Andrej Pázman; Jean-Baptiste Denis

Bias of LS estimators in nonlinear regression models with constraints. Part I: General case

Applications of Mathematics, Vol. 44 (1999), No. 5, 359–374

Persistent URL: `http://dml.cz/dmlcz/133891`

BIAS OF LS ESTIMATORS IN NONLINEAR REGRESSION MODELS WITH CONSTRAINTS. PART I: GENERAL CASE

Andrej Pázman, Bratislava and Jean-Baptiste Denis, Versailles

(Received July 13, 1998)

Abstract. We derive expressions for the asymptotic approximation of the bias of the least squares estimators in nonlinear regression models with parameters which are subject to nonlinear equality constraints.

The approach suggested modifies the normal equations of the estimator, and approximates them up to $o_p(N^{-1})$, where N is the number of observations. The "bias equations" so obtained are solved under different assumptions on constraints and on the model. For functions of the parameters the invariance of the approximate bias with respect to reparametrisations is demonstrated. Singular models are considered as well, in which case the constraints may serve either to identify the parameters, or eventually to restrict the parameter space.

Keywords: nonlinear least squares, maximum likelihood, asymptotic bias, nonlinear constraints, transformation of parameters

MSC 2000: 62J02

1. Introduction

Approximate bias formulae for least squares estimators in nonlinear regression models were first derived by Cox and Snell [2] and Box [1] for regular models without constraints. Here we generalize these formulæ to the case where parameters are subject to nonlinear constraints. To do so we use some classical results on the first order asymptotics of maximum likelihood estimation with constraints, as presented in Silvey [8, 9], in combination with the methods of bias approximation in models without constraints (cf. Chapter 6 of Pázman [5]). This can be done only when the LS and the ML estimators coincide, e.g. when the errors are normally distributed.

This work was supported by INRA (France) and by the Slovak Grant Agency (grant N°1/4196/97).

In Appendix we show how these results can be extended to the LS estimation under non-normal errors under supplementary assumptions. Of course, in classical nonlinear models with a few parameters it is always possible to reparametrize the model to obtain a full rank model without constraints. This is not the case for some bilinear analysis of variance type models (cf. Denis and Pázman [4]), where a large number of parameters and constraints considered would imply big numerical difficulties, due also to the loss of symmetry produced by the reparametrization. Indeed, this was the first motivation for this work. Nevertheless, we believe that in many situations, the possibility of constraints in defining models is important because a proper interpretation can be given to the parameters, allowing easier interpretations.

The paper is organized in the following manner. After presenting the model in Section 2, in Section 3 we modify the normal equations of the constrained least squares and approximate them up to the order $o_p(N^{-1})$. Without changing the order of approximation we substitute the estimators by their first order approximations at some places. This leads to linear equations for the approximate bias, called *bias equations*. The way to solve these equations is discussed in Section 4, first for the case where constraints are restrictive in a regular (i.e. full rank) model, then where they are identifying in a singular model, and finally, where they are restrictive and identifying in a singular model. In the first two cases an explicit expression for the bias is obtained, while in the last case the existence and uniqueness of the solution of the bias equation is proved. We also show that the general formulae proposed here for the approximate bias are consistent with the formulae already well known in non-constrained models. In Section 5 the approximate bias of an arbitrary function of the parameters is obtained, and it is shown that this bias is invariant with respect to reparametrization of the model.

2. The model

Let us observe a vector $y^* = (y_1^*, y_2^*, \dots, y_n^*)^T$ of n observations modelled by

$$y^* = \eta(\theta) + \varepsilon^* \qquad (1)$$
$$E[\varepsilon^*] = \mathbf{0}_n; \quad \mathrm{Var}[\varepsilon^*] = \sigma^2 \mathbf{I}_n$$

where the p-vector of parameters θ is subject to q independent constraints

$$\varphi(\theta) = \mathbf{0}_q, \qquad (2)$$

which ensures that the parameters θ can be locally identified. Usual regularity conditions are supposed: the parameter space Θ is convex and compact, $\overline{\theta}$, the true value of θ, is an interior point of Θ, $\eta(\theta)$ and $\varphi(\theta)$ are twice continuously differentiable in θ. $\mathbf{0}_r$ denotes the null r-vector and $\mathbf{I}_n$ is the identity matrix of size n.

The errors ε^* are supposed to be normally distributed. (See Appendix for a more general case.) To evaluate the accuracy of approximations below we consider N independent replications of observation vectors: $y^{*(1)}, y^{*(2)}, \dots, y^{*(N)}$. Let us denote by $\varepsilon^{*(i)}$ the error vector associated with the ith replication. As is well known, the arithmetic mean

$$y = \frac{1}{N}\sum_{i=1}^{N} y^{*(i)}$$

is a sufficient statistics for the parameter vector θ. So instead of N replications, we have the model

$$\begin{aligned} & y = \eta(\theta) + \varepsilon \\ & E[\varepsilon] = \mathbf{0}_n; \quad \operatorname{Var}[\varepsilon] = \frac{\sigma^2}{N}\mathbf{I}_n \\ & \varepsilon \text{ is normally distributed} \end{aligned} \tag{3}$$

where $\varepsilon = \frac{1}{N}\sum\limits_{i=1}^{N} \varepsilon^{*(i)}$, and the p-vector of parameters θ is still subject to the same q independent constraints $\varphi(\theta) = \mathbf{0}$. We see from (3) that $N \to \infty$ has the same effect on the estimators as $\sigma \to 0$. In particular, the bias approximation considered holds either for a large N or for a small σ.

3. Modification and approximation of the normal equations

The normal equations for the estimator

$$\widehat{\theta} \in \arg \min_{\theta:\ \varphi(\theta)=0} [y - \eta(\theta)]^T [y - \eta(\theta)] \tag{4}$$

are (cf. also Silvey [8], p. 390)

$$\begin{aligned} & J^T\big(\widehat{\theta}\big)\big[\eta\big(\widehat{\theta}\big) - y\big] + L^T\big(\widehat{\theta}\big)\widehat{\lambda} = \mathbf{0} \\ & \varphi\big(\widehat{\theta}\big) = \mathbf{0} \end{aligned} \tag{5}$$

where $J(\theta) = \frac{\partial \eta(\theta)}{\partial \theta^T}$, $L(\theta) = \frac{\partial \varphi(\theta)}{\partial \theta^T}$ are the Jacobians of the response and constraint functions, and $\widehat{\lambda}$ is the estimated q-vectorial Lagrangian multiplicator. We suppose that the constraints are independent, that means that $L(\theta)$ is of full rank q. On the other hand the model itself may be singular, and in that case $J(\theta)$ is of rank r smaller than p. However, it is supposed that the total rank satisfies

$$\operatorname{rk}\begin{pmatrix} J(\theta) \\ L(\theta) \end{pmatrix} = p$$

at every interior point of Θ.

Let us denote by $M(\theta) = J^T(\theta)J(\theta)$ the Fisher information matrix when $\sigma = 1$, and by

$$P(\theta) = J(\theta)M^-(\theta)J^T(\theta)$$

an orthogonal projector. Here $M^-(\theta)$ is any g-inverse of $M(\theta)$, and it will be specified later. Let us denote $P_\perp(\theta) = \mathbf{I} - P(\theta)$.

Premultiply the first equation of (5) by $J(\widehat{\theta})M^-(\widehat{\theta})$. We obtain

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

where $U(\theta) = J(\theta)M^-(\theta)L^T(\theta)$. It follows that

$$P(\widehat{\theta})y = P(\widehat{\theta})\left[\eta(\widehat{\theta}) + U(\widehat{\theta})\widehat{\lambda}\right].$$

Using this we obtain

$$\begin{aligned}
\varepsilon &= P(\widehat{\theta})\varepsilon + P_\perp(\widehat{\theta})\varepsilon \\
&= P(\widehat{\theta})\left[y - \eta(\overline{\theta})\right] + P_\perp(\widehat{\theta})\varepsilon \\
&= P(\widehat{\theta})\left[\eta(\widehat{\theta}) - \eta(\overline{\theta}) + U(\widehat{\theta})\widehat{\lambda}\right] + P_\perp(\widehat{\theta})\varepsilon \\
&= \left[\eta(\widehat{\theta}) - \eta(\overline{\theta}) + U(\widehat{\theta})\widehat{\lambda}\right] + P_\perp(\widehat{\theta})\left[\varepsilon - \eta(\widehat{\theta}) + \eta(\overline{\theta})\right]
\end{aligned}$$

since $P_\perp(\widehat{\theta})U(\widehat{\theta}) = \mathbf{0}_{n\times q}$. Now we can write $P_\perp(\widehat{\theta}) = \Omega(\widehat{\theta})\Omega^T(\widehat{\theta})$ where the columns of $\Omega(\widehat{\theta})$ form an orthonormal basis of the projection space of $P_\perp(\widehat{\theta})$. So we obtain

$$\varepsilon = [\eta(\widehat{\theta}) - \eta(\overline{\theta}) + U(\widehat{\theta})\widehat{\lambda}] + \Omega(\widehat{\theta})\Omega^T(\widehat{\theta})\left[\varepsilon - \eta(\widehat{\theta}) + \eta(\overline{\theta})\right]. \tag{6}$$

For sake of brevity, let us denote

$$\Delta = \widehat{\theta} - \overline{\theta}.$$

and denote 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 Taylor expansions

$$\begin{aligned}
\eta_k(\widehat{\theta}) - \eta_k(\overline{\theta}) &= J_{k.}(\overline{\theta})\Delta + \frac{1}{2}\Delta^T H^k_{..}(\overline{\theta})\Delta + o_p\Big(\frac{1}{N}\Big) \\
U_{k.}(\widehat{\theta})\widehat{\lambda} &= U_{k.}(\overline{\theta})\widehat{\lambda} + \sum_i \Delta_i \frac{\partial U_{k.}(\overline{\theta})}{\partial\theta_i}\widehat{\lambda} + o_p\Big(\frac{1}{N}\Big) \\
\Omega_{kl}(\widehat{\theta}) &= \Omega_{kl}(\overline{\theta}) + \sum_i \Delta_i \frac{\partial\Omega_{kl}(\overline{\theta})}{\partial\theta_i} + o_p\Big(\frac{1}{\sqrt{N}}\Big)
\end{aligned} \tag{7}$$

which hold since $\sqrt{N}\Delta$ and $\sqrt{N}\widehat{\lambda}$ are asymptotically normally distributed with zero expectations (i.e. they converge in distribution to some normal variables, cf. Silvey [8], p. 401). Obviously the same holds also for the components of the vector $\sqrt{N}\varepsilon$. Notice that Δ, ε and $\widehat{\lambda}$ depend on N, and that a sequence of random variables $\{\zeta_N\}_{N_0}^{\infty}$ is called $o_p(N^{-k})$ when the sequence $\{N^k\zeta_N\}_{N_0}^{\infty}$ converges to zero in probability. As is well known, a sequence converging in the distribution function to zero converges also in probability to zero, and a product of two sequences converging in probability to zero converges in probability as well (theorem of Slutsky, cf. Cramér [3] p. 255).

The last expression in (7) allows us to write

$$\text{(8)}\quad \Omega(\widehat{\theta})\Omega^T(\widehat{\theta}) = \left[\Omega(\overline{\theta})\Omega^T(\overline{\theta}) + \sum_i \Delta_i (\Omega(\overline{\theta})\frac{\partial \Omega^T(\overline{\theta})}{\partial \theta_i} + \frac{\partial \Omega(\overline{\theta})}{\partial \theta_i}\Omega^T(\overline{\theta}))\right] + o_p\Big(\frac{1}{\sqrt{N}}\Big).$$

Now let us put the expressions from (7) and (8) into (6) and premultiply the result by $J^T(\overline{\theta})$. We obtain

$$\begin{aligned}\text{(9)}\qquad J^T(\overline{\theta})\varepsilon = {} & M(\overline{\theta})\Delta + \frac{1}{2}\sum_i J^T(\overline{\theta})\,[\Delta^T H^i_{\cdot\cdot}(\overline{\theta})\Delta] \\ & + J^T(\overline{\theta})U(\overline{\theta})\widehat{\lambda} + \sum_i J^T(\overline{\theta})\Big[\Delta_i \frac{\partial U(\overline{\theta})}{\partial \theta_i}\widehat{\lambda}\Big] \\ & + \sum_i \Delta_i J^T(\overline{\theta})\frac{\partial \Omega(\overline{\theta})}{\partial \theta_i}\Omega^T(\overline{\theta})\varepsilon + o_p\Big(\frac{1}{N}\Big)\end{aligned}$$

where we have used the definition of $M(\overline{\theta})$ and the fact that $J^T(\theta)\Omega(\theta) = 0$ for every interior point of Θ. This implies further that $J^T(\overline{\theta})\frac{\partial \Omega(\overline{\theta})}{\partial \theta_i} = -\frac{\partial J^T(\overline{\theta})}{\partial \theta_i}\Omega(\overline{\theta})$. We use this in (9), as well as the evident equality $J^T(\overline{\theta})U(\overline{\theta}) = M(\overline{\theta})M^-(\overline{\theta})L^T(\overline{\theta})$. The term $o_p(\frac{1}{N})$ is not modified if we replace $\widehat{\lambda}$ and Δ in the terms of (9), which are quadratic in $\hat{\lambda}$, Δ, ε, by their first order asymptotic approximations; they will be denoted by $\widehat{\lambda}^{(1)}$ and $\Delta^{(1)}$. So finally we obtain

$$\begin{aligned}\text{(10)}\qquad J^T(\overline{\theta})\varepsilon = {} & M(\overline{\theta})\Delta + \frac{1}{2}J^T(\overline{\theta})[[\Delta^{(1)}]^T H(\overline{\theta})\Delta^{(1)}] \\ & + M(\overline{\theta})M^-(\overline{\theta})L^T(\overline{\theta})\widehat{\lambda} + \sum_i J^T(\overline{\theta})\Big[\Delta_i^{(1)}\frac{\partial U(\overline{\theta})}{\partial \theta_i}\widehat{\lambda}^{(1)}\Big] \\ & + \sum_i \Delta_i^{(1)}\frac{\partial J^T(\overline{\theta})}{\partial \theta_i}P_\perp(\overline{\theta})\varepsilon + o_p\Big(\frac{1}{N}\Big)\end{aligned}$$

Denote by $K(\theta)$ the Hessian operator with components $K^k_{r,s}(\theta) = \frac{\partial^2 \varphi_k(\theta)}{\partial \theta_r \partial \theta_s}$.

Proposition 1. *The bias of $\widehat{\theta}$ and of $\hat{\lambda}$, $E[\Delta]$ and $E[\hat{\lambda}]$, satisfy the equation*

$$\begin{pmatrix} M(\overline{\theta}) & M(\overline{\theta})M^{-}(\overline{\theta})L^{T}(\overline{\theta}) \\ L(\overline{\theta}) & \mathbf{0}_{q\times q} \end{pmatrix} \begin{pmatrix} E[\Delta] \\ E[\widehat{\lambda}] \end{pmatrix} \\ = -\frac{1}{2}\begin{pmatrix} J^{T}(\overline{\theta})\operatorname{Tr}\{H(\overline{\theta})\operatorname{Var}[\Delta^{(1)}]\} \\ \operatorname{Tr}\{K(\overline{\theta})\operatorname{Var}[\Delta^{(1)}]\} \end{pmatrix} + E\Big[o_p\Big(\frac{1}{N}\Big)\Big]$$

where $\operatorname{Tr}\{H(\overline{\theta})\operatorname{Var}[\Delta^{(1)}]\}$ *is an n-vector whose kth component is* $\sum_{r,s} H^{k}_{r,s}(\overline{\theta}) \times \operatorname{Cov}[\Delta^{(1)}_r, \Delta^{(1)}_s]$ *and similarly* $\operatorname{Tr}\{K(\overline{\theta})\operatorname{Var}[\Delta^{(1)}]\}$ *is a q-vector whose ith component is equal to* $\{\sum_{r,s} K^{i}_{r,s}(\overline{\theta})\operatorname{Cov}[\Delta^{(1)}_r, \Delta^{(1)}_s]\}$.

Proof. Let us take the expectation on both sides of Equation (10). We use the fact that $\widehat{\lambda}^{(1)}$ and $\Delta^{(1)}$ are independent random vectors (Silvey [8], p. 401) as are $\Delta^{(1)}$ and the vector of residuals $P_{\perp}(\widehat{\theta})\varepsilon$. We obtain

$$\mathbf{0}_p = M(\overline{\theta})E[\Delta] + \frac{1}{2}J^{T}(\overline{\theta})\operatorname{Tr}\{H(\overline{\theta})\operatorname{Var}[\Delta^{(1)}]\} \\ + M(\overline{\theta})M^{-}(\overline{\theta})L^{T}(\overline{\theta})E[\widehat{\lambda}] + E\Big[o_p\Big(\frac{1}{N}\Big)\Big]. \tag{11}$$

From the second equality in (5), after using the Taylor formula we obtain

$$\mathbf{0}_q = \varphi(\widehat{\theta}) = \varphi(\overline{\theta}) + L(\overline{\theta})\Delta + \frac{1}{2}\Delta^{T}K(\overline{\theta})\Delta + o_p\Big(\frac{1}{N}\Big).$$

We have $\varphi(\overline{\theta}) = \mathbf{0}$ and we can replace again Δ by $\Delta^{(1)}$ in the quadratic term. So by taking the expectation, we obtain

$$L(\overline{\theta})E[\Delta] + \frac{1}{2}\operatorname{Tr}\{K(\overline{\theta})\operatorname{Var}[\Delta^{(1)}]\} + E\Big[o_p\Big(\frac{1}{N}\Big)\Big] = \mathbf{0}_q. \tag{12}$$

Combining both (11) and (12) one obtains the required equation. □

Denote by $b(\hat{\theta})$ and $b(\hat{\lambda})$ the approximate bias of $\hat{\theta}$ and $\hat{\lambda}$. They are solutions of the *bias equations*, which are obtained from the above proposition by neglecting the $o_p(\frac{1}{N})$ terms, and by substituting $b(\hat{\theta})$ and $b(\hat{\lambda})$ instead of $E[\Delta]$ and of $E[\widehat{\lambda}]$:

$$\begin{pmatrix} M(\overline{\theta}) & M(\overline{\theta})M^{-}(\overline{\theta})L^{T}(\overline{\theta}) \\ L(\overline{\theta}) & \mathbf{0}_{q\times q} \end{pmatrix} \begin{pmatrix} b(\hat{\theta}) \\ b(\hat{\lambda}) \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} J^{T}(\overline{\theta})\operatorname{Tr}\{H(\overline{\theta})\operatorname{Var}[\Delta^{(1)}]\} \\ \operatorname{Tr}\{K(\overline{\theta})\operatorname{Var}[\Delta^{(1)}]\} \end{pmatrix}. \tag{13}$$

4. Solution of the bias equations

As indicated in the introduction, three cases are distinguished.

4.1. Restricting constraints in regular models

When $M(\overline{\theta})$ is regular, the parameters θ have estimators also in the model without constraints, which means that the constraints just restrict somehow the parametric space Θ. In this case we have $M(\overline{\theta})^{-1} = M^{-}(\overline{\theta})$, and the first matrix on the left hand side of (13) can be inverted (cf. Silvey [9], p. 177), so $b(\hat{\theta})$ can be expressed explicitly:

$$
\begin{aligned}
b(\hat{\theta}) = & -\frac{1}{2} M(\overline{\theta})^{-1}(\mathbf{I}_p - L^T(\overline{\theta})T(\overline{\theta})L(\overline{\theta})M(\overline{\theta})^{-1}) \\
& \times J^T(\overline{\theta}) \operatorname{Tr}\{H(\overline{\theta}) \operatorname{Var}[\Delta^{(1)}]\} \\
& - \frac{1}{2} M(\overline{\theta})^{-1} L^T(\overline{\theta}) T(\overline{\theta}) \operatorname{Tr}\{K(\overline{\theta}) \operatorname{Var}[\Delta^{(1)}]\}
\end{aligned} \tag{14}
$$

where $T(\overline{\theta}) = (L(\overline{\theta})M(\overline{\theta})^{-1}L^T(\overline{\theta}))^{-1}$. Notice that in the regular model $\operatorname{Var}[\Delta^{(1)}] = \sigma^2 M(\overline{\theta})^{-1}(\mathbf{I}_p - L^T(\overline{\theta})T(\overline{\theta})L(\overline{\theta})M(\overline{\theta})^{-1})$ (cf. Silvey [9], p. 177). Hence the first additive term in (14) corresponds to the known formula of Box [1] which has been derived for the regular model without constraints (cf. also Cox and Snell, [2]).

4.2. Identifying constraints in singular models

In some cases, when $M(\theta)$ is singular of a constant rank $r < p$, one introduces $p - r$ constraints, which serves just to identify the parameters. So let us suppose that $q = p - r$. We recall that $\operatorname{rk}\begin{pmatrix} J(\theta) \\ L(\theta) \end{pmatrix} = p$. One can find r linearly independent rows of $M(\overline{\theta})$ which are linearly independent of the rows of the matrix $L(\overline{\theta})$. Let us denote by $\psi(\overline{\theta})$ the matrix formed from these rows. Because of independence, one can always find a positive definite matrix C such that we have the orthogonality relation

$$\psi(\overline{\theta}) C L^T(\overline{\theta}) = \mathbf{0}_{r\times q}.$$

This means that there exists a g-inverse $M^{-}(\overline{\theta})$ such that $M(\overline{\theta})M^{-}(\overline{\theta})L^T(\overline{\theta}) = \mathbf{0}_{p\times q}$ since $M(\overline{\theta})M^{-}(\overline{\theta})$ is a projector onto $\mathcal{M}[M(\overline{\theta})]$, the column space of $M(\overline{\theta})$, and since one can find a g-inverse such that $M(\overline{\theta})M^{-}(\overline{\theta})$ is an C-orthogonal projector (cf. Rao and Mitra [10], lemma 5.3.1).

From Equation (13) we obtain

$$\begin{pmatrix} M(\overline{\theta}) \\ L(\overline{\theta}) \end{pmatrix} b(\hat{\theta}) = -\frac{1}{2} \begin{pmatrix} J^T(\overline{\theta}) \operatorname{Tr}\{H(\overline{\theta}) \operatorname{Var}[\Delta^{(1)}]\} \\ \operatorname{Tr}\{K(\overline{\theta}) \operatorname{Var}[\Delta^{(1)}]\} \end{pmatrix}.$$

Premultiplying it by $(M(\overline{\theta}) + L^T(\overline{\theta})L(\overline{\theta}))^{-1}\,(\,\mathbf{I}_p, \quad L^T(\overline{\theta})\,)$ we obtain

Proposition 2. *When* $\mathrm{rk}(J(\overline{\theta})) + \mathrm{rk}(L(\overline{\theta})) = p$ *and* $L(\overline{\theta})$ *is of full rank, then*

$$
\begin{aligned}
(15) \qquad b(\hat{\theta}) = \; & -\frac{1}{2}(M(\overline{\theta}) + L^T(\overline{\theta})L(\overline{\theta}))^{-1} \\
& (J^T(\overline{\theta})\,\mathrm{Tr}\{H(\overline{\theta})\,\mathrm{Var}[\Delta^{(1)}]\} + L^T(\overline{\theta})\,\mathrm{Tr}\{K(\overline{\theta})\,\mathrm{Var}[\Delta^{(1)}]\}).
\end{aligned}
$$

Here (cf. Silvey [9])

$$
\mathrm{Var}[\Delta^{(1)}] = \sigma^2\{Q(\overline{\theta})^{-1} - Q(\overline{\theta})^{-1}L(\overline{\theta})^T[L(\overline{\theta})Q(\overline{\theta})^{-1}L(\overline{\theta})^T]^{-1}L(\overline{\theta})Q(\overline{\theta})^{-1}\}
$$

with $Q(\overline{\theta}) = M(\overline{\theta}) + L(\overline{\theta})^T L(\overline{\theta})$.

We note that the bias in singular models without constraints has been considered in Pázman [6].

4.3. Restricting and identifying constraints in singular models

This is a more general case, which encompasses the two previous ones. Nevertheless it was worthwhile considering them because they give simpler solutions.

Again, the matrix $M(\theta)$ is supposed to be singular with a constant rank r. Equation (13) can be written in the form

$$
\begin{aligned}
(16) \qquad & \begin{pmatrix} M(\overline{\theta})M^-(\overline{\theta}) & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} M(\overline{\theta}) & L^T(\overline{\theta}) \\ L(\overline{\theta}) & \mathbf{0}_{q\times q} \end{pmatrix} \begin{pmatrix} b(\hat{\theta}) \\ b(\widehat{\lambda}) \end{pmatrix} \\
& = -\frac{1}{2}\begin{pmatrix} J^T(\overline{\theta})\,\mathrm{Tr}\{H(\overline{\theta})\,\mathrm{Var}[\Delta^{(1)}]\} \\ \mathrm{Tr}\{K(\overline{\theta})\,\mathrm{Var}[\Delta^{(1)}]\} \end{pmatrix}.
\end{aligned}
$$

The first of these matrices is singular. Hence it is not possible to identify all components of $b(\hat{\theta})$ and $b(\widehat{\lambda})$. We will show, however, that the vector $b(\hat{\theta})$ will be obtained in a unique way. To this purpose we modify (16). We use geometrical arguments. Since $\mathrm{rk}\,(\,M(\overline{\theta}), \quad L^T(\overline{\theta})\,) = p$ and $\mathrm{rk}(M(\overline{\theta})) = r$, $\mathrm{rk}(L^T(\overline{\theta})) = q$, we obtain that the dimension of the linear space $\mathcal{M}[L^T(\overline{\theta})] \cap \mathcal{M}[M(\overline{\theta})]$ is equal to $(q + r) - p := s$. Denote by $L^T_{re}(\overline{\theta})$ the $s \times p$ matrix with rows equal to a linear basis of this space. Denote further by $L_{\mathrm{id}}(\overline{\theta})$ arbitrary $p - r$ rows of $L(\overline{\theta})$ which are linearly independent of the rows of $M(\overline{\theta})$, and let $L^*(\overline{\theta}) := (L^T_{\mathrm{id}}(\overline{\theta}), L^T_{re}(\overline{\theta}))^T$.

We put $L^*(\overline{\theta})$ instead of $L(\overline{\theta})$ into the bias equation. This substitution can be interpreted as a change of the original constraints $\varphi(\theta) = \mathbf{0}_q$ to equivalent constraints $\Psi(\theta) = \mathbf{0}_q$ such that $\frac{\partial\Psi(\theta)}{\partial\theta^T} = L^*(\theta)$. The relation between the two sets of constraints is linear: $\Psi(\theta) = L^*(\overline{\theta})L^T(\overline{\theta})[L(\overline{\theta})L^T(\overline{\theta})]^{-1}\varphi(\theta)$. Further, we have a freedom in the choice of the g-inverse $M^-(\overline{\theta})$. Since the columns of $L^T_{\mathrm{id}}(\overline{\theta})$ and of $M(\overline{\theta})$ are linearly independent, one can choose $M^-(\overline{\theta})$ such that $M(\overline{\theta})M^-(\overline{\theta})L^T_{\mathrm{id}}(\overline{\theta}) = \mathbf{0}_{p\times(p-r)}$. Finally, since $M(\overline{\theta})M^-(\overline{\theta})$ is a projector onto $\mathcal{M}[M(\overline{\theta})]$, one has $M(\overline{\theta})M^-(\overline{\theta})L^T_{re}(\overline{\theta}) = L^T_{re}(\overline{\theta})$, according to the definition of $L^T_{re}(\overline{\theta})$. Consequently, from Equation (13) we obtain

$$\begin{pmatrix} M(\overline{\theta}) & \mathbf{0}_{p\times(p-r)} & L^T_{re}(\overline{\theta}) \\ L^*(\overline{\theta}) & \mathbf{0}_{q\times(p-r)} & \mathbf{0}_{q\times(q+r-p)} \end{pmatrix} \begin{pmatrix} b(\hat{\theta}) \\ b(\widehat{\lambda}^*_{\mathrm{id}}) \\ b(\widehat{\lambda}^*_{\mathrm{re}}) \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} J^T(\overline{\theta})\,\mathrm{Tr}\{H(\overline{\theta})\,\mathrm{Var}[\Delta^{(1)}]\} \\ \mathrm{Tr}\{K^*(\overline{\theta})\,\mathrm{Var}[\Delta^{(1)}]\} \end{pmatrix},$$

which gives

$$\begin{pmatrix} M(\overline{\theta}) & L^T_{re}(\overline{\theta}) \\ L^*(\overline{\theta}) & \mathbf{0}_{q\times(q+r-p)} \end{pmatrix} \begin{pmatrix} b(\hat{\theta}) \\ b(\widehat{\lambda}^*_{re}) \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} J^T(\overline{\theta})\,\mathrm{Tr}\{H(\overline{\theta})\,\mathrm{Var}[\Delta^{(1)}]\} \\ \mathrm{Tr}\{K^*(\overline{\theta})\,\mathrm{Var}[\Delta^{(1)}]\} \end{pmatrix} \tag{17}$$

where $K^*(\theta) = \frac{\partial L^*(\theta)}{\partial \theta}$ and λ^*_{id} and λ^*_{re} are the modified Lagrangian multipliers resulting from the modification of $L(\theta)$ to $L^*(\theta)$.

Lemma 3.
a) *The matrix* $M(\theta) + L^{*T}(\theta)L^*(\theta)$ *is nonsingular.*
b) *The matrix*

$$\begin{pmatrix} M(\theta) & L^{*T}(\theta) \\ L^*(\theta) & \mathbf{0}_{q\times q} \end{pmatrix}$$

is nonsingular.
c) *The matrix*

$$D := \begin{pmatrix} M(\overline{\theta}) & L^{*T}_{re}(\overline{\theta}) \\ L^*(\overline{\theta}) & \mathbf{0}_{q\times(q+r-p)} \end{pmatrix}$$

is of the full rank.

Proof. a) If we have $x^T(M(\theta) + L^{*T}(\theta)L^*(\theta))x = 0$ for some $x \neq \mathbf{0}_p$, then $x^T M(\theta)x = x^T L^{*T}(\theta)L^*(\theta)x = 0$. This implies that $M(\theta)x = \mathbf{0}_p$ and $L^*(\theta)x = \mathbf{0}_q$ and consequently $\begin{pmatrix} M(\theta) \\ L^*(\theta) \end{pmatrix} x = \begin{pmatrix} M(\theta)x \\ L^*(\theta)x \end{pmatrix} = \mathbf{0}_{p+q}$, which is not possible since the matrix $\begin{pmatrix} M(\theta) \\ L^*(\theta) \end{pmatrix}$ is of the full rank.

b) We can verify that the inverse of the matrix has the form

$$\begin{pmatrix} A & B^T \\ B & C \end{pmatrix}$$

where

$$\begin{aligned} B &= (L^*(\theta)(M(\theta) + L^{*T}(\theta)L^*(\theta))^{-1}L^{*T}(\theta))^{-1}L^*(\theta)(M(\theta) + L^{*T}(\theta)L^*(\theta))^{-1} \\ A &= (M(\theta) + L^{*T}(\theta)L^*(\theta))^{-1}(\mathbf{I}_p - L^{*T}(\theta)B) \\ C &= \mathbf{I}_q - (L^*(\theta)L^{*T}(\theta))^{-1}L^*(\theta)(M(\theta) + L^{*T}(\theta)L^*(\theta))B^T. \end{aligned}$$

These three matrices exist due to the statement a).

c) Columns of the matrix D are columns of the matrix considered in b), so they are linearly independent. $\square$

Note that further details on the matrices involved are in Pázman and Denis [7].

Theorem 4. *Equation* (17) *has a unique solution.*

Proof. Due to Proposition (3) we can premultiply both sides of the equation by $(D^T D)^{-1} D^T$ and obtain the solution of (17). The solution is unique, since the matrix D is of the full rank. $\square$

5. Bias of functions of parameters

Proposition 5. *Let $h(\theta)$ be any twice differentiable parameter function. The approximate bias of $h(\widehat{\theta})$ is given by*

$$b(h) = \frac{\partial h(\overline{\theta})}{\partial \theta^T} b(\hat{\theta}) + \frac{1}{2} \operatorname{tr}\Big(\frac{\partial^2 h(\overline{\theta})}{\partial \theta \partial \theta^T} \operatorname{Var}(\Delta^{(1)}) \Big). \tag{18}$$

Proof. Consider the Taylor formula

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

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

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

$$y = \nu(\beta) + \varepsilon, \qquad \kappa(\beta) = 0 \tag{19}$$

where by definition $\nu(\beta) := \eta[\theta(\beta)]$, $\kappa(\beta) := \varphi[\theta(\beta)]$. Let $h(\theta)$ be a given parametric function. Let $l(\beta) := h[\theta(\beta)]$, and denote by $b_{\overline{\beta}}(l)$ the approximative bias of $l(\widehat{\beta})$ in Model (19).

Theorem 6. *The approximate bias is parametrically invariant, i.e.*

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

P r o o f. Denote by $b_{\bar\theta}$ and $b_{\bar\beta}$ the solutions of the bias equations in Model (1) and Model (19), respectively. First of all it is necessary to verify that the solutions of the two bias equations are related by (18), i.e. to prove that

$$(b_{\bar\beta})_i = \frac{\partial \beta_i}{\partial \theta} b_{\bar\theta} + \frac{1}{2}\operatorname{tr}\Big(\frac{\partial^2 \beta_i}{\partial\theta\partial\theta^T} V^{(\bar\theta)}\Big) \tag{20}$$

where $V^{(\bar\theta)} = \mathrm{Var}[\Delta^{(1)}]$. Similarly we define $V^{(\bar\beta)}$ in Model (19). For the sake of brevity we write $\frac{\partial\beta}{\partial\theta}$ and $\frac{\partial\theta}{\partial\beta}$ instead of $\frac{\partial\beta(\theta)}{\partial\theta}|_{\theta=\bar\theta}$ and $\frac{\partial\theta(\beta)}{\partial\beta}|_{\beta=\overline{\beta}}$, etc. We have the following evident rules:

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

We put (20) into the bias equation of Model (19), apply the transformation rules presented and as a result we obtain the bias equation of Model (1). This proves that (20) is correct. According to (18)

$$b_{\bar\theta}(h) = \frac{\partial h}{\partial\theta^T} b_{\bar\theta} + \frac{1}{2}\operatorname{tr}\Big(\frac{\partial^2 h}{\partial\theta\partial\theta^T}\Big) V^{(\bar\theta)} \tag{21}$$

and

$$b_{\bar\beta}(l) = \frac{\partial l}{\partial\beta^T} b_{\bar\beta} + \frac{1}{2}\operatorname{tr}\Big(\frac{\partial^2 l}{\partial\beta\partial\beta^T}\Big) V^{(\bar\beta)}. \tag{22}$$

From $l(\beta) = h[\theta(\beta)]$ we obtain the derivatives of $l(\beta)$ in a standard way. Inserting them together with (20) into (22) we obtain the right-hand side of (21). Hence $b_{\bar\theta}(h) = b_{\bar\beta}(l)$. □

Appendix

We show here that the bias equations can be obtained without assuming normal errors. We prefered to present the case with normal errors in the main text because of having the possibility to refer to well established results by Silvey [8, 9]. Since now the LS estimators are no more equal to the ML estimators, one cannot use the results of Silvey [8, 9], and one has to proceed differently.

Moreover, instead of considering approximations for large N, we consider them for small σ, which is equivalent in our case.

Let us restart our investigation with Model (1) but writing y and ε instead of y^* and ε^*, i.e.

$$\begin{aligned} y &= \eta(\theta) + \varepsilon \\ E[\varepsilon] &= \mathbf{0}_n; \quad \mathrm{Var}[\varepsilon] = \sigma^2 \mathbf{I}_n \\ \varphi(\theta) &= \mathbf{0}_q. \end{aligned}$$

We make a supplementary assumption on the parameter space Θ and on $\eta(\theta)$, namely we suppose that $\eta(\theta)$ is differentiable also on the boundary of the compact set Θ, and that the errors are sufficiently small to neglect the probability that the estimator $\hat\theta$ is on the boundary of Θ. (In fact it also means that $\eta(\theta)$ is defined on an open neighborhood of Θ). The LS estimator $\widehat{\theta}$ is still defined by Equation (4), hence the normal equations for $\hat\theta$ are still given by (5).

Lemma 7. *We have*

$$\begin{aligned} \widehat{\lambda} = \left[L(\widehat{\theta}) L^T(\widehat{\theta})\right]^{-1} L(\widehat{\theta}) J^T(\widehat{\theta}) \left[y - \eta(\widehat{\theta})\right] &= O_p(\sigma) \\ y - \eta(\widehat{\theta}) &= O_p(\sigma). \end{aligned}$$

P r o o f. Multiply (5) from the left by $L(\widehat{\theta})$ to obtain

$$\widehat{\lambda} = A(\widehat{\theta}) \left[y - \eta(\widehat{\theta})\right]$$

with $A(\theta) := [L(\theta)L^T(\theta)]^{-1} L(\theta) J^T(\theta)$. Since $A(\theta)$ is continuous in θ and Θ is compact, we have that $A(\theta)$ is bounded. Consequently,

$$\|\widehat{\lambda}\|^2 \leqslant c_1^2 \|y - \eta(\widehat{\theta})\|^2 \leqslant c_1^2 \|\varepsilon\|^2$$

for some $c_1 \in \mathbb{R}$. The last inequality follows from (4). From the Markov inequalities from probability theory, for any $c_2 \in \mathbb{R}$ we have

$$P\left\{\left\|\frac{\widehat{\lambda}}{\sigma}\right\|^2 > c_2^2\right\} \leqslant P\{\|\varepsilon\|^2 > \frac{c_2^2 \sigma^2}{c_1^2}\} \leqslant n \frac{c_1^2}{c_2^2}.$$

Hence for any $\delta > 0$ there is $C \in R$ such that $P\left\{\left\|\frac{\widehat{\lambda}}{\sigma}\right\| > C\right\} < \delta$. In symbols $\widehat{\lambda} = O_p(\sigma)$. We proceed similarly for $y - \eta(\widehat{\theta})$. $\square$

Corollary 8.

$$\widehat{\theta}\left[\eta(\overline{\theta})\right] = \overline{\theta}; \qquad \widehat{\lambda}\left[\eta(\overline{\theta})\right] = 0.$$

The left hand side of (5) can be regarded as a function of $\widehat{\theta}$, $\widehat{\lambda}$, y, and will be denoted briefly by $F\big(\widehat{\theta},\widehat{\lambda},y\big)$. So $\widehat{\lambda}(y)$, $\widehat{\theta}(y)$ are implicitly defined by the equation

$$.F(\theta,\lambda,y)|_{\substack{\theta=\widehat{\theta}(y)\\ \lambda=\widehat{\lambda}(y)}}=0.$$

This holds in a neighborhood $\mathcal{U}$ of the point $\eta(\bar{\theta})$, since evidently $F(\bar{\theta},0,\eta(\bar{\theta}))=0$. By the implicit function theorem we can write for any $y\in\mathcal{U}$

$$\begin{pmatrix}\frac{\partial\widehat{\theta}}{\partial y^T}\\ \frac{\partial\widehat{\lambda}}{\partial y^T}\end{pmatrix}=-\Big\{\Big(\frac{\partial F(\theta,\lambda,y)}{\partial\theta^T},\frac{\partial F(\theta,\lambda,y)}{\partial\lambda^T}\Big)^{-1}\frac{\partial F(\theta,\lambda,y)}{\partial y^T}\Big\}_{\substack{\theta=\widehat{\theta}(y)\\ \lambda=\widehat{\lambda}(y)}}. \tag{23}$$

When we derive this with respect to y, we obtain expressions for $\partial^2\widehat{\theta}/\partial y_i\partial y_j$, $\partial^2\widehat{\lambda}/\partial y_i\partial y_j$, etc. One can obtain in a straightforward way that

$$\begin{aligned}\Big(\frac{\partial F}{\partial\theta^T},\frac{\partial F}{\partial\lambda^T}\Big)&=\begin{pmatrix}M(\theta)+\frac{\partial J^T(\theta)}{\partial\theta^T}[\eta(\theta)-y]+\frac{\partial L^T(\theta)}{\partial\theta^T}\lambda, & L^T(\theta)\\ L(\theta) & \mathbf{0}\end{pmatrix}\\ &:=G(\theta,\lambda,y),\\ \frac{\partial F}{\partial y^T}&=\begin{pmatrix}-J^T(\theta)\\ \mathbf{0}\end{pmatrix}:=-Z(\theta).\end{aligned}$$

Hence

$$\begin{aligned}&\begin{pmatrix}\frac{\partial^2\widehat{\theta}}{\partial y_i\partial y_j}\\ \frac{\partial^2\widehat{\lambda}}{\partial y_i\partial y_j}\end{pmatrix}=G_{i.}^{-1}(\theta,\lambda,y)\times\\ &\Big\{-\Big[\frac{\partial G}{\partial\theta^T}\frac{\partial\widehat{\theta}}{\partial y_j}+\frac{\partial G}{\partial\lambda^T}\frac{\partial\widehat{\lambda}}{\partial y_j}+\frac{\partial G}{\partial y_j}\Big]G^{-1}(\theta,\lambda,y)Z(\theta)+\frac{\partial Z(\theta)}{\partial\theta^T}\frac{\partial\widehat{\theta}}{\partial y_j}\Big\}_{\substack{\theta=\widehat{\theta}(y)\\ \lambda=\widehat{\lambda}(y)}}.\end{aligned} \tag{24}$$

Evidently

$$G(\overline{\theta},0,\eta(\overline{\theta}))=\begin{pmatrix}M(\overline{\theta}) & L^T(\overline{\theta})\\ L(\overline{\theta}) & \mathbf{0}\end{pmatrix}.$$

Since Θ is compact and the matrix functions used are continuous, we obtain from Lemma 7

$$G\big(\widehat{\theta},\widehat{\lambda},y\big)=\begin{pmatrix}M\big(\widehat{\theta}\big)+O_p(\sigma), & L^T\big(\widehat{\theta}\big)\\ L\big(\widehat{\theta}\big) & \mathbf{0}\end{pmatrix}. \tag{25}$$

Lemma 9. *We have*

$$\widehat{\theta}-\overline{\theta}=O_p(\sigma).$$

P r o o f. By the Taylor formula we have

$$\widehat{\theta}_i-\overline{\theta}_i=\big(\partial\widehat{\theta}_i/\partial y^T\big)_{y=y^*}\varepsilon$$

with $\varepsilon = y - \eta(\bar{\theta})$, where $y^{(*)}$ is a point between y and $\eta(\bar{\theta})$. Hence from (23) and (25) we obtain

$$\widehat{\theta} - \overline{\theta} = S_\sigma\big(\widehat{\theta}, y^{(*)}\big)\varepsilon \tag{26}$$

where

$$S_\sigma\big(\widehat{\theta}, y^{(*)}\big) := \big(\, M\big(\widehat{\theta}\big) + O_p(\sigma), \quad L^T\big(\widehat{\theta}\big)\,\big) \begin{pmatrix} J^T\big(\widehat{\theta}\big) \\ \mathbf{0} \end{pmatrix}.$$

Since Θ is compact and the matrix functions considered are continuous, we see that for every $\delta > 0$ there is a σ_0 such that for $\sigma < \sigma_0$

$$\mathcal{P}\{y\colon\ S_\sigma\big(\widehat{\theta}(y), y\big) \text{ is bounded}\} > 1 - \delta.$$

Consequently (26) implies that $\|\widehat{\theta} - \overline{\theta}\|^2/\sigma^2$ is bounded for any given probability < 1, if σ is sufficiently small. Hence $\widehat{\theta} - \overline{\theta} = O_p(\sigma)$. $\square$

Now, let us define

$$\widehat{\theta}^{(2)} = \overline{\theta} + \frac{\partial\widehat{\theta}}{\partial y^T}\Big|_{y=\eta(\overline{\theta})}[y - \eta(\bar{\theta})],$$

$$\widehat{\lambda}^{(2)} = \frac{\partial\widehat{\lambda}}{\partial y^T}\Big|_{y=\eta(\overline{\theta})}[y - \eta(\bar{\theta})].$$

Evidently $E_{\overline{\theta}}\big[\widehat{\theta}^{(2)}\big] = 0$, $E_{\overline{\theta}}\big[\widehat{\lambda}^{(2)}\big] = 0$. We shall show that also other properties are similar to those of the asymptotic approximations $\widehat{\theta}^{(1)}$, $\widehat{\lambda}^{(1)}$ considered in the main text.

Lemma 10. *We have*

(a) $\quad \widehat{\theta} - \widehat{\theta}^{(2)} = o_p(\sigma)$

(b) $\quad \widehat{\lambda} - \widehat{\lambda}^{(2)} = o_p(\sigma)$

(c) $\quad \operatorname{Cov}\big[\widehat{\theta}^{(2)}, \widehat{\lambda}^{(2)}\big] = 0$

(d) $\quad \operatorname{Var}\big[\widehat{\theta}^{(2)}\big] = \operatorname{Var}\big[\widehat{\Delta}^{(1)}\big]$ *as considered in the main text.*

P r o o f. From the Taylor formula we have

$$\widehat{\theta}_k(y) = \widehat{\theta}_k^{(2)}(y) + \varepsilon^T \frac{\partial^2\widehat{\theta}_k}{\partial y \partial y^T}\Big|_{y^{(\natural)}}\varepsilon$$

$$\widehat{\lambda}_l(y) = \widehat{\lambda}_l^{(2)}(y) + \varepsilon^T \frac{\partial^2\widehat{\lambda}_l}{\partial y \partial y^T}\Big|_{z^{(\natural)}}\varepsilon$$

for some points $y^{(\natural)}$, $z^{(\natural)} \in \mathbb{R}^n$ which are between y and $\eta(\overline{\theta})$. Hence to prove (a) and (b), it is sufficient to prove that $\partial^2\widehat{\theta}_k/\partial y_i \partial y_j$, $\partial^2\widehat{\theta}_l/\partial y_i \partial y_j$ given by (24) are bounded with a given probability when σ is sufficiently small. This is done essentially in the same way as in the proof of (9): we use (24), express $\partial G/\partial\theta^T$, $\partial G/\partial\lambda^T$, and bound with a required probability the terms of order $O_p(\sigma)$.

To prove (c) and (d) we write according to (23)

$$\begin{pmatrix} \widehat{\theta}^{(2)} - \overline{\theta} \\ \widehat{\lambda}^{(2)} \end{pmatrix} = \begin{pmatrix} M & L^T \\ L & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} J^T \\ \mathbf{0} \end{pmatrix} \varepsilon$$

where we omit $\overline{\theta}$ in terms like $M(\overline{\theta})$, etc. Hence

$$\begin{aligned} \begin{pmatrix} \operatorname{Var}\big[\widehat{\theta}^{(2)}\big] & \operatorname{Cov}\big[\widehat{\theta}^{(2)}, \widehat{\lambda}^{(2)}\big] \\ \operatorname{Cov}\big[\widehat{\theta}^{(2)}, \widehat{\lambda}^{(2)}\big] & \operatorname{Var}\big[\widehat{\lambda}^{(2)}\big] \end{pmatrix} &= \begin{pmatrix} M & L^T \\ L & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} M & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} M & L^T \\ L & \mathbf{0} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} RMR & RMQ \\ Q^T MR & Q^T MQ \end{pmatrix} \end{aligned}$$

where

$$\begin{pmatrix} R & Q \\ Q^T & S \end{pmatrix} := \begin{pmatrix} M & L^T \\ L & \mathbf{0} \end{pmatrix}^{-1} = \begin{pmatrix} M + L^T L & L^T \\ L & \mathbf{0} \end{pmatrix}^{-1}.$$

From Silvey [9], p. 178 we have

$$\begin{aligned} R &= U^{-1} - U^{-1}L^T[LU^{-1}L^T]^{-1}LU^{-1} \\ Q^T &= [LU^{-1}L^T]^{-1}LU^{-1} \end{aligned}$$

where $U := [M + L^T L]$. Evidently $RL^T = U$, hence $RMQ = R(M + L^T L)Q = RUQ = RL^T[LU^{-1}L^T]^{-1} = \mathbf{0}$. Further, $\operatorname{Var}\big[\widehat{\theta}^{(2)}\big] = RMR$, which is equal to $\operatorname{Var}\big[\widehat{\theta}^{(1)}\big]$ in the asymptotic investigation of Silvey [8]. □

Proposition 11. *Under the assumptions of Appendix the statement in Proposition* 1 *holds.*

P r o o f. We use Lemmas 7 and 9 to obtain (7) without refering to Silvey [8]. We set $\widehat{\theta}^{(2)}$ and $\hat{\lambda}^{(2)}$ instead of $\widehat{\theta}^{(1)}$ and $\hat{\lambda}^{(1)}$ and apply Lemma 10 to obtain (10). Then we proceed in the same way as in the main part of the paper. □

Acknowledgment. The first author is grateful to INRA, France for supporting his visits in Unité de Biométrie, Versailles.

References

[1] *M.J. Box*: Bias in nonlinear estimation. Journal of the Royal Statistical Society B *33* (1971), 171–201.

[2] *D.R. Cox and E.J. Snell*: A general definition of residuals. Journal of the Royal Statistical Society B *30* (1968), 248–275.

[3] *H. Cramér*: Mathematical Methods of Statistics, 13th edition. Princeton University Press, Princeton, 1974.

[4] *J.-B. Denis and A. Pázman*: Bias of LS estimators in nonlinear regression models with constraints. Part II: Biadditive models. Appl. Math. *44* (1999), 375–403.

[5] *A. Pázman*: Nonlinear statistical models. Kluwer Academic Publishers, Dordrecht, 1993.

[6] *A. Pázman*: Bias of the MLE in singular nonlinear regression models. To appear in Proc. of ProbaStat'98, Tatra-Mountains Math. Publ., Bratislava.

[7] *A. Pázman and J.-B. Denis*: Bias in nonlinear regression models with constrained parameters. Technical report n° 4. Unité de biométrie INRA, Versailles, 1997.

[8] *S.D. Silvey*: The Lagrangian multiplier test. Annals of Mathematical Statistics.

[9] *S.D. Silvey*: Statistical Inference, 3rd edition. Chapman and Hall, London, 1979.

[10] *C.R. Rao and S.K. Mitra*: Generalized inverse of matrices and its applications. John Wiley, New York, 1971.

Authors' addresses: *Andrej Pázman*, Dept. of Probability and Statistics, Faculty of Mathematics and Physics, Commenius University, Mlynská dolina, 842 15, Bratislava, Slovakia, `pazman@fmph.uniba.sk`; *Jean-Baptiste Denis*, Unité de Biométrie, INRA, F-78026 Versailles Cedex, France, `denis@versailles.inra.fr`.