Tatra Mt. Math. Publ. **17** (1999), 55–63

TATRA MOUNTAINS
Mathematical Publications

BIAS OF THE M. L. E. IN SINGULAR NONLINEAR REGRESSION MODELS

ANDREJ PÁZMAN

ABSTRACT. In a singular nonlinear regression model with normal and independent errors we derive expressions for the approximation of the bias of the maximum likelihood estimators of estimable parameter functions. So we extend results known in regular models to singular ones. This is done by putting equality constraints on the parameters, to make all parameters identifiable, and by approximating the normal equations of the estimator up to $o_p(N^{-1})$. Invariance of the approximate bias under reparametrizations and under the choice of identifying constraints is also proved.

1. Introduction

A statistical model is called singular if its information matrix is singular. A way of removing this singularity is to reparametrize the model. However, to find a reparametrization with given properties in a nonlinear model is usually a very difficult task. Therefore, in general, better it is to maintain the original parameters but use the structure of the model.

One way of regularizing a singular model maintaining the original parameters is to put some "identifying" constraints on the parameters. In applications we find such models which are singular but regularized by nonlinear constraints, for example the ANOVA bilinear model. The original parametrization, which implies singularity, should be maintained here because of its symmetry and good interpretation (cf. [2]).

First order asymptotic properties of maximum likelihood estimators in nonlinear models with constraints (i.e., the asymptotic normality of the estimators) have been investigated in the classical work by Silvey, cf. [7, 8]. However, to approximate the bias of the M. L. E. one needs a second order approximation.

2000 Mathematics Subject Classification: 62J02, 62F12.
Keywords: nonlinear least squares, second order asymptotics, nonlinear constraints.

Prepared as a contribution to a common project with INRA Versailles, France. Supported by the VEGA grant of the Slovak Grant Agency No. 1/4196/97.

For regular models without constraints this approximation is well known, and has been presented for example in [1]. In the present paper, in singular models with constraints we derive the "bias equation" and its solution, we prove the invariance of the approximate bias under a regular reparametrization of the model, and we present the approximate bias of those parameter functions which were estimable in the singular model without constraints, and prove that this bias is invariant with respect to the (auxiliary) choice of constraints.

2. The model

Let us consider the nonlinear regression model with normal errors

$$y_{ij} = \eta(\theta, x_i) + \varepsilon_{ij}\,, \qquad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)\,. \tag{1}$$

The observations y_{ij} are performed independently at n points $x_1, x_2, \dots, x_n$ according to the design $\xi(x_1), \xi(x_2), \dots, \xi(x_n)$. The meaning of the design is that when N is the total number of observations, then $k_i = N\xi(x_i)$ gives the number of replications of the observations at the point x_i. We shall suppose that the design ξ does not change when N increases.

The usual regularity conditions on the model are assumed: the parameter space $\Theta \in \mathbb{R}^p$ is convex and compact, $\bar\theta$, the true value of θ, is an interior point of Θ, $\eta(\theta, x_i)$ is twice continuously differentiable at any $\theta \in \Theta$. We put $\eta(\theta) = \big(\eta(\theta, x_1), \dots, \eta(\theta, x_n)\big)^T$, and by y, resp. ε, we denote the n-vectors of arithmetic means of replicated observations, resp. errors, i.e., $y_i = \frac{1}{k_i}\sum\limits_{r=1}^{k_i} y_{i,r}$, resp. $\varepsilon_i = \frac{1}{k_i}\sum\limits_{r=1}^{k_i} \varepsilon_{i,r}$, $i = 1, \dots, n$.

We also put $D = \operatorname{diag}\big(\xi(x_1), \xi(x_2), \dots, \xi(x_n)\big)$. When we are interested only in the estimator of θ (and not of σ), the regression model after N observations is equivalent to

$$y = \eta(\theta) + \varepsilon\,, \qquad \varepsilon \sim \mathcal{N}\big(0, (\sigma^2/N)D^{-1}\big)\,. \tag{2}$$

This model is supposed to be singular,which means that the Jacobian

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

has a rank less than the number of parameters. Consequently, not all the parameters and parametric functions are identifiable, and the M. L. estimates are not defined uniquely. It is supposed that the true value $\bar\theta$ has a neighbourhood U with a constant rank $\operatorname{rk}\big[J(\theta)\big] = r < p$.

To impose identifying constraints means to find a $p-r$ dimensional, twice continuously differentiable vector function $\phi(\theta)$ whose Jacobian

$$L(\theta) = \frac{\partial \phi(\theta)}{\partial \theta^T}$$

has rank $p-r$ on U, and such that the composite Jacobian $\left(J^T(\theta), L^T(\theta)\right)$ has full rank p. The parameter constraints under question are then

$$\phi(\theta) = 0\,,$$

and they ensure the local uniqueness of the maximum likelihood estimator

$$\widehat{\theta} = \arg \min_{\theta:\ \phi(\theta)=0} \left[y - \eta(\theta)\right]^T D \left[y - \eta(\theta)\right].$$

Thus in this paper we consider the model (1) together with the above identifying constraints.

3. The approximate bias of the M. L. E. in models with constraints

The normal equations for the estimator $\widehat{\theta}$ are (cf. [7], p. 390)

$$\begin{aligned} J^T(\widehat{\theta}) D\left[\eta(\widehat{\theta}) - y\right] + L^T(\widehat{\theta})\widehat{\lambda} &= \mathbf{0}\,, \\ \phi(\widehat{\theta}) &= \mathbf{0}\,, \end{aligned} \tag{3}$$

where $\widehat{\lambda}$ is the estimated vectorial Lagrangean multiplicator. Let us denote by $M(\theta) = J^T(\theta) D J(\theta)$ the Fisher information matrix when $\sigma = 1$, and by $P(\theta) = J(\theta) M^-(\theta) J^T(\theta) D$ a D-orthogonal projector (i.e., orthogonal with respect to the inner product defined by the matrix D). Here $M^-(\theta)$ is an arbitrary g-inverse of $M(\theta)$ and it will be specified later. To see that such a choice does not influence $P(\theta)$ bear in mind that $J^T(\theta)$ and $M(\theta)$ have the same column spaces. So, by the use of the Moore-Penrose g-inverse we show that $P(\theta)P(\theta) = P(\theta)$, and the symmetry with respect to the inner product defined by D is straightforward. Let us denote $P_\perp(\theta) = I - P(\theta)$, the projector D-orthogonal to $P(\theta)$.

Notation. For the sake of brevity we shall omit the argument $\bar{\theta}$ in some expressions such as $\eta(\theta)$, $J(\theta)$, $M(\theta)$, $P(\theta)$ etc. taken at $\bar{\theta}$, the true parameter value.

Premultiply the first Equation of (3) by $J(\widehat{\theta}) M^-(\widehat{\theta})$. We obtain

$$P(\widehat{\theta})\left[\eta(\widehat{\theta}) - y\right] + U(\widehat{\theta})\widehat{\lambda} = \mathbf{0}\,,$$

where $U(\theta) = J(\theta)M^{-}(\theta)L^T(\theta)$. It follows that $P(\widehat{\theta})y = P(\widehat{\theta})\big[\eta(\widehat{\theta}) + U(\widehat{\theta})\widehat{\lambda}\big]$. Using this we obtain

$$\varepsilon = P(\widehat{\theta})\varepsilon + P_\perp(\widehat{\theta})\varepsilon = \big[\eta(\widehat{\theta}) - \eta + U(\widehat{\theta})\widehat{\lambda}\big] + P_\perp(\widehat{\theta})\big[\varepsilon - \eta(\widehat{\theta}) + \eta\big]$$

since $P_\perp(\widehat{\theta})U(\widehat{\theta}) = \mathbf{0}$. Further, we can write $P_\perp(\widehat{\theta}) = \Omega(\widehat{\theta})\Omega^T(\widehat{\theta})D$, where the columns of $\Omega(\widehat{\theta})$ form a D-orthonormal basis of the projection space of $P_\perp(\widehat{\theta})$. So we obtain

$$\varepsilon = \big[\eta(\widehat{\theta}) - \eta + U(\widehat{\theta})\widehat{\lambda}\big] + \Omega(\widehat{\theta})\Omega^T(\widehat{\theta})D\big[\varepsilon - \eta(\widehat{\theta}) + \eta\big]\,. \tag{4}$$

For sake of brevity, let us denote $\Delta = \widehat{\theta} - \bar{\theta}$. Denote further by $H(\theta)$ the Hessian operator with components $H^k_{r,s}(\theta) = \frac{\partial^2 \eta_k(\theta)}{\partial\theta_r \partial\theta_s}$. We consider the following Taylor expansions $\eta_k(\widehat{\theta}) - \eta_k = J_{k.}\Delta + \frac{1}{2}\Delta^T H^k_{..}\Delta + o_p\big(\frac{1}{N}\big)$, $U_{k.}(\widehat{\theta})\widehat{\lambda} = U_{k.}\widehat{\lambda} + \sum_i \Delta_i \frac{\partial U_{k.}(\bar{\theta})}{\partial\theta_i}\widehat{\lambda} + o_p\big(\frac{1}{N}\big)$, $\Omega_{kl}(\widehat{\theta}) = \Omega_{kl} + \sum_i \Delta_i \frac{\partial \Omega_{kl}(\bar{\theta})}{\partial\theta_i} + o_p\big(\frac{1}{\sqrt{N}}\big)$. The o_p terms are justified by the fact that $\sqrt{N}\Delta$ and $\sqrt{N}\widehat{\lambda}$ are asymptotically normally distributed with zero expectations. (cf. [7], p. 401), and that obviously, from the definition of ε, it follows that the same also holds for the components of the vector $\sqrt{N}\varepsilon$.

Now put the above Taylor expansions into (4) and premultiply the result by $J^T D$. We obtain

$$\begin{aligned} J^T D\varepsilon = M\Delta &+ \frac{1}{2}\sum_i J^T D_{.i}\big[\Delta^T H^i_{..}\Delta\big] + J^T DU\widehat{\lambda} + \sum_i J^T D\left[\Delta_i \frac{\partial U}{\partial\theta_i}\widehat{\lambda}\right] \\ &+ \sum_i \Delta_i J^T D\frac{\partial\Omega}{\partial\theta_i}\Omega^T\varepsilon + o_p\Big(\frac{1}{N}\Big)\,, \end{aligned} \tag{5}$$

where we used the fact that $J^T(\theta)D\,\Omega(\theta) = 0$. Taking the derivative of this we obtain that $J^T D\frac{\partial\Omega}{\partial\theta_i} = -\frac{\partial J^T}{\partial\theta_i}D\Omega$. We put this in (5), together with the equality $J^T DU = MM^{-}L^T$. The order of approximation $o_p\big(\frac{1}{N}\big)$ is not modified if we replace $\widehat{\lambda}$ and Δ in the quadratic terms of (5) by their first order asymptotic approximations; they will be denoted by $\widehat{\lambda}^{(1)}$ and $\Delta^{(1)}$. We obtain

$$\begin{aligned} J^T D\varepsilon = M\Delta &+ \frac{1}{2}J^T D\Big[\big[\Delta^{(1)}\big]^T H\Delta^{(1)}\Big] + MM^{-}L^T\widehat{\lambda} + \sum_i J^T D\Big[\Delta^{(1)}_i \frac{\partial U}{\partial\theta_i}\widehat{\lambda}^{(1)}\Big] \\ &+ \sum_i \Delta^{(1)}_i \frac{\partial J^T}{\partial\theta_i}DP_\perp\varepsilon + o_p\Big(\frac{1}{N}\Big)\,. \end{aligned}$$

Let us neglect the term $o_p\big(\frac{1}{N}\big)$ and in fact obtain approximate normal equations. Then we take the expectation on both sides. We use the fact that $\widehat{\lambda}^{(1)}$

and $\Delta^{(1)}$ are uncorrelated random vectors (cf. [7], p401) as are also $\Delta^{(1)}$ and the vector of residuals $P_{\perp}\varepsilon$. We obtain

$$\mathbf{0} = ME[\Delta] + \frac{1}{2}J^T D \operatorname{Tr}\left\{H \operatorname{Var}\left[\Delta^{(1)}\right]\right\} + MM^- L^T E\left[\widehat{\lambda}\right], \tag{6}$$

where $\operatorname{Tr}\left\{H \operatorname{Var}\left[\Delta^{(1)}\right]\right\}$ is a n-vector whose kth component is $\sum_{r,s} H^k_{r,s} \operatorname{Cov}\left[\Delta^{(1)}_r, \Delta^{(1)}_s\right]$. From the second equality in (3), after using Taylor formula we obtain

$$\phi(\bar{\theta}) + L\Delta + \frac{1}{2}\Delta^T K \Delta + o_p\left(\frac{1}{N}\right) = \phi(\widehat{\theta}) = 0,$$

where $K(\theta)$ is the Hessian operator with components $K^k_{r,s}(\theta) = \frac{\partial^2 \phi_k(\theta)}{\partial\theta_r \partial\theta_s}$. We have $\phi(\bar{\theta}) = \mathbf{0}$ and we can again replace Δ by $\Delta^{(1)}$ in the quadratic term. So by neglecting $o_p\left(\frac{1}{N}\right)$ and taking the expectation, we obtain approximately

$$LE[\Delta] + \frac{1}{2}\operatorname{Tr}\left\{K \operatorname{Var}\left[\Delta^{(1)}\right]\right\} = 0, \tag{7}$$

where $\operatorname{Tr}\left\{K \operatorname{Var}\left[\Delta^{(1)}\right]\right\}$ is a q-vector whose ith component is $\sum_{r,s} K^i_{r,s} \operatorname{Cov}\left[\Delta^{(1)}_r, \Delta^{(1)}_s\right]$. Combining both (6) and (7) we obtain the following approximate "bias equation"

$$\begin{pmatrix} M & MM^- L^T \\ L & \mathbf{0} \end{pmatrix} \begin{pmatrix} E[\Delta] \\ E[\widehat{\lambda}] \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} J^T D \operatorname{Tr}\left\{H \operatorname{Var}\left[\Delta^{(1)}\right]\right\} \\ \operatorname{Tr}\left\{K \operatorname{Var}\left[\Delta^{(1)}\right]\right\} \end{pmatrix}. \tag{8}$$

PROPOSITION 1. *The approximate bias of $\widehat{\theta}$ which is a solution of* (8) *is equal to*

$$b_{\bar{\theta}} = -\frac{1}{2}\left(M + L^T L\right)^{-1}\left(J^T D \operatorname{Tr}\left\{H \operatorname{Var}\left[\Delta^{(1)}\right]\right\} + L^T \operatorname{Tr}\left\{K \operatorname{Var}\left[\Delta^{(1)}\right]\right\}\right), \tag{9}$$

where

$$\begin{aligned} &\operatorname{Var}\left[\Delta^{(1)}\right] = \sigma^2\left\{Q^{-1} - Q^{-1}L^T\left[LQ^{-1}L^T\right]^{-1}LQ^{-1}\right\}, \\ &\textit{with} \qquad Q = M + L^T L. \end{aligned}$$

P r o o f. We recall that $\operatorname{rk}\left(J^T(\theta),\ L^T(\theta)\right) = p$. One can find r linearly independent rows of M which are linearly independent of the rows of the matrix L. Let us denote by ψ a matrix formed from these rows. Because of the independence, one can always find a positive definite matrix C satisfying the orthogonality relation $\psi C L^T = \mathbf{0}$. This means that there exist a g-inverse M^- such that $MM^- L^T = \mathbf{0}$ since MM^- is a projector onto $\mathcal{M}[M]$, the column space of M, and since for any matrix C one can find a g-inverse such that MM^-

is an C-orthogonal projector (as proved in Rao & Mitra, 1971, Lemma 5.3.1). From (8) we obtain

$$\begin{pmatrix} M \\ L \end{pmatrix} b_{\bar{\theta}} = -\frac{1}{2} \begin{pmatrix} J^T D \operatorname{Tr}\left\{ H \operatorname{Var}\left[\Delta^{(1)}\right] \right\} \\ \operatorname{Tr}\left\{ K \operatorname{Var}\left[\Delta^{(1)}\right] \right\} \end{pmatrix} . \tag{10}$$

Premultiplying this by $\left(M + L^T L\right)^{-1} \left(I_p, L^T\right)$ we obtain the required expression. The expression for $\operatorname{Var}\left[\Delta^{(1)}\right]$ is given in [7]. □

4. Bias of parametric functions

4.1 Bias in models with constraints.

Let $h(\theta)$ be an arbitrary twice differentiable parameter function. For the approximate bias of $h(\widehat{\theta})$ one can use the simple formula

$$b_{\bar{\theta}}(h) = \frac{\partial h(\bar{\theta})}{\partial \theta^T} b_{\bar{\theta}} + \frac{1}{2} \operatorname{tr}\left(\frac{\partial^2 h(\bar{\theta})}{\partial \theta \, \partial \theta^T} \operatorname{Var}\left(\Delta^{(1)}\right) \right) . \tag{11}$$

To prove this, consider the Taylor formula

$$h(\widehat{\theta}) - h(\bar{\theta}) = \frac{\partial h(\bar{\theta})}{\partial \theta^T} \Delta + \frac{1}{2} \Delta^T \frac{\partial^2 h(\bar{\theta})}{\partial \theta \, \partial \theta^T} \Delta + o_p\left(\frac{1}{N}\right)$$

substitute $\Delta^{(1)}$ for Δ in the quadratic term, neglect $o_p\left(\frac{1}{N}\right)$, and take the expectation.

Let $\beta = \beta(\theta)$ be a regular one-to-one reparametrisation of the model (1), and denote by $\theta = \theta(\beta)$ its inverse. The reparametrized model with constraints has the form

$$y = \nu(\beta) + \varepsilon \,, \qquad \kappa(\beta) = 0 \,, \tag{12}$$

where by definition $\nu(\beta) = \eta\left[\theta(\beta)\right]$, $\kappa(\beta) = \phi\left[\theta(\beta)\right]$. Let $h(\theta)$ be a given parametric function. Let $l(\beta) = h\left[\theta(\beta)\right]$, and denote by $b_{\bar{\beta}}(l)$ the approximative bias of $l(\widehat{\beta})$ in the reparametrized model (12).

Proposition 2. *The approximate bias is parametrically invariant, i.e.,*

$$b_{\bar{\theta}}(h) = b_{\bar{\beta}}(l) \,.$$

P r o o f. Denote by $b_{\bar{\theta}}$, respectively $b_{\bar{\beta}}$, the solutions of the bias equations in model (1) and model (12) respectively. First of all it is necessary to verify that the solutions of the two bias equation are related by (11), i.e., to prove that

$$(b_{\bar{\beta}})_i = \frac{\partial \beta_i}{\partial \theta} b_{\bar{\theta}} + \frac{1}{2} \operatorname{tr}\left(\frac{\partial^2 \beta_i}{\partial \theta \partial \theta^T} V^{(\theta)} \right) , \tag{13}$$

where $V^{(\theta)} = \mathrm{Var}\big[\Delta_\theta^{(1)}\big]$. Similarly we define $V^{(\beta)}$ in Model (12). For the sake of brevity we write $\frac{\partial \beta}{\partial \theta}$ and $\frac{\partial \theta}{\partial \beta}$ instead of $\frac{\partial \beta(\theta)}{\partial \theta}\Big|_{\theta=\bar\theta}$ and $\frac{\partial \theta(\beta)}{\partial \beta}\Big|_{\beta=\bar\beta}$, etc. We have the following evident rules:

$$\frac{\partial \theta}{\partial \beta^T}\frac{\partial \beta}{\partial \theta^T} = I_p\,, \qquad J(\bar\beta) = J(\bar\theta)\frac{\partial \theta}{\partial \beta^T}\,, \qquad V^{(\beta)} = \frac{\partial \beta}{\partial \theta^T}V^{(\theta)}\frac{\partial \beta^T}{\partial \theta}\,,$$

$$H(\bar\beta) = \frac{\partial \theta^T}{\partial \beta}H(\bar\theta)\frac{\partial \theta}{\partial \beta^T} + J(\bar\theta)\frac{\partial^2 \theta}{\partial \beta\,\partial \beta^T}, \qquad M^-(\bar\beta) = \frac{\partial \beta}{\partial \theta^T}M^-(\bar\theta)\frac{\partial \beta^T}{\partial \theta}\,, \quad \text{etc.}$$

We put (13) into the bias equation of model (12), apply the presented transformation rules and as a result we obtain the bias equation of model (1). This proves that (13) is correct. According to (11)

$$b_{\bar\theta}(h) = \frac{\partial h}{\partial \theta^T}b_{\bar\theta} + \frac{1}{2}\,\mathrm{tr}\bigg(\frac{\partial^2 h}{\partial \theta\,\partial \theta^T}\bigg)V^{(\theta)} \tag{14}$$

and

$$b_{\bar\beta}(l) = \frac{\partial l}{\partial \beta^T}b_{\bar\beta} + \frac{1}{2}\,\mathrm{tr}\bigg(\frac{\partial^2 l}{\partial \beta\,\partial \beta^T}\bigg)V^{(\beta)}\,. \tag{15}$$

From $l(\beta) = h\big[\theta(\beta)\big]$ we obtain the derivatives of $l(\beta)$ in a standard way. Substituting them with (13) into (15) we obtain the right-hand side of (14). Hence $b_{\bar\theta}(h) = b_{\bar\beta}(l)$. □

4.2 Bias in the singular model without constraints.

In the singular model (1), taken however without constraints, not all parametric functions of the form $h(\theta)$ are estimable, a necessary and sufficient condition of estimability being that $h(\theta)$ is a function of the observed vector $\eta(\theta)$ (cf. [3]). In symbols, there should be a mapping F such that $h(\theta) = F\big[\eta(\theta)\big]$.

Proposition 3. *The approximate bias of the vector* $\eta(\hat\theta) = \big(\eta(\hat\theta, x_1), \dots, \eta(\hat\theta, x_n)\big)^T$ *is equal to*

$$b_{\bar\theta}[\eta] = \frac{1}{2}[I-P]\,\mathrm{Tr}\,\Big\{H\,\mathrm{Var}\big[\Delta^{(1)}\big]\Big\}\,.$$

P r o o f. From (10) we obtain

$$Mb_{\bar\theta} = -\frac{1}{2}J^T D\,\mathrm{Tr}\,\Big\{H\,\mathrm{Var}\big[\Delta^{(1)}\big]\Big\}\,.$$

The column space of J is equal to the column space of M, $\mathcal{M}[J] = \mathcal{M}[M]$. Hence the Moore-Penrose g-inverse M^+ has the property that $JM^+M = J$. We use this when we multiply the above equality from the left by JM^+ to obtain

$$Jb_{\bar\theta} = -\frac{1}{2}P\,\mathrm{Tr}\,\Big\{H\,\mathrm{Var}\big[\Delta^{(1)}\big]\Big\}\,.$$

On the other hand, from (11) we obtain that the approximate bias of $\eta(\hat{\theta})$ is equal to $Jb_{\bar{\theta}} + \frac{1}{2}\operatorname{Tr}\big(H\operatorname{Var}(\Delta^{(1)})\big)$. Combining both equalities we obtain the required result for $b\big[\eta(\hat{\theta})\big]$. □

By the use of (11) we obtain

COROLLARY 1. *The approximate bias of $F\big[\eta(\hat{\theta})\big]$ is equal to*

$$b_{\bar{\theta}}\big(F[\eta]\big) = \frac{\partial F}{\partial t}\Big|_{\eta(\hat{\theta})} b_{\bar{\theta}}[\eta] + \operatorname{Tr}\left\{\frac{\partial^2 F}{\partial t\,\partial t^T}\Big|_{\eta(\hat{\theta})} \operatorname{Var}\big[\hat{\eta}^{(1)}\big]\right\},$$

where $\operatorname{Var}\big[\hat{\eta}^{(1)}\big] = J\operatorname{Var}\big[\Delta^{(1)}\big]J^T$.

PROPOSITION 4. *The expressions for $b_{\bar{\theta}}[\eta]$ and $b_{\bar{\theta}}\big(F[\eta]\big)$ are invariant with respect to the choice of identifying constraints.*

Proof. Since $\mathcal{M}(M)$, and $\mathcal{M}(L^TL)$ are complementary subspaces of $\mathbb{R}^p$ of dimension r and $p-r$ respectively, there is a regular matrix Z such that $ZMZ^T = \begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix}$, $ZL^TLZ = \begin{pmatrix} 0 & 0 \\ 0 & I_{(p-r)\times(p-r)} \end{pmatrix}$. It follows that $MQ^{-1}L^T = 0$, and that Q^{-1} is a g-inverse of M, and also of L^TL. Consequently, using the expression for $\operatorname{Var}\big[\Delta^{(1)}\big]$ from Proposition 1, we obtain

$$J\operatorname{Var}\big[\Delta^{(1)}\big]J^T = JM^-J^T,$$
$$(I-P)\operatorname{Tr}\left\{H\operatorname{Var}\big[\Delta^{(1)}\big]\right\} = (I-P)\operatorname{Tr}\left\{HM^-\right\},$$

and the statement follows from Proposition 3 and Corollary 1. □

REFERENCES

[1] BOX, M. J.: *Bias in nonlinear estimation*, J. Roy. Statist. Soc. Ser. B **33** (1971), 171–201.

[2] DENIS, J. B.—GOWER, J. G.: *Asymptotic covariance for the parameters of biadditive models*, Utilitas Math. **46** (1994), 193–205.

[3] KOUTKOVÁ, H.: *On estimable and locally estimable functions in the nonlinear regression model*, Kybernetika **28** (1992), 120–128.

[4] PÁZMAN, A.—DENIS, J. B.: *Bias in nonlinear models with constrained parameters*, report No. 4, Unité de biometrie, INRA, Versailles, 1997.

[5] DENIS, J. B.—PÁZMAN, A.: *Biadditive* ANOVA *models: reminders and asymptotic bias*, report No. 4, Unité de biometrie, INRA, Versailles, 1998.

[6] PÁZMAN, A.: *Nonlinear Statistical Models*, Kluwer Academic Publishers, Dordrecht, 1995, p. 259.

[7] SILVEY, S. D.: *The Lagrangian multiplier test*, Ann. Math. Statist. **30** (1959), 389–407.

[8] SILVEY, S. D.: *Statistical Inference*, Chapman and Hall, London, 1975, p. 191.

[9] RAO, C. R.—MITRA, S. K.: *Generalized Inverse of Matrices and its Applications*, John Wiley & sons, New York, 1971, p. 240.

Received April 27, 1998

Faculty of Mathematics and Physics
Commenius University
Mlynská dolina
SK–842 15, Bratislava
SLOVAKIA

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