$\hat{s}^2 = \frac{\left\| [I - P^o][y - \eta(\theta^o)] \right\|^2}{N - m} ,$$
$$\tilde{s}^2 = \frac{\left\| [I - P^*][y - E_\pi(\eta)] \right\|^2}{N - m} ,$$

where

$$P^o = X^o (X^{oT} X^o)^{-1} X^{oT}, \quad P^* = X^* (X^{*T} X^*)^{-1} X^{*T}$$

Since $I - P^o$ and $I - P^*$ are projectors of rank $N - m$, under the additional assumption that the error vector $\varepsilon = y - \eta(\theta)$ is normal, one obtains straightforwardly

$$\begin{aligned} E_\theta(\hat{s}^2) &= \sigma^2 + \frac{\left\| [I - P^o][\eta(\theta) - \eta(\theta^o)] \right\|^2}{N - m} , \\ E_\theta(\tilde{s}^2) &= \sigma^2 + \frac{\left\| [I - P^*][\eta(\theta) - E_\pi(\eta)] \right\|^2}{N - m} . \end{aligned} \tag{4.2}$$

Hence the mean absolute bias is equal to

$$E_\pi \left[\left| E_\theta(\hat{s}^2) - \sigma^2 \right| \right] = \frac{\mathrm{tr}\{[I - P^o]\,\mathrm{Var}_\pi(\eta)\}}{N - m} + \frac{\left\| [I - P^o][E_\pi(\eta) - \eta(\theta^o)] \right\|^2}{N - m} ,$$

$$E_\pi \left[\left| E_\theta(\tilde{s}^2) - \sigma^2 \right| \right] = \frac{\mathrm{tr}\{[I - P^*]\,\mathrm{Var}_\pi(\eta)\}}{N - m} ,$$

which indicates again that the linearization by smoothing may be better. In particular, for $\eta(\theta)$ a quadratic polynomial in θ, under the assumptions of Theorem 3, one obtains

$$E_\pi \left[\left| E_\theta(\hat{s}^2) - \sigma^2 \right| \right] = E_\pi \left[\left| E_\theta(\tilde{s}^2) - \sigma^2 \right| \right] + \frac{\left\| [I - P^o][E_\pi(\eta) - \eta(\theta^o)] \right\|^2}{N - m}$$

so the estimator $\tilde{s}^2$ is better.

5. Related results

5.1 Quadratic approximation of $\eta(\theta)$

To evaluate $E_\theta(\tilde{\theta})$ or $E_\theta(\tilde{s}^2)$ (for a fixed θ) one needs a quadratic approximation of $\eta(\theta)$.

We proceed in the spirit of Lemma 1. Let us denote $\Delta = \theta - E_\pi(\theta)$.

THEOREM 5. *Let all third central moments of the prior π be zero. Then the minimization problem*

$$\min_{\substack{f\in\mathbb{R}^N, X\in\mathbb{R}^{N\times m}\\ Q\in\mathbb{R}^{m\times N\times m}}} E_\pi\left\{\left\|\eta(\theta) - [f + X\Delta + \Delta^T Q\Delta]\right\|^2\right\}$$

has the solution

$$\begin{aligned} f_t &= E_\pi(\eta_t) - \operatorname{Cov}_\pi\big(\eta_t, \Delta^T\otimes\Delta^T\big)\operatorname{Var}_\pi^{-1}[\Delta\otimes\Delta]\ \operatorname{Vec}\big[\operatorname{Var}_\pi(\theta)\big]\,, \\ \operatorname{Vec}(Q^t) &= \operatorname{Cov}_\pi\big(\eta_t, \Delta^T\otimes\Delta^T\big)\operatorname{Var}_\pi^{-1}[\Delta\otimes\Delta]\,; \quad t = 1,\dots,N\,, \\ X &= X^* = \operatorname{Cov}_\pi\big[\eta, \theta^T\big]\operatorname{Var}_\pi^{-1}(\theta)\,. \end{aligned}$$

Proof. Denote $D = \eta - E_\pi(\eta)$. The symbol Q is a 3-dimensional array with components $Q^t_{i,j}$; $i,j = 1,\dots,m$, $t = 1,\dots,N$. By $\Delta^T Q\Delta$ we denote the vector in $\mathbb{R}^N$ with components $\Delta^T Q^t\Delta$. We can write

$$\begin{aligned} &E_\pi\left\{\left\|\eta(\theta) - [f + X\Delta + \Delta^T Q\Delta]\right\|^2\right\} \\ &= E_\pi\left\{\left\|[E_\pi(\eta) - f - E_\pi(\Delta^T Q\Delta)] + D - X\Delta - [\Delta^T Q\Delta - E_\pi(\Delta^T Q\Delta)]\right\|^2\right\} \\ &= E_\pi\left\{\left\|D - X\Delta - [\Delta^T Q\Delta - E_\pi(\Delta^T Q\Delta)]\right\|^2\right\} \\ &\quad + \left\|[E_\pi(\eta) - f - E_\pi(\Delta^T Q\Delta)]\right\|^2 \\ &\geq E_\pi\left\{\left\|D - X\Delta - [\Delta^T Q\Delta - E_\pi(\Delta^T Q\Delta)]\right\|^2\right\} \end{aligned}$$

with equality iff

$$f_t = E_\pi(\eta_t) - E_\pi(\Delta^T Q^t\Delta)\,; \quad t = 1,\dots,N\,. \tag{5.1}$$

Taking such f, we have to minimize

$$\min_{X\in\mathbb{R}^{N\times m}} E_\pi\left\{\left\|[D - \Delta^T Q\Delta + E_\pi(\Delta^T Q\Delta)] - X\Delta\right\|^2\right\}.$$

Exactly like in the proof of Lemma 1, but taking $[D - \Delta^T Q\Delta]$ instead of $\eta(\theta)$, we obtain that the minimum is attained for

$$X = \operatorname{Cov}_\pi\big([D - \Delta^T Q\Delta]\,,\ \Delta^T\big)\operatorname{Var}_\pi^{-1}(\Delta)\,.$$

However, the third moments of Δ are zero, hence

$$\mathrm{Cov}_\pi\big(\Delta^T Q \Delta, \Delta^T\big) = 0\,.$$

Thus

$$\begin{aligned} X &= \mathrm{Cov}_\pi\big(D, \theta^T\big)\,\mathrm{Var}_\pi^{-1}(\Delta) \\ &= \mathrm{Cov}_\pi\big(\eta, \theta^T\big)\,\mathrm{Var}_\pi^{-1}(\theta)\,, \end{aligned}$$

which is the same expression as in Lemma 1. Finally, we have to solve

$$\begin{aligned} &\min_Q E_\pi\Big\{\big\|\,[D - \Delta^T Q\Delta + E_\pi\big(\Delta^T Q\Delta\big)] - \mathrm{Cov}_\pi\big(\eta, \theta^T\big)\,\mathrm{Var}_\pi^{-1}(\theta)\Delta\big\|^2\Big\} \\ &= \min_Q E_\pi\Bigg\{\sum_t\Big\{\big[D - \mathrm{Cov}_\pi(\eta, \theta^T)\,\mathrm{Var}_\pi^{-1}(\theta)\Delta\big]_t \\ &\qquad\qquad - (\mathrm{Vec}\,Q^t)^T\big[\Delta\otimes\Delta - E_\pi(\Delta\otimes\Delta)\big]\Big\}^2\Bigg\}\,. \end{aligned}$$

This can be solved again in the same way as for X in Lemma 1, just we have $\big[D - \mathrm{Cov}_\pi\big(\eta, \theta^T\big)\,\mathrm{Var}_\pi^{-1}(\theta)\Delta\big]$ instead of $\eta(\theta)$, $\mathrm{Vec}\,Q^t$ instead of $X_{t,\cdot}$, and $\Delta\otimes\Delta$ instead of θ. So we obtain

$$\begin{aligned} \mathrm{Vec}\,Q^t &= \mathrm{Cov}_\pi\Big\{\big[D_t - \mathrm{Cov}_\pi\big(\eta_t, \theta^T\big)\,\mathrm{Var}_\pi^{-1}(\theta)\Delta\big],\, \Delta^T\otimes\Delta^T\Big\}\ \mathrm{Var}_\pi^{-1}(\Delta\otimes\Delta) \\ &= \mathrm{Cov}_\pi\big(\eta_t, \Delta^T\otimes\Delta^T\big)\ \mathrm{Var}_\pi^{-1}(\Delta\otimes\Delta)\,. \end{aligned}$$

Finally from (5.1) we have

$$\begin{aligned} f_t &= E_\pi(\eta_t) - \big(\mathrm{Vec}\,Q^t\big)^T E_\pi(\Delta\otimes\Delta) \\ &= E_\pi(\eta_t) - (\mathrm{Vec}\,Q^t)^T\,\mathrm{Vec}\big[\mathrm{Var}_\pi(\theta)\big]\,. \end{aligned}$$

□

5.1.1 Evaluation of the bias

From (4.1, 4.2) one obtains straightforwardly

$$E_\theta(\tilde\theta) - \theta = Z^*\big[\Delta^T Q\Delta - E_\pi\big(\Delta^T Q\Delta\big)\big]$$

and

$$E_\theta(\tilde s^2) - \sigma^2 = \frac{\big\|[I - P^*][\Delta^T Q\Delta - E_\pi(\Delta^T Q\Delta)]\big\|^2}{N - m}\,.$$

5.2 Linearization of parametric functions

In the standard approach one linearizes a parametric function $h(\theta)$ by the linear Taylor formula, and considers as an estimator of $h(\theta)$ the expression

$$h(\theta^o) + \frac{\partial h(\theta^o)}{\partial \theta^T}(\hat{\theta} - \theta^o).$$

By the smoothed linearization one has to linearize according to

$$\min_{c\in\mathbb{R},\, l\in\mathbb{R}^m} E_\pi\left\{ \left\| h(\theta) - [l\theta + c] \right\|^2 \right\}$$

which is solved like in Lemma 1, to obtain the linear estimator for $h(\theta)$

$$E_\pi(h) + \mathrm{Cov}_\pi\left(h, \theta^T\right) \mathrm{Var}_\pi^{-1}(\theta)\left[\tilde{\theta} - E_\pi(\theta)\right].$$

It is easy to compute its mean and variance to see the quality of the estimator.

REFERENCES

[1] BATES, D. M.—WATTS, D. G.: *Relative curvature measures of nonlinearity*, J. Roy. Statist. Soc. B **42** (1980), 1–25.

[2] JENČOVÁ, A.: *A comparison of linearization and quadratization domains*, Appl. Math. **42** (1997), 279–291.

[3] KUBÁČEK, L.: *On the general problem of adjustment of measured values*, Mat. Čas. **19** (1969), 270–275.

[4] KUBÁČEK, L.: *About the generalization of an orthogonal regression*, Apl. Mat. **20** (1975), 87–95.

[5] KUBÁČEK, L.: *Universal model for adjusting observed values*, Stud. Geophys. Geod. **20** (1976), 103–113.

[6] KUBÁČEK, L.: *Foundations of Estimation Theory*, Elsevier, Amsterdam, 1988, pp. 141–206.

[7] KUBÁČEK, L.: *On a linearization of regression models*, Appl. Math. **40** (1995), 61–78.

[8] KUBÁČEK, L.: *Quadratic regression models*, Math. Slovaca **46** (1996), 111–126.

[9] KUBÁČEK, L.—PÁZMAN, A.: *Optimum nets*, Stud. Geophys. Geod. **18** (1974), 91–102.

[10] LINNIK, J. W.: *Methode der kleinsten Quadrate in moderner Darstellung*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1961.

[11] PÁZMAN, A.: *The analysis of the error from a linear approximation in the statistical elaboration of results of measurement*, Avtometriya **3** (1966), 76–84. (in Russian)

[12] PÁZMAN, A.: *A sufficient statistic and a nonstandard linearization in nonlinear regression*, Kybernetika **25** (1989), 441–452.

[13] RAO, C. R.: *Linear Statistical Inference and Its Applications*, J.Wiley, New York, 1965.

[14] SERFLING, R. J.: *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

Received September 21, 2000

Department of Probability
and Mathematical Statistics
Faculty of Mathematics,
Physics and Informatics
Comenius University
Mlynská dolina
SK–842 48 Bratislava
SLOVAKIA

E-mail: Andrej.Pazman@fmph.uniba.sk