Jour. Ind. Soc. Ag. Statistics
49 (Golden Jub. No.), 1996-97 : 277–288

Optimal Design for An Inverse Prediction Problem under Random Coefficient Regression Models

Erkki P. Liski, Arto Luoma, Nripes K. Mandal[1] and Bikas K. Sinha[2]
University of Tampere, Tampere, Finland

SUMMARY

In this paper we consider the problem of estimating the value of the non-stochastic regressor at which the response attains a pre-specified level. We assume some prior knowledge, based on previous experience, about the distribution of such a target point, and derive the optimum design for the estimation problem in the case of linear and quadratic regression under fixed and random coefficient models.

Key words : Linear regression, Quadratic regression, Optimum design.

1. Introduction

Let Y be a random response variable whose expected value $\eta(x)$ depends on the level of a non-stochastic regressor x. The problem considered here is the determination of the value of $x = x_0$ at which the expected response $\eta(x)$ attains a specified value η_0. In general, $\eta(x)$ is unknown. We assume that, in the region of interest, $\eta(x)$ can be approximated by a polynomial of a given degree p.

Marking for bucking is the problem of converting tree stems into smaller logs optimally according to a given utility function (Liski and Nummi [6]). Its solution is important for the planning and complishing of forest harvesting affecting directly the conditions for lumber production. An admissible cutting pattern is a set of stem points $0 \le x_1 < \ldots < x_k$ such that the logs satisfy the conditions

$$x_{k+1} - x_k \in [\, l_{min}, l_{max} \,] \text{ and } y_k \ge y_0$$

for $k = 1, 2, \ldots, K$, where x_1 is at the butt of a tree, l_{min} is the minimum and l_{max} the maximum length of an admissible log and y_0 is the minimum acceptable log diameter and y_k is the observed diameter at the point x_k. Then, in fact, x_k is the distance of the kth cutting point from the butt end of a stem. Now

1 University of Calcutta, Calcutta-700019
2 Indian Statistical Institute, Calcutta-700035

$E(Y \mid x) = \eta(x)$ is the expected diameter value at the stem height x. We assume here that $\eta(x)$ is a continuous non-increasing function of x and $\eta(x) = 0$ for $x \geq x_H$, where x_H is the height of a tree. If y_0 is attained at the point x_0, then the length of the exploitable part of a tree is x_0. It is of great practical importance to know the height at which the diameter attains the value y_0. For example, the number of logs K depends on x_0. We assume that there are data available on previously measured trees, which can be utilized when estimating x_0 on current trees. Other applications of inverse prediction are considered in Liski and Nummi ([5], [7]). The extensive calibration literature contains further applications (see Osborne [8]).

When predicting future observations on a statistical unit, as on a tree stem, given past measurements on the same and other similar units, it may not be optimal to use all available data. Then the question arises: How to select the best subset of observations? Rao [9] considered one aspect of this problem in considering prediction under growth curve models. He observed that the required information for forecasting is often contained in the previous few measurements. Liski and Nummi [4] made the same finding when studying the behaviour of their growth curve predictor. One opportunity is to borrow strength in estimation by incorporating data from similar units or parallel circumstances (see Brillinger [1]). Observations may also be available on some explanatory variables, which make it possible to find "similar" units. Another approach is to use given measurements (the initial measurement, selection index) as a covariate, as in Liski and Nummi ([4], [5], [6]). These authors also considered the problem of selecting the estimation set or training set, which is the term used in neural network literature (Liski and Nummi [5], [6]).

Let y_a be the observation at the initial value $x = a$, say, for the tree under consideration. It could be any available diameter measurement like the diameter at the breast height ($a = 1.3$ m), which is a standard measurement in practice. Then we can select a subset of trees whose Y values at $x = 0$ fall in $(y_a \pm \varepsilon)$. In this paper we restrict ourselves to a consideration of this simple way of selecting a subset. Note that the theory can be straightforwardly generalized for more general constraints when selecting a subset. For the fixed effects model, using the usual least squares predictor (LSP), we try to find the optimum design in Section 2. A random effects regression model will be considered in Section 3.

2. *Fixed Regression Coefficient Model*

In this section, we consider a fixed effects regression model.

2.1 *Linear regression*

Let us start with a linear regression set-up :

$$Y_x = \beta_0 + \beta_1 x + e_x \tag{2.1}$$

where Y_x is a random response variable with $E(Y \mid x) = \beta_0 + \beta_1 x$, x is a non-stochastic regressor, β_0, β_1 are fixed regression coefficients and e_x is a random error with mean 0 and variance σ_e^2. We further assume that x varies in some domain χ. The problem is to find the value of $x = x_0$ for which $E(Y_x \mid x = x_0) = \eta(x_0) = \eta_0$, a specified value.

If β_0 and β_1 are known, then we can automatically get the value of x_0 by

$$x_0 = (\eta_0 - \beta_0)/\beta_1$$

Since β_0 and β_1 are unknown, we replace them by their least squares estimates (LSE) to obtain

$$\hat{x}_0 = (\eta_0 - \hat{\beta}_0)/\hat{\beta}_1 \tag{2.2}$$

Since x_0 is a non-linear function of β_0 and β_1, any measure of the accuracy of x_0 will depend on unknown parameters. Here we assume that, from past experiences, we have some prior knowledge regarding x_0. For the linear case we require only the first two moments of x_0:

$$\varepsilon(x_0) = \mu \text{ and } \nu(x_0) = v \tag{2.3}$$

From the practical point of view, it appears that the response values can be made available at the beginning of the range of values of the regressor x. Thus, specifically, we will take the experimental region to be

$$0 \le x \le h \tag{2.4}$$

Now let us make the following transformation :

$$x^* = (x - \mu)/v^{1/2} \text{ and } x_0^* = (x_0 - \mu)/v^{1/2} \tag{2.4a}$$

Then (2.1) reduces to

$$E(Y \mid x) = \beta_0 + \beta_1 (\mu + x^* v^{1/2})$$

$$= \beta_0^* + \beta_1^* x^* \tag{2.5}$$

where $\beta_0^* = \beta_0 + \beta_1 \mu$ and $\beta_1^* = \beta_1 v^{1/2}$. Moreover,

$$\mathcal{E}(x_0^*) = 0 \text{ and } \mathcal{V}(x_0^*) = 1 \tag{2.6}$$

and the experimental region (2.4) reduces to

$$l \leq x^* \leq u, \; l = -\frac{\mu}{\sqrt{v}}, \; u = \frac{h - \mu}{\sqrt{v}} \tag{2.7}$$

In the following, we work with the set-up (2.5)-(2.7) and drop the asterisks in x^*, x_0^*, β_0^* and β_1^* throughout.

Since x_0 given by (2.2) is non-linear in the parameters, we apply the δ – method and a large sample approximation to find $V(\hat{x}_0)$:

$$V(\hat{x}_0) \cong \left(\frac{\partial x_0}{\partial \beta_0} \; \frac{\partial x_0}{\partial \beta_1}\right) D(\hat{\beta}) \left(\frac{\partial x_0}{\partial \beta_0} \; \frac{\partial x_0}{\partial \beta_1}\right)' \tag{2.8}$$

where $D(\hat{\beta})$ is the dispersion matrix of $\hat{\beta}$, the LSE of β_0 and β_1 based on the observations $y_1, y_2, \ldots, y_n$ corresponding to the n-points $x = (x_1, x_2, \ldots, x_n)'$.

The problem is that of finding n points in (2.7) in order to estimate $\hat{x}_0$ most efficiently. However, for obvious reasons, we work with the continuous (approximate) design set-up (c.f. Kiefer [3], Fedorov [2], Silvey [10]). It is easy to see that

$$\left(\frac{\partial x_0}{\partial \beta_0}, \frac{\partial x_0}{\partial \beta_1}\right) = -\frac{1}{\beta_1}(1, x_0)$$

Hence (2.8) reduces to

$$V(\hat{x}_0) \cong \frac{1}{\beta_1^2}(1, x_0) D(\hat{\beta}) \begin{pmatrix} 1 \\ x_0 \end{pmatrix} \tag{2.9}$$

Since $V(\hat{x}_0)$ given by (2.9) involves unknown x_0, we take

$$\mathcal{E}[V(\hat{x}_0)] \cong \frac{1}{\beta_1^2} \mathcal{E} \operatorname{tr} D(\hat{\beta}) \begin{pmatrix} 1 \\ x_0 \end{pmatrix} (1, x_0) \tag{2.10}$$

as a criterion for comparing different designs. Now we have

$$D(\hat{\beta}) = \sigma_e^2 (X'X)^{-1} \tag{2.10a}$$

where $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$. Using (2.6), it can be shown that (2.10) reduces to

$$\mathcal{E}[V(\hat{x}_0)] = \frac{\sigma_e^2}{\beta_1^2 n}\left\{1 + \frac{1}{\mu'_2 - \mu'^2_1}[1 + \mu'^2_1]\right\}$$

where $\mu'_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\mu'_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$. For fixed n, it is equivalent to minimizing

$$1 + \frac{1}{\mu'_2 - \mu'^2_1}[1 + \mu'^2_1] = \frac{1 + \mu'_2}{\mu'_2 - \mu'^2_1}$$

The minimum is attained when $\mu_2 = \mu'_2 - \mu'^2_1$ is maximized and μ'_2 is minimized. It turns out that to achieve the above objective, observations are to be taken at the extremes, i.e. at $x^* = l$ and at $x^* = u$. In terms of the original factor space (2.4), the optimum design consists of the points at the two extremes of the factor space (2.4). The optimal distribution of total mass (under the approximate design set-up) at the extremes is given by : p (at zero) and $q = 1 - p$ (at h), where

$$p_{opt} = \frac{-(1+u^2) + \sqrt{(1+u^2)(1+l^2)}}{l^2 - u^2} \tag{2.11}$$

This result is obtained by routine work of differentiation regarding the objective function as a function of p. The minimum value of $\mathcal{E}[V(\hat{x}_0)]$ is given by

$$\left(\frac{\sigma_e^2}{n\beta_1^2}\right)\left(\frac{1 + \mu'_2}{\mu_2}\right)$$

evaluated at p_{opt} for the optimal two-point design.

Consider for example the experimental region [0,5] with $\mu = 20$, $v = 4$. Then in terms of the transformed variable we have $l = -10$, $u = -7.5$. The minimization problem yields $p_{opt} = 0.4295$. Therefore, the optimal design is given by :

Point :	0	5
Mass :	0.43	0.57

2.2 *Quadratic regression*

We now consider the quadratic regression (in the transformed factor space)

$$\eta(x) = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{2.12a}$$

with $\beta_2 < 0$. Then the problem is to estimate

$$x_0 = \{ -\beta_1 \pm [\beta_1^2 - 4\beta_2 (\beta_0 - \eta_0)]^{1/2} \} / 2\beta_2 \tag{2.12b}$$

for given η_0 with all the parameters in the model (2.12a) unknown. As before, we propose the "plug-in" estimate

$$\hat{x}_0 = \{ -\hat{\beta}_1 - [\hat{\beta}_1^2 - 4\hat{\beta}_2 (\hat{\beta}_0 - \eta_0)]^{1/2} \} / 2\hat{\beta}_2 \tag{2.13}$$

where $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are the ordinary least squares estimates of the regression coefficients. Here again $V(\hat{x}_0)$ will depend on the unknown parameters. Using (2.12b) it can be shown that

$$\left(\frac{\partial x_0}{\partial \beta_0}, \frac{\partial x_0}{\partial \beta_1}, \frac{\partial x_0}{\partial \beta_2} \right) = - \frac{1}{(\beta_1 + 2\beta_2 x_0)} (1, x_0, x_0^2)$$

and consequently

$$V(\hat{x}_0) \cong \frac{1}{(\beta_1 + 2\beta_2 x_0)^2} \mathbf{x}'_0 D(\hat{\beta}) \mathbf{x}_0 \tag{2.14}$$

where $\mathbf{x}_0 = (1, x_0, x_0^2)'$ and $\hat{\beta}$ denotes the LSE of $\beta = (\beta_0, \beta_1, \beta_2)'$. Let us write

$$D(\hat{\beta}) = \frac{\sigma^2}{n} \mathbf{M}^{-1}$$

where $\mathbf{M}$ is the moment matrix. Then from (2.14) we have

$$V(\hat{x}_0) \cong \frac{1}{(\beta_1 + 2\beta_2 x_0)^2} \frac{\sigma^2}{n} \mathbf{x}'_0 \mathbf{M}^{-1} \mathbf{x}_0$$

Thus, for a fixed n, we have to find an optimum design for which

$$\mathcal{E}[V(\hat{x}_0)] \cong \frac{1}{\mathcal{E}(\beta_1 + 2\beta_2 x_0)^2} \frac{\sigma^2}{n} \operatorname{tr} \mathbf{M}^{-1} \mathcal{E}\, \mathbf{x}_0 \mathbf{x}'_0 \tag{2.15}$$

attains its minimum. From (2.15) it is clear that we only need a priori assumption on x_0 and there is no need for any assumption on β_0, β_1 or β_2 separately. We

use the same assumption (2.3) as in Section 2.1. Besides, we require a further assumption on the third and the fourth moments of x_0. The purpose of the transformation (2.4 a) was to validate (2.6) and (2.7). In addition, we further assume

$$\mathcal{E}(x_0^{*3}) = 0, \quad \mathcal{E}(x_0^{*4}) = \tau \tag{2.16}$$

Hence (2.15) simplifies to

$$\frac{\sigma^2 L(\mathbf{M}^{-1})}{n(\beta_1^2 + 4\beta_2^2)} \tag{2.17}$$

where $$L(\mathbf{M}^{-1}) = \text{tr}\, \mathbf{M}^{-1} \mathcal{E}\, \mathbf{x}_0 \mathbf{x}_0' \tag{2.18}$$

Here again, for obvious reasons, instead of n-point designs we consider the class Ξ of approximate (continuous) designs.

Writing

$$\mathbf{M} = \begin{pmatrix} 1 & \mu_1' & \mu_2' \\ \mu_1' & \mu_2' & \mu_3' \\ \mu_2' & \mu_3' & \mu_4' \end{pmatrix} = ((s_{ij}))$$

and $\mathbf{M}^{-1} = ((s^{ij}))$, it turns out that (2.18) simplifies to

$$L(\mathbf{M}^{-1}) = \text{tr}\left[\begin{pmatrix} s^{00} & s^{02} \\ s^{20} & s^{22} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & \tau \end{pmatrix}\right] + s^{11} \tag{2.19}$$

Next observe that

$$\det(\mathbf{M}) = \mu_2' \mu_4' - \mu_3'^2 - \mu_2'^3 - \mu_4' \mu_1'^2 + 2\mu_1' \mu_2' \mu_3' \tag{2.20}$$

Further, $$s^{11} = \frac{\mu_4' - \mu_2'^2}{\det(\mathbf{M})} \tag{2.21}$$

and $$\begin{pmatrix} s^{00} & s^{02} \\ s^{20} & s^{22} \end{pmatrix} = \frac{1}{\det(\mathbf{M})}\begin{pmatrix} \mu_2'\mu_4' - \mu_3'^2 & \mu_1'\mu_3' - \mu_2'^2 \\ \mu_1'\mu_3' - \mu_2'^2 & \mu_2' - \mu_1'^2 \end{pmatrix} \tag{2.22}$$

Thus finally,

$$L(\mathbf{M}^{-1}) = \frac{1}{\det(\mathbf{M})}\left[(\mu_2'\mu_4' - \mu_3'^2) + 2(\mu_1'\mu_3' - \mu_2'^2) + \tau(\mu_2' - \mu_1'^2) + (\mu_4' - \mu_2'^2)\right] \tag{2.23}$$

where det (M) is given in (2.20).

To find the support points of the optimal design note that the criterion function $L(\mathbf{M}^{-1})$ given by (2.18) is a linear optimality criterion. The equivalence theorem for this criterion states that a design ξ_0 is optimum in the above sense *iff*

$$f'(\mathbf{x})\,\mathbf{M}_0^{-1}\,\mathbf{B}\mathbf{M}_0^{-1}\,f(\mathbf{x}) \leq \operatorname{tr}\mathbf{M}_0^{-1}\,\mathbf{B} \tag{2.24}$$

for all $\mathbf{x}$ with equality at the support points of the design (see Silvey [10] or Fedorov [2].)

Here $f(\mathbf{x}) = (1 \quad x \quad x^2)'$ and $\mathbf{M}_0$ corresponds to $\mathbf{M}$ for the optimal design. Now the left hand side of (2.24) is a quartic in x. Thus it can not have more than three maxima in [l, u] and if it has three, two of these must be at the extremes $x = l$ and $x = u$, since there must be a local minima between every two maxima. Hence this design cannot have more than three support points. But it must have at least three in order that the information matrix is nonsingular. It, therefore, has exactly three support points and two of them must be at the extremes. As to the choice of the intermediate point and the choice of probability mass at these three points, we may proceed as follows :

We start with the set-up

point:	l	m	u
mass:	p	r	q(= 1 – p – r)

so that

$$\mathbf{M}_0 = p\,f(\mathbf{l})\,f'(\mathbf{l}) + rf(\mathbf{m})\,f'(\mathbf{m}) + qf(\mathbf{u})\,f'(\mathbf{u})$$

Let $\gamma(\mathbf{l})$ be orthogonal to $f(\mathbf{m})$ and $f(\mathbf{u})$. Then it is readily seen that

$$\mathbf{M}_0^{-1} f(\mathbf{l}) = \frac{\gamma(\mathbf{l})}{pf'(\mathbf{l})\,\gamma(\mathbf{l})}$$

Similarly, with analogous properties of $\gamma(\mathbf{m})$ and $\gamma(\mathbf{u})$, we have

$$\mathbf{M}_0^{-1} f(\mathbf{m}) = \frac{\gamma(\mathbf{m})}{rf'(\mathbf{m})\,\gamma(\mathbf{m})}$$

and

$$\mathbf{M}_0^{-1} f(\mathbf{u}) = \frac{\gamma(\mathbf{u})}{qf'(\mathbf{u})\,\gamma(\mathbf{u})}$$

Now, an examination of "=" in (2.24) yields

$$p_{opt} = \frac{\sqrt{A_l}}{\sqrt{A_l} + \sqrt{A_m} + \sqrt{A_u}},\ r_{opt} = \frac{\sqrt{A_m}}{\sqrt{A_l} + \sqrt{A_m} + \sqrt{A_u}},\ q_{opt} = \frac{\sqrt{A_u}}{\sqrt{A_l} + \sqrt{A_m} + \sqrt{A_u}},$$

where

$$A_l = \frac{\gamma'(l)\, B\gamma(l)}{(f'(l)\,\gamma(l))^2}, \quad A_m = \frac{\gamma'(m)\, B\gamma(m)}{(f'(m)\,\gamma(m))^2}, \quad A_u = \frac{\gamma'(u)\, B\gamma(u)}{(f'(u)\,\gamma(u))^2}.$$

Finally, to determine m (only unknown quantity in the interior of the interval [l, u]), we obtain the following equation:

$$\sqrt{A_l} + \sqrt{A_m} + \sqrt{A_u} = \sqrt{\mathrm{tr}\,(M_0^{-1} B)} = \sqrt{L\,(M_0^{-1})}$$

Exact analytic solution for m is quite difficult. Extensive numerical computations indicate that m does not differ much from $\frac{l+u}{2}$. We, therefore, take the middle point viz., $\frac{l+u}{2}$ and derive the nature of optimal 3-point design. We adopt

point:	l	$\frac{l+u}{2}$	u
mass:	p	r	q(= 1 – p – r)

For given l, u and τ, we can now find p_{opt}, q_{opt} and r_{opt} from the expressions given above.

2.3 *Numerical computations and major findings*

Extensive numerical computations show that the optimal 3-point design is quite robust against possible variation in the values of τ. However, it depends on the extreme values of the reduced factor space. In the following table we show some results.

Table 1. Optimal weights p, r and q for some combinations of l, u and τ

		$\tau = 2$			$\tau = 3$		
l	u	p (0)	r (h / 2)	q (h)	p(0)	r(h /2)	q (h)
–2	–1	0.1804	0.5227	0.2969	0.1841	0.5215	0.2944
–3	–1	0.0600	0.5233	0.4167	0.0615	0.5234	0.4151
–3	–2	0.0843	0.7303	0.1854	0.0851	0.7282	0.1867
–4	–3.5	0.0685	0.8417	0.0898	0.0691	0.8403	0.0906
–5	–4.8	0.0803	0.8323	0.0874	0.0816	0.8307	0.0877

3. *Random Coefficients Regression Model*

3.1 *Linear regression*

Let us start with a linear structure

$$Y_x = (\beta_0 + b_0) + (\beta_1 + b_1)\, x + e_x \qquad (3.1)$$

where β_0 and β_1 are as before. The random variables b_0, b_1 and e_x are uncorrelated, with zero means and variances σ_0^2, σ_1^2 and σ_e^2 respectively, b_0 and b_1 represent random components associated with β_0 and β_1. Here the error term e_x is defined as $e_x = Y_x - E(Y_x \mid b_0, b_1)$. The problem is, as before, that of estimating the value of $x = x_0$ at which $E(Y_x) = \eta(x)$ attains the pre-specified value η_0.

Suppose that we have n observations $y_1, y_2, \ldots, y_n$ corresponding to n points $x_1, x_2, \ldots, x_n$ in same domain χ. The objective is to find an optimal design in χ so that we can estimate,

$$x_0 = (\eta_0 - \beta_0) / \beta_1$$

as accurately as possible. Here, again, we estimate x_0 by

$$\hat{x}_0 = (\eta_0 - \hat{\beta}_0) / \hat{\beta}_1$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the LSEs of β_0 and β_1, respectively. Now, in model (3.1), the dispersion matrix of $\mathbf{Y} = (y_1, y_2, \ldots, y_n)'$, the n observation vector corresponding to the n points $x_1, x_2, \ldots, x_n$ in χ, is given by

$$\Sigma = (\sigma_\varepsilon^2 \mathbf{I}_n + \sigma_0^2 \mathbf{J}_n) + \sigma_1^2 \mathbf{x}\mathbf{x}', \tag{3.2}$$

where $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}'_n$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$, $\mathbf{1}_n$ being the $n \times 1$ vector of ones. To solve the problem, as in the fixed effects case (Section 2), here also we make the same prior assumption (2.3) on x_0 and the same experimental region (2.4). We work with the transformed set-up (2.6)-(2.7). To compare the performance of the designs, we use $\mathcal{E}[V(\hat{x}_0)]$, where $V(\hat{x}_0)$ is given by (2.8) and $\mathcal{E}$ is the expectation operator with respect to the prior distribution of x_0. The main difference with the fixed effects model is that here

$$D(\hat{\beta}) = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \tag{3.3}$$

where Σ is given by (3.2) while in the fixed coefficient regression model it is given by (2.10a).

However, under the covariance structure (3.2), the ordinary least squares estimator (OLSE) and the GLSE are equal (c.f. Rao [9]). Thus their covariance matrices are also equal and (3.3) can be written as

$$D(\hat{\beta}) = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \Sigma \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \tag{3.4}$$

If we denote the covariance matrix of **b** by **D**, then $\Sigma = \mathbf{XDX}' + \sigma_\varepsilon^2 \mathbf{I}_n$, and (3.4) further simplifies to

$$D(\hat{\beta}) = \mathbf{D} + \sigma_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1} \tag{3.5}$$

Since **D** is constant with respect to the design matrix **X**, minimizing (2.10) is equivalent to minimizing

$$\text{tr } \mathbf{M}^{-1} \mathcal{E}_{\mathbf{x}_0 \mathbf{x}'_0}$$

and we see that the optimal design is the same as in the fixed model case. The same argument also remains valid when we adopt a random coefficient quadratic regression model. In that case we tactly assume that $L(\mathbf{M}^{-1})$ is invariant with respect to sign changes. This is explained in the APPENDIX. Thus we do *not* encounter any new problems under random coefficient models.

REFERENCES

[1] Brillinger, D.R., 1987. Comment on "Prediction of future observations in growth curve models" by C.R. Rao. *Statist. Sci.*, **2**, 448-450.

[2] Fedorov, V.V., 1972. *Theory of Optimal Experiments*. Academic Press, New York.

[3] Kiefer, J.C., 1974. General equivalence theory for optimum designs (approximate theory). *Ann. Statist.*, **2**, 849-879.

[4] Liski, Erkki P. and Nummi, Tapio, 1990. Prediction in growth curve models using the EM algorithm. *Comput. Statist. Data Anal.*, **10**, 99-108.

[5] Liski, Erkki P. and Nummi, Tapio, 1995a. Prediction and inverse estimation in repeated-measures models. *J. Statist. Plann. Inf.*, **47**, 141-151.

[6] Liski, Erkki P. and Nummi, Tapio, 1995b. Prediction of tree stems to improve efficiency in automatized harvesting of forests. *Scand. J. Statist.*, **22**, 255-269.

[7] Liski, Erkki P. and Nummi, Tapio, 1996. Prediction in repeated-measures models with engineering applications. *Technometrics*, **38**, 25-36.

[8] Osborne, Christine, 1991. Statistical calibration : A review. *Internat. Statist. Rev.*, **59**, 309-336.

[9] Rao, C.R., 1987. Prediction of future observations in growth curve models. *Statist. Sci.*, **2**, 434-471.

[10] Silvey, S.D., 1980. *Optimal Design*. Chapman and Hall, London.

APPENDIX

In Section 3.2 we assume that L (M) is invariant with respect to sign changes. To see this, note that, under the transformation $x \rightarrow -x$, all the expressions remain the same except b and e, which assume precisely the same absolute value but with the opposite sign. Moreover, because of the same transformation, x_0 also changes to $-x_0$ resulting in $\mathcal{E}(x_0^3) \rightarrow -\mathcal{E}(x_0^3)$. Thus to show that the criterion function

$$\text{tr} \begin{pmatrix} a & b & c \\ & d & e \\ & & f \end{pmatrix}^{-1} \mathcal{E}\, x_0 x'_0$$

with $\mathbf{x}_0 = (1, x_0, x_0^2)'$ remains invariant under the transformation $x \rightarrow -x$, it would be sufficient to show that

$$\text{tr} \begin{pmatrix} a & -b & c \\ & d & -e \\ & & f \end{pmatrix}^{-1} \mathcal{E}\, x_0 x'_0$$

$$= \text{tr} \begin{pmatrix} a & b & c \\ & d & e \\ & & f \end{pmatrix}^{-1} \mathcal{E}\, x_0 x'_0 \qquad \text{(A.1)}$$

But

$$\begin{pmatrix} a & -b & c \\ & d & -e \\ & & f \end{pmatrix} = \mathbf{I}_{(-2)} \begin{pmatrix} a & b & c \\ & d & e \\ & & f \end{pmatrix} \mathbf{I}_{(-2)}$$

where

$$\mathbf{I}_{(-2)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1' & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Moreover $\mathbf{I}_{(-2)}^2 = \mathbf{I}$ so that $\mathbf{I}_{(-2)}^{-1} = \mathbf{I}_{(-2)}$. Hence

$$\begin{pmatrix} a & -b & c \\ & d & -e \\ & & f \end{pmatrix}^{-1} = \mathbf{I}_{(-2)} \begin{pmatrix} a & b & c \\ & d & e \\ & & f \end{pmatrix}^{-1} \mathbf{I}_{(-2)} \qquad \text{(A.2)}$$

Substituting (A.2) into the first expression of (A.1) yields the result.