



## Small area estimation using reduced rank regression models

Tatjana von Rosen & Dietrich von Rosen

To cite this article: Tatjana von Rosen & Dietrich von Rosen (2020) Small area estimation using reduced rank regression models, Communications in Statistics - Theory and Methods, 49:13, 3286-3297, DOI: [10.1080/03610926.2019.1586946](https://doi.org/10.1080/03610926.2019.1586946)

To link to this article: <https://doi.org/10.1080/03610926.2019.1586946>



© 2019 Tatjana von Rosen and Dietrich von Rosen. Published with license by Taylor & Francis Group, LLC.



Published online: 26 Apr 2019.



Submit your article to this journal [↗](#)



Article views: 530



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



# Small area estimation using reduced rank regression models

Tatjana von Rosen<sup>a</sup> and Dietrich von Rosen<sup>b,c</sup>

<sup>a</sup>Department of Statistics, Stockholm University, Stockholm, Sweden; <sup>b</sup>Department of Engineering and Technology, Swedish University of Agricultural Sciences, Uppsala, Sweden; <sup>c</sup>Department of Mathematics, Linköping University, Linköping, Sweden

## ABSTRACT

Small area estimation techniques have got a lot of attention during the last decades due to their important applications in survey studies. Mixed linear models and reduced rank regression analysis are jointly used when considering small area estimation. Estimates of parameters are presented as well as prediction of random effects and unobserved area measurements.

## ARTICLE HISTORY

Received 8 November 2018  
Accepted 14 February 2019

## KEYWORDS

Growth curve model; mixed linear model; reduced rank regression

## 1. Introduction

Small area estimation is an active research area in statistics with many applications. Today there exist several approaches to deal with small area problems. For recent contributions to the topic when considering discrete data (e.g. see Chandra et al. 2018), and for continuous data (see Baldermann 2018). In both articles many works of other authors are presented and discussed. For those who want to get a solid introduction to the subject the book (Rao and Molina 2015) can be recommended. Usually a common basis for small area estimation problems is that a survey study (finite population case) has been conducted and based on the survey one intends to extract information concerning small domains by adding specific local information which was not accounted for in the comprehensive survey.

In this article a case is considered where units of a survey sample are investigated several times, leading to the existence of dependent observations. It is assumed that observed background information exist which then is implemented via a linear model, giving a model which includes variance components. Moreover, it is assumed that latent processes exist which often show up when different types of systems are studied, for example when weather impact is studied. In this case often a huge number of variables are studied but we do not know the effects of the individual variables, for example we cannot directly relate temperature to plant growth. This leads us to exploit reduced rank regression models. Thus, the model which will be discussed in this article will include time dependent fix effects, i.e. structured mean responses, random parameters and rank restrictions on

**CONTACT** Dietrich von Rosen  [dietrich.von.rosen@slu.se](mailto:dietrich.von.rosen@slu.se)  Department of Energy and Technology, Box 7032, SE-750 07 Uppsala, Sweden.

© 2019 Tatjana von Rosen and Dietrich von Rosen. Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

parameters, which to our knowledge is a new type of model. In [Section 2](#) more motivating details are given, including basic references, and in [Section 3](#) the assumptions are summarized together with all technical details. Briefly, in [Section 4](#) the estimation of all unknown parameters is considered and finally two prediction results are presented in [Section 5](#).

## 2. Background

In this section, via an example, we will introduce our way of thinking about extracting information from a survey sample. In particular, a number of different sources of uncertainty are presented which then will be included in a statistical model. It turns out that the model will comprise three different types of uncertainty. [Section 3](#) includes all details of the model and in the subsequent section likelihood-based estimators are derived. In the last section these estimators are utilized when finding predictors.

**Example 2.1.** Let there be a national survey which collects data about some specific production from a certain type of companies. The survey is followed-up once per year and it is repeated, say, five years.

Altogether there exist 20 regions and each region consists of 10 subregions. The survey sample (units) consists of four companies from each subregion. Thus, in the survey, in total 800 companies are followed five years. With the models presented in this article the aim is to obtain information about the companies in the subregions. Note that in reality the companies should be classified into strata and then at least one company should be sampled from each stratum and each subregion. To simplify the presentation we will only have one stratum. However, in each subregion we assume that there is a lot of background information which we would like to take into account when predicting production in subregions or predicting production of unobserved units.

Small area models, also called small domain models, can be classified into two categories. One category is the area-level models category (Fay and Herriot 1979) and the other category is the unit-level models category (Battese et al. 1988). Both model categories are discussed in detail in a well-written book (Rao and Molina 2015). In this article we are thinking of area levels but if unit-level information would be available the proposed model can also be used.

Example 2.1 indicates the following use of notation. Let

$$\hat{\theta}_{ij}, \quad i = 1, \dots, m, \quad j = 1, 2, \dots, n_i, \quad n = \sum_{i=1}^m n_i$$

be the direct estimates from the survey for the  $j$ th subarea within the  $i$ th area. The estimates are based on the survey design where units have been sampled on the subarea level. Note that we will not discuss the underlying survey in this article, only use the outcome of the survey, i.e. the direct estimators of the subareas.

Statistics is partly about quantifying uncertainty. Often this is carried out by assuming that an estimate is a representation of an estimator which follows some distribution. However, we also have cases where uncertainty is not described via distributions. Examples are rounding errors, specific types of truncations, identification of influential observations, outlier detection, model comparisons, the number of observations in a

sample, etc. One choice of distribution for the estimator  $\hat{\theta}_{ij}$  would be the one generated by the survey design, i.e. how many and under what model the units have been sampled would determine the distribution. However, for our purposes we have difficulties to work with the probability space generated by the survey design. Instead, we now think of the direct estimates  $\hat{\theta}_{ij}$  being misspecified quantities since the survey sample distribution does not take into account those covariables we are going to use in the adjustment of  $\hat{\theta}_{ij}$ . The covariables can be used to correct for bias as well as diminish the uncertainty in the estimate. For example, returning to our example, if the survey design does not take into account the size of companies small companies can be overrepresented in the sample and then  $\theta_{ij}$  is wrongly estimated.

In order to derive an applicable model lets start with

$$y_{ij} = \theta_{ij} + \epsilon_{ij} \quad (2.1)$$

where for notational convenience  $\hat{\theta}_{ij}$  has been replaced by  $y_{ij}$ . The joint distribution for  $\{\epsilon_{ij}\}$  is specified via

$$\epsilon \sim N_n(\mathbf{0}, \mathbf{V}) \quad (2.2)$$

where  $\epsilon = (\epsilon_{11}, \dots, \epsilon_{1,n_1}, \dots, \epsilon_{m,n_m})$  is a row vector,  $\mathbf{V}$  is supposed to be known and determined by the survey design. Often  $\mathbf{V}$  is solely a function of the number of sampled units. Moreover,  $\epsilon$  is defined on another probability space than the one generated by the survey design. However, the uncertainty in the direct estimates, due to the survey, is incorporated in the dispersion of  $\epsilon$ .

Furthermore, one basic assumption in this article is that some of the  $\theta_{ij}$  will be the same, i.e. the survey estimators are the same for clusters of subregions. To make such an assumption is often realistic but has to be checked in practise. Under this assumption, instead of (2.1), the model can be written in matrix notation

$$\mathbf{y} = \boldsymbol{\beta}'\mathbf{C} + \epsilon \quad (2.3)$$

where  $\mathbf{y} = (y_{11}, \dots, y_{1,n_1}, \dots, y_{m,n_m})$ ,  $(\theta_{11}, \dots, \theta_{1,n_1}, \dots, \theta_{m,n_m}) = \boldsymbol{\beta}'\mathbf{C}$ , and  $\boldsymbol{\beta}: k \times 1$  consists of unknown parameters and  $\mathbf{C}: k \times n$  is a usual design matrix describing the relations among the elements  $\{\theta_{ij}\}$ . Usually  $\mathbf{C}$  consists of blocks of one-vectors but in principle it can be arbitrary as long as it fits together with the covariate structures which will be presented later. For example, with three areas and four subareas in each area  $\mathbf{C}$  equals

$$\mathbf{C} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The error in Equation (2.3) is defined by Equation (2.2). The model means that a probability space is only connected to the modeling of  $\{\hat{\theta}_{ij}\}$ , i.e. the set of estimated  $\theta_{ij}$ , but we include in the model extra variation due to the survey design.

In the introductory example it was stated that the survey was repeated five times. When survey studies are repeated it will for this article be assumed that when there exists a linear trend it is of the same form for all subareas. In this case there will be a matrix  $\mathbf{Y}$  where each row in the matrix follows Equation (2.3), besides an unknown scalar multiplier. Thus, Equation (2.3) implies

$$Y = \mathbf{ABC} + E \quad (2.4)$$

where  $A$  is a design matrix which models  $p$  repeated surveys (within subarea design matrix),  $B$  is a matrix of unknown parameters and  $E \sim N_{p,n}(\mathbf{0}, \Sigma, V)$ , i.e.  $E$  is matrix normally distributed with a separable dispersion structure  $D[E] = D[\text{vec}E] = V \otimes \Sigma$ , where  $\otimes$  stands for the Kronecker product and  $\Sigma$  is an unknown positive definite matrix. The parameter  $\Sigma$  is used to model the dependency due to the repeated measurement which exist because the observed units of the design have been repeatedly measured. No particular structure in  $\Sigma$  is assumed since it would only make sense if we would have precise knowledge about the mean structure, which is not the case in this article. For example, it does not make sense to assume an autocorrelated structure if the mean has not been established. The model in Equation (2.4) is often called Growth Curve model, GMANOVA and recently bilinear regression model (see von Rosen 2018, for basic results and references).

As an example of  $A$  we can have

$$A = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \\ 1 & t_5 & t_5^2 \end{pmatrix}$$

meaning that there is a second order trend which is observed over five years.

When expressing  $\theta_{ij}$  as a linear function of an unknown parameter vector  $\beta$ , as it was done in Equation (2.3), we usually cannot say that the model is completely true for all subareas. Instead we believe that there is some random variation around a true imaginary model, i.e. we introduce a variance component. Thus, instead of Equation (2.4), the following model is set up:

$$Y = \mathbf{ABC} + \mathbf{UZ} + E \quad (2.5)$$

where  $U \sim N_{p,l}(\mathbf{0}, \Sigma_u, I_l)$  is independent of  $E$  and  $Z$  is related to  $C$  because, in the model, usually via  $C$ , the observations are linked to the subareas and  $Z$  helps to describe the random variation (errors) in the subareas. Therefore, formally,  $\mathcal{C}(Z') \subseteq \mathcal{C}(C')$ , where  $\mathcal{C}(\bullet)$  denotes the column vector space. Additionally we will standardize these random effects by assuming  $ZV^{-1}Z' = I_l$ , meaning that it has been adjusted for a different number of subareas within an area, taking into an account the uncertainty in the survey estimates. This has two important implications. One is that it simplifies the mathematical treatment of the model and the second is that it becomes easier to compare predicted random effects.

Now observed covariables will be introduced in the model. There exist different types of covariables but we are not going to distinguish between different types. For example one type of covariables are those which are accounted for in the survey design (e.g. variables defining different strata) and another type can be covariables observed in registries and a third type can be baseline data which were available when a survey study with repeated measurements on units was initiated. The effect of the covariate will be modeled by a term  $B_1C_1$  where  $B_1$  stands for the unknown effect and  $C_1$  collects the observed covariables leading to the model

$$Y = \mathbf{ABC} + \mathbf{B}_1\mathbf{C}_1 + \mathbf{UZ} + \mathbf{E} \quad (2.6)$$

Moreover, suppose that there are a number of latent variables. It can be large clusters of variables, e.g. weather variables, socio-econometric variables, chemical characteristics of soil and water, etc. A typical phenomena of these clusters are that one knows that there should be an effect on the survey estimates but it is difficult to express this via some functional relationship. Instead we will model the relationship via rank restrictions on parameter matrices and will consider the following model:

$$Y = \mathbf{ABC} + \mathbf{B}_1\mathbf{C}_1 + \mathbf{\Psi F} + \mathbf{UZ} + \mathbf{E} \quad (2.7)$$

where  $F: t \times n$  is known and  $\mathbf{\Psi}$  is unknown of rank  $r(\mathbf{\Psi}) = r \leq \min(p, t)$ . The idea is that all clustered variables appear in  $F$  and these variables are governed by  $r$  unobserved latent variables (processes) which are incorporated in the model via the rank restrictions on  $\mathbf{\Psi}$ . It is important to note that  $p$  has to be larger than  $r$  meaning that modeling the number of underlying latent variables depends on how many times the survey has been repeated. Thus, we really see an advantage of performing repeated surveys when latent variables are supposed to exist, which indeed we believe is a common phenomena.

### 3. Detailed model specification

Let  $Y$  consist of the direct estimates of the subareas where each column corresponds to a subarea including  $p$  estimators from the  $p$  repeated measurements. When performing likelihood inference the direct estimates are used but in the presentation we will not distinguish between estimators and estimates.

**Definition 3.1.** *Let*

$$Y = \mathbf{ABC} + \mathbf{B}_1\mathbf{C}_1 + \mathbf{\Psi F} + \mathbf{UZ} + \mathbf{E}$$

where  $A: p \times q$ ,  $C: k \times n$ ,  $C_1: k_1 \times n$ ,  $F: t \times n$ , are all known matrices,  $\mathcal{C}(F') \subseteq \mathcal{C}(C')$ ,  $r(\mathbf{\Psi}) = r < \min(p, t)$ ,  $Z: l \times n$ ,  $p \leq l < n$ ,  $ZV^{-1}Z' = I_l$ ,  $U \sim N_{p,l}(\mathbf{0}, \mathbf{\Sigma}_u, I_l)$ ,  $E \sim N_{p,n}(\mathbf{0}, \mathbf{\Sigma}_e, V)$  where  $U$  and  $E$  are independently distributed. The parameters  $\mathbf{\Sigma}_u$  and  $\mathbf{\Sigma}_e$  are supposed to be unknown and positive definite whereas  $V$  is known and determined from the survey.

When in Definition 3.1  $\mathbf{B}_1\mathbf{C}_1 = \mathbf{0}$ ,  $\mathbf{\Psi F} = \mathbf{0}$  and  $\mathbf{UZ} = \mathbf{0}$  then we have the classical growth curve model (GMANOVA) (see Potthoff and Roy 1964; von Rosen 2018). When  $\mathbf{\Psi F} = \mathbf{0}$  and  $\mathbf{UZ} = \mathbf{0}$  then the growth curve model with covariate information appears which sometimes is identified as a mixture of the GMANOVA and MANOVA models. Some references to this model, as well as more general models (see Chinchilli and Elswick 1985; Verbyla and Venables 1988; von Rosen 1989; and Bai and Shi 2007). If  $\mathbf{B}_1\mathbf{C}_1 = \mathbf{0}$  and  $\mathbf{\Psi F} = \mathbf{0}$  we say that we have the growth curve model with random effects (see Ip et al. 2007) whereas if only  $\mathbf{\Psi F} = \mathbf{0}$  holds (see Yokoyama and Fujikoshi 1992; Yokoyama 1995) where similar models are considered and where references to earlier works can be found. However, usually in these works one puts structures on the covariance matrices which will not be discussed in this article.

### 4. Estimation of parameters

Let  $Q$  be a matrix of basis vectors such that  $C(Q) = C(V^{-1/2}C' : V^{-1/2}C_1' : V^{-1/2}Z')$ , where  $V^{1/2}$  is a symmetric square root of  $V$  in  $E \sim N_{p,n}(\mathbf{0}, \Sigma, V)$ . Further, let  $Q^o$  be any matrix of full rank satisfying  $C(Q^o) = C(C' : C_1' : Z')^\perp$ . We assume  $Q'Q = I_\nu$ , where  $\nu = r(C' : C_1' : Z')$ ,  $\nu > l$ .

A one-one transformation of the model in Definition 3.1, using  $Q$ , yields

$$YV^{-1/2}Q = ABCV^{-1/2}Q + B_1C_1V^{-1/2}Q + \Psi FV^{-1/2}Q + UZV^{-1/2}Q + EV^{-1/2}Q, \tag{4.8}$$

$$YV^{-1/2}Q^o = EV^{-1/2}Q^o \tag{4.9}$$

The matrix  $G = Q'V^{-1/2}Z'ZV^{-1/2}Q$  of size  $\nu \times \nu$  is idempotent since  $C(V^{-1/2}Z') \subseteq C(Q)$  and  $ZV^{-1}Z' = I$ , and therefore

$$G = \Gamma \begin{pmatrix} I_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \Gamma'$$

where  $\Gamma : \nu \times \nu$  is a known orthogonal matrix. The identity in Equation (4.8) is post-multiplied by  $\Gamma$  leading to the model

$$YV^{-1/2}Q\Gamma = ABCV^{-1/2}Q\Gamma + B_1C_1V^{-1/2}Q\Gamma + \Psi FV^{-1/2}Q\Gamma + UZV^{-1/2}Q\Gamma + EV^{-1/2}Q\Gamma. \tag{4.10}$$

However, since  $\Gamma'Q'V^{-1/2}Z'ZV^{-1/2}Q\Gamma = I$ ,

$$D[UZV^{-1/2}Q\Gamma] = \begin{pmatrix} I_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \otimes \Sigma_u$$

and this leads to that the model in Equation (4.10) will be split into two models. Let  $\Gamma = (\Gamma_1 : \Gamma_2) : \nu \times l, \nu \times (\nu - l)$ . Then we have three models which will be used when finding estimators:

$$(i) \quad YV^{-1/2}Q\Gamma_1 = ABCV^{-1/2}Q\Gamma_1 + B_1C_1V^{-1/2}Q\Gamma_1 + \Psi FV^{-1/2}Q\Gamma_1 + UZV^{-1/2}Q\Gamma_1 + EV^{-1/2}Q\Gamma_1, \\ UZV^{-1/2}Q\Gamma_1 \sim N_{p,l}(\mathbf{0}, \Sigma_u, I_l), \quad EV^{-1/2}Q\Gamma_1 \sim N_{p,l}(\mathbf{0}, \Sigma_e, I_l), \tag{4.11}$$

$$(ii) \quad YV^{-1/2}Q\Gamma_2 = ABCV^{-1/2}Q\Gamma_2 + B_1C_1V^{-1/2}Q\Gamma_2 + \Psi FV^{-1/2}Q\Gamma_2 + EV^{-1/2}Q\Gamma_2, \quad EV^{-1/2}Q\Gamma_2 \sim N_{p,\nu-l}(\mathbf{0}, \Sigma_e, I_{\nu-l}), \tag{4.12}$$

$$(iii) \quad YV^{-1/2}Q^o = EV^{-1/2}Q^o, \quad EV^{-1/2}Q^o \sim N_{p,n-\nu}(\mathbf{0}, \Sigma_e, I_{n-\nu}). \tag{4.13}$$

The idea is to utilize Equations (4.12) and (4.13) when estimating  $B, B_1, \Psi$  and  $\Sigma_e$ . To estimate  $\Sigma_u$  these estimators are inserted in Equation (4.11). Suppose that the estimators  $\hat{B}, \hat{B}_1, \hat{\Psi}$  and  $\hat{\Sigma}_e$  have been obtained, and let

$$Y_0 = XV^{-1/2}Q\Gamma_1 - \hat{A}\hat{B}CV^{-1/2}Q\Gamma_1 - \hat{B}_1C_1V^{-1/2}Q\Gamma_1 - \hat{\Psi}FV^{-1/2}Q, \\ \Omega = \Sigma_u + \Sigma_e.$$

Under the assumption of no randomness in  $\hat{\mathbf{B}}$ ,  $\hat{\mathbf{B}}_1$  and  $\hat{\Psi}$ ,

$$\mathbf{Y}_0 = \tilde{\mathbf{E}}, \quad \tilde{\mathbf{E}} \sim N_{p,l}(0, \mathbf{\Omega}, \mathbf{I}_l)$$

Thus, if  $p \leq l$ ,  $\hat{\mathbf{\Omega}} = \frac{1}{l} \mathbf{Y}_0 \mathbf{Y}'_0$ , implying

$$\hat{\mathbf{\Sigma}}_u = \frac{1}{l} \mathbf{Y}_0 \mathbf{Y}'_0 - \hat{\mathbf{\Sigma}}_e.$$

Here it is assumed that the difference is positive definite. If  $\hat{\mathbf{\Sigma}}_u$  is not positive definite, which should occur with some positive probability, this indicates that data does not support the model presented in Definition 3.1 or the estimation procedure has some deficiencies. It can be noted that the problem with negative variance components in mixed linear models has a long history and an early reference is Nelder (1954). A more recent contribution to the discussion of the problem is provided by Molenberghs and Verbeke (2011). If we would start with the model in Equation (2.5) and only consider  $D[\mathbf{Y}] = \mathbf{Z}\mathbf{Z}' \otimes \mathbf{\Sigma}_u + \mathbf{V} \otimes \mathbf{\Sigma}_e$  we could avoid to assume  $\mathbf{\Sigma}_u$  to be positive definite. However, in our model derivation, via  $\mathbf{U} \sim N_{p,l}(\mathbf{0}, \mathbf{\Sigma}_u, \mathbf{I}_l)$ ,  $\mathbf{\Sigma}_u$  has to be positive definite. Thus, if obtaining a non-positive definite  $\hat{\mathbf{\Sigma}}_u$  we have to either reformulate the model or change the estimator. One way to modify  $\hat{\mathbf{\Sigma}}_u$  is to work with the eigenvalues of  $\hat{\mathbf{\Sigma}}_u$ . If there are a few “small” (by absolute value) non-positive eigenvalues these eigenvalues can be replaced by small positive quantities. This should of course take place with caution and common sense but as far as we know there is no general recipe of how to handle a non-positive definite  $\hat{\mathbf{\Sigma}}_u$ .

To estimate  $\hat{\mathbf{B}}$ ,  $\hat{\mathbf{B}}_1$ ,  $\hat{\Psi}$  and  $\hat{\mathbf{\Sigma}}_e$  Equations (4.12) and (4.13) are merged:

$$\begin{aligned} \mathbf{Y}\mathbf{V}^{-1/2}(\mathbf{Q}\mathbf{\Gamma}_2 : \mathbf{Q}^o) &= \mathbf{A}\mathbf{B}(\mathbf{C}\mathbf{V}^{-1/2}\mathbf{Q}\mathbf{\Gamma}_2 : \mathbf{0}) + \mathbf{B}_1(\mathbf{C}_1\mathbf{V}^{-1/2}\mathbf{Q}\mathbf{\Gamma}_2 : \mathbf{0}) \\ &+ \mathbf{\Psi}(\mathbf{F}\mathbf{V}^{-1/2}\mathbf{Q}\mathbf{\Gamma}_2 : \mathbf{0}) + \tilde{\mathbf{E}}, \end{aligned} \quad (4.14)$$

where  $\tilde{\mathbf{E}} \sim N_{p,n-l}(\mathbf{0}, \mathbf{\Sigma}_e, \mathbf{I}_{n-l})$ . Put

$$\begin{aligned} \mathbf{X} &= \mathbf{Y}\mathbf{V}^{-1/2}(\mathbf{Q}\mathbf{\Gamma}_2 : \mathbf{Q}^o) : p \times (n-l), \\ \mathbf{D}_1 &= (\mathbf{C}\mathbf{V}^{-1/2}\mathbf{Q}\mathbf{\Gamma}_2 : \mathbf{0}) : k \times (n-l), \\ \mathbf{D}_2 &= (\mathbf{C}_1\mathbf{V}^{-1/2}\mathbf{Q}\mathbf{\Gamma}_2 : \mathbf{0}) : k_1 \times (n-l), \\ \mathbf{D}_3 &= (\mathbf{F}\mathbf{V}^{-1/2}\mathbf{Q}\mathbf{\Gamma}_2 : \mathbf{0}) : t \times (n-l). \end{aligned}$$

Equation (4.14) is identical to

$$\mathbf{X} = \mathbf{A}\mathbf{B}\mathbf{D}_1 + \mathbf{B}_1\mathbf{D}_2 + \mathbf{\Psi}\mathbf{D}_3 + \tilde{\mathbf{E}} \quad (4.15)$$

which is an extended Growth Curve model (GMANOVA + MANOVA) with a reduced rank regression component. Furthermore, the likelihood function which corresponds to the model in Equation (4.15) equals

$$\begin{aligned} L(\mathbf{B}, \mathbf{B}_1, \mathbf{\Psi}, \mathbf{\Sigma}_e) &= (2\pi)^{-\frac{1}{2}p(n-l)} |\mathbf{\Sigma}_e|^{-\frac{1}{2}(n-l)} \\ &\times \exp\left(-\frac{1}{2} \text{tr}\{\mathbf{\Sigma}_e^{-1}(\mathbf{X} - \mathbf{A}\mathbf{B}\mathbf{D}_1 - \mathbf{B}_1\mathbf{D}_2 - \mathbf{\Psi}\mathbf{D}_3)(\cdot)'\}\right), \end{aligned} \quad (4.16)$$



where we have used the convention that instead of  $(\mathbf{H})(\mathbf{H})'$  it is written  $(\mathbf{H})(\cdot)'$  for any arbitrary matrix expression  $\mathbf{H}$ . Our first observation is that the likelihood in Equation (4.16) is smaller or equal to

$$(2\pi)^{-\frac{1}{2}p(n-l)} |\boldsymbol{\Sigma}_e|^{-\frac{1}{2}(n-l)} \times \exp\left(-\frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}_e^{-1}\left((\mathbf{X}-\mathbf{A}\mathbf{B}\mathbf{D}_1)(\mathbf{I}-\mathbf{P}_{D'_2})-\boldsymbol{\Psi}\mathbf{D}_3(\mathbf{I}-\mathbf{P}_{D'_2})(\cdot)'\right)\right\}\right) \tag{4.17}$$

with equality if and only if

$$\mathbf{B}_1\mathbf{D}_2 = \mathbf{X}\mathbf{P}_{D'_2} - \mathbf{A}\mathbf{B}\mathbf{D}_1\mathbf{P}_{D'_2} - \boldsymbol{\Psi}\mathbf{D}_3\mathbf{P}_{D'_2}$$

which, under some full rank conditions on  $\mathbf{D}_2$ , determines  $\mathbf{B}_1$  as a function of  $\mathbf{B}$  and  $\boldsymbol{\Psi}$ . Thus, if  $\mathbf{B}$  and  $\boldsymbol{\Psi}$  can be estimated the covariate effect described by  $\mathbf{B}_1\mathbf{D}_2$  can be estimated. The density in Equation (4.17) corresponds to the model

$$\mathbf{X}(\mathbf{I}-\mathbf{P}_{D'_2}) = \mathbf{A}\mathbf{B}\mathbf{D}_1(\mathbf{I}-\mathbf{P}_{D'_2}) + \boldsymbol{\Psi}\mathbf{D}_3(\mathbf{I}-\mathbf{P}_{D'_2}) + \tilde{\mathbf{E}}, \tag{4.18}$$

$$\tilde{\mathbf{E}} \sim N_{p,n-l}(\mathbf{0}, \boldsymbol{\Sigma}_e, \mathbf{I}_{n-l})$$

which is a Growth Curve model with latent variables (reduced rank effect). The model has been considered (Reinsel and Velu 1998; von Rosen and von Rosen 2017), where also other references can be found. Now it is shortly described how  $\mathbf{B}$  and  $\boldsymbol{\Psi}$  can be estimated. For notational conveniences the model in Equation (4.18) is written

$$\mathbf{X}_1 = \mathbf{A}\mathbf{B}\tilde{\mathbf{D}}_1 + \boldsymbol{\Psi}\tilde{\mathbf{D}}_3 + \tilde{\mathbf{E}}, \quad \tilde{\mathbf{E}} \sim N_{p,n-l}(\mathbf{0}, \boldsymbol{\Sigma}_e, \mathbf{I}_{n-l}) \tag{4.19}$$

Note that  $\mathcal{C}(\tilde{\mathbf{D}}'_3) \subseteq \mathcal{C}(\tilde{\mathbf{D}}'_1)$  because  $\mathcal{C}(\mathbf{F}') \subseteq \mathcal{C}(\mathbf{C}')$  which is essential for being able to obtain explicit estimators.

Let

$$\mathbf{W} = (\mathbf{X}_1 - \mathbf{A}\mathbf{B}\tilde{\mathbf{D}}_1 - \boldsymbol{\Psi}\tilde{\mathbf{D}}_3)(\cdot)'$$

and then the likelihood function for model Equation (4.19) equals

$$L(\mathbf{B}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}_e) = (2\pi)^{-(n-l)p/2} |\boldsymbol{\Sigma}_e|^{-(n-l)/2} \exp\left(-\frac{1}{2}\text{tr}\{\boldsymbol{\Sigma}_e^{-1}\mathbf{W}\}\right) \tag{4.20}$$

The upper bound of the likelihood function in Equation (4.20) is well known (see Srivastava and Khatri 1979, Theorem 1.10.4):

$$L(\mathbf{B}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}_e) \leq (2\pi)^{-(n-l)p/2} e^{\{-\frac{1}{2}(n-l)p\}} \left| \frac{1}{n-l} \mathbf{W} \right|^{-(n-l)/2} \tag{4.21}$$

and in Equation (4.21) equality holds if and only if  $(n-l)\boldsymbol{\Sigma}_e = \mathbf{W}$ . Hence, if  $\mathbf{B}$  and  $\boldsymbol{\Psi}$  are estimated then  $\boldsymbol{\Sigma}_e$  is also estimated. Maximizing the log-likelihood function is equivalent to minimizing the determinant  $|\mathbf{W}|$  which now will take place.

It is worth noting that it is used, since  $r(\boldsymbol{\Psi}) = r$ , that the matrix  $\boldsymbol{\Psi}$  can be factored into two terms, i.e.  $\boldsymbol{\Psi} = \boldsymbol{\Psi}_1\boldsymbol{\Psi}_2$ , where  $\boldsymbol{\Psi}_1$  and  $\boldsymbol{\Psi}_2$  are of size  $p \times r$  and size  $r \times t$ , respectively. Depending on the knowledge about the model  $\boldsymbol{\Psi}_1$  and  $\boldsymbol{\Psi}_2$  can be interpreted but there are also cases where the factorization has no clear interpretation.

In the subsequent, let  $\mathbf{P}_Q = \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^- \mathbf{Q}'$  and  $\mathbf{P}_{Q,V} = \mathbf{Q}(\mathbf{Q}'\mathbf{V}^{-1}\mathbf{Q})^- \mathbf{Q}'\mathbf{V}^{-1}$  with  $(\cdot)^-$  denoting an arbitrary generalized inverse.

Following the approach presented in (Kollo and von Rosen 2005, Chapter 4) we have

$$|\mathbf{W}| \geq |\mathbf{S}_1 + \mathbf{T}_1 (\mathbf{X}\mathbf{P}_{\tilde{D}'_1} - \mathbf{\Psi}_1 \mathbf{\Psi}_2 \tilde{\mathbf{D}}_3) (\mathbf{X}\mathbf{P}_{\tilde{D}'_1} - \mathbf{\Psi}_1 \mathbf{\Psi}_2 \tilde{\mathbf{D}}_3)' \mathbf{T}'_1| \tag{4.22}$$

where

$$\mathbf{S}_1 = \mathbf{X}_1 (\mathbf{I}_{n-l} - \mathbf{P}_{\tilde{D}'_1}) \mathbf{X}'_1, \tag{4.23}$$

$$\mathbf{T}_1 = \mathbf{I}_p - \mathbf{P}_{\mathbf{A}, \mathbf{S}_1} \tag{4.24}$$

and equality in Equation (4.22) holds if and only if  $\mathbf{A}\mathbf{B}\tilde{\mathbf{D}}_1 = \mathbf{P}_{\mathbf{A}, \mathbf{S}_1} (\mathbf{X}\mathbf{P}_{\tilde{D}'_1} - \mathbf{\Psi}_1 \mathbf{\Psi}_2 \tilde{\mathbf{D}}_3)$ . Throughout the article it is supposed that  $\mathbf{S}_1$  is positive definite which holds with probability 1. Thus,  $\mathbf{B}$  as a function of  $\mathbf{\Psi} = \mathbf{\Psi}_1 \mathbf{\Psi}_2$  can be obtained. Moreover,

$$\begin{aligned} & |\mathbf{S}_1 + \mathbf{T}_1 (\mathbf{X}\mathbf{P}_{\tilde{D}'_1} - \mathbf{\Psi}_1 \mathbf{\Psi}_2 \tilde{\mathbf{D}}_3) (\mathbf{X}\mathbf{P}_{\tilde{D}'_1} - \mathbf{\Psi}_1 \mathbf{\Psi}_2 \tilde{\mathbf{D}}_3)' \mathbf{T}'_1| \\ & \geq |\mathbf{S}_2 + \mathbf{T}_2 \mathbf{T}_1 \mathbf{X}\mathbf{P}_{\tilde{D}'_3} \mathbf{X}' \mathbf{T}'_1 \mathbf{T}'_2|, \end{aligned} \tag{4.25}$$

where

$$\mathbf{S}_2 = \mathbf{S}_1 + \mathbf{T}_1 \mathbf{X} (\mathbf{P}_{\tilde{D}'_1} - \mathbf{P}_{\tilde{D}'_3}) \mathbf{X}' \mathbf{T}'_1, \tag{4.26}$$

$$\mathbf{T}_2 = \mathbf{I}_p - \mathbf{P}_{\mathbf{T}_1 \mathbf{\Psi}_1, \mathbf{S}_2},$$

and the equality in Equation (4.25) holds if and only if

$$\mathbf{T}_1 \mathbf{\Psi}_1 \mathbf{\Psi}_2 \tilde{\mathbf{D}}_3 = \mathbf{P}_{\mathbf{T}_1 \mathbf{\Psi}_1, \mathbf{S}_2} \mathbf{X}\mathbf{P}_{\tilde{D}'_3} \tag{4.27}$$

Given  $\mathbf{\Psi}_1$ , the linear system of equations in Equation (4.27) is consistent and hence  $\mathbf{\Psi}_2$  can always be estimated as a function of  $\mathbf{\Psi}_1$  :

$$\mathbf{\Psi}_2 = (\mathbf{\Psi}'_1 \mathbf{T}'_1 \mathbf{S}_2^{-1} \mathbf{T}_1 \mathbf{\Psi}_1)^{-1} \mathbf{\Psi}'_1 \mathbf{T}'_1 \mathbf{S}_2^{-1} \mathbf{X}\tilde{\mathbf{D}}_3 (\tilde{\mathbf{D}}_3 \tilde{\mathbf{D}}_3')^{-1}, \tag{4.28}$$

under the assumption that  $r(\tilde{\mathbf{D}}_3) = k_1$  and  $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{\Psi}_1) = \{0\}$ . The only parameter in the right-hand side of Equation (4.25) which is left to estimate is  $\mathbf{\Psi}_1$ . The right-hand side of Equation (4.25) can be factored as

$$|\mathbf{S}_2 + \mathbf{T}_2 \mathbf{T}_1 \mathbf{X}\mathbf{P}_{\tilde{D}'_3} \mathbf{X}' \mathbf{T}'_1 \mathbf{T}'_2| = |\mathbf{S}_2| |\mathbf{R}| |\mathbf{M}' \mathbf{U} \mathbf{M}|,$$

where

$$\mathbf{R} = \mathbf{I}_n + \mathbf{P}_{\tilde{D}'_3} \mathbf{X}' \mathbf{T}'_1 \mathbf{S}_2^{-1} \mathbf{T}_1 \mathbf{X}\mathbf{P}_{\tilde{D}'_3},$$

$$\mathbf{M} = \mathbf{H}' \mathbf{\Psi}_1 (\mathbf{\Psi}'_1 \mathbf{H} \mathbf{H}' \mathbf{\Psi}_1)^{-1/2},$$

$$\mathbf{U} = \mathbf{I}_{p-r(\mathbf{A})} - \mathbf{H}' \mathbf{X}\mathbf{P}_{\tilde{D}'_3} \mathbf{R}^{-1} \mathbf{P}_{\tilde{D}'_3} \mathbf{X}' \mathbf{H},$$

with  $\mathbf{H} : p \times r(\mathbf{A})$  is such that  $\mathbf{T}'_1 \mathbf{S}_2^{-1} \mathbf{T}_1 = \mathbf{H} \mathbf{H}'$ ,  $\mathbf{M} : (p-r(\mathbf{A})) \times r$  is semi-orthogonal of rank  $r$  and  $\mathbf{U}$  is a positive definite matrix. The square root in  $\mathbf{M}$  is supposed to be symmetric. Due to the Poincaré separation theorem (see Rao 1979)

$$|\mathbf{S}_2| |\mathbf{R}| |\mathbf{M}' \mathbf{U} \mathbf{M}| \geq |\mathbf{S}_2| |\mathbf{R}| \prod_{i=1}^r \lambda_{p-r(\mathbf{A})-r+i},$$

where  $\lambda_1 \geq \dots \geq \lambda_{p-r(\mathbf{A})}$  are the eigenvalues of  $\mathbf{U}$  which do not depend on any unknown parameter. The lower bound is then attained if  $\mathbf{M}$  consists of the

corresponding eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$ . Let  $\tilde{\mathbf{M}} = (\mathbf{v}_1, \dots, \mathbf{v}_r)$ . Therefore, a maximum likelihood estimator of  $\Psi_1$  is found if we can find  $\Psi_1$  satisfying the following equality:

$$\mathbf{H}'\Psi_1(\Psi_1\mathbf{H}\mathbf{H}'\Psi_1)^{-1/2} = \tilde{\mathbf{M}}. \tag{4.29}$$

Since  $\tilde{\mathbf{M}}'\tilde{\mathbf{M}} = \mathbf{I}_r$ , the following choice of  $\Psi_1$  is appropriate:

$$\hat{\Psi}_1 = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\tilde{\mathbf{M}}. \tag{4.30}$$

There remains an unresolved problem of presenting all solutions to Equation (4.29), but this is not considered here.

Together with Equation (4.28) this yields that  $\hat{\Psi}\tilde{\mathbf{D}}_3 = \hat{\Psi}_1\hat{\Psi}'_1T'_1S_2^{-1}\mathbf{X}\mathbf{P}_{\tilde{\mathbf{D}}'_3}$ . Now, the obtained results will be stated in the following theorem.

**Theorem 4.1.** *Let the direct estimates of a survey follow the model given in Definition 3.1. Moreover, all matrices in the statements have earlier been defined in this section.*

- (i)  $\hat{\Psi}_1 = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\tilde{\mathbf{M}}_1$ ;
- (ii) If  $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\Psi) = \{\mathbf{0}\}$ ,  $\hat{\Psi}\tilde{\mathbf{D}}_3 = \hat{\Psi}_1\hat{\Psi}'_1T'_1S_2^{-1}\mathbf{X}_1\mathbf{P}_{\tilde{\mathbf{D}}'_3}$ ;  
 If additionally  $r(\tilde{\mathbf{D}}'_3) = t$ ,  $\hat{\Psi} = \hat{\Psi}_1\hat{\Psi}'_1T'_1S_2^{-1}\mathbf{X}_1\tilde{\mathbf{D}}'_3(\tilde{\mathbf{D}}_3\tilde{\mathbf{D}}'_3)^{-1}$ ;
- (iii) If  $\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\Psi) = \{\mathbf{0}\}$ ,  $\mathbf{A}\hat{\mathbf{B}}\tilde{\mathbf{D}}_1 = \mathbf{A}(\mathbf{A}'\mathbf{S}_1^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}_1^{-1}(\mathbf{X}_1\mathbf{P}_{\tilde{\mathbf{D}}'_3} - \hat{\Psi}\tilde{\mathbf{D}}_3)$ ;  
 If additionally  $r(\tilde{\mathbf{D}}_1) = k$  and  $r(\mathbf{A}) = q$

$$\hat{\mathbf{B}} = (\mathbf{A}'\mathbf{S}_1^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}_1^{-1}\left(\mathbf{Y}_1\tilde{\mathbf{D}}'_1(\tilde{\mathbf{D}}_1\tilde{\mathbf{D}}'_1)^{-1} - \hat{\Psi}\tilde{\mathbf{D}}_3\tilde{\mathbf{D}}'_1(\tilde{\mathbf{D}}_1\tilde{\mathbf{D}}'_1)^{-1}\right);$$

- (iv)  $\hat{\mathbf{B}}\mathbf{D}_2 = \mathbf{X}\mathbf{P}_{\mathbf{D}'_2} - \mathbf{A}\hat{\mathbf{B}}_1\mathbf{D}_1\mathbf{P}_{\mathbf{D}'_2} - \hat{\Psi}\mathbf{D}_3\mathbf{P}_{\mathbf{D}'_2}$ ;  
 If additionally  $r(\mathbf{D}_2) = k_2$

$$\hat{\mathbf{B}}_1 = \mathbf{X}\mathbf{D}'_2(\mathbf{D}_2\mathbf{D}'_2)^{-1} - \mathbf{A}\hat{\mathbf{B}}\mathbf{D}_1\mathbf{D}'_2(\mathbf{D}_2\mathbf{D}'_2)^{-1} - \hat{\Psi}\mathbf{D}_3\mathbf{D}'_2(\mathbf{D}_2\mathbf{D}'_2)^{-1};$$

- (v)  $(n-l)\hat{\Sigma}_e = \mathbf{S}_2 + \hat{\mathbf{T}}_2\mathbf{T}_1\mathbf{Y}_1\mathbf{P}_{\tilde{\mathbf{D}}'_3}\mathbf{Y}'_1\mathbf{T}'_1\hat{\mathbf{T}}'_2$ , where

$$\hat{\mathbf{T}}_2 = \mathbf{I}_p - \mathbf{T}_1\hat{\Psi}_1\left(\hat{\Psi}'_1\mathbf{T}'_1\mathbf{S}_2^{-1}\mathbf{T}_1\hat{\Psi}_1\right)^{-1}\hat{\Psi}'_1\mathbf{T}'_1\mathbf{S}_2^{-1};$$

- (vi) Suppose  $\frac{1}{l}\mathbf{Y}_0\mathbf{Y}'_0 - \hat{\Sigma}_e$  is positive definite. Then

$$\hat{\Sigma}_u = \frac{1}{l}\mathbf{Y}_0\mathbf{Y}'_0 - \hat{\Sigma}_e.$$

### 5. Prediction

Small area estimation is very often about prediction of unobserved units (subareas). For our model, due to normality assumptions, it is relatively straight forward to construct predictors, in particular if the explicit estimators of the previous section are utilized. The strategy will be to first predict  $\mathbf{U}$  which will be based on the observed data and thereafter the unobserved units are predicted.

For prediction of  $\mathbf{U}$  the conditional mean  $E[\mathbf{U}|\mathbf{Y}]$  is crucial and thus the joint distribution of  $\mathbf{U}$  and  $\mathbf{Y}$  will be derived. According to Definition 3.1,  $\mathbf{U}$  and  $\mathbf{E}$  are independently distributed ( $\text{vec}\mathbf{U}, \text{vec}\mathbf{Y}$ ) is normally distributed, i.e.

$$\begin{pmatrix} \text{vec}U \\ \text{vec}Y \end{pmatrix} \sim N_{p(n+l)}(E[\bullet], D[\bullet]),$$

where

$$E[\bullet] = (\mathbf{0}, \text{vec}'(\mathbf{ABC} + \mathbf{B}_1\mathbf{C}_1 + \mathbf{\Psi F}))',$$

$$D[\bullet] = \begin{pmatrix} \mathbf{I} \otimes \mathbf{\Sigma}_u & \mathbf{Z} \otimes \mathbf{\Sigma}_u \\ \mathbf{Z}' \otimes \mathbf{\Sigma}_u & \mathbf{Z}'\mathbf{Z} \otimes \mathbf{\Sigma}_u + \mathbf{V} \otimes \mathbf{\Sigma}_e \end{pmatrix}.$$

Thus,

$$E[\text{vec}U|\text{vec}Y] = (\mathbf{Z} \otimes \mathbf{\Sigma}_u)(\mathbf{Z}'\mathbf{Z} \otimes \mathbf{\Sigma}_u + \mathbf{V} \otimes \mathbf{\Sigma}_e)^{-1} \text{vec}(Y - \mathbf{ABC} - \mathbf{B}_1\mathbf{C}_1 - \mathbf{\Psi F})$$

and the next proposition can be stated.

**Proposition 5.1.** *Let the model be given in Definition 3.1 and use the notation introduced earlier in Section 4.*

- i. *The predicted value  $\hat{U}$  is given by*

$$\text{vec}\hat{U} = (\mathbf{Z} \otimes \hat{\mathbf{\Sigma}}_u)(\mathbf{Z}'\mathbf{Z} \otimes \hat{\mathbf{\Sigma}}_u + \mathbf{V} \otimes \hat{\mathbf{\Sigma}}_e)^{-1} \text{vec}(Y - \hat{\mathbf{A}}\hat{\mathbf{B}}\hat{\mathbf{C}} - \hat{\mathbf{B}}_1\hat{\mathbf{C}}_1 - \hat{\mathbf{\Psi}}\mathbf{F}),$$

where  $\hat{\mathbf{\Sigma}}_u$ ,  $\hat{\mathbf{\Sigma}}_e$ ,  $\hat{\mathbf{A}}\hat{\mathbf{B}}\hat{\mathbf{C}}$ ,  $\hat{\mathbf{B}}_1\hat{\mathbf{C}}_1$  and  $\hat{\mathbf{\Psi}}\mathbf{F}$  are presented in Theorem 4.1 and  $Y$  consists of subarea estimates.

- ii. *Let  $Y_o$  consist of the unobserved direct estimates, corresponding to the non-sampled units. The corresponding covariables are available which are denoted  $\mathbf{C}_o$ ,  $\mathbf{C}_{1o}$ ,  $\mathbf{F}_o$  and  $\mathbf{Z}_o$ . Then*

$$\text{vec}\hat{Y}_o = \text{vec}(\hat{\mathbf{A}}\hat{\mathbf{B}}\mathbf{C}_o + \hat{\mathbf{B}}_1\mathbf{C}_{1o} + \hat{\mathbf{\Psi}}\mathbf{F}_o) + (\mathbf{Z}'_o \otimes \mathbf{I})\text{vec}\hat{U},$$

where  $\text{vec}\hat{U}$  is given in (i).

## Funding

This research has been supported by The Swedish Foundation for Humanities and Social Sciences (P14-0641:1) and The Swedish Natural Research Council (2017-03003).

## References

- Bai, P., and L. Shi. 2007. Exact distributions of MLEs of regression coefficients in GMANOVA-MANOVA model. *Journal of Multivariate Analysis* 98 (9):1840–52. doi:10.1016/j.jmva.2007.06.001.
- Battese, G. E., R. M. Harter, and W. A. Fuller. 1988. An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* 83 (401):28–36. doi:10.1080/01621459.1988.10478561.
- Chinchilli, V., and R. Elswick. 1985. A mixture of the Manova and Gmanova models. *Communication in Statistics- Theory and Methods* 14 (12):3075–89. doi:10.1080/03610928508829096.
- Fay, R. E., and R. A. Herriot. 1979. Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74 (366a):269–77. doi:10.2307/2286322.
- Ip, W.-C., M.-X. Wu, S.-G. Wang, and H. Wong. 2007. Estimation for parameters of interest in random effects growth curve models. *Journal of Multivariate Analysis* 98 (2):317–27. doi:10.1016/j.jmva.2006.04.001.

- Kollo, T., and D. von Rosen. 2005. *Advanced Multivariate Statistics with Matrices*. Dordrecht: Springer.
- Molenberghs, G., and G. Verbeke. 2011. A note on a hierarchical interpretation for negative variance components. *Statistical Modelling: An International Journal* 11 (5):389–408. doi:[10.1177/1471082X1001100501](https://doi.org/10.1177/1471082X1001100501).
- Nelder, J. A. 1954. The interpretation of negative components of variance. *Biometrika* 41 (3–4): 544–8. doi:[10.2307/2332734](https://doi.org/10.2307/2332734).
- Potthoff, R. F., and S. N. Roy. 1964. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51 (3–4):313–26. doi:[10.2307/2334137](https://doi.org/10.2307/2334137).
- Rao, C. R. 1979. Separation theorems for singular values of matrices and their applications in multi-variate analysis. *Journal of Multivariate Analysis* 9 (3):362–77. doi:[10.1016/0047-259X\(79\)90094-0](https://doi.org/10.1016/0047-259X(79)90094-0).
- Rao, J. N. K., and I. Molina. 2015. *Small Area Estimation*. 2nd ed. Hoboken, NJ: Wiley.
- Reinsel, G. C., and R. P. Velu. 1998. *Multivariate Reduced-Rank regression. Theory and applications*. *Lect. Notes stat.* 136, New York: Springer-Verlag.
- von Rosen, D. 1989. Maximum likelihood estimators in multivariate linear normal models. *Journal of Multivariate Analysis* 31 (2):187–200. doi:[10.1016/0047-259X\(89\)90061-4](https://doi.org/10.1016/0047-259X(89)90061-4).
- von Rosen, D. 2018. *Bilinear Regression Analysis: An Introduction*. *Lect. Notes stat.* 220, New York: Springer-Verlag.
- von Rosen, T., and D. von Rosen. 2017. On estimation in some reduced rank extended growth curve models. *Mathematical Methods of Statistics* 26 (4):299–310. doi:[10.3103/S1066530717040044](https://doi.org/10.3103/S1066530717040044).
- Srivastava, M. S., and C. G. Khatri. 1979. *An Introduction to Multivariate Statistics*. New York: North-Holland.
- Verbyla, A. P., and W. N. Venables. 1988. An extension of the growth curve model. *Biometrika* 75 (1):129–38. doi:[10.1093/biomet/75.1.129](https://doi.org/10.1093/biomet/75.1.129).
- Yokoyama, T. 1995. Statistical inference on some mixed MANOVA - GMANOVA models with random effects. *Hiroshima Mathematical Journal* 25 (3):441–74. doi:[10.32917/hmj/1206127624](https://doi.org/10.32917/hmj/1206127624).
- Yokoyama, T., and Y. Fujikoshi. 1992. Tests for random-effects covariance structures in the growth curve model with covariates. *Hiroshima Mathematical Journal* 22 (1):195–202. doi:[10.32917/hmj/1206128618](https://doi.org/10.32917/hmj/1206128618).