UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 05-40
Statistics and Econometrics Series 07
June 2005

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

MEAN SQUARED ERRORS OF SMALL AREA ESTIMATORS UNDER A UNIT-LEVEL MULTIVARIATE MODEL

Amparo Baíllo and Isabel Molina*

Abstract

This work deals with estimating the vector of means of characteristics of small areas. In this context, a unit level multivariate model with correlated sampling errors is considered. An approximation is obtained for the mean squared and cross product errors of the empirical best linear unbiased predictors of the means. This approach has been implemented on a Monte Carlo study using economic data observed for a sample of Australian farms.

Keywords: Small area; maximum likelihood; multivariate mixed linear model; best linear unbiased predictor; mean squared error.

*Baíllo, Statistics Department, University Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: abaillo@est-econ.uc3m.es; Molina, Statistics Department, University Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: imolina@est-econ.uc3m.es. Research partially supported by Spanish grants MTM2004-00098 and 06/HSE/0181/2004 from Comunidad de Madrid.

Mean Squared Errors of Small Area Estimators under a Unit-Level Multivariate Model *

Amparo Baíllo and Isabel Molina†

Universidad Carlos III de Madrid

28903 Getafe (Madrid), Spain

Abstract

This work deals with estimating the vector of means of characteristics of small areas. In this context, a unit level multivariate model with correlated sampling errors is considered. An approximation is obtained for the mean squared and cross product errors of the empirical best linear unbiased predictors of the means. This approach has been implemented on a Monte Carlo study using economic data observed for a sample of Australian farms.

Keywords: Small area; maximum likelihood; multivariate mixed linear model; best linear unbiased predictor; mean squared error.
AMS Classification 2000: 62D05, 62J05.

1 Introduction

Assume a subpopulation (like a county or an age domain) of a global population (e.g., a state or a country) from which a sample has been drawn. In order to estimate certain characteristics of the subpopulation (poverty counts, per capita income, ...), it is possible to use direct estimators, constructed only with observations coming from that specific subpopulation. The definition of "small area", although somewhat diffuse, could be a subpopulation where estimates with higher precision can be obtained by incorporating information from

*Research partially supported by Spanish grants MTM2004-00098 and BMF2003-04820, and by 06/HSE/0181/2004 from Comunidad de Madrid.

†Address for correspondence: Isabel Molina, Depto. Estadística, Univ. Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), Spain. e-mail: imolina@est-econ.uc3m.es

outside this target subpopulation. This is the aim of the procedures and techniques comprised in the discipline called small area estimation. Typically the precision of the estimators is characterized by its mean squared error (MSE).

The interest in small area estimation has increased in the last decades. On the one hand, the public sector finds it useful for planning regional policies and allocating government funds. On the other, the private sector has a growing demand of models that account for local socio-economic conditions that affect businesses and industries. For instance, in U.S.A. this interest has given rise to the Small Area Income and Poverty Estimates program of the Census Bureau or the Local Area Unemployment Statistics program from the Bureau of Labor Statistics. In Europe the project called EURAREA (Enhancing Small Area Estimation Techniques to Meet European Needs) produces small-area socio-economic measures, like unemployment rates or annual average gross family income. A large part of the most important recent advances and references on small area estimation can be found in the monograph by Rao (2003).

In this framework models represent a powerful tool, since they "borrow strength" of related areas in the global population and take into account auxiliary information (such as that provided by census and administrative records). In particular, mixed models (see, e.g., McCulloch and Searle, 2001) that include area-specific random effects usually provide estimators with better precision. This is because these models allow to separate and estimate the variation between areas that is not due to auxiliary variables. But it is necessary to point out that their strength relies on a careful identification and fitting of the model, and on the information provided by the available explanatory variables.

For instance, Gaussian mixed linear models have been used to estimate the income or poverty counts in small places in U.S.A (Fay and Herriot, 1979; National Research Council, 2000), the census undercount in the decennial census of U.S.A. (Ericksen and Kadane, 1985) and Canada (Dick, 1995) and the extension of county crop areas using satellite information (Battese *et al.*, 1988).

When the aim is to estimate a multidimensional characteristic $\boldsymbol{\tau}$ depending on several correlated response variables $\mathbf{Y}$, then the natural extension is to use multivariate linear mixed models. For example, Datta *et al.* (1999) considered the estimation of crop areas under corn and soybeans (dimension $r = 2$) in different counties using a multivariate nested error regression model. These authors implemented empirical best linear predictors (EBLUP) and empirical Bayes estimators and an approximation to their MSE. Fay (1987) and Datta *et al.* (1991) proposed a multivariate Fay-Herriot model, and showed its improvement in precision over the univariate modelling. Datta *et al.* (1996) developed hierarchical Bayes estimators of median income in the context of a multivariate Fay-Herriot model.

Under a multivariate model we might also be interested in estimating a quantity which is a function $h(\boldsymbol{\tau})$ of an r-dimensional ($r > 1$) characteristic $\boldsymbol{\tau}$. If we estimate $h(\boldsymbol{\tau})$ by $h(\hat{\mathbf{t}})$, where $\hat{\mathbf{t}} = (\hat{t}_1, \ldots, \hat{t}_r)'$ is an estimator of $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_r)'$, then an approximation to the mean squared error $E(h(\boldsymbol{\tau}) - h(\hat{\mathbf{t}}))^2$ will most likely require an approximation to the mean cross products $E[(\hat{t}_i - \tau_i)(\hat{t}_j - \tau_j)]$, $i, j = 1, \ldots, r$. This work deals with this last problem when $\boldsymbol{\tau}$ is a quantity related to the vector of means of the response variables in a certain small area.

The approximation to the MSE has been tackled in several small-area models. For univariate Gaussian mixed models and estimating the dispersion parameters by a method of moments, Prasad and Rao (1990) provided an approximation to the MSE of an EBLUP up to order $o(D^{-1})$, where D denotes the number of small areas in the global population. This approximation was proved to be of the same order under maximum likelihood (ML) or restricted ML (RML) estimation by Das *et al.* (2004). Here we focus on a unit-level multivariate model and estimate the unknown parameters by ML. In Section 3 we obtain an $o(D^{-1})$ approximation to the matrix $E[(\hat{\mathbf{t}} - \boldsymbol{\tau})(\hat{\mathbf{t}} - \boldsymbol{\tau})']$ of mean squared errors and mean crossed product terms. Finally, in Section 4, the results are checked in a Monte Carlo experiment with real economic data observed for Australian farms.

2 Description of the model

Assume that an r-dimensional response vector is available, together with certain auxiliary variables, for population elements in D small areas. Denote by N_d the population size in the d-th area, $d = 1, \ldots, D$. We assume that the following general linear model relates the response variables to the auxiliary ones

$$\mathbf{y}_{dj} = \mathbf{X}_{dj}\boldsymbol{\beta} + u_d\mathbf{1}_r + \mathbf{e}_{dj}, \quad j = 1, \ldots, N_d,\ d = 1, \ldots, D, \tag{1}$$

where $\mathbf{y}_{dj}$ is the response corresponding to the j-th individual from the d-th area, $\mathbf{X}_{dj}$ is an $r \times p$ matrix containing the values of the auxiliary variables in that same individual, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown parameters and $\mathbf{1}_r$ denotes the $r \times 1$ vector of ones. Here the area-specific and individual effects follow Gaussian distributions with

$$u_d \overset{\text{iid}}{\sim} N(0, \sigma_u^2), \quad \mathbf{e}_{dj} \overset{\text{iid}}{\sim} N_r(\mathbf{0}, \boldsymbol{\Sigma}), \quad u_d \text{ and } \mathbf{e}_{dj} \text{ independent}, \tag{2}$$

where the covariance matrix of the errors is given by

$$\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \varrho & \cdots & \varrho \\ \varrho & 1 & \cdots & \varrho \\ \vdots & \vdots & \ddots & \vdots \\ \varrho & \varrho & \cdots & 1 \end{pmatrix}_{r \times r}, \tag{3}$$

and $\sigma^2 > 0$, ϱ are unknown parameters. From now on, when $r > 1$, we will assume that the condition $-(r-1)^{-1} < \varrho < 1$ holds, since it guarantees the positive definiteness of $\boldsymbol{\Sigma}$.

Observe that if we define the parameter $\phi = \sigma^2 \varrho$, then $\boldsymbol{\Sigma}$ can be expressed as

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_r + \phi(\mathbf{J}_r - \mathbf{I}_r), \tag{4}$$

where $\mathbf{I}_r$ denotes the $r \times r$ identity matrix and $\mathbf{J}_r = \mathbf{1}_r \mathbf{1}_r'$. In the new reparameterization, the dispersion parameter space under the model is

$$\Theta = \{\boldsymbol{\theta} = (\sigma_u^2, \sigma^2, \phi)' : \sigma_u^2, \sigma^2 \in (0, \infty),\ -\sigma^2 (r-1)^{-1} < \phi < \sigma^2\}.$$

We are interested in computing the small area vectors of means

$$\bar{\mathbf{Y}}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} \mathbf{y}_{dj} \tag{5}$$

In order to estimate these parameters of interest, a sample s_d of size n_d is taken in the d-th area, for all $d = 1, \ldots, D$. For the sake of simplicity we reorder the population elements in such a way that the first n_d individuals in the d-th area are those of the sample s_d. We assume that the sample elements comply with the model given by (1)–(3), that is,

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_D \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_D \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_D \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_D \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_D \end{bmatrix},$$

where $\mathbf{y}_d = (\mathbf{y}_{d1}', \ldots, \mathbf{y}_{dn_d}')'$, $\mathbf{e}_d = (\mathbf{e}_{d1}', \ldots, \mathbf{e}_{dn_d}')'$,

$$\mathbf{X}_d = \begin{bmatrix} \mathbf{X}_{d1} \\ \vdots \\ \mathbf{X}_{dn_d} \end{bmatrix}, \qquad \mathbf{Z}_d = \begin{bmatrix} \mathbf{0}_{r\,n_d \times (d-1)} & \mathbf{1}_{r\,n_d} & \mathbf{0}_{r\,n_d \times (D-d)} \end{bmatrix}$$

and $\mathbf{u}$ and $\mathbf{e}$ are independent Gaussian distributed random vectors. From now on we will denote $n = r \sum_{d=1}^{D} n_d$.

3 Outline of main results

Observe that each area mean can be split into sampled and nonsampled elements as follows

$$\bar{\mathbf{Y}}_d = \frac{1}{N_d}\sum_{j\in s_d}\mathbf{y}_{dj} + \frac{1}{N_d}\sum_{j\notin s_d}\mathbf{X}_{dj}\ \boldsymbol{\beta} + \left(1-\frac{n_d}{N_d}\right)u_d\mathbf{1}_r + \frac{1}{N_d}\sum_{j\notin s_d}\mathbf{e}_{dj}.$$

Thus it suffices to predict

$$\boldsymbol{\tau}_d = \frac{1}{N_d}\sum_{j\notin s_d}\mathbf{X}_{dj}\ \boldsymbol{\beta} + \left(1-\frac{n_d}{N_d}\right)u_d\mathbf{1}_r. \tag{6}$$

Following Henderson (1975), the best linear unbiased predictor (BLUP) of $\boldsymbol{\tau}_d$ is given by

$$\mathbf{t}_d = \frac{1}{N_d}\sum_{j\notin s_d}\mathbf{X}_{dj}\ \tilde{\boldsymbol{\beta}} + \left(1-\frac{n_d}{N_d}\right)\tilde{u}_d\mathbf{1}_r \tag{7}$$

(see also Rao 2003, p. 110). In this expression,

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

is the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$,

$$\tilde{\mathbf{u}} = (\tilde{u}_1,\ldots,\tilde{u}_D)' = \sigma_u^2\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})$$

is the vector of predicted values of the area effects, and $\mathbf{V}$ is the covariance matrix of $\mathbf{y}$, $\mathrm{Var}(\mathbf{y})$. Observe that $\mathbf{V} = \mathrm{diag}(\mathbf{V}_1,\ldots,\mathbf{V}_D)$ with

$$\mathbf{V}_d = \mathrm{Var}(\mathbf{y}_d) = \sigma_u^2\mathbf{J}_{rn_d} + \sigma^2\mathbf{I}_{rn_d} + \phi\ \mathrm{diag}(\mathbf{J}_r-\mathbf{I}_r,\ldots,\mathbf{J}_r-\mathbf{I}_r). \tag{8}$$

Let us denote $\mathbf{R} = \mathrm{Var}(\mathbf{e}) = \mathrm{diag}\{\boldsymbol{\Sigma},\ldots,\boldsymbol{\Sigma}\}$ and $\mathbf{G} = \mathrm{Var}(\mathbf{u}) = \sigma_u^2\mathbf{I}_D$. Observe that $\boldsymbol{\tau}_d$ is a vector of type $\mathbf{K}\boldsymbol{\beta}+\mathbf{M}\mathbf{u}$, where

$$\mathbf{K} = \frac{1}{N_d}\sum_{j\notin s_d}\mathbf{X}_{dj} \qquad \text{and} \qquad \mathbf{M} = \left(1-\frac{n_d}{N_d}\right)\left[\begin{array}{ccc}\mathbf{0}_{r\times(d-1)} & \mathbf{1}_r & \mathbf{0}_{r\times(D-d)}\end{array}\right]. \tag{9}$$

Then the mean squared error (MSE) of the BLUP, $\mathbf{t}_d = \mathbf{K}\tilde{\boldsymbol{\beta}}+\mathbf{M}\tilde{\mathbf{u}}$, is given by

$$\begin{aligned}\mathrm{MSE}(\mathbf{t}_d) &= E[(\mathbf{t}_d-\boldsymbol{\tau}_d)(\mathbf{t}_d-\boldsymbol{\tau}_d)'] \\ &= \mathbf{M}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{M}+\boldsymbol{\Lambda}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\boldsymbol{\Lambda}',\end{aligned} \tag{10}$$

where

$$\boldsymbol{\Lambda} = \mathbf{K}-\mathbf{M}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}$$

(see, e.g., Datta *et al.*, 1999).

Let $\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02}, \theta_{03})' \in \operatorname{int}(\Theta)$ denote the true, unknown value of the parameter $\boldsymbol{\theta}$ in the population. As the BLUP estimator $\mathbf{t}_d$ depends on $\boldsymbol{\theta}_0$, we replace this parameter by its MLE estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{y}) = (\hat{\sigma}_u^2, \hat{\sigma}^2, \hat{\phi})'$. Thus we obtain the empirical best linear unbiased predictor (EBLUP) $\hat{\mathbf{t}}_d = \mathbf{t}_d(\hat{\boldsymbol{\theta}})$, whose mean squared error is

$$\mathrm{MSE}(\hat{\mathbf{t}}_d) = \mathrm{MSE}(\mathbf{t}_d) + E[(\hat{\mathbf{t}}_d - \mathbf{t}_d)(\hat{\mathbf{t}}_d - \mathbf{t}_d)'], \tag{11}$$

since $\hat{\boldsymbol{\theta}}$ is translation invariant (see Kackar and Harville, 1981).

If we substitute $\boldsymbol{\beta}$ by $\tilde{\boldsymbol{\beta}}$ in the loglikelihood

$$l(\boldsymbol{\theta}) = c - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

we obtain the profile loglikelihood

$$l_P(\boldsymbol{\theta}) = c - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where c is a constant and

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}.$$

Observe that the MLE $\hat{\boldsymbol{\theta}}$ is also a solution to the equation $\partial l_P/\partial\boldsymbol{\theta} = \mathbf{0}$.

In Theorem 1, conditions on the model (1)–(3) are given, under which the second term in (11) can be expressed as

$$E[(\hat{\mathbf{t}}_d - \mathbf{t}_d)(\hat{\mathbf{t}}_d - \mathbf{t}_d)'] = \left[E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})]\right]_{k,\ell=1}^{r} + [o(D^{-1})]_{r\times r}, \tag{12}$$

where, for $\mathbf{t}_d = (t_{d1}, \ldots, t_{dr})'$,

$$\mathbf{h}_{dk} = \left.\frac{\partial t_{dk}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0}, \quad \mathbf{s} = \left.\frac{\partial l_P}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0} \quad \text{and} \quad \mathcal{I} = E\left(-\left.\frac{\partial^2 l_P}{\partial\boldsymbol{\theta}^2}\right|_{\boldsymbol{\theta}_0}\right). \tag{13}$$

The last two quantities, $\mathbf{s}$ and $\mathcal{I}$, are the scores vector and the Fisher information matrix respectively. Plugging expression (12) into (11) yields the following decomposition

$$\mathrm{MSE}(\hat{\mathbf{t}}_d) = \mathrm{MSE}(\mathbf{t}_d) + \left[E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})]\right]_{k,\ell=1}^{r} + [o(D^{-1})]_{r\times r}. \tag{14}$$

Finally the following spelled-out formula is proved in Theorem 2

$$E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})] = \operatorname{tr}\left.\left(\frac{\partial\boldsymbol{\gamma}'_d}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_d}{\partial\boldsymbol{\theta}}\mathcal{I}^{-1}\right)\right|_{\boldsymbol{\theta}_0} + o(D^{-1}), \qquad k,\ell = 1,\ldots,r,$$

with $\boldsymbol{\gamma}_d = (1 - n_d/N_d)\sigma_u^2\mathbf{V}^{-1}\mathbf{m}_d$, where $\mathbf{m}_d$ denotes the d-th column of $\mathbf{Z}$. As a consequence, we have

$$\mathrm{MSE}(\hat{\mathbf{t}}_d) = \mathrm{MSE}(\mathbf{t}_d) + \mathbf{J}_r \operatorname{tr}\left.\left(\frac{\partial\boldsymbol{\gamma}'_d}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_d}{\partial\boldsymbol{\theta}}\mathcal{I}^{-1}\right)\right|_{\boldsymbol{\theta}_0} + [o(D^{-1})]_{r\times r}. \tag{15}$$

4 Theorems and proofs

In the following, we use the notation $\lambda_{\min}(\mathbf{B})$ and $\lambda_{\max}(\mathbf{B})$ for the minimum and maximum eigenvalues respectively of a square matrix $\mathbf{B}$, $\|\mathbf{B}\| = \lambda_{\max}^{1/2}(\mathbf{B}'\mathbf{B})$, $\|\mathbf{B}\|_2 = \mathrm{tr}^{1/2}(\mathbf{B}'\mathbf{B})$, $\boldsymbol{\Delta}_i = \partial \mathbf{V}/\partial\theta_i$ and $\boldsymbol{\Delta}_{id} = \partial \mathbf{V}_d/\partial\theta_i$, for $i = 1,2,3$, $d = 1,\ldots,D$. Throughout this work we will make use of the hypotheses stated below, where $\mathbf{V}_0$ denotes $\mathbf{V}$ evaluated at $\boldsymbol{\theta}_0$.

(A1) $p < \infty$ and $r < \infty$.

(A2) $\liminf_{D\to\infty} D^{-1}\lambda_{\min}(\mathbf{X}'\mathbf{X}) > 0$.

(A3) All the possible values of the elements of $\mathbf{X}$ are uniformly bounded.

(A4) $\limsup_{D\to\infty} \max_{1\leq d\leq D} n_d < \infty$ and $\liminf_{D\to\infty} \min_{1\leq d\leq D} n_d > 0$.

(A5) $\liminf_{D\to\infty} D^{-1}\lambda_{\min}(\mathcal{I}) > 0$.

The following result provides the convergence rate of $\hat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_0$ as $D \to \infty$, together with an asymptotic representation of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ which will be used in the proof of Theorem 1.

Lemma 1 : *Let the multivariate model (1)–(3) satisfy assumptions (A1), (A4) and (A5). Then, for any $0 < \eta < 1$, there exists a set $\mathcal{B}$ on which, for large D, it holds that $|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < D^{-\eta/2}$ and $|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathcal{I}^{-1}\mathbf{s}| \leq D^{-\eta}u$, with $E(u^g)$ bounded for $g > 0$. Furthermore $P(\mathcal{B}^c) = O(D^{-\zeta g/2})$, where $\zeta = \min(1/4, 1-\eta)$.*

Proof: It suffices to check that the hypotheses in Theorem 2.1 of Das *et al.* (2004) are satisfied for the profile loglikelihood l_P. More concretely, we have to see if conditions (i) and (ii) below hold:

(i) $\liminf_{D\to\infty} D^{-1}\lambda_{\min}(\mathcal{I}) < \infty$.

(ii) The g-th moments ($g > 0$) of the following variables are bounded,

$$\frac{1}{\sqrt{D}}\left|\frac{\partial l_P}{\partial\theta_i}\bigg|_{\boldsymbol{\theta}_0}\right|,\quad \frac{1}{D}\left|\frac{\partial^2 l_P}{\partial\theta_i\partial\theta_j}\bigg|_{\boldsymbol{\theta}_0} - E\left(\frac{\partial^2 l_P}{\partial\theta_i\partial\theta_j}\bigg|_{\boldsymbol{\theta}_0}\right)\right|,\quad \frac{1}{D}\sup_{\boldsymbol{\theta}\in S_\delta}\left|\frac{\partial^3 l_P}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|,$$

where the expectations are taken at $\boldsymbol{\theta}_0$ and $S_\delta = S_\delta(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} : |\theta_i - \theta_{0i}| \leq \delta,\ i = 1,2,3\}$, for some $\delta > 0$.

The first order partial derivatives of l_P are given by

$$\begin{aligned}
\frac{\partial l_P}{\partial \theta_i} &= -\frac{1}{2}\,\mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_i) + \frac{1}{2}\,\mathbf{y}'\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\mathbf{y}, && (16)\\
\frac{\partial^2 l_P}{\partial \theta_i \partial \theta_j} &= \frac{1}{2}\,\mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_i\mathbf{V}^{-1}\boldsymbol{\Delta}_j) - \mathbf{y}'\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\mathbf{y}, && (17)\\
\frac{\partial^3 l_P}{\partial \theta_i \partial \theta_j \partial \theta_k} &= -\mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_k\mathbf{V}^{-1}\boldsymbol{\Delta}_j\mathbf{V}^{-1}\boldsymbol{\Delta}_i) + \mathbf{y}'\mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\mathbf{y} \\
&\quad + \mathbf{y}'\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\mathbf{y} + \mathbf{y}'\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}\mathbf{y}, && (18)
\end{aligned}$$

for $i, j, k = 1, 2, 3$. Thus the (i, j)-th element of $\mathcal{I}$ is

$$\mathcal{I}_{ij} = E\left(-\left.\frac{\partial^2 l_P}{\partial \theta_i \partial \theta_j}\right|_{\boldsymbol{\theta}_0} \right) = -\frac{1}{2}\mathrm{tr}\left(\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i\mathbf{V}_0^{-1}\boldsymbol{\Delta}_j\right) + \mathrm{tr}\left(\mathbf{P}_0\boldsymbol{\Delta}_i\mathbf{P}_0\boldsymbol{\Delta}_j\right), \qquad (19)$$

where $\mathbf{P}_0$ is matrix $\mathbf{P}$ evaluated at $\boldsymbol{\theta}_0$.

To prove condition (i) it is enough to see that the following stronger condition holds

$$\liminf_{D\to\infty} \left(D^{-1}\mathcal{I}_{ii}\right) < \infty, \qquad i = 1, 2, 3. \qquad (20)$$

In order to do this observe that

$$|\mathcal{I}_{ii} - \frac{1}{2}\mathrm{tr}(\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i)^2| \le |\mathrm{tr}(\mathbf{W}\Delta_i\mathbf{V}_0^{-1}\Delta_i)| + |\mathrm{tr}(\mathbf{W}\Delta_i\mathbf{P}_0\Delta_i)|, \qquad (21)$$

where $\mathbf{W} = \mathbf{V}_0^{-1} - \mathbf{P}_0$. Now let us prove that the two terms on the right-hand side of last inequality are bounded. Indeed,

$$\begin{aligned}
|\mathrm{tr}(\mathbf{W}\boldsymbol{\Delta}_i\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i)| &= |\mathrm{tr}(\mathbf{V}_0^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i)| \\
&\le p\|\mathbf{V}_0^{-1/2}\boldsymbol{\Delta}_i\mathbf{V}_0^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1/2}\|^2 \\
&\le p\,\lambda_{\min}^{-2}(\mathbf{V}_0)\,\|\boldsymbol{\Delta}_i\|^2\,\|\mathbf{V}_0^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1/2}\|^2
\end{aligned}$$

By assumptions (A1) and (A4), all the terms in this last product are bounded. In a similar way it can be proved that $|\mathrm{tr}\,(\mathbf{W}\boldsymbol{\Delta}_i\mathbf{P}_0\boldsymbol{\Delta}_i)|$ is bounded. Therefore

$$\lim_{D\to\infty} D^{-1}\left(\mathcal{I}_{ii} - \frac{1}{2}\mathrm{tr}(\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i)^2\right) = 0.$$

Taking advantage of the block-diagonal structure of the matrices $\mathbf{V}_0^{-1}$ and $\boldsymbol{\Delta}_i$, it is easy to see that, under assumptions (A1) and (A4), $\mathrm{tr}(\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i)^2$ is of exact order D. This implies the statement in (20).

Let us now prove condition (ii). It suffices to consider $g \geq 2$. By (16) and Minkowski inequality it can be seen that

$$E\left(\frac{1}{\sqrt{D}}\left|\frac{\partial l_P}{\partial \theta_i}\right|\right)^g = \frac{1}{D^{g/2}2^g}E|\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi} - \mathrm{tr}(\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i)|^g \tag{22}$$

$$\leq \frac{1}{D^{g/2}2^g}\,[(E|\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi} - E(\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})|^g)^{1/g} + |\mathrm{tr}(\mathbf{P}_0\boldsymbol{\Delta}_i) - \mathrm{tr}(\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i)|\,]^g, \tag{23}$$

where $\mathbf{M} = \mathbf{V}_0^{1/2}\mathbf{P}_0\boldsymbol{\Delta}_i\mathbf{P}_0\mathbf{V}_0^{1/2}$ and $\boldsymbol{\xi}$ is a random vector with distribution $N_n(\mathbf{0}, \mathbf{I}_n)$. For $i = 1, 2$ the term in (23) can be bounded in the following way

$$|\mathrm{tr}(\mathbf{P}_0\boldsymbol{\Delta}_i) - \mathrm{tr}(\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i)| \leq p\,\|\mathbf{B}_i\|^2\lambda_{\min}^{-1}(\mathbf{V}_0)\,\|\mathbf{V}_0^{-1/2}\mathbf{X}(\mathbf{X}\mathbf{V}_0^{-1}\mathbf{X})^{-1/2}\|^2, \tag{24}$$

with $\mathbf{B}_i$ given by $\boldsymbol{\Delta}_i = \mathbf{B}_i'\mathbf{B}_i$. For $i = 3$ the proof is analogous. Using (A1) and (A4) we have that all the terms appearing in the product on the right-hand side of (24) are bounded. On the other hand, by Lemma 5.1 in Das *et al.* (2004), we have that, for some constant c depending only on g,

$$E|\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi} - E(\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})|^g \leq c\|\mathbf{M}\|_2^g = c\;\mathrm{tr}^{g/2}(\boldsymbol{\Delta}_i\mathbf{P}_0)^2.$$

This, together with the fact that $\mathrm{tr}(\boldsymbol{\Delta}_i\mathbf{P}_0)^2$ is $O(D)$, implies that the left-hand side of (22) is bounded.

Concerning the second-order derivatives of the profile log-likelihood, remark that, by (17) and (19), we intend to prove that the term

$$\frac{1}{D^g}E|\mathbf{y}'\mathbf{P}_0\boldsymbol{\Delta}_i\mathbf{P}_0\boldsymbol{\Delta}_j\mathbf{P}_0\mathbf{y} - \mathrm{tr}(\mathbf{P}_0\boldsymbol{\Delta}_i\mathbf{P}_0\boldsymbol{\Delta}_j)|^g, \tag{25}$$

is bounded. Indeed, denoting $\mathbf{M} = \mathbf{V}_0^{1/2}\mathbf{P}_0\boldsymbol{\Delta}_j\mathbf{P}_0\boldsymbol{\Delta}_i\mathbf{P}_0\mathbf{V}_0^{1/2}$ the expression in (25) can be rewritten as

$$\frac{1}{D^g}E|\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi} - E(\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})|^g \leq \frac{1}{D^g}c\|\mathbf{M}\|_2^g,$$

where once more we have used Lemma 5.1 in Das *et al.* (2004). By applying Lemma 5.2 of the same authors it is easy to show that $\|\mathbf{M}\|_2 \leq \|\boldsymbol{\Delta}_i\mathbf{P}_0\|_2\|\boldsymbol{\Delta}_j\mathbf{P}_0\|_2 = O(D)$ and this finishes the proof that (25) is bounded.

Finally it just remains to prove the claim, stated in (ii), about the third-order partial derivatives of l_P. By (18) and Minkowski inequality, it suffices to bound the following two terms, for $i, j, k = 1, 2, 3$,

$$\frac{1}{D^g}E\left(\sup_{\boldsymbol{\theta}\in S_\delta}|\mathbf{y}'\mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\mathbf{y}|\right)^g \tag{26}$$

and

$$\frac{1}{D}\sup_{\boldsymbol{\theta}\in S_\delta}|\mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_k\mathbf{V}^{-1}\boldsymbol{\Delta}_j\mathbf{V}^{-1}\boldsymbol{\Delta}_i)|. \tag{27}$$

The second term can be bounded noting that

$$|\mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_k\mathbf{V}^{-1}\boldsymbol{\Delta}_j\mathbf{V}^{-1}\boldsymbol{\Delta}_i)| \le r\sum_{d=1}^{D} n_d\|\mathbf{V}_d^{-1}\|^3\|\|\boldsymbol{\Delta}_{di}\|\|\boldsymbol{\Delta}_{dj}\|\|\boldsymbol{\Delta}_{dk}\|.$$

We know that $\|\boldsymbol{\Delta}_{di}\| = O(1)$, $i = 1,2,3$ and $\|\mathbf{V}_d^{-1}\| \le \lambda_{\min}^{-2}(\mathbf{V}_d)$. When $\phi \ge 0$, it holds that $\lambda_{\min}(\mathbf{V}_d) \ge \sigma^2 - \phi$. By taking $\delta > 0$ small we can get $\sigma^2 - \phi > c > 0$, for a constant c independent of θ. For $\phi < 0$, the term $\lambda_{\min}(\mathbf{V}_d)$ can be proved to be bounded away from zero analogously. This means that (27) is $O(1)$.

Concerning the term (26), there exists an $n \times (n-p)$ matrix $\mathbf{F}$, not depending on $\boldsymbol{\theta}$, such that $\mathbf{F}'\mathbf{X} = \mathbf{0}$, $\mathrm{rg}(\mathbf{F}) = n - p$ and $\mathbf{P} = \mathbf{F}(\mathbf{F}'\mathbf{V}\mathbf{F})^{-1}\mathbf{F}'$ (see Searle *et al.* 1992, p. 451). Let us define $\mathbf{z} = \mathbf{F}'\mathbf{y}$, $\mathbf{H} = (\mathbf{F}'\mathbf{V}\mathbf{F})^{-1}$ and $\mathbf{K}_i = \mathbf{F}'\boldsymbol{\Delta}_i\mathbf{F}$, for $i = 1,2,3$. Then, we have

$$\mathbf{y}'\mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\mathbf{y} = \mathbf{z}'\mathbf{H}\mathbf{K}_k\mathbf{H}\mathbf{K}_j\mathbf{H}\mathbf{K}_i\mathbf{H}\mathbf{z} \le |\mathbf{z}'\mathbf{H}_0^{1/2}|^2\|\mathbf{H}_0^{-1/2}\mathbf{H}\mathbf{K}_k\mathbf{H}\mathbf{K}_j\mathbf{H}\mathbf{K}_i\mathbf{H}\mathbf{H}_0^{-1/2}\|.$$

Therefore,

$$\begin{aligned} E&\left(\sup_{\boldsymbol{\theta}\in S_\delta}|\mathbf{y}'\mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\mathbf{y}|\right)^g \\ &\le \left(\sup_{\boldsymbol{\theta}\in S_\delta}\|\mathbf{H}_0^{-1/2}\mathbf{H}\mathbf{K}_k\mathbf{H}\mathbf{K}_j\mathbf{H}\mathbf{K}_i\mathbf{H}\mathbf{H}_0^{-1/2}\|\right)^g E\left(|\mathbf{z}'\mathbf{H}_0\mathbf{z}|^g\right). \end{aligned} \tag{28}$$

To see that $E(|\mathbf{z}'\mathbf{H}_0\mathbf{z}|^g) = O(D^g)$, it suffices to observe that $\mathbf{z}'\mathbf{H}_0\mathbf{z}$ follows a $\mathcal{X}^2_{n-p}$ distribution. Finally it just remains to check that the first term on the right-hand side of (28) is bounded. Indeed, observe that

$$\|\mathbf{H}_0^{-1/2}\mathbf{H}\mathbf{K}_k\mathbf{H}\mathbf{K}_j\mathbf{H}\mathbf{K}_i\mathbf{H}\mathbf{H}_0^{-1/2}\| = \|\mathbf{V}_0\mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\| \le \|\mathbf{V}_0\|\|\mathbf{P}\boldsymbol{\Delta}_k\|\|\mathbf{P}\boldsymbol{\Delta}_j\|\mathbf{P}\boldsymbol{\Delta}_i\|\|\mathbf{P}\|.$$

By assumptions (A1) and (A4), we know that $\|\mathbf{V}_0\|$ is bounded. It can also be seen that

$$\|\mathbf{P}\boldsymbol{\Delta}_i\| \le \lambda_{\min}^{-1}(\mathbf{V})\,\lambda_{\max}^{1/2}(\boldsymbol{\Delta}_i^2),$$

which is bounded for δ sufficiently small. □

Lemma 2 : *Let the multivariate model (1)–(3) satisfy assumptions (A1)–(A5) and let the parameter space be*

$$\tilde{\Theta} = \Theta \cap \{\boldsymbol{\theta} = (\sigma_u^2, \sigma^2, \phi)\ :\ \sigma_u^2 \le C_u,\ \sigma^2 \le C_e,\ c_e - \sigma^2(r-1)^{-1} \le \phi \le \sigma^2 - c_e'\}$$

for some fixed constants $C_u, C_e, c_e, c'_e > 0$. Assume that $\boldsymbol{\theta}_0 \in int(\tilde{\Theta})$. Then there exists a constant $\delta > 0$ such that each of the components of the BLUP $\mathbf{t}_d = (t_{d1}, \ldots, t_{dr})'$, given in (7) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, can be expressed in the form

$$t_{dk} = t_{dk}(\boldsymbol{\theta}_0, \mathbf{y}) = \sum_{q=1}^{K} \lambda_{kq}(\boldsymbol{\theta}_0) W_q(\mathbf{y}), \qquad k = 1, \ldots, r, \tag{29}$$

where $K = O(D)$, and the following terms are bounded

$$\max_{1\leq q\leq K} E|W_q(\mathbf{y})|^b, \quad \max_{1\leq q\leq K} \sup_{\boldsymbol{\theta}\in\tilde{\Theta}} |\lambda_{kq}(\boldsymbol{\theta})|, \quad \sum_{q=1}^{K} \left| \frac{\partial \lambda_{kq}}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}_0} \right|, \quad \sum_{q=1}^{K} \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} \left\| \frac{\partial^2 \lambda_{kq}}{\partial \boldsymbol{\theta}^2} \right\|, \tag{30}$$

for all $b > 0$.

Proof: Let $\mathbf{x}_{djk}$ denote the k-th row of $\mathbf{X}_{dj}$ and $\mathbf{Q} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. Then

$$t_{dk}(\boldsymbol{\theta}, \mathbf{y}) = \mathbf{a}'(\boldsymbol{\theta})\, \mathbf{y} + \mathbf{b}'(\boldsymbol{\theta})\, \mathbf{y}_d + \mathbf{c}'(\boldsymbol{\theta})\, \mathbf{y},$$

where $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are $(n \times 1)$, $(rn_d \times 1)$ and $(n \times 1)$ vectors respectively given by

$$\begin{aligned}
\mathbf{a}'(\boldsymbol{\theta}) &= \frac{1}{N_d} \sum_{j\notin s_d} \mathbf{x}_{djk}\, \mathbf{Q}\mathbf{X}'\mathbf{V}^{-1}, \\
\mathbf{b}'(\boldsymbol{\theta}) &= \left(1 - \frac{n_d}{N_d}\right) \sigma_u^2 \mathbf{1}'_{r\,n_d} \mathbf{V}_d^{-1}, \\
\mathbf{c}'(\boldsymbol{\theta}) &= -\left(1 - \frac{n_d}{N_d}\right) \sigma_u^2 \mathbf{1}'_{r\,n_d} \mathbf{V}_d^{-1}\mathbf{X}_d\mathbf{Q}\mathbf{X}'\mathbf{V}^{-1}.
\end{aligned}$$

Since the terms λ_{kq} in (29) are actually given by the components of vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, we have $K = 2n + rn_d$, which is $O(D)$ under assumptions (A1) and (A4). On the other hand, each $W_q(\mathbf{y})$ in (29) is just a component of $\mathbf{y}$. The normality assumption of the model yields the boundedness of the first term in (30).

It is clear that the second term in (30) is bounded if the suprema over $\boldsymbol{\theta}$ of $|\mathbf{a}|$, $|\mathbf{b}|$ and $|\mathbf{c}|$ respectively are bounded. To check this point, observe that

$$\begin{aligned}
|\mathbf{a}| &\leq \frac{1}{N_d} \sum_{j\notin s_d} |\mathbf{x}_{djk}|\, \lambda_{\max}^{1/2}(\mathbf{Q})\, \lambda_{\min}^{-1/2}(\mathbf{V}), \\
|\mathbf{b}| &\leq \sigma_u^2\, r\, n_d\, \lambda_{\min}^{-1}(\mathbf{V}_d) \\
|\mathbf{c}| &\leq \sigma_u^2\, r\, n_d\, \lambda_{\min}^{-3/2}(\mathbf{V}_d) \|\mathbf{X}_d\|\, \lambda_{\max}^{1/2}(\mathbf{Q}),
\end{aligned}$$

with $\lambda_{\min}(\mathbf{V}_d) \geq \lambda_{\min}(\mathbf{V})$, and where $\lambda_{\min}(\mathbf{V})$ is bounded away from zero over $\tilde{\Theta}$. Further,

$$\lambda_{\max}(\mathbf{Q}) = \left(\min_{\mathbf{v}} \frac{\mathbf{v}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{v}}{\mathbf{v}'\mathbf{v}} \right)^{-1} \leq \lambda_{\max}(\mathbf{V})\lambda_{\min}^{-1}(\mathbf{X}'\mathbf{X}). \tag{31}$$

This means that, by assumptions (A1)–(A3), $|\mathbf{a}|$, $|\mathbf{b}|$ and $|\mathbf{c}|$ are uniformly bounded over $\boldsymbol{\theta} \in \tilde{\Theta}$.

In order to see that the third term in (30) is bounded, it is enough to show that

$$\left| \left.\frac{\partial a_m}{\partial \theta_i}\right|_{\boldsymbol{\theta}_0} \right| = O(D^{-1}), \quad \left| \left.\frac{\partial b_m}{\partial \theta_i}\right|_{\boldsymbol{\theta}_0} \right| = O(1), \quad \left| \left.\frac{\partial c_m}{\partial \theta_i}\right|_{\boldsymbol{\theta}_0} \right| = O(D^{-1}),$$

for $i = 1, 2, 3$ and any value of m, where $\mathbf{a} = (a_m)_{m=1}^{n}$, $\mathbf{b} = (b_m)_{m=1}^{rn_d}$ and $\mathbf{c} = (c_m)_{m=1}^{n}$. Indeed, if we denote by $\mathbf{e}_m$ the $(n \times 1)$ unit vector in the direction m, we have

$$\left| \left.\frac{\partial a_m}{\partial \theta_i}\right|_{\boldsymbol{\theta}_0} \right| \leq \sup_{j \notin s_d} |\mathbf{x}_{djk}| \, \|\mathbf{Q}_0\| |\mathbf{X}'\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i\mathbf{P}_0\mathbf{e}_m|,$$

where $\mathbf{Q}_0 = (\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1}$. By the definition of $\mathbf{P}$, we have

$$|\mathbf{X}'\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i\mathbf{P}_0\mathbf{e}_m| \leq \|\mathbf{X}_d'\mathbf{V}_{0d}^{-1}\boldsymbol{\Delta}_{id}\mathbf{V}_{0d}^{-1}\| + \|\mathbf{X}'\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i\mathbf{V}_0^{-1}\mathbf{X}\|\|\mathbf{Q}_0\|\|\mathbf{X}_d'\mathbf{V}_{0d}^{-1}\|.$$

But $\|\mathbf{X}_d'\mathbf{V}_{0d}^{-1}\boldsymbol{\Delta}_{id}\mathbf{V}_{0d}^{-1}\|$ and $\|\mathbf{X}_d'\mathbf{V}_{0d}^{-1}\|$ are bounded under assumptions (A1), (A3) and (A4), while $\|\mathbf{Q}_0\| = O(D^{-1})$ by assumption (A2) and inequality (31). Further,

$$\|\mathbf{X}'\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i\mathbf{V}_0^{-1}\mathbf{X}\| \leq \sum_{d=1}^{D} \|\mathbf{X}_d'\mathbf{V}_{0d}^{-1}\boldsymbol{\Delta}_{id}\mathbf{V}_{0d}^{-1}\mathbf{X}_d\| = O(D),$$

and this implies the desired result. The derivatives of b_m and c_m can be bounded following similar arguments.

Finally we will focus on the last term of (30). Observe that it suffices to see that, for $i, \ell = 1, 2, 3$ and any m,

$$\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} \left| \frac{\partial^2 a_m}{\partial \theta_i \partial \theta_\ell} \right| = O(D^{-1}), \quad \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} \left| \frac{\partial^2 b_m}{\partial \theta_i \partial \theta_\ell} \right| = O(1) \quad \text{and} \quad \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} \left| \frac{\partial^2 c_m}{\partial \theta_i \partial \theta_\ell} \right| = O(D^{-1}).$$

Since the second derivative of $\mathbf{a}$ is given by

$$\frac{\partial^2 \mathbf{a}'}{\partial \theta_i \partial \theta_\ell} = \frac{1}{N_d} \sum_{j \notin s_d} \mathbf{x}_{djk}\mathbf{Q}\mathbf{X}'\mathbf{V}^{-1}(\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_\ell\mathbf{P} + \boldsymbol{\Delta}_\ell\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}),$$

we have

$$\left| \frac{\partial^2 a_m}{\partial \theta_i \partial \theta_\ell} \right| \leq \sup_{j \notin s_d} |\mathbf{x}_{djk}| \, \|\mathbf{Q}\| \left(|\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_\ell\mathbf{P}\mathbf{e}_m| + |\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\Delta}_\ell\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\mathbf{e}_m| \right).$$

On the one hand, under assumptions (A1)-(A4), $\sup_{j\notin s_d}|\mathbf{x}_{djk}|$ is bounded and the supremum of $\|\mathbf{Q}\|$ is $O(D^{-1})$. On the other hand, we have

$$\begin{aligned}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_\ell\mathbf{P} &= \boldsymbol{\Delta}_i\mathbf{V}^{-1}\boldsymbol{\Delta}_\ell\mathbf{V}^{-1}-\boldsymbol{\Delta}_i\mathbf{V}^{-1}\boldsymbol{\Delta}_\ell\mathbf{V}^{-1}\mathbf{X}\mathbf{Q}\mathbf{X}'\mathbf{V}^{-1}-\boldsymbol{\Delta}_i\mathbf{V}^{-1}\mathbf{X}\mathbf{Q}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\Delta}_\ell\mathbf{V}^{-1}\\ &\quad+\boldsymbol{\Delta}_i\mathbf{V}^{-1}\mathbf{X}\mathbf{Q}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\Delta}_\ell\mathbf{V}^{-1}\mathbf{X}\mathbf{Q}\mathbf{X}'\mathbf{V}^{-1}.\end{aligned}$$

Let us study each term. For the first one observe that

$$|\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\Delta}_i\mathbf{V}^{-1}\boldsymbol{\Delta}_\ell\mathbf{V}^{-1}\mathbf{e}_m|\le\|\mathbf{X}_d'\mathbf{V}_d^{-1}\boldsymbol{\Delta}_{id}\mathbf{V}_d^{-1}\boldsymbol{\Delta}_{\ell d}\mathbf{V}_d^{-1}\|=O(1).$$

for some d. By similar arguments, and taking into account that terms like $\|\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\Delta}_\ell\mathbf{V}^{-1}\mathbf{X}\|$ are $O(D)$, the remaining terms in $|\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_\ell\mathbf{P}\mathbf{e}_m|$ can be bounded.

Regarding $\mathbf{b}$, when $i\neq 1$ and $\ell\neq 1$, its second derivative has the following expression

$$\frac{\partial^2\mathbf{b}'}{\partial\theta_i\partial\theta_\ell}=\left(1-\frac{n_d}{N_d}\right)\sigma_u^2\mathbf{1}_{rn_d}'\mathbf{V}_d^{-1}(\boldsymbol{\Delta}_{id}\mathbf{V}_d^{-1}\boldsymbol{\Delta}_{\ell d}+\boldsymbol{\Delta}_{\ell d}\mathbf{V}_d^{-1}\boldsymbol{\Delta}_{id})\mathbf{V}_d^{-1},$$

which can be bounded as before. For $i=1$ or $\ell=1$, and for the second derivative of $\mathbf{c}$ the proofs are analogous. □

The following result provides an approximation up to $[o(D^{-1})]_{r\times r}$ to the second term on the right-hand side of (11). Consequently, the decomposition given in (14) for the MSE of the EBLUP, $\hat{\mathbf{t}}_d$, holds.

Theorem 1 : *Under the hypotheses of Lemma 2, for* $\mathbf{h}_{dk}$, $\mathbf{s}$ *and* $\mathcal{I}$ *as defined in (13),* $\mathbf{t}_d=(t_{d1},\dots,t_{dr})'$ *defined in (7) and* $\hat{\mathbf{t}}_d=(\hat{t}_{d1},\dots,\hat{t}_{dr})'$*, it holds that*

$$E[(\hat{t}_{dk}-t_{dk})(\hat{t}_{d\ell}-t_{d\ell})]=E[(\mathbf{h}_{dk}'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_{d\ell}'\mathcal{I}^{-1}\mathbf{s})]+o(D^{-1}),\qquad k,\ell=1,\dots,r. \tag{32}$$

Proof: For any fixed $0<\eta<1$, the term on the left-hand side of (32) can be decomposed as

$$E[(\hat{t}_{dk}-t_{dk})(\hat{t}_{d\ell}-t_{d\ell})]=E[(\hat{t}_{dk}-t_{dk})(\hat{t}_{d\ell}-t_{d\ell})1_{\mathcal{B}}]+E[(\hat{t}_{dk}-t_{dk})(\hat{t}_{d\ell}-t_{d\ell})1_{\mathcal{B}^c}], \tag{33}$$

where $\mathcal{B}$ was the set introduced in Lemma 1 and $1_{\mathcal{B}}$ denotes the indicator function of $\mathcal{B}$. Using the expression of t_{dk} stated in (29), bounding $|\lambda_{kq}|$ uniformly in $\boldsymbol{\theta}\in\tilde{\Theta}$, for $q=1,\dots,K$ and $k=1,\dots,r$ by a constant $M>0$, and applying Hölder inequality, we get

$$E[(\hat{t}_{dk}-t_{dk})(\hat{t}_{d\ell}-t_{d\ell})1_{\mathcal{B}^c}]\le 4M^2\left[\sum_{q=1}^{K}E^{1/2}(W_q^2(\mathbf{y})1_{\mathcal{B}^c})\right]^2. \tag{34}$$

Let $b > 2$ be a constant. By Lemma 2 there exists a constant $\mathcal{W} > 0$ bounding $E(|W_q(\mathbf{y})|^b)$ for all $q = 1, \ldots, K$. Then Hölder inequality and Lemma 1 with $\eta = 2/3$ and $\zeta = 1/4$ yield

$$E(W_q^2(\mathbf{y})1_{\mathcal{B}^c}) \leq \mathcal{W}^{2/b}(P(\mathcal{B}^c))^{1-2/b} = O(D^{-\frac{g}{8}\left(1-\frac{2}{b}\right)}).$$

Plugging this expression into (34) and using the fact that $K = O(D)$, we arrive to

$$E[(\hat{t}_{dk} - t_{dk})(\hat{t}_{d\ell} - t_{d\ell})1_{\mathcal{B}^c}] = O(D^{2-\frac{g}{8}(1-\frac{2}{b})}).$$

Observe that $g > 0$ and $b > 2$ can be taken as large as desired so that this term is $o(D^{-1})$.

Regarding the first term on the right-hand side of (33), consider the following Taylor series expansion on the set $\mathcal{B}$

$$\hat{t}_{dk} - t_{dk} = \mathbf{h}'_{dk}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + r_k, \tag{35}$$

where, by Lemma 1, r_k satisfies $|r_k| \leq u_k|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}|^2 \leq D^{-\eta}u_k$ with

$$u_k = \frac{1}{2}\sum_{q=1}^{K} \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<D^{-\eta/2}} \left\| \frac{\partial^2\lambda_{kq}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2} \right\| |W_q(\mathbf{y})|.$$

By Lemma 2, $E(u_k^2)$ is bounded. Now, applying Lemma 1, we get

$$\hat{t}_{dk} - t_{dk} = \mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s} + r_k^*,$$

where $|r_k^*| \leq D^{-\eta}(|\mathbf{h}_{dk}|u + u_k) = D^{-\eta}u_k^*$, with $E[(u_k^*)^2]$ bounded. This holds for $k = 1, \ldots, r$. Thus,

$$\begin{aligned} E[(\hat{t}_{dk} - t_{dk})(\hat{t}_{d\ell} - t_{d\ell})1_{\mathcal{B}}] &= E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})1_{\mathcal{B}}] \\ &+E(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s}\, r_\ell^*\, 1_{\mathcal{B}}) + E(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s}\, r_k^*\, 1_{\mathcal{B}}) + E(r_k^*\, r_\ell^*\, 1_{\mathcal{B}}). \end{aligned} \tag{36}$$

It is clear that $E(r_k^* r_\ell^*) = O(D^{-2\eta})$. On the other hand, for the second and third terms on the right-hand side of (36), Hölder inequality leads to

$$E(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s}\, r_\ell^* 1_{\mathcal{B}}) \leq E^{1/4}|\mathbf{h}_{dk}|^4\ E^{1/4}|\mathcal{I}^{-1}\mathbf{s}|^4\ E^{1/2}|r_\ell^*|^2, \tag{37}$$

for any $k, \ell = 1, \ldots, r$, where we know that $E^{1/2}(r_\ell^*)^2 = O(D^{-\eta})$. Now by assumption (A5), Hölder inequality applied to $E|D^{-1/2}\mathbf{s}|^4$ and Lemma 2, we get

$$E^{1/4}|\mathcal{I}^{-1}\mathbf{s}|^4 \leq D^{1/2}\,\|\mathcal{I}^{-1}\|\ E^{1/4}|D^{-1/2}\mathbf{s}|^4 = O(D^{-1/2}).$$

The first expectation on the right of (37) can be bounded applying Hölder inequality and Lemma 2. Thus

$$E(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s}\, r_\ell^* 1_{\mathcal{B}}) = O(D^{-1/2-\eta}), \qquad k, \ell = 1, \ldots, r. \tag{38}$$

Finally, the first term on the right-hand side of (36) can be expressed in the form

$$E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})1_{\mathcal{B}}] = E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})] - E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})1_{\mathcal{B}^c}]. \qquad (39)$$

For the last term on the right-hand side of (39), proceeding as before, we have

$$E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})1_{\mathcal{B}^c}] \leq E^{1/2}[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})^2 1_{\mathcal{B}^c}]\, E^{1/2}[(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})^2 1_{\mathcal{B}^c}] = O(D^{-1-g/16}),$$

which implies that

$$E[(\hat{t}_{dk} - t_{dk})(\hat{t}_{d\ell} - t_{d\ell})1_{\mathcal{B}}] = E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})] + o(D^{-1}). \qquad \square$$

The following result is a technical lemma to be used in the proof of Theorem 2.

Lemma 3 : *Let* $\mathbf{u} \sim N_k(\mathbf{0}, \mathbf{\Sigma})$, $w_{ij} = \boldsymbol{\lambda}'_{ij}\mathbf{u}$, $q_j = \mathbf{u}'\mathbf{A}_j\mathbf{u}$, $i = 1, 2$, $j = 1, \dots, s$, *where* $\boldsymbol{\lambda}_{ij}$ *and* $\mathbf{A}_j$ *are constant* $k \times 1$ *vectors and* $k \times k$ *symmetric matrices respectively, and* $\mathbf{\Sigma}$ *is positive definite. Then, for* $\mathbf{w}_i = (w_{i1}, \dots, w_{is})'$ *and* $\mathbf{q} = (q_1, \dots, q_s)'$, *the following equalities hold*

(i) $E[\mathbf{w}'_1(\mathbf{q} - E\mathbf{q})\mathbf{w}'_2\mathbf{a}] = 2(\sum_{\ell=1}^{s} a_l \boldsymbol{\lambda}'_{2\ell})\mathbf{\Sigma}(\sum_{j=1}^{s} \mathbf{A}_j\mathbf{\Sigma}\boldsymbol{\lambda}_{1j})$, *where* $\mathbf{a} = (a_1, \dots, a_s)'$ *is any constant vector;*

(ii) $E[\mathbf{w}'_1(\mathbf{q} - E\mathbf{q})\mathbf{w}'_2(\mathbf{q} - E\mathbf{q})] = \mathrm{tr}[\mathrm{Cov}(\mathbf{w}_1, \mathbf{w}_2)\mathrm{Var}(\mathbf{q})] + 4\sum_{j=1}^{s}\sum_{\ell=1}^{s} \boldsymbol{\lambda}'_{1j}\mathbf{\Sigma}(\mathbf{A}_j\mathbf{\Sigma}\mathbf{A}_\ell + \mathbf{A}_\ell\mathbf{\Sigma}\mathbf{A}_j)\mathbf{\Sigma}\boldsymbol{\lambda}_{2\ell}$.

Proof:

(i) It is a direct consequence of Lemma 3.1(i) in Das *et al.* (2004).

(ii) It can be seen that

$$\begin{aligned}
&E[\mathbf{w}'_1(\mathbf{q} - E\mathbf{q})\mathbf{w}'_2(\mathbf{q} - E\mathbf{q})] \\
&\quad = \sum_{j=1}^{s}\sum_{\ell=1}^{s} E[w_{1j}(q_j - Eq_j)w_{2\ell}(q_\ell - Eq_\ell)] \\
&\quad = \sum_{j=1}^{s}\sum_{\ell=1}^{s} \boldsymbol{\lambda}'_{1j} E[\mathbf{u}(\mathbf{u}'\mathbf{A}_j\mathbf{u} - E(\mathbf{u}'\mathbf{A}_j\mathbf{u}))(\mathbf{u}'\mathbf{A}_\ell\mathbf{u} - E(\mathbf{u}'\mathbf{A}_\ell\mathbf{u}))\mathbf{u}']\boldsymbol{\lambda}_{2\ell} \\
&\quad = \sum_{j=1}^{s}\sum_{\ell=1}^{s} 2\mathrm{tr}(\mathbf{A}_j\mathbf{\Sigma}\mathbf{A}_\ell\mathbf{\Sigma})\boldsymbol{\lambda}'_{1j}\mathbf{\Sigma}\boldsymbol{\lambda}_{2\ell} + 4\sum_{j=1}^{s}\sum_{\ell=1}^{s} \boldsymbol{\lambda}'_{1j}\mathbf{\Sigma}(\mathbf{A}_j\mathbf{\Sigma}\mathbf{A}_\ell + \mathbf{A}_\ell\mathbf{\Sigma}\mathbf{A}_j)\mathbf{\Sigma}\boldsymbol{\lambda}_{2\ell}
\end{aligned}$$

where we have used Lemma 3.1 (iii) in Das *et al.* (2004). $\square$

The following result provides an approximation to the second term on the right-hand side of (14).

Theorem 2 : *Under the same hypotheses of Theorem 1, the following equality holds*

$$E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})] = \mathrm{tr}\left(\frac{\partial\boldsymbol{\gamma}'_d}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_d}{\partial\boldsymbol{\theta}}\mathcal{I}^{-1}\right)\Bigg|_{\boldsymbol{\theta}_0} + o(D^{-1}), \qquad k,\ell = 1,\dots r, \tag{40}$$

where $\boldsymbol{\gamma}_d = (1-n_d/N_d)\sigma_u^2\mathbf{V}^{-1}\mathbf{m}_d$ *and* $\mathbf{m}_d$ *is the* d*-th column of* $\mathbf{Z}$*.*

Proof: Let us first compute the expression for $\mathbf{h}_{dk}$. Observe that we can write

$$\begin{aligned} t_{dk} \quad &= \mathbf{K}_k\tilde{\boldsymbol{\beta}} + \boldsymbol{\gamma}'_d(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}) \\ &= \mathbf{K}_k\boldsymbol{\beta} + \mathbf{K}_k(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{v} + \boldsymbol{\gamma}'_d\mathbf{V}\mathbf{P}\mathbf{v}, \end{aligned}$$

where $\mathbf{K}_k$ is the k-th row of matrix $\mathbf{K}$ in (9) and $\mathbf{v} = \mathbf{Z}\mathbf{u} + \mathbf{e}$. This yields

$$\frac{\partial t_{dk}}{\partial\theta_j}\bigg|_{\boldsymbol{\theta}_0} = \left(\mathbf{f}'_{kj} + \frac{\partial\boldsymbol{\gamma}'_d}{\partial\theta_j}\bigg|_{\boldsymbol{\theta}_0}\right)\mathbf{v},$$

with

$$\mathbf{f}'_{kj} = -(\mathbf{K}_k - \boldsymbol{\gamma}'_d|_{\boldsymbol{\theta}_0}\mathbf{X})(\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_0^{-1}\boldsymbol{\Delta}_j\mathbf{P}_0 - \frac{\partial\boldsymbol{\gamma}'_d}{\partial\theta_j}\bigg|_{\boldsymbol{\theta}_0}\mathbf{X}(\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_0^{-1},$$

for $k = 1,\dots,r$, $j = 1,2,3$, and

$$\frac{\partial\boldsymbol{\gamma}'_d}{\partial\theta_j} = \mathbf{m}'_d\mathbf{V}^{-1}\left(\frac{\partial\sigma_u^2}{\partial\theta_j}\mathbf{I}_n - \sigma_u^2\boldsymbol{\Delta}_j\mathbf{V}^{-1}\right)\left(1-\frac{n_d}{N_d}\right), \qquad j = 1,2,3.$$

Thus we have obtained that $\mathbf{h}_{dk} = \mathbf{D}_k\mathbf{v} = (\mathbf{F}_k + \partial\boldsymbol{\gamma}'_d/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}_0})\mathbf{v}$, where $\mathbf{F}_k$ is the matrix whose j-th row is given by $\mathbf{f}'_{kj}$. Observe also that $\mathbf{s} = (\mathbf{q}-E\mathbf{q})/2 + \boldsymbol{\nu}$, with $q_i = \mathbf{v}'\mathbf{P}_0\boldsymbol{\Delta}_i\mathbf{P}_0\mathbf{v}$ and $\nu_i = (\mathrm{tr}(\mathbf{P}_0\boldsymbol{\Delta}_i) - \mathrm{tr}(\mathbf{V}_0^{-1}\boldsymbol{\Delta}_i))/2$, $i = 1,2,3$.

If we denote $\mathbf{w}_k = \mathcal{I}^{-1}\mathbf{h}_{dk} = \mathcal{I}^{-1}\mathbf{D}_k\mathbf{v}$, the left-hand side of (40) can be written as

$$\begin{aligned} E[(\mathbf{h}'_{dk}\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}'_{d\ell}\mathcal{I}^{-1}\mathbf{s})] &= \frac{1}{4}E[\mathbf{w}'_k(\mathbf{q}-E\mathbf{q})\mathbf{w}'_\ell(\mathbf{q}-E\mathbf{q})] \\ &+\frac{1}{2}E[\mathbf{w}'_k\boldsymbol{\nu}\ \mathbf{w}'_\ell(\mathbf{q}-E\mathbf{q})] + \frac{1}{2}E[\mathbf{w}'_k(\mathbf{q}-E\mathbf{q})\mathbf{w}'_\ell\boldsymbol{\nu}] + E[\mathbf{w}'_k\boldsymbol{\nu}\ \mathbf{w}'_\ell\boldsymbol{\nu}]. \end{aligned} \tag{41}$$

Now we apply Lemma 3(ii) to the first term on the right-hand side of (41)

$$\begin{aligned} E[\mathbf{w}'_k(\mathbf{q}-E\mathbf{q})\mathbf{w}'_\ell(\mathbf{q}-E\mathbf{q})] &= \mathrm{tr}[\mathrm{Cov}(\mathbf{w}_k,\mathbf{w}_\ell)\mathrm{Var}(\mathbf{q})] \\ &+4\sum_{i=1}^{3}\sum_{j=1}^{3}\boldsymbol{\lambda}'_{ki}\mathbf{V}_0\mathbf{P}_0(\boldsymbol{\Delta}_i\mathbf{P}_0\mathbf{V}_0\mathbf{P}_0\boldsymbol{\Delta}_j + \boldsymbol{\Delta}_j\mathbf{P}_0\mathbf{V}_0\mathbf{P}_0\boldsymbol{\Delta}_i)\mathbf{P}_0\mathbf{V}_0\boldsymbol{\lambda}_{\ell j}. \end{aligned} \tag{42}$$

where $\boldsymbol{\lambda}_{kj}$ is in this case the j-th row of $\mathcal{I}^{-1}\mathbf{D}_k$. Under the hypotheses of the theorem we have that $\|\mathcal{I}^{-1}\| = O(D^{-1})$, $\|\mathbf{V}_0^{-1}\| = O(1)$, $\|\mathbf{D}_k\| = O(1)$ for all k and $\|\boldsymbol{\Delta}_i\| = O(1)$ for all i. This implies that the last term in (42) is $O(D^{-2})$. On the other hand, by (19) and (21), it is easy to check that

$$\frac{1}{4}\mathrm{tr}[\mathrm{Cov}(\mathbf{w}_k, \mathbf{w}_\ell)\mathrm{Var}(\mathbf{q})] = \mathrm{tr}(\mathcal{I}^{-1}\mathbf{D}_k\mathbf{V}_0\mathbf{D}'_\ell) + O(D^{-2})$$

$$= \mathrm{tr}\left(\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}'_d}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_d}{\partial\boldsymbol{\theta}}\right)\bigg|_{\boldsymbol{\theta}_0} + \mathrm{tr}\left(\mathcal{I}^{-1}\mathbf{F}_k\mathbf{V}_0\mathbf{F}'_\ell\right) \tag{43}$$

$$+\mathrm{tr}\left(\mathcal{I}^{-1}\mathbf{F}_k\mathbf{V}\frac{\partial\boldsymbol{\gamma}_d}{\partial\boldsymbol{\theta}}\right)\bigg|_{\boldsymbol{\theta}_0} + \mathrm{tr}\left(\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}'_d}{\partial\boldsymbol{\theta}}\mathbf{V}\mathbf{F}'_\ell\right)\bigg|_{\boldsymbol{\theta}_0} + O(D^{-2}). \tag{44}$$

By the assumptions in the theorem and using that $\|\mathbf{F}_k\| = O(D^{-1/2})$, the second term in (43) is bounded as

$$\mathrm{tr}\left(\mathcal{I}^{-1}\mathbf{F}_k\mathbf{V}_0\mathbf{F}'_\ell\right) \leq 3\|\mathcal{I}^{-1}\mathbf{F}_k\mathbf{V}_0\mathbf{F}'_\ell\| \leq 3\lambda_{\max}(\mathbf{V}_0)\,\|\mathbf{F}_k\|\,\|\mathbf{F}'_\ell\|\,\|\mathcal{I}^{-1}\|.$$

On the other hand, since $\|\partial\boldsymbol{\gamma}_d/\partial\boldsymbol{\theta}\|$ is bounded for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, it can be seen that each of the two terms in (44) is $o(D^{-1})$. Thus

$$\frac{1}{4}\mathrm{tr}[\mathrm{Cov}(\mathbf{w}_k, \mathbf{w}_\ell)\mathrm{Var}(\mathbf{q})] = \mathrm{tr}\left(\frac{\partial\boldsymbol{\gamma}'_d}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_d}{\partial\boldsymbol{\theta}}\mathcal{I}^{-1}\right)\bigg|_{\boldsymbol{\theta}_0} + o(D^{-1}).$$

To check that the last three terms in (41) are $O(D^{-2})$ it suffices to apply Lemma 3(i) and and proceed as before. $\square$

5 A Monte Carlo Experiment with Real Data: Australian Farms

From the Australian Agricultural and Grazing Industries Survey (AAGIS) there are data available for 1652 farms. The type of small areas considered in this study are within-state regions (corresponding to different farming areas), which will be indexed by $d \in \{1, \ldots, D\}$, with $D = 29$. The regions are located in seven states of Australia. Each farm in the sample was assigned a weight depending on the number of farms it represented in the region.

The variables recorded consisted of financial and production aspects of the observed farm. In particular, here we have considered the vector of responses $\mathbf{y} = (y_1, y_2, y_3)'$, where y_1 is the logarithm of the variable "equity", the difference (in A\$) between the value of the farm business and its debt, and y_2 and y_3 are the logarithm of the total cash costs and receipts respectively (in A\$) of the farm over the surveyed year. The four auxiliary variables will be the logarithm of the total area (in hectares) of the farm, the logarithm of its cultivated area, of the number of beef cattle and of the number of sheep in the farm.

The AAGIS data were used to generate, via bootstrap, a global population of 81982 farms. The sampling was performed with replacement and with probability proportional to the sample weight of the farm within each region. From the global population we have extracted 1000 independently selected stratified Monte Carlo samples, each of size 1652. The small area sample sizes n_i, $i = 1, \ldots, 29$, were chosen equal to those in the original AAGIS data set and are displayed in Table 1.

For each of the Monte Carlo samples the goal was to predict the regional mean as given in (5), to implement the estimation of the MSE obtained in the previous section and to compare it with the real squared error which can be computed using the global population data. The results appearing in Table 1 give, for each region, the average values (over the Monte Carlo samples) of the difference between the estimated MSE and the real squared error. If we denote by $\mathbf{A}$ this (3×3) matrix of averaged differences in region d, the d-th row of Table 1 contains the number of individuals sampled in that region (n_d), the proportion of farms in the global (bootstrapped) population that belong to the region (weight) and components A_{11} A_{12}, A_{13}, A_{22}, A_{23} and A_{33} of matrix $\mathbf{A}$. It can be seen that the regions with smaller weight in the global population are the ones yielding larger differences, that is, worse estimation of the squared error. But, in general terms, the approximation provided by (15) is satisfactory.

Acknowledgements. We are grateful to Prof. R. L. Chambers for providing us with the data used in this work.

References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28–36.

Das, K., Jiang, J. and Rao, J. N. K. (2004), Mean squared error of empirical predictor. *Annals of Statistics*, **32**, 818–840.

Datta, G. S., Day, B. and Basawa, I. (1999), Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation. *Journal of Statistical Planning and Inference*, **75**, 269–279.

Datta, G. S., Fay, R. E. and Ghosh, M. (1991). Hierarchical and Empirical Bayes Multivariate Analysis in Small Area Estimation. *Proceedings of Bureau of the Census 1991 Annual Research Conference*, U. S. Bureau of the Census, Washington, DC, 63–79.

Datta, G. S., Ghosh, M., Nangia, N., and Natarajan, K. (1996). Estimation of median income of four-person families: a Bayesian approach. *Bayesian Analysis in Statistics and Econometrics*, (D. A. Berry, K. M. Chaloner and J. M. Geweke Eds.), Wiley, New York, 129–140.

Dick, P. (1995). Modelling net undercoverage in the 1991 Canadian census. *Survey Methodology*, **21**, 45–54.

Ericksen, E. P. and Kadane, J. B. (1985). Estimating the population in census year: 1980 and beyond (with discussion). *Journal of the American Statistical Association*, **80**, 98–131.

Fay, R. E. (1987). Application of multivariate regression to small domain estimation. In *Small Area Statistics*, R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh (Eds.), Wiley, New York, pp. 91-102.

Fay, R. E. and Herriot, R. A. (1979). Estimation of income from small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–447.

Kackar, R. N. and Harville, D. A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics, Series A*, **10**, 1249-1261.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear and Mixed Models.* Wiley, New York.

National Research Council (2000). *Small Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology.* C. F. Citro and G. Kalton (Eds.), Committee on National Statistics. National Academy Press, Washington, DC.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163–171.

Rao, J. N. K. (2003). *Small Area Estimation*. Wiley, New Jersey.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components*. Wiley, New York.

Table 1: Difference between estimated MSE and real squared error in AAGIS data

d	n_d	Weight	A_{11}	A_{12}	A_{13}	A_{22}	A_{23}	A_{33}
1	55	0.024	0.0475	-0.0863	-0.1101	-0.0382	-0.0247	-0.0158
2	103	0.067	-0.0020	-0.1494	-0.1461	-0.0736	-0.0711	-0.0753
3	60	0.033	0.1470	-0.2876	-0.2168	-0.0108	-0.0530	-0.0829
4	60	0.033	0.2081	-0.2244	-0.2209	-0.0123	-0.0081	-0.0121
5	34	0.035	1.5828	-1.0776	-0.8660	0.3955	0.2816	0.1840
6	80	0.055	0.0730	-0.1678	-0.1570	-0.0436	-0.0419	-0.0499
7	62	0.022	0.1178	-0.2355	-0.2362	-0.0603	-0.0572	-0.0621
8	74	0.037	0.0603	-0.1768	-0.2204	-0.0899	-0.0695	-0.0473
9	81	0.085	0.1973	-0.2855	-0.1959	0.0255	-0.0282	-0.0650
10	79	0.031	0.1379	-0.2217	-0.2421	-0.0582	-0.0441	-0.0370
11	123	0.079	-0.0737	-0.1066	-0.1115	-0.0948	-0.0884	-0.0880
12	77	0.133	0.5723	-0.4431	-0.4119	0.0820	0.0687	0.0507
13	51	0.038	0.2907	-0.4109	-0.2843	0.0429	-0.0338	-0.0864
14	73	0.048	0.5061	-0.3964	-0.3773	0.0618	0.0554	0.0443
15	95	0.056	-0.1091	-0.1414	-0.1450	-0.1270	-0.1218	-0.1238
16	36	0.005	-0.0176	-0.1676	-0.1859	-0.0749	-0.0592	-0.0621
17	117	0.055	-0.1224	-0.1556	-0.1706	-0.1468	-0.1377	-0.1335
18	30	0.012	0.1470	-0.2708	-0.3131	-0.0863	-0.0595	-0.0476
19	83	0.065	-0.0660	-0.1019	-0.1069	-0.0896	-0.0786	-0.0796
20	19	0.006	0.2386	-0.3635	-0.4143	-0.1289	-0.1085	-0.0987
21	51	0.019	-0.0977	-0.1217	-0.1091	-0.1052	-0.0989	-0.1061
22	30	0.009	-0.1670	-0.1708	-0.1493	-0.0899	-0.1846	-0.1438
23	25	0.004	0.2710	-0.3664	-0.2821	0.0192	-0.0089	-0.0698
24	47	0.027	0.2421	-0.2775	-0.3155	-0.0470	-0.0155	0.0005
25	36	0.007	0.2757	-0.4046	-0.2174	0.0755	-0.0404	-0.1008
26	30	0.002	0.9564	-0.7305	-0.8043	0.1499	0.1661	0.1247
27	10	0.001	-0.1721	-0.2399	-0.3072	-0.2366	-0.2314	-0.2346
28	40	0.009	0.3979	-0.3995	-0.4421	-0.0074	0.0133	0.0194
29	6	0.001	0.6029	-1.4085	-0.3500	0.6448	-0.4141	-0.4076