Tatra Mt. Math. Publ. **26** (2003), 195–199

Tatra Mountains
Mathematical Publications

VARIANCE COMPONENTS ESTIMABILITY IN THE EXTENDED GROWTH CURVE MODEL

Ivan Žežula

ABSTRACT. The extended growth curve model (multivariate linear model) with one fixed and several random effects is considered. Some important observations about the estimability of its variance components are presented. The results are applied to some special cases.

1. Introduction

There are different motivations leading to the general extended growth curve model (ECGM):

$$Y = \sum_{i=0}^{k-1} X_i B_i Z_i + \varepsilon\,,$$

$$\mathrm{E}\varepsilon = 0\,, \qquad \operatorname{var}(\operatorname{vec}\varepsilon) = \Sigma_k \otimes I\,,$$

$$\mathrm{E}B_i = 0\,, \quad \operatorname{var}(\operatorname{vec}B_i) = \Sigma_i \otimes I, \quad i = 1,\dots,k-1\,,$$

ε and $B_1,\dots,B_{k-1}$ being independent random variables (dimension of Y, ε, X_i, B_i and Z_i are $n\times p$, $n\times p$, $n\times m_i$, $m_i\times r_i$, and $r_i\times p$, respectively). We shall denote by P_G the orthogonal projector on column space $\mathcal{R}(G)$ of a matrix G and by $M_G = I - P_G$ the orthogonal projector on its orthogonal complement.

If there are more non-random terms, they can still be expressed in a single term using a structured (blocked) matrix B_0. We will not discuss complications arising in such situation.

Different motivations lead to different simplifications or additional assumptions to make the model tractable (see, e.g., [1], [4], [6], [7], [8], [10]).

Clearly,

$$\operatorname{var}(\operatorname{vec}Y) = \sum_{i=1}^{k-1} Z_i'\Sigma_i Z_i \otimes X_i X_i' + \Sigma_k \otimes I\,.$$

2000 Mathematics Subject Classification: Primary 62J10; Secondary 62H15.
Keywords: extended growth curve model, variance components, multivariate regression.

Supported by grant VEGA MŠ SR No. 1/7471/20.

In order to make inference in the model, we need to estimate this matrix as well as a whole and/or its parts, i.e., variance components.

Can we tell something about estimability of the variance components without additional assumptions?

2. Quadratic estimators

Since quadratic estimators are the optimal ones in the case of normal distribution and usually the only tractable ones in other cases, we shall restrict ourselves to the estimators of the form $Y'AY$, $A_{n\times n}$.

According to [2], we have

$$\mathrm{E}\, Y'AY = \sum_{i=1}^{k-1} Z_i'\Sigma_i Z_i \cdot \mathrm{Tr}(AX_iX_i') + \Sigma_k \cdot \mathrm{Tr}(A) + Z_0'B_0'X_0'AX_0B_0Z_0 .$$

In order to exclude unknown B_0, let us take $A = M_{(X_0,P)}$, where P is an another matrix. Thus,

$$\mathrm{E}\, Y'M_{(X_0,P)}Y = \sum_{i=1}^{k-1} Z_i'\Sigma_i Z_i \cdot \mathrm{Tr}\big(M_{(X_0,P)}X_iX_i'\big) + \Sigma_k \cdot \mathrm{Tr}\big(M_{(X_0,P)}\big) .$$

By choosing different matrices P we can obtain possibly different linear combinations of variance components. Estimators of their estimable functions can be obtained by solving the system of unbiased estimating equations

$$Y'M_{(X_0,P)}Y = \widehat{\mathrm{E}\, Y'M_{(X_0,P)}Y}$$

with the suitable different P's.

In particular, choosing $P = \emptyset,\ X_1,\ (X_1,X_2),\ \dots,\ (X_1,X_2,\dots,X_{k-1})$, we get

$$Y'M_{X_0}Y = \sum_{i=1}^{k-1} Z_i'\widehat{\Sigma}_i Z_i \cdot \Big(\|X_i\|^2 - \|P_{X_0}X_i\|^2\Big) + \widehat{\Sigma}_k\big(n - r(X_0)\big) ,$$

$$Y'M_{(X_0,X_1)}Y = \sum_{i=2}^{k-1} Z_i'\widehat{\Sigma}_i Z_i \cdot \Big(\|X_i\|^2 - \|P_{(X_0,X_1)}X_i\|^2\Big) + \widehat{\Sigma}_k\big(n - r(X_0,X_1)\big) ,$$

$$\vdots$$

$$Y'M_{(X_0,X_1,\dots,X_{k-1})}Y = 0 + \widehat{\Sigma}_k\big(n - r(X_0,X_1,\dots,X_{k-1})\big) .$$

This is a triangular system. If we denote $A = \{a_{ij}\}_{i,j=1}^{k}$ and $A^{-} = \{\alpha_{ij}\}_{i,j=1}^{k}$, where

$$\begin{aligned}
a_{ij} &= \operatorname{Tr}\big(X_j' M_{(X_0,X_1,\ldots,X_{i-1})} X_j\big) \\
&= \|X_j\|^2 - \big\|P_{(X_0,X_1,\ldots,X_{i-1})} X_j\big\|^2, \quad i = 1,\ldots,k-1,\ j = i, i+1,\ldots,k-1\,, \\
a_{ij} &= 0\,, \qquad i = 1,\ldots,k,\ j = 1,\ldots,i-1\,, \\
a_{ik} &= \operatorname{Tr}\big(M_{(X_0,X_1,\ldots,X_{i-1})}\big) = n - r(X_0, X_1, \ldots, X_{i-1})\,, \qquad i = 1,\ldots,k\,,
\end{aligned}$$

and

$$Q_j = Y' M_{(X_0,X_1,\ldots,X_{j-1})} Y, \qquad j = 1,\ldots,k\,,$$

then we can see that all solutions of the (potentially singular) system are

$$Z_i' \widehat{\Sigma}_i Z_i = \sum_{j=1}^{k} \alpha_{ij} Q_j$$

(if A is regular, $A^{-} = A^{-1}$ is also triangular and the sum goes from $j = i$ to k).

We see that Σ_k is always estimable, unless $r(X_0, \ldots, X_{k-1}) = n$ (but this does not happen in practically important situations). Matrices Z_i are regression constants matrices and usually are of full rank. Thus, estimability of Σ_i depends on the fact whether $\mathcal{R}(X_i) \subset \mathcal{R}(X_0, \ldots, X_{i-1})$ or not, because

$$\mathcal{R}(X_i) \subset \mathcal{R}(X_0, \ldots, X_{i-1}) \implies \|X_i\|^2 - \big\|P_{(X_0,\ldots,X_{i-1})} X_i\big\|^2 = 0\,.$$

Also, there is a question of the ordering of these matrices.

We say that a natural ordering of X_i's exists if (after suitable reindexing) there holds

$$\mathcal{R}(X_i) \cap \mathcal{R}(X_0, \ldots, X_{i-1})^c \neq \{0\} \quad \forall i\,.$$

In such a case, A^{-1} exists and all variance components are estimable. Note that modified von Rosen's condition

$$\mathcal{R}(X_0) \supsetneqq \mathcal{R}(X_1) \supsetneqq \cdots \supsetneqq \mathcal{R}(X_{k-1})$$

is a sufficient condition for the existence of a natural ordering. Also note that if a natural ordering exists and we do not use it, a singular matrix A can arise even if all variance components are estimable. Thus, use of natural orderings for the whole sequence of X_i and its subsequences is a crucial thing.

It is also worth noting that when the modified von Rosen's condition is fulfilled, then

$$M_{(X_0,\ldots,X_{i-1})} = M_{X_{i-1}}\,, \qquad i = 1,\ldots,k\,,$$

and the whole system becomes substantially simpler.

If no such natural ordering exists, it is an open question whether changing of the ordering of X_i's or using another P's can change our ability to estimate Σ_i's.

However, the system suggests that existence of a natural ordering of X_i's could be not only a sufficient but also a necessary condition for the unbiased estimability of all variance components.

3. Special cases

(**1**) Consider $k = 2$ (one fixed and one random effect) with $\mathcal{R}(X_1) \not\subseteq \mathcal{R}(X_0)$. Then, both variance components are estimable:

$$\widehat{\Sigma}_2 = \frac{1}{n - r(X_0, X_1)} Y' M_{(X_0,X_1)} Y,$$

and

$$Z_1' \widehat{\Sigma}_1 Z_1 = \frac{1}{\operatorname{Tr}(X_1' M_{X_0} X_1)} \left[Y' M_{X_0} Y - \frac{n - r(X_0)}{n - r(X_0, X_1)} Y' M_{(X_0,X_1)} Y \right].$$

Unfortunately, explicit form of the plug-in estimator is not available. We have to use the vectorized form of the estimator

$$\operatorname{vec} \widehat{\widehat{B}}_0 = \left[(Z_0 \otimes X_0') \Big(\widehat{\Sigma}_2 \otimes I_n + Z_1' \widehat{\Sigma}_1 Z_1 \otimes X_1 X_1' \Big)^{-1} (Z_0' \otimes X_0) \right]^{-1} \times (Z_0 \otimes X_0') \Big(\widehat{\Sigma}_2 \otimes I_n + Z_1' \widehat{\Sigma}_1 Z_1 \otimes X_1 X_1' \Big)^{-1} \operatorname{vec} Y$$

and de-vectorize it after the computation.

(**2**) Consider the case when there exists $\gamma_i \geq 0$ such that

$$P_{X_0} X_i X_i' P_{X_0} = \gamma_i P_{X_0} \qquad \forall i = 1, \dots, k-1.$$

Then, according to [10], the LSE of B_0 is

$$\widehat{B}_0 = (X_0' X_0)^{-1} X_0' Y \Gamma^{-1} Z_0' (Z_0 \Gamma^{-1} Z_0')^{-1},$$

where

$$\Gamma = \sum_{j=1}^{k-1} \gamma_j Z_j' \Sigma_j Z_j + \Sigma_k .$$

When A is regular, Γ is estimable and we have

$$\widehat{\Gamma} = \sum_{i=1}^{k-1} \sum_{j=i}^{k} \gamma_i \alpha_{ij} Q_j + \alpha_{kk} Q_k .$$

The plug-in estimator has then an explicit form

$$\widehat{\widehat{B}}_0 = (X_0' X_0)^{-1} X_0' Y \widehat{\Gamma}^{-1} Z_0' (Z_0 \widehat{\Gamma}^{-1} Z_0')^{-1}.$$

Acknowledgement. The author would like to thank to anonymous referee whose suggestions substantially improved the quality of the paper.

REFERENCES

[1] FUJIKOSHI, Y.: *Simultaneous confidence intervals in an extended growth curve model*, Comm. Statist. Theory Methods **28** (1999), 671–682.

[2] GHAZAL, G. A.—NEUDECKER, H.: *On second-order and fourth-order moments of jointly distributed random matrices: a survey*, Lin. Algebra Appl. **321** (2000), 61–93.

[3] KHATRI, C. G.—SHAH, K. R.: *On the unbiased estimation of fixed effects in a mixed model for growth curves*, Comm. Statist. Theory Methods **10** (1981), 401–406.

[4] TIMM, N. H.: *The* CGMANOVA *model*, Comm. Statist. Theory Methods **26** (1997), 1083–1098.

[5] VON ROSEN, D.: *Multivariate linear normal models with special reference to the growth curve model*, PhD-thesis, University of Stockholm.

[6] VON ROSEN, D.: *Maximum likelihood estimators in multivariate linear normal models*, J. Multivariate Anal. **31** (1989), 187–200.

[7] WESOŁOWSKA-JANCZAREK, M.: *Zastosowanie krzywych wzrostu w rolnictwie*, Fragmenta Agronomica **XIII** (1996), 1–53.

[8] YOKOYAMA, T.: *Extended growth curve models with random-effects covariance structures*, Comm. Statist. Theory Methods **25** (1996), 571–584.

[9] ŽEŽULA, I.: *A note on* MANOVA *in the extended growth curve model*, Tatra Mt. Math. Publ. **17** (1999), 319–323.

[10] ŽEŽULA, I.: *A remark on the extended growth curve model*, in: New trends in Prob. and Stat., Vol. 5, Multivariate Statistics (Proceedings of the 6$^{\text{th}}$ Tartu Conference), VSP Publ., 2000, pp. 219–223.

Received March 27, 2002

Institute of Mathematics
Šafárik University
Jesenná 5
SK–041 54 Košice
SLOVAKIA
E-mail: zezula@kosice.upjs.sk