



Working Paper 05-49 (10)
Statistics and Econometrics Series 061
September 2005

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

ANALYTIC AND BOOTSTRAP APPROXIMATIONS OF PREDICTION ERRORS UNDER A MULTIVARIATE FAY-HERRIOT MODEL¹

Wenceslao González-Manteiga², María J. Lombardía², Isabel Molina³, Domingo
Morales⁴ and Laureano Santamaría⁴.

Abstract

A Multivariate Fay-Herriot model is used to aid the prediction of small area parameters of dependent variables with sample data aggregated to area level. The empirical best linear unbiased predictor of the parameter vector is used, and an approximation of the elements of the mean cross product error matrix is obtained by an extension of the results of Prasad and Rao (1990) to the multiparameter case. Three different bootstrap approximations of those elements are introduced, and a simulation study is developed in order to compare the efficiency of all presented approximations, including a comparison under lack of normality. Further, the number of replications needed for the bootstrap procedures to get stabilized are studied.

Keywords: Resampling methods, Bootstrap, superpopulation model, linear mixed regression model, small area, mean squared error.

¹ Supported by grants 06/HSE/0181/2004, BMF2003-04820, BFM2002-03213, GV04B-670 and PGIDIT03PX\-IC20702PN.

² Departamento de Estadística e Investigación Operativa, Universidad de Santiago de Compostela.

³ Departamento de Estadística, Universidad Carlos III de Madrid.

⁴ Centro de Investigación Operativa, Universidad Miguel Hernández de Elche.

Analytic and Bootstrap Approximations of Prediction Errors under a Multivariate Fay-Herriot Model

González-Manteiga, W. and Lombardía, M.J.

*Departamento de Estadística e Investigación Operativa,
Universidad de Santiago de Compostela.*

Molina, I.*

Departamento de Estadística, Universidad Carlos III de Madrid.

Morales, D. and Santamaría, L.

*Centro de Investigación Operativa,
Universidad Miguel Hernández de Elche.*

Abstract: A Multivariate Fay-Herriot model is used to aid the prediction of small area parameters of dependent variables with sample data aggregated to area level. The empirical best linear unbiased predictor of the parameter vector is used, and an approximation of the elements of the mean cross product error matrix is obtained by an extension of the results of Prasad and Rao (1990) to the multiparameter case. Three different bootstrap approximations of those elements are introduced, and a simulation study is developed in order to compare the efficiency of all presented approximations, including a comparison under lack of normality. Further, the number of replications needed for the bootstrap procedures to get stabilized are studied.

Key Words: Resampling methods, Bootstrap, superpopulation model, linear mixed regression model, small area, mean squared error.

AMS Classification (2000): 62D05, 62J05.

1 Introduction

In small area estimation, the sample is extracted from a (large) finite population, but estimates of parameters of small areas contained in the population are required. Direct estimates fail to

*Corresponding author: Isabel Molina, Departamento de Estadística, Universidad Carlos III de Madrid, C/Madrid, 126, 28903 Getafe (Madrid), Tf: +34 916249887, Fax: +34 916249849, Email: isabel.molina@uc3m.es

provide reliable values due to the small number of observations within areas. Small-area models improve the accuracy of direct area estimators by including via the model the information of all sample observations and not just the ones within the corresponding area, and by making use of auxiliary information. However, often individual information is not available because of privacy reasons, but data aggregated to some geographical level can be found in public registers. Fay and Herriot (1979) proposed a model for data aggregated to area level, in order to estimate average per-capita income for small areas in U.S.

Multivariate models are useful when the interest is to estimate some descriptive measures of several correlated variables, or a function of those descriptives. A multivariate Fay-Herriot model was already used by Fay (1987) and Datta, Fay and Ghosh (1991). They compared the small area estimators obtained from univariate models for each response variable with the ones obtained by a multivariate model, showing that the precision was improved by using the multivariate model. Datta, Ghosh, Nangia and Natarajan (1996) used the same type of model for calculating hierarchical Bayes estimates of median income of four-person families for U.S. states.

In the univariate set-up with a scalar parameter, Prasad and Rao (1990) gave an approximation of the mean squared error of the empirical best linear unbiased predictor (EBLUP) of the parameter, and proposed an estimator based on the plug-in principle. But estimators of the mean squared error obtained by resampling techniques are currently competing with analytical approximations. For instance, Jiang, Lahiri and Wan (2002) provided jackknife estimators with reduced bias. Butar and Lahiri (2003) used a parametric bootstrap for calculating an estimator of the mean squared error under a mixed linear model. Pfefferman and Tiller (2002) propose a similar estimator, but under the set-up of State-Space Models, and they propose a second estimator based on nonparametric bootstrap. Further references on this topic can be found in Lahiri (2003).

Here we use a multivariate extension of the Fay-Herriot model to assist the estimation of the mean vectors of a multidimensional response in small areas (Section 2). In this case the mean squared error becomes a matrix, referred here as mean crossed product error matrix (MCPEM). We propose two different estimators of the variance component involved, with spelled out expressions and good properties (Section 3). We extend the Prasad-Rao results to the approximation of the MCPEM (Section 4). Further, in Section 5 we propose three different bootstrap estimators. The first one is obtained applying directly the bootstrap method described in that section. The second one uses bootstrap to estimate just the term of the MCPEM that cannot be explicitly calculated in practice. The third estimator is just a bias-correction of the latter. The consistency of the proposed bootstrap estimators, as the number of areas tends to

infinity, is shown via the consistency of the Prasad-Rao-type estimator (Section 6).

In a simulation study (Section 7) we compare the finite sample properties of the Prasad-Rao extension and the three bootstrap estimators, checking their robustness to the lack of normality and the influence of the number of small areas in the inference problem. Finally, the number of bootstrap iterations needed for the bootstrap estimators to become stable is analyzed.

2 Multivariate Fay-Herriot model

Let P be a finite population of size N , partitioned into D subpopulations called small areas P_d of sizes N_d , $d = 1, \dots, D$. Let $\mathbf{Y} = (Y_1, \dots, Y_r)'$ be the random vector of interest, whose values in the population units are

$$\mathbf{y}_{dj} = (y_{dj1}, \dots, y_{djr})', \quad j = 1, \dots, N_d, \quad d = 1, \dots, D.$$

Let $\mu_{dk} = N_d^{-1} \sum_{j=1}^{N_d} y_{dj k}$ denote the mean of variable Y_k for area d , $k = 1, \dots, r$, and let us denote the vector of means for the d -th area by $\boldsymbol{\mu}_d = (\mu_{d1}, \dots, \mu_{dr})'$, with direct estimator $\bar{\mathbf{y}}_d = (\bar{y}_{d1}, \dots, \bar{y}_{dr})'$, $d = 1, \dots, D$. The model is established by assumptions (M_1) and (M_2) below.

(M_1) The direct estimators $\bar{\mathbf{y}}_d$, given $\boldsymbol{\mu}_d$, are independent, with

$$\bar{\mathbf{y}}_d | \boldsymbol{\mu}_d \sim N(\boldsymbol{\mu}_d, \Sigma_d), \quad d = 1, \dots, D,$$

where the $r \times r$ matrices Σ_d are known.

Assumption (M_1) can be restated in the following way,

$$\bar{\mathbf{y}}_d = \boldsymbol{\mu}_d + \mathbf{e}_d, \quad \mathbf{e}_d \sim N(\mathbf{0}, \Sigma_d), \quad d = 1, \dots, D,$$

where the random errors, $\mathbf{e}_1, \dots, \mathbf{e}_D$, are independent. Furthermore, it is assumed that the means μ_{dk} are linearly related to p explanatory variables. Let $\mathbf{x}_{dk} = (x_{dk1}, \dots, x_{dkp})$ be the vector of values of the explanatory variables associated to the k -th variable Y_k , for d -th area. Let $\boldsymbol{\beta}_k$ be a column vector of size p . Let us define the $r \times rp$ matrix $X_d = \text{diag}\{\mathbf{x}_{d1}, \dots, \mathbf{x}_{dr}\}$ and the rp vector of coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_r)'$. Let $\mathbf{1}_r$ and I_D denote a column vector of ones of size r and the identity matrix of order D respectively. Finally, let us introduce scalar random effects u_d representing the (random) variations between areas not explained by X_1, \dots, X_D .

(M_2) The vectors of area means satisfy

$$\boldsymbol{\mu}_d = X_d \boldsymbol{\beta} + \mathbf{1}_r u_d, \quad u_d \sim N(0, \sigma_u^2), \quad d = 1, \dots, D, \quad (1)$$

where $\sigma_u^2 \in (0, \infty)$ is unknown, the random effects u_1, \dots, u_D are independent, and also independent of the random errors $\mathbf{e}_1, \dots, \mathbf{e}_D$.

Now let us define the vectors and matrices

$$\mathbf{y} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_D \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_D \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_D \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_D \end{pmatrix}.$$

Then the model described by (M_1) and (M_2) can be rewritten in the general form of a linear mixed model

$$\mathbf{y} = X\boldsymbol{\beta} + Z\mathbf{u} + \mathbf{e}, \quad \mathbf{u} \sim N(\mathbf{0}, \Sigma_u), \quad \mathbf{e} \sim N(\mathbf{0}, \Sigma_e), \quad (2)$$

where

$$Z = Z_{rD \times D} = \text{diag}\{\mathbf{1}_r, \dots, \mathbf{1}_r\}, \quad \Sigma_u = \sigma_u^2 I_D, \quad \Sigma_e = \text{diag}\{\Sigma_1, \dots, \Sigma_D\}.$$

The mean vector and covariance matrix of \mathbf{y} are

$$E[\mathbf{y}] = X\boldsymbol{\beta}, \quad \text{Var}[\mathbf{y}] = Z\Sigma_u Z' + \Sigma_e \triangleq V, \quad (3)$$

where V is block diagonal; more explicitly,

$$V = \text{diag}\{V_1, \dots, V_D\}, \quad V_d = \sigma_u^2 \mathbf{1}_r \mathbf{1}_r' + \Sigma_d, \quad d = 1, \dots, D.$$

We are interested in estimating the realized value of the mean μ_{dk} of variable Y_k in small area d , for all $k = 1, \dots, r$ and all $d = 1, \dots, D$, that is, the target parameter is the realized value of the random vector

$$\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_D)' = X\boldsymbol{\beta} + Z\mathbf{u}.$$

When the covariance matrix V is completely known, Henderson (1975) calculated the BLUP of any linear combination of fixed effects $\boldsymbol{\beta}$ and random effects \mathbf{u} in a general model of the form (2). Following his results, the BLUP of $\boldsymbol{\mu}$ is given by

$$\hat{\boldsymbol{\mu}}_B = \hat{\boldsymbol{\mu}}(\sigma_u^2, \mathbf{y}) = X\hat{\boldsymbol{\beta}}_B + Z\hat{\mathbf{u}}_B, \quad (4)$$

where

$$\hat{\boldsymbol{\beta}}_B = \hat{\boldsymbol{\beta}}(\sigma_u^2, \mathbf{y}) = (X^t V^{-1} X)^{-1} X^t V^{-1} \mathbf{y}, \quad \hat{\mathbf{u}}_B = \hat{\mathbf{u}}(\sigma_u^2, \mathbf{y}) = \Sigma_u Z' V^{-1} (\mathbf{y} - X\hat{\boldsymbol{\beta}}_B).$$

But as indicated by the notation, $\hat{\boldsymbol{\beta}}_B$ and $\hat{\mathbf{u}}_B$ depend on σ_u^2 (through V) that is unknown. If we calculate an estimator $\hat{\sigma}_u^2$ of σ_u^2 and we replace it in (4), we get the Empirical BLUP (EBLUP)

$$\hat{\boldsymbol{\mu}}_E = \hat{\boldsymbol{\mu}}(\hat{\sigma}_u^2, \mathbf{y}) = X\hat{\boldsymbol{\beta}}_E + Z\hat{\mathbf{u}}_E,$$

where $\hat{\boldsymbol{\beta}}_E = \hat{\boldsymbol{\beta}}(\hat{\sigma}_u^2, \mathbf{y})$ and $\hat{\mathbf{u}}_E = \hat{\mathbf{u}}(\hat{\sigma}_u^2, \mathbf{y})$. In the following section we describe two possible estimators of σ_u^2 , and we show some asymptotic properties.

3 Estimation of σ_u^2

We use two methods for estimating the variance σ_u^2 , both providing spelled out expressions of the estimators, something useful for analyzing their order of consistency.

3.1 Method of moments

Let $\tilde{\boldsymbol{\beta}}$ be the ordinary least squares estimator of $\boldsymbol{\beta}$ in the model obtained by omitting the random effects, that is, $\tilde{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$, and let $\tilde{\mathbf{u}}$ be the vector of residuals for the same model, that is, $\tilde{\mathbf{u}} = \mathbf{y} - X\tilde{\boldsymbol{\beta}}$. Then,

$$\tilde{\mathbf{u}}'\tilde{\mathbf{u}} = \mathbf{y}'P_1\mathbf{y}, \quad P_1 = I_{Dr} - X(X'X)^{-1}X',$$

where P_1 is symmetric, idempotent and satisfies $P_1X = \mathbf{0}$. Taking expectation, we get

$$E[\tilde{\mathbf{u}}'\tilde{\mathbf{u}}] = E[\mathbf{y}'P_1\mathbf{y}] = \sigma_u^2 \text{tr}\{P_1ZZ'\} + \text{tr}\{P_1\Sigma_e\}.$$

From here, we get the following unbiased estimator of σ_u^2 ,

$$\hat{\sigma}_{u1}^2 = \hat{\sigma}_{u1}^2(\mathbf{y}) = \frac{\mathbf{y}'P_1\mathbf{y} - \text{tr}\{P_1\Sigma_e\}}{\text{tr}\{P_1ZZ'\}},$$

where $\text{tr}\{A\}$ denotes the trace of A . A more operative expression is obtained after some algebra,

$$\hat{\sigma}_{u1}^2 = \frac{\tilde{\mathbf{u}}'\tilde{\mathbf{u}} - \sum_{d=1}^D (\text{tr}\{\Sigma_d\} - \text{tr}\{X'_d\Sigma_d X_d(X'X)^{-1}\})}{Dr - \sum_{d=1}^D \mathbf{1}'_r X_d(X'X)^{-1} X'_d \mathbf{1}_r}. \quad (5)$$

The variance of $\hat{\sigma}_{u1}^2$ is given by

$$\text{Var}[\hat{\sigma}_{u1}^2] = \frac{\text{Var}[\tilde{\mathbf{u}}'\tilde{\mathbf{u}}]}{[Dr - \sum_{d=1}^D \mathbf{1}'_r X_d(X'X)^{-1} X'_d \mathbf{1}_r]^2},$$

where

$$\begin{aligned} \text{Var}[\tilde{\mathbf{u}}'\tilde{\mathbf{u}}] &= \sum_{d=1}^D (\text{tr}\{\Sigma_d^2\} - 2\text{tr}\{X'_d\Sigma_d^2 X_d(X'X)^{-1}\} + \text{tr}\{(X'_d\Sigma_d X_d(X'X)^{-1})^2\}) \\ &\quad + 2\sigma_u^2 \sum_{d=1}^D (\mathbf{1}'_r \Sigma_d \mathbf{1}_r - 2\mathbf{1}'_r X_d(X'X)^{-1} X'_d \mathbf{1}_r + \mathbf{1}'_r X_d(X'X)^{-1} X' \Sigma_e X(X'X)^{-1} X'_d \mathbf{1}_r) \\ &\quad + 2(\sigma_u^2)^2 \left[r^2 D - r \sum_{d=1}^D (\mathbf{1}'_r X_d(X'X)^{-1} X'_d \mathbf{1}_r + \mathbf{1}'_r X_d(X'X)^{-1} X' Z Z' X(X'X)^{-1} X'_d \mathbf{1}_r) \right]. \end{aligned}$$

The following proposition shows the order of consistency of the proposed estimator. We use the notation $f(D) = O(g(D))$ for two functions $f(D)$ and $g(D)$ satisfying $\lim_{D \rightarrow \infty} |f(D)/g(D)| < \infty$. The notation $f(D) = \underline{O}(g(D))$ is used for the more accurate relation $\lim_{D \rightarrow \infty} |f(D)/g(D)| \in$

$(0, \infty)$, and $f(D) = o(g(D))$ is used when the same limit is zero. Further, $f(D) = O_p(g(D))$ and $f(D) = o_p(g(D))$ denote respectively boundedness and convergence to zero in probability of $f(D)/g(D)$. When $f(D)$ is a $m \times n$ matrix whose elements are $O(g(D))$, we write $f(D) = [O(g(D))]_{m \times n}$, and the same brackets notation is used with the rest of symbols of asymptotic order.

Proposition 1. *Under a model defined by (M_1) and (M_2) satisfying the following assumptions*

(H_1) $0 < p < \infty$ and $0 < r < \infty$,

(H_2) $|x_{dk\ell}| \leq x < \infty$, $\ell = 1, \dots, p$, $k = 1, \dots, r$, $d = 1, \dots, D$,

(H_3) the covariance matrices Σ_d , $d = 1, \dots, D$, are positive definite and their elements are uniformly bounded,

(H_4) $X'X = [Q(D)]_{pr \times pr}$,

it holds

$$|\hat{\sigma}_{u1}^2 - \sigma_u^2| = O_p(D^{-1/2}). \quad (6)$$

Proof. By studying the order of each term in the expression of $Var[\hat{\sigma}_{u1}^2]$ when D tends to infinity, it can be proved that $DVar[\hat{\sigma}_{u1}^2] = O(1)$, and this result implies (6). \square

3.2 Henderson Method 3

Using Henderson method 3 (see e.g. Section 5.5 in Searle et al., 1992), an unbiased estimator of σ_u^2 is

$$\hat{\sigma}_{u2}^2 = \hat{\sigma}_{u2}^2(\mathbf{y}) = \frac{\mathbf{y}'P_2\mathbf{y} - \sigma_e^2(n - r(X))}{\text{tr}\{Z'P_2Z\}}, \quad (7)$$

where $r(A)$ denotes the rank of A ,

$$P_2 = \Sigma_e^{-1} - \Sigma_e^{-1}X(X'\Sigma_e^{-1}X)^{-1}X'\Sigma_e^{-1}$$

and it holds $P_2X = \mathbf{0}$. Let us denote $Q = (X'\Sigma_e^{-1}X)^{-1}$. The variance of $\hat{\sigma}_{u2}^2$ is

$$Var[\hat{\sigma}_{u2}^2] = \frac{Var[\mathbf{y}'P_2\mathbf{y}]}{\left[\sum_{d=1}^D(\mathbf{1}'_r\Sigma_d^{-1}\mathbf{1}_r - \mathbf{1}'_r\Sigma_d^{-1}X_dQX'_d\Sigma_d^{-1}\mathbf{1}_r)\right]^2},$$

where

$$\begin{aligned}
\text{Var}[\mathbf{y}'P_2\mathbf{y}] &= 2(\sigma_u^2)^2 \left[\sum_{d=1}^D (\mathbf{1}'_r \Sigma_d^{-1} \mathbf{1}_r)^2 - 2 \sum_{d=1}^D \mathbf{1}'_r \Sigma_d^{-1} \mathbf{1}_r \mathbf{1}'_r \Sigma_d^{-1} X_d Q X_d' \Sigma_d^{-1} \mathbf{1}_r \right. \\
&\quad \left. + \sum_{d=1}^D \mathbf{1}'_r \Sigma_d^{-1} X_d Q \left(\sum_{d=1}^D X_d' \Sigma_d^{-1} \mathbf{1}_r \mathbf{1}'_r \Sigma_d^{-1} X_d \right) Q X_d' \Sigma_d^{-1} \mathbf{1}_r \right] \\
&\quad + 4\sigma_u^2 \left(\sum_{d=1}^D \mathbf{1}'_r \Sigma_d^{-1} \mathbf{1}_r - \sum_{d=1}^D \mathbf{1}'_r \Sigma_d^{-1} X_d Q X_d' \Sigma_d^{-1} \mathbf{1}_r \right) + 2Dr - 2 \sum_{d=1}^D \text{tr}\{X_d' \Sigma_d^{-1} X_d Q\}.
\end{aligned}$$

Proposition 2. Under a model defined by (M_1) and (M_2) satisfying assumptions (H_1) – (H_3) and

$$(H_5) \quad X' \Sigma_e^{-1} X = [Q(D)]_{pr \times pr},$$

$$(H_6) \quad \sum_{d=1}^D \mathbf{1}'_r \Sigma_d^{-1} \mathbf{1}_r = Q(D),$$

it holds

$$|\hat{\sigma}_{u2}^2 - \sigma_u^2| = O_p(D^{-1/2}).$$

Proof. It follows the same steps of the proof of Proposition 1. \square

An estimator $\hat{\theta} = \hat{\theta}(\mathbf{y})$ of a parameter θ is called *translation invariant* when for any $\mathbf{a} \in \mathbb{R}^p$, $\hat{\theta}(\mathbf{y} + X\mathbf{a}) = \hat{\theta}(\mathbf{y})$.

Proposition 3. The estimators $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{u2}^2$ are translation invariant.

Proof. The proof is direct by observing that $P_i X = \mathbf{0}$ and $X' P_i = \mathbf{0}$, $i = 1, 2$. \square

4 Analytic approximation of the mean crossed product error

The mean crossed product error matrix of $\hat{\boldsymbol{\mu}}_E$ is defined as $\mathbf{MCPE}(\hat{\boldsymbol{\mu}}_E) = E[(\hat{\boldsymbol{\mu}}_E - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}}_E - \boldsymbol{\mu})']$, and can be written in the form

$$\begin{aligned}
\mathbf{MCPE}(\hat{\boldsymbol{\mu}}_E) &= \mathbf{MCPE}(\hat{\boldsymbol{\mu}}_B) + E[(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)'] \\
&\quad + E[(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)(\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu})'] + E[(\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)'].
\end{aligned}$$

In this relation, $\mathbf{MCPE}(\hat{\boldsymbol{\mu}}_B)$ accounts for the error in predicting $\boldsymbol{\mu}$ when σ_u^2 is known, and the following term accounts for the increase in error due to estimating σ_u^2 . Further, following Kackar and Harville (1984), it can be proved (using the normality assumption) that for any unbiased and translation invariant estimator of σ_u^2 , the elements of the matrices corresponding to the last

two terms are equal to zero. The mean crossed product error of $\hat{\boldsymbol{\mu}}_B$ can be obtained from the results of Henderson (1975), and it can be expressed in the form

$$MCPE(\hat{\boldsymbol{\mu}}_B) = \mathbf{G}_1(\sigma_u^2) + \mathbf{G}_2(\sigma_u^2),$$

where

$$\mathbf{G}_1(\sigma_u^2) = ZTZ', \quad \mathbf{G}_2(\sigma_u^2) = (X - ZTZ'\Sigma_e^{-1}X)P(X' - X'\Sigma_e^{-1}ZTZ'), \quad (8)$$

and

$$P = (X'V^{-1}X)^{-1}, \quad T = \Sigma_u - \Sigma_u Z'V^{-1}Z\Sigma_u.$$

Prasad and Rao (1990) obtained an approximation of $E[(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)']$ for univariate models ($r = 1$) when predicting a single area mean, but their results are easily generalizable to the multivariate case for the parameter vector $\boldsymbol{\mu}$. For this, let $\hat{\mu}_{B,di} = \hat{\mu}_{di}(\sigma_u^2, \mathbf{y})$ and $\hat{\mu}_{E,di} = \hat{\mu}_{di}(\hat{\sigma}_u^2, \mathbf{y})$ be the di -th components of vectors $\hat{\boldsymbol{\mu}}_B = \hat{\boldsymbol{\mu}}(\sigma_u^2, \mathbf{y})$ and $\hat{\boldsymbol{\mu}}_E = \hat{\boldsymbol{\mu}}(\hat{\sigma}_u^2, \mathbf{y})$ respectively. The approximation is obtained by a first order Taylor expansion of the function $\hat{\mu}_{di}(\omega, \mathbf{y})$, being ω an admissible value of σ_u^2 , around the true value σ_u^2 at point $\omega = \hat{\sigma}_u^2$, that is,

$$\hat{\mu}_{E,di} - \hat{\mu}_{B,di} = s_{di}(\hat{\sigma}_u^2 - \sigma_u^2) + o_p(|\hat{\sigma}_u^2 - \sigma_u^2|), \quad \text{where } s_{di} = \partial \hat{\mu}_{di}(\omega, \mathbf{y}) / \partial \omega |_{\omega = \sigma_u^2}.$$

It is easy to see that $s_{di} = O_p(1)$, $i = 1, \dots, r$, $d = 1, \dots, D$. Applying the Taylor formula to the ℓj -th element, and multiplying both expressions, we get

$$(\hat{\mu}_{E,di} - \hat{\mu}_{B,di})(\hat{\mu}_{E,\ell j} - \hat{\mu}_{B,\ell j}) = s_{di}s_{\ell j}(\hat{\sigma}_u^2 - \sigma_u^2) + o_p((\hat{\sigma}_u^2 - \sigma_u^2)^2).$$

Defining the vector $\mathbf{s} = (s_{11}, \dots, s_{Dr})'$, we can write

$$(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)' = \mathbf{s}\mathbf{s}'(\hat{\sigma}_u^2 - \sigma_u^2)^2 + [o_p((\hat{\sigma}_u^2 - \sigma_u^2)^2)]_{Dr \times Dr}. \quad (9)$$

The following theorem provides an approximation of order $o(D^{-1})$ of the right-hand side term in expression (9). It is a slight modification of Theorem A.1 of Prasad and Rao (1990). The details of the proof can be found in Appendix.

Theorem 1. *Suppose that model (M_1) – (M_2) satisfies (H_1) – (H_3) along with*

$$(H_7) \quad (X'V^{-1}X)^{-1} = [O(D^{-1})]_{pr \times pr},$$

(H₈) $\hat{\sigma}_u^2 = k + \mathbf{y}'C\mathbf{y}$ is an unbiased, consistent and translation invariant estimator of σ_u^2 , where

$$k = O(1) \quad \text{and} \quad C = \text{diag} \{ [O(D^{-1})]_{r \times r}, \dots, [O(D^{-1})]_{r \times r} \} + [O(D^{-2})]_{Dr \times Dr}.$$

Then, for $L = \partial(Z\Sigma_u Z'V^{-1}) / \partial \sigma_u^2$, it holds

$$E[\mathbf{s}\mathbf{s}'(\hat{\sigma}_u^2 - \sigma_u^2)^2] = LVL'Var[\hat{\sigma}_u^2] + [o(D^{-1})]_{Dr \times Dr}.$$

Remark 1. Under the conditions of Proposition 1 (Proposition 2), the estimator $\hat{\sigma}_{u1}^2$ ($\hat{\sigma}_{u2}^2$) satisfies assumption (H_8).

From (9) and Theorem 1, the Prasad-Rao approximation of $MCPE(\hat{\boldsymbol{\mu}}_E)$ is

$$MCPE^{PR}(\hat{\boldsymbol{\mu}}_E) = \mathbf{G}_1(\sigma_u^2) + \mathbf{G}_2(\sigma_u^2) + \mathbf{G}_3(\sigma_u^2), \quad (10)$$

where $\mathbf{G}_1(\sigma_u^2)$ and $\mathbf{G}_2(\sigma_u^2)$ are given by (8), and

$$\mathbf{G}_3(\sigma_u^2) = LVL'Var[\hat{\sigma}_u^2].$$

By extension of Prasad and Rao (1990), an estimator of $MCPE^{PR}(\hat{\boldsymbol{\mu}}_E)$ is given by

$$mcpe^{PR}(\hat{\boldsymbol{\mu}}_E) = \mathbf{G}_1(\hat{\sigma}_u^2) + \mathbf{G}_2(\hat{\sigma}_u^2) + 2\mathbf{G}_3(\hat{\sigma}_u^2),$$

where $\hat{\sigma}_u^2$ is an estimator of σ_u^2 , for instance $\hat{\sigma}_{u1}^2$ or $\hat{\sigma}_{u2}^2$.

5 Bootstrap approximations of the mean crossed product error

In this section we introduce three alternative ways of approximating $MCPE(\hat{\boldsymbol{\mu}}_E)$ by a simple bootstrap procedure. Although the normality assumption is needed in order to prove the consistency of the approximations, the application of the method does not require an imposed probability distribution. In *Steps 1-5* we describe the procedure for computing the *direct* bootstrap estimator of $MCPE(\hat{\boldsymbol{\mu}}_E)$, and later we describe how to obtain the other two approximations.

Step 1. Calculate estimates $\hat{\sigma}_u^2 = \hat{\sigma}_u^2(\mathbf{y})$ and $\hat{\boldsymbol{\beta}}_E = \hat{\boldsymbol{\beta}}(\hat{\sigma}_u^2, \mathbf{y})$ of σ_u^2 and $\boldsymbol{\beta}$ respectively; for example, using the method of moments or the Henderson method 3.

Step 2. Generate D independent copies of a variable W_1 with $E[W_1] = 0$ and $E[W_1^2] = 1$. Construct the vector $\mathbf{u}^* = (u_1^*, \dots, u_D^*)'$, with elements $u_d^* = \hat{\sigma}_u W_1$. The mean vector and covariance matrix of \mathbf{u}^* are respectively $\mathbf{0}_D$ and $\hat{\Sigma}_u = \hat{\sigma}_u^2 I_D$.

Step 3. Generate $r \times D$ independent copies of a random variable W_2 with $E[W_2] = 0$ and $E[W_2^2] = 1$, independent of W_1 . Construct $\mathbf{e}^* = \Sigma_e^{1/2} \mathbf{W}_2$ with $\mathbf{W}_2 = (\mathbf{W}'_{21}, \dots, \mathbf{W}'_{2D})'$ and $\mathbf{W}_{2d} = (W_{2d1}, \dots, W_{2dr})'$ for all $d = 1, \dots, D$. The mean vector of \mathbf{e}^* is $\mathbf{0}_{Dr}$ and its covariance matrix is $\Sigma_e = \text{diag}\{\Sigma_1, \dots, \Sigma_D\}$.

Step 4. Construct the bootstrap model $\bar{\mathbf{y}}_d^* = \boldsymbol{\mu}_d^* + \mathbf{e}_d^*$, $d = 1, \dots, D$, with $\boldsymbol{\mu}_d^* = X_d \hat{\boldsymbol{\beta}}_E + \mathbf{1}_r u_d^*$, which can be rewritten in general form as

$$\mathbf{y}^* = X \hat{\boldsymbol{\beta}}_E + Z \mathbf{u}^* + \mathbf{e}^*. \quad (11)$$

The remaining step requires further notation. Let us denote by E_* , Var_* , Cov_* and $MCPE_*$ the expectation, variance, covariance and mean crossed product error matrix under the probability distribution supplied by bootstrap model (11), given the initial sample.

We define the bootstrap parameter $\boldsymbol{\mu}^* = X\hat{\boldsymbol{\beta}}_E + Z\mathbf{u}^*$ by analogy with the parameter $\boldsymbol{\mu} = X\boldsymbol{\beta} + Z\mathbf{u}$. Under bootstrap model (11), the BLUP of $\boldsymbol{\mu}^*$ is

$$\hat{\boldsymbol{\mu}}_B^* = \hat{\boldsymbol{\mu}}(\hat{\sigma}_u^2, \mathbf{y}^*) = X\hat{\boldsymbol{\beta}}_B^* + Z\hat{\mathbf{u}}_B^*, \quad (12)$$

where $\hat{\boldsymbol{\beta}}_B^* = \hat{\boldsymbol{\beta}}(\hat{\sigma}_u^2, \mathbf{y}^*)$ and $\hat{\mathbf{u}}_B^* = \hat{\mathbf{u}}(\hat{\sigma}_u^2, \mathbf{y}^*) = (\hat{u}_{B1}^*, \dots, \hat{u}_{BD}^*)'$ are the bootstrap versions of $\hat{\boldsymbol{\beta}}_B$ and $\hat{\mathbf{u}}_B$. Now let $\hat{\sigma}_u^{2*} = \hat{\sigma}_u^2(\mathbf{y}^*)$ be the bootstrap estimator of σ_u^2 , obtained from model (11). Replacing such estimator in (12) we get the bootstrap EBLUP

$$\hat{\boldsymbol{\mu}}_E^* = \hat{\boldsymbol{\mu}}(\hat{\sigma}_u^{2*}, \mathbf{y}^*) = X\hat{\boldsymbol{\beta}}_E^* + Z\hat{\mathbf{u}}_E^*, \quad (13)$$

where $\hat{\boldsymbol{\beta}}_E^* = \hat{\boldsymbol{\beta}}(\hat{\sigma}_u^{2*}, \mathbf{y}^*)$ and $\hat{\mathbf{u}}_E^* = \hat{\mathbf{u}}(\hat{\sigma}_u^{2*}, \mathbf{y}^*) = (\hat{u}_{E1}^*, \dots, \hat{u}_{ED}^*)'$. In the same way, the bootstrap MCPPEM of $\hat{\boldsymbol{\mu}}_E^*$ is given by

$$MCPE_*(\hat{\boldsymbol{\mu}}_E^*) = E_*[(\hat{\boldsymbol{\mu}}_E^* - \boldsymbol{\mu}^*)(\hat{\boldsymbol{\mu}}_E^* - \boldsymbol{\mu}^*)'].$$

The *direct* bootstrap estimator of $MCPE(\hat{\boldsymbol{\mu}}_E)$ proposed here is $MCPE^{*1}(\hat{\boldsymbol{\mu}}_E^*) = MCPE_*(\hat{\boldsymbol{\mu}}_E^*)$. In practice, this estimator is approximated via Monte Carlo as described in *Step 5*:

Step 5. Calculate $\boldsymbol{\mu}^{*(b)}$ and $\hat{\boldsymbol{\mu}}_E^{*(b)}$, $b = 1, \dots, B$. The Monte Carlo approximation of the matrix $MCPE^{*1}(\hat{\boldsymbol{\mu}}_E^*)$ is

$$mcppe^{*1}(\hat{\boldsymbol{\mu}}_E^*) = B^{-1} \sum_{b=1}^B (\hat{\boldsymbol{\mu}}_E^{*(b)} - \boldsymbol{\mu}^{*(b)})(\hat{\boldsymbol{\mu}}_E^{*(b)} - \boldsymbol{\mu}^{*(b)})'. \quad (14)$$

Remark 2. Observe that as B tends to infinity, $mcppe^{*1}(\hat{\boldsymbol{\mu}}_E^*)$ is a consistent estimator of $MCPE_*(\hat{\boldsymbol{\mu}}_E^*)$.

The bootstrap method is specially useful for approximating unknown quantities. Thus, it is reasonable to apply this method for estimating only the term of $MCPE(\hat{\boldsymbol{\mu}}_E)$ that cannot be calculated in practice, namely $E[(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)(\hat{\boldsymbol{\mu}}_E - \hat{\boldsymbol{\mu}}_B)']$. Thus, the *term-to-term* bootstrap estimator is defined as

$$MCPE^{*2}(\hat{\boldsymbol{\mu}}_E^*) = \mathbf{G}_1(\hat{\sigma}_u^2) + \mathbf{G}_2(\hat{\sigma}_u^2) + E_*[(\hat{\boldsymbol{\mu}}_E^* - \hat{\boldsymbol{\mu}}_B^*)(\hat{\boldsymbol{\mu}}_E^* - \hat{\boldsymbol{\mu}}_B^*)'].$$

It is known that the quantity $\mathbf{G}_1(\hat{\sigma}_u^2) + \mathbf{G}_2(\hat{\sigma}_u^2)$ is a biased estimator of $\mathbf{G}_1(\sigma_u^2) + \mathbf{G}_2(\sigma_u^2)$. Thus, we define the bias-corrected bootstrap estimator as

$$MCPE^{*3}(\hat{\boldsymbol{\mu}}_E^*) = 2[\mathbf{G}_1(\hat{\sigma}_u^2) + \mathbf{G}_2(\hat{\sigma}_u^2)] - E_*[\mathbf{G}_1(\hat{\sigma}_u^{2*}) + \mathbf{G}_2(\hat{\sigma}_u^{2*})] + E_*[(\hat{\boldsymbol{\mu}}_E^* - \hat{\boldsymbol{\mu}}_B^*)(\hat{\boldsymbol{\mu}}_E^* - \hat{\boldsymbol{\mu}}_B^*)'].$$

The Monte Carlo approximations $mcppe^{*2}(\hat{\boldsymbol{\mu}}_E^*)$ and $mcppe^{*3}(\hat{\boldsymbol{\mu}}_E^*)$ of the bootstrap quantities $MCPE^{*2}(\hat{\boldsymbol{\mu}}_E^*)$ and $MCPE^{*3}(\hat{\boldsymbol{\mu}}_E^*)$ respectively are obtained in a similar way as (14).

6 Consistency of the Bootstrap Estimators

In this section we show that if the extension (10) of the Prasad-Rao approximation is consistent, then the bootstrap estimators 2 and 3 proposed in Section 5 are also consistent. More concretely, following the arguments of Section 5, the analogue Prasad-Rao approximation $MCPE_*^{PR}(\hat{\boldsymbol{\mu}}_E^*)$ of the matrix $MCPE_*(\hat{\boldsymbol{\mu}}_E^*)$ can be obtained. Here we prove that $MCPE_*^{PR}(\hat{\boldsymbol{\mu}}_E^*)$ is a consistent estimator of $MCPE^{PR}(\hat{\boldsymbol{\mu}}_E)$ by the method of *imitation* (see Shao and Tu, 1995, p.76), where the consistency property is defined with respect to the probability distribution provided by bootstrap model (11). Thus, if $MCPE^{PR}(\hat{\boldsymbol{\mu}}_E)$ is a consistent estimator of $MCPE(\hat{\boldsymbol{\mu}}_E)$, analogously $MCPE_*^{PR}(\hat{\boldsymbol{\mu}}_E^*)$ is consistent for $MCPE_*(\hat{\boldsymbol{\mu}}_E^*)$, and this implies that $MCPE_*(\hat{\boldsymbol{\mu}}_E^*)$ is consistent for $MCPE(\hat{\boldsymbol{\mu}}_E)$.

The consistency is provided by the good properties of the bootstrap model (11). Observe that the vectors \mathbf{u}^* and \mathbf{e}^* generated in *Steps 2* and *3* are independent. Furthermore, the expectation and the covariance matrix of \mathbf{y}^* imitate those of \mathbf{y} ,

$$\begin{aligned} E_*[\mathbf{y}^*] &= E_*[X\hat{\boldsymbol{\beta}}_E + Z\mathbf{u}^* + \mathbf{e}^*] = X\hat{\boldsymbol{\beta}}_E, \\ Var_*[\mathbf{y}^*] &= ZE_*[\mathbf{u}^*\mathbf{u}^{*'}]Z' + E_*[\mathbf{e}^*\mathbf{e}^{*'}] = \hat{\sigma}_u^2 ZZ' + \Sigma_e = \hat{V} = \text{diag}\{\hat{V}_d; d = 1, \dots, D\}. \end{aligned} \quad (15)$$

Proposition 4. *Under assumptions (M_1) , (M_2) and (H_1) – (H_6) , it holds that*

$$E_*[\mathbf{y}^*] - E[\mathbf{y}] = [o_p(1)]_{Dr \times 1} \quad \text{and} \quad Var_*[\mathbf{y}^*] - Var[\mathbf{y}] = [o_p(1)]_{Dr \times Dr}.$$

Proof. For $\hat{\sigma}_{u1}^2$, the result is immediate by applying (3), (15) and Proposition 1. In the case of $\hat{\sigma}_{u2}^2$, the result follows from (3), (15) and Proposition 2. \square

Proposition 5 establishes the results needed for obtaining the approximation of the mean crossed product error matrix (Theorem 2), following the arguments of Prasad and Rao (1990). Its proof is analogous to Proposition 1 or Proposition 2 depending on the considered estimator of σ_u^2 , but taking into account that the probability distribution is conditional on the initial sample. Thus, symbols O_{p^*} and o_{p^*} indicate respectively boundedness and convergence to zero in probability, under the probability distribution given by model (11).

Proposition 5. *Let model (11) satisfies assumptions (M_1) , (M_2) and (H_1) – (H_4) . Then*

- (i) $|\hat{\sigma}_{u1}^{*2} - \hat{\sigma}_{u1}^2| = O_{p^*}(D^{-1/2})$;
- (ii) *assuming additionally (H_5) and (H_6) , it holds $|\hat{\sigma}_{u2}^{*2} - \hat{\sigma}_{u2}^2| = O_{p^*}(D^{-1/2})$;*
- (iii) *the estimators $\hat{\sigma}_{ui}^{*2}$ ($i = 1, 2$) are unbiased, consistent and translation invariant for σ_u^2 , and can be expressed as $\hat{\sigma}_{ui}^{*2} = k + \mathbf{y}^{*'}C\mathbf{y}^*$, where*

$$k = O(1) \quad \text{and} \quad C = \text{diag}\{[O(D^{-1})]_{r \times r}, \dots, [O(D^{-1})]_{r \times r}\} + [O(D^{-2})]_{Dr \times Dr}.$$

Proof. The proofs of (i) and (ii) follow the same steps of that of Propositions 1 and 2 respectively. Note that the normality assumption is required. Finally, (iii) follows from the construction of both estimators. \square

Theorem 2. *Let us suppose that model (11) has been generated as described in Steps 1–4 with $W_i \sim N(0, 1)$ ($i = 1, 2$) and with an estimator $\hat{\sigma}_u^2 \in (0, \infty)$ satisfying (H_8) . Assume further assumptions (H_1) – (H_6) , and (H_7) for the matrix V evaluated at any admissible value of σ_u^2 . Then,*

$$E[\mathbf{s}^* \mathbf{s}^{*'}(\hat{\sigma}_u^{*2} - \hat{\sigma}_u^2)] = \hat{L} \hat{V} \hat{L}' \text{Var}_*[\hat{\sigma}_u^{*2}] + [o_P(D^{-1})]_{Dr \times Dr},$$

where $\mathbf{s}^* = (s_{11}^*, \dots, s_{Dr}^*)'$ with components $s_{di}^* = \partial \hat{\mu}_{di}(\omega, \mathbf{y}^*) / \partial \omega|_{\omega = \hat{\sigma}_u^2}$, $i = 1, \dots, r$, $d = 1, \dots, D$, $\hat{V} = V(\hat{\sigma}_u^2)$, $\hat{L} = \partial(Z \hat{\Sigma}_u Z' \hat{V}^{-1}) / \partial \hat{\sigma}_u^2$, and where $\text{Var}_*[\hat{\sigma}_u^{*2}]$ is the bootstrap version of $\text{Var}[\hat{\sigma}_u^2]$ (see Section 4).

Proof. Under the imposed conditions, taking into account the normality of W_i and the fact that $\hat{\sigma}_u^{*2} = \hat{\sigma}_u^2(\mathbf{y}^*)$ satisfies (H_8) , the proof follows similar steps to that of Theorem 1 (Appendix), but now under bootstrap model (11). \square

By Theorem 2, the quantity $\mathbf{MCPE}_*(\hat{\boldsymbol{\mu}}_E^*)$ can be approximated in the same sense as in Prasad and Rao (1990), by

$$\mathbf{MCPE}_*^{PR}(\hat{\boldsymbol{\mu}}_E^*) = \mathbf{G}_1(\hat{\sigma}_u^2) + \mathbf{G}_2(\hat{\sigma}_u^2) + \mathbf{G}_3(\hat{\sigma}_u^2),$$

where

$$\mathbf{G}_1(\hat{\sigma}_u^2) = Z \hat{T} Z', \quad \mathbf{G}_2(\hat{\sigma}_u^2) = (X - Z \hat{T} Z' \Sigma_e^{-1} X) \hat{P} (X - X \Sigma_e^{-1} Z \hat{T} Z'), \quad \mathbf{G}_3(\hat{\sigma}_u^2) = \hat{L} \hat{V} \hat{L}' \text{Var}_*[\hat{\sigma}_u^{*2}].$$

and where \hat{T} and \hat{P} are the empirical versions of T and P , that is,

$$\hat{T} = \hat{\Sigma}_u - \hat{\Sigma}_u Z' \hat{V}^{-1} Z \hat{\Sigma}_u, \quad \hat{P} = (X' \hat{V}^{-1} X)^{-1}.$$

In the following theorems we use notation $\text{MCPE}_{\ell m}^{PR}(\hat{\boldsymbol{\mu}}_E)$ and $\text{MCPE}_{*\ell m}^{PR}(\hat{\boldsymbol{\mu}}_E^*)$ for the (ℓ, m) -th element of the matrices $\mathbf{MCPE}^{PR}(\hat{\boldsymbol{\mu}}_E)$ and $\mathbf{MCPE}_*^{PR}(\hat{\boldsymbol{\mu}}_E^*)$ respectively.

Theorem 3. *Under the assumptions of Theorem 2 with $\hat{\sigma}_u^2 = \hat{\sigma}_{u1}^2$ or $\hat{\sigma}_u^2 = \hat{\sigma}_{u2}^2$, it holds that*

$$|\text{MCPE}_{*\ell m}^{PR}(\hat{\boldsymbol{\mu}}_E^*) - \text{MCPE}_{\ell m}^{PR}(\hat{\boldsymbol{\mu}}_E)| = O_p(D^{-1/2}), \quad \ell, m = 1, \dots, r.$$

Proof. By (10) and Theorem 2, we know that

$$\text{MCPE}_{\ell m}^{PR}(\hat{\boldsymbol{\mu}}_E) = \sum_{k=1}^3 G_{k\ell m}(\sigma_u^2) \quad \text{and} \quad \text{MCPE}_{*\ell m}^{PR}(\hat{\boldsymbol{\mu}}_E^*) = \sum_{k=1}^3 G_{k\ell m}(\hat{\sigma}_u^2),$$

where $G_{k\ell m}(\sigma_u^2)$ denote the (ℓ, m) -th element of $\mathbf{G}_k(\sigma_u^2)$, $k = 1, 2, 3$. Taking into account Proposition 1 if $\hat{\sigma}_u^2 = \hat{\sigma}_{u1}^2$ and Proposition 2 if $\hat{\sigma}_u^2 = \hat{\sigma}_{u2}^2$, and then subtracting term to term we get, for any pair (ℓ, m) ,

$$|G_{k\ell m}(\hat{\sigma}_u^2) - G_{k\ell m}(\sigma_u^2)| = O_p(D^{-1/2}), \quad k = 1, 2, 3. \quad \square$$

Remark 3. Under normality, the consistency of $MCPE^{*2}(\hat{\boldsymbol{\mu}}_E^*)$ and $MCPE^{*3}(\hat{\boldsymbol{\mu}}_E^*)$ are clear because the term $E_*[(\hat{\boldsymbol{\mu}}_E^* - \hat{\boldsymbol{\mu}}_B^*)(\hat{\boldsymbol{\mu}}_E^* - \boldsymbol{\mu}_B^*)']$ is consistently approximated by $\mathbf{G}_3(\hat{\sigma}_u^2)$.

7 Simulation Study

In this section we describe a simulation study designed for analyzing the accuracy of the four presented estimators of the mean crossed product error matrix, namely the extension of the Prasad-Rao estimator and the three bootstrap-based estimators, when the number of areas D is finite.

We start describing how the data were simulated. The sample vectors $\mathbf{y}_d = (y_{d1}, y_{d2})'$, $d = 1, \dots, D$ were generated following a bivariate normal mixed model with random effects associated to areas. Two covariates were considered, each one affecting one of the dependent variables. The values $(x_{d1}, x_{d2})'$ of the two covariates in the areas were generated from a bivariate normal distribution, with means $\mu_1 = \mu_2 = 10$, variances $\sigma_{xd11} = 1 + \delta_x(d-1)/(D-1)$ and $\sigma_{xd22} = 2 + 2\delta_x(d-1)/(D-1)$ for $\delta_x = 0, 1$, and covariance $\sigma_{xd12} = \rho_x \sqrt{\sigma_{xd11}\sigma_{xd22}}$, for $\rho_x = 0, 1/2$. Random effects associated to areas u_d were generated independently from a normal distribution with zero mean and variance $\sigma_u^2 = 2$. The vector of values of the random errors $(e_{d1}, e_{d2})'$ was simulated from a bivariate normal distribution with zero mean vector, and in a multivariate heteroskedastic framework, where the variances and the covariance are proportional to the values of the two covariates. More explicitly, we define weights $w_d = (x_{d1}^2 + x_{d2}^2)^{-\ell/2}$, $\ell = 0, 1/2$, and then the elements of the covariance matrices Σ_d are $\sigma_{dij} = \sigma_e^2 r_{ij}/w_d$ ($i, j = 1, 2$), for $\sigma_e^2 = 1$, $r_{11} = 1$, $r_{22} = 2$, $r_{12} = \rho_e \sqrt{r_{11}r_{22}}$ and $\rho_e = 1/2$. Finally, the vector of values of the two response variables were generated, taking $\beta_1 = \beta_2 = 1$, from the model

$$y_{dk} = \beta_k x_{dk} + u_d + e_{dk}, \quad k = 1, 2, \quad d = 1, \dots, D. \quad (16)$$

For the sake of brevity, we only present the numerical results obtained with constant variances ($\delta_x = 0$), dependent covariates ($\rho_x = 0.5$) and in the heteroskedastic framework ($\ell = 0.5$).

A preliminary simulation study was carried out in order to analyze the precision of model parameter estimates, with $I = 1000$ simulations of samples with sizes $D = 50, 100, 200, 500, 1000$. The Monte Carlo approximations of the mean squared errors of the estimators of model param-

eters β_1, β_2 and σ_u^2 , calculated using both the method of moments ($\hat{\sigma}_{u1}^2$) and Henderson method 3 ($\hat{\sigma}_{u2}^2$), are listed respectively in Tables 1 and 2 below.

D	50	100	200	500	1000
$EMSE(\hat{\beta}_{E1})$	0.001395	0.000706	0.000366	0.000151	0.000076
$EMSE(\hat{\beta}_{E2})$	0.002249	0.001018	0.000507	0.000211	0.000112
$EMSE(\hat{\sigma}_{u1}^2)$	2.960361	1.469272	0.723034	0.297036	0.155238

Table 1. Mean squared errors of model parameter estimators using the method of moments, with $\delta_x = 0$, $\rho_x = 0.5$ and $\ell = 0.5$.

D	50	100	200	500	1000
$MSE(\hat{\beta}_{E1})$	0.001396	0.000706	0.000366	0.000151	0.000076
$MSE(\hat{\beta}_{E2})$	0.002249	0.001018	0.000507	0.000211	0.000112
$MSE(\hat{\sigma}_{u2}^2)$	2.971057	1.435667	0.689207	0.291275	0.152444

Table 2. Mean squared errors of model parameter estimators using Henderson method 3, with $\delta_x = 0$, $\rho_x = 0.5$ and $\ell = 0.5$.

It can be seen within both tables that the estimates of model coefficients β_1, β_2 are much more precise than the estimate of the variance σ_u^2 . But as it has been theoretically shown, the precision of the variance estimator improves when the number of areas D increases. The two estimation methods present almost the same results, but $\hat{\sigma}_{u2}^2$ is slightly better for moderate number of areas ($D = 100, 200$).

The main simulation study, designed for comparing the proposed estimators of the mean crossed product error matrix, followed the scheme described by items 1-5 below.

1. Generate independently $I = 1000$ samples $\mathbf{y}^{(i)} = (y_{11}^{(i)}, y_{12}^{(i)}, \dots, y_{D1}^{(i)}, y_{D2}^{(i)})'$ as described at the beginning of current section, and calculate mean vectors $\boldsymbol{\mu}^{(i)} = (\mu_{11}^{(i)}, \mu_{12}^{(i)}, \dots, \mu_{D1}^{(i)}, \mu_{D2}^{(i)})'$, $i = 1, \dots, I$.
2. Calculate estimators $\hat{\sigma}_u^{2(i)} = \hat{\sigma}_u^2(\mathbf{y}^{(i)})$ and $\hat{\boldsymbol{\beta}}_E^{(i)} = \hat{\boldsymbol{\beta}}(\hat{\sigma}_u^{2(i)}, \mathbf{y}^{(i)})$ $i = 1, \dots, I$.
3. For $i = 1, \dots, I$, compute predictors $\hat{\boldsymbol{\mu}}_{Ed}^{(i)} = (\hat{\mu}_{Ed1}^{(i)}, \hat{\mu}_{Ed2}^{(i)})'$, and Prasad-Rao-type estimators $\mathbf{mcp}e_d^{PR(i)} = \mathbf{mcp}e_d^{PR}(\hat{\boldsymbol{\mu}}_{Ed}^{(i)})$, $d = 1, \dots, D$.
4. From estimators $\hat{\sigma}_u^{2(i)}$ and $\hat{\boldsymbol{\beta}}_E^{(i)}$, generate B bootstrap samples as described in Section 5 with W_1 and W_2 generated from standard normal distributions (see *Step 2-3* in Section 5), and compute the bootstrap approximations $\mathbf{mcp}e_d^{*a(i)} = \mathbf{mcp}e_d^{*a}(\hat{\boldsymbol{\mu}}_{Ed}^{(i)})$, for $a = 1, 2, 3$.

5. For each area d , calculate the Monte Carlo approximations of theoretical mean crossed product error matrix $MCPE_d$ and of the four matrix estimators, by formulas

$$MCPE_d = \frac{1}{I} \sum_{i=1}^I (\hat{\boldsymbol{\mu}}_{Ed}^{(i)} - \boldsymbol{\mu}_d^{(i)})(\hat{\boldsymbol{\mu}}_{Ed}^{(i)} - \boldsymbol{\mu}_d^{(i)})',$$

$$mcp_e_d^{PR} = \frac{1}{I} \sum_{i=1}^I mcp_e_d^{PR(i)}, \quad mcp_e_d^{*a} = \frac{1}{I} \sum_{i=1}^I mcp_e_d^{*a(i)}, \quad a = 1, 2, 3.$$

Calculate also the mean squared error over simulations of each element (ℓ, m) of the four matrices; that is, calculate for each $\ell, m = 1, 2$ and for $a = 1, 2, 3$,

$$E_{d\ell m}^{PR} = \frac{1}{I} \sum_{i=1}^I (mcp_e_{d\ell m}^{PR(i)} - MCPE_{d\ell m})^2, \quad E_{d\ell m}^{*a} = \frac{1}{I} \sum_{i=1}^I (mcp_e_{d\ell m}^{*a(i)} - MCPE_{d\ell m})^2.$$

In small area estimation problems, the number of areas may differ considerably from one application to another. In order to compare the estimators for different number of areas, the algorithm described above has been run for all $D \in \mathcal{D}$, where $\mathcal{D} = \{50, 100, 200, 300, 400, 500\}$.

In Figure 1, the line with label $MCPE_{d11}$ represents the median of $\{MCPE_{d11}, d = 1, \dots, D\}$ for each $D \in \mathcal{D}$, being $MCPE_{d11}$ the element $(1, 1)$ of the $MCPE_d$ matrix. The remaining lines correspond to the medians over areas of the Prasad-Rao estimators $mcp_e_{d11}^{PR}$, and of the three bootstrap estimators $mcp_e_{d11}^{*1}$, $mcp_e_{d11}^{*2}$ and $mcp_e_{d11}^{*3}$. Figures 2 and 3 represent the same quantities for the elements $(1, 2)$ and $(2, 2)$ respectively of the same matrices. The results were obtained by fitting the model using the method of moments estimator of σ_u^2 . The results for Henderson method 3 are omitted due to the great similarity.

Relative patterns are similar in the three pictures. The medians over areas of the Prasad-Rao estimator $mcp_e_{d11}^{PR}$ and the direct bootstrap estimator $mcp_e_{d11}^{*1}$ show a negative bias for all D . The other two estimators behave almost the same for $D \geq 100$, but it is observed that the bias-uncorrected bootstrap estimator works better than the bias-corrected estimator in median for $D = 50$.

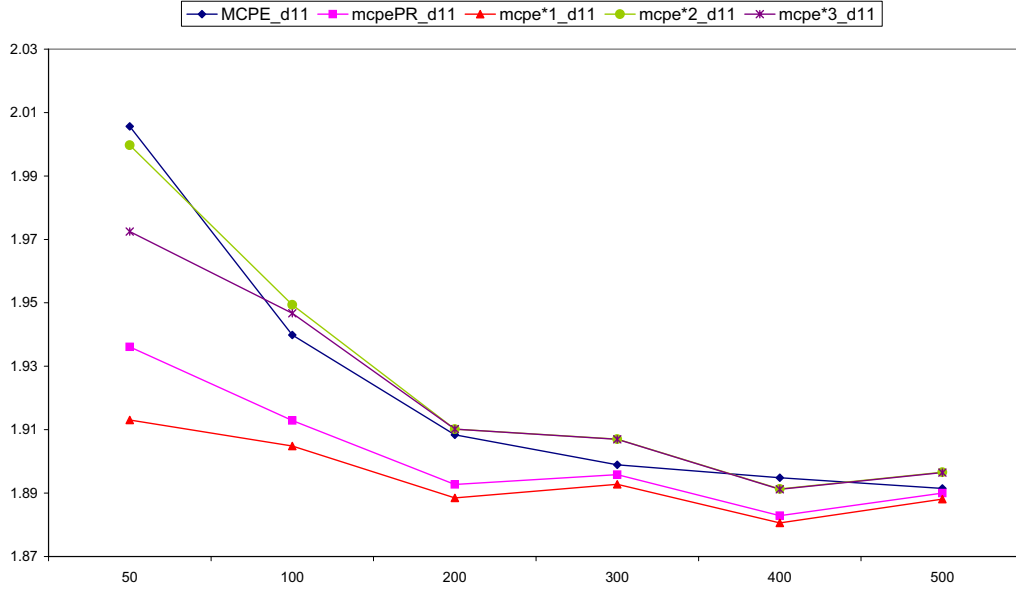


Figure 1. Medians over areas of $MCPE_{d11}$, $mcpe_{d11}^{PR}$, and $mcpe_{d11}^{*a}$, $a = 1, 2, 3$, versus D .

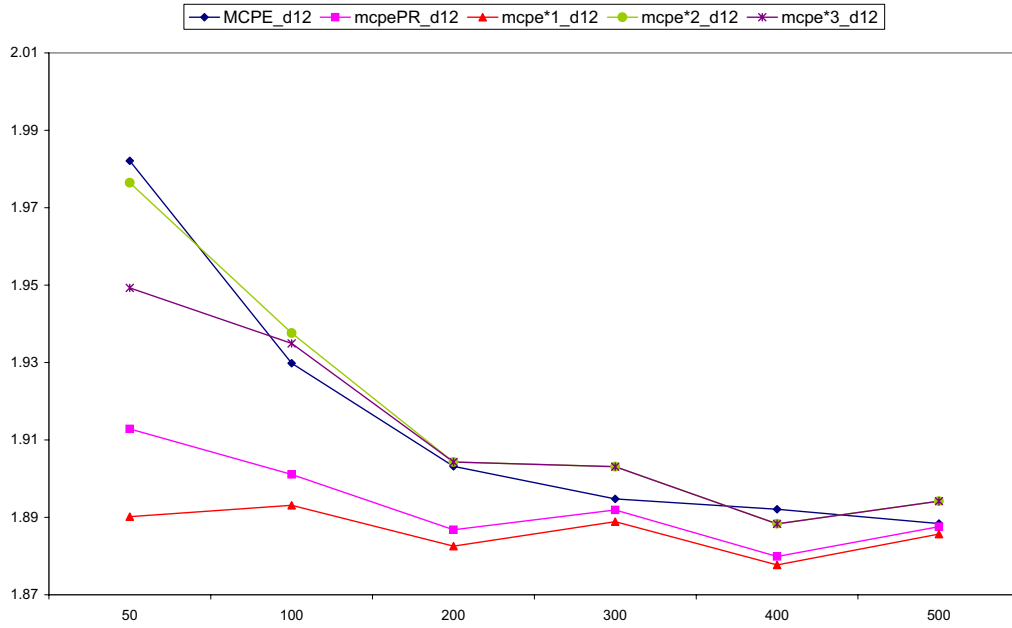


Figure 2. Medians over areas of $MCPE_{d12}$, $mcpe_{d12}^{PR}$, and $mcpe_{d12}^{*a}$, $a = 1, 2, 3$ versus D .

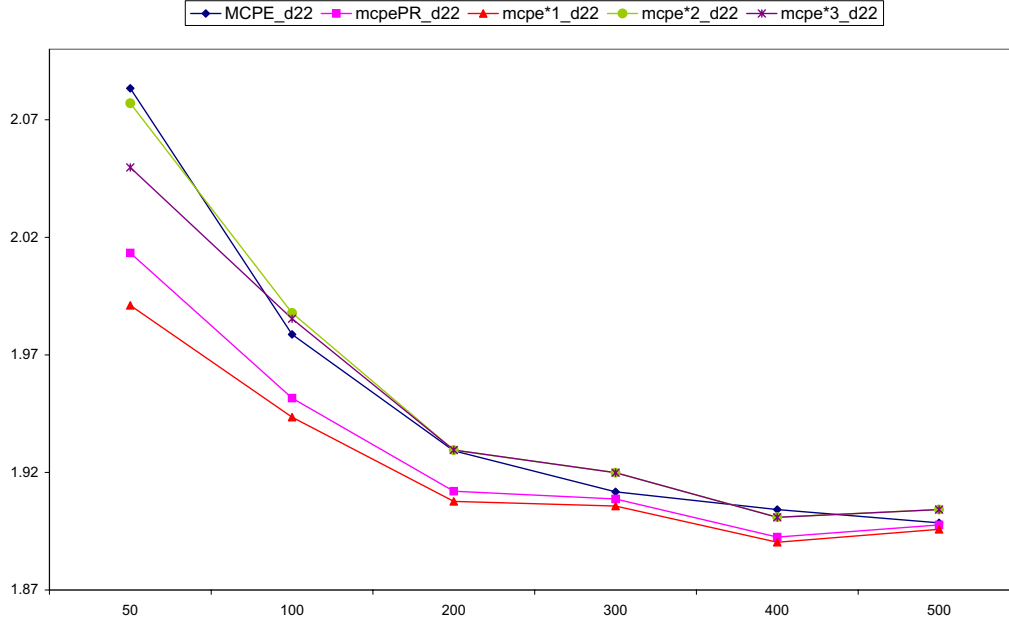


Figure 3. Medians over areas of $MCPE_{d22}$, $mcpe_{d22}^{PR}$ and $mcpe_{d22}^{*a}$, $a = 1, 2, 3$, versus D .

In order to get more insight into the behavior of the estimators along iterations, we have calculated, for each pair (ℓ, m) and for each $D \in \mathcal{D}$, the medians over the D areas of the mean squared errors $E_{d\ell m}^{PR}$, $E_{d\ell m}^{*1}$, $E_{d\ell m}^{*2}$ and $E_{d\ell m}^{*3}$ of the estimators $mcpe_{d\ell m}^{PR}$, $mcpe_{d\ell m}^{*1}$, $mcpe_{d\ell m}^{*2}$ and $mcpe_{d\ell m}^{*3}$ respectively. The results are listed in Table 3. In this table we can see how for $D = 50$, the Prasad-Rao-type estimator and the term-to-term bootstrap estimators have similar result, while for $D > 50$ the bootstrap estimators $mcpe_{d\ell m}^{*2}$ and $mcpe_{d\ell m}^{*3}$ present slightly lower mean squared error.

D	E_{d11}^{PR}	E_{d11}^{*1}	E_{d11}^{*2}	E_{d11}^{*3}	E_{d12}^{PR}	E_{d12}^{*1}	E_{d12}^{*2}	E_{d12}^{*3}	E_{d22}^{PR}	E_{d22}^{*1}	E_{d22}^{*2}	E_{d22}^{*3}
50	0.176	0.198	0.173	0.218	0.176	0.198	0.173	0.217	0.183	0.206	0.180	0.225
100	0.087	0.095	0.082	0.086	0.087	0.095	0.082	0.085	0.090	0.098	0.085	0.088
200	0.041	0.048	0.039	0.039	0.041	0.048	0.039	0.039	0.042	0.049	0.041	0.041
300	0.028	0.035	0.028	0.028	0.028	0.035	0.028	0.028	0.029	0.036	0.028	0.028
400	0.023	0.030	0.023	0.023	0.023	0.030	0.023	0.023	0.024	0.031	0.023	0.023
500	0.019	0.026	0.019	0.019	0.019	0.027	0.019	0.019	0.020	0.027	0.020	0.020

Table 3. Medians over areas of mean squared errors $E_{d\ell m}^{PR}$, $E_{d\ell m}^{*1}$, $E_{d\ell m}^{*2}$ and $E_{d\ell m}^{*3}$.

A further analysis has been done attending to the percentage of areas where the Prasad-Rao-type estimator presents smaller mean squared error than each bootstrap estimator. Thus, for each D , we define the quantities $P_{\ell m}^{*1}$, $P_{\ell m}^{*2}$ and $P_{\ell m}^{*3}$ as the percentage of the D areas where the respective differences $E_{d\ell m}^{PR} - E_{d\ell m}^{*1}$, $E_{d\ell m}^{PR} - E_{d\ell m}^{*2}$ and $E_{d\ell m}^{PR} - E_{d\ell m}^{*3}$ are negative. The results are listed in Table 4. We can see how the Prasad-Rao-type estimator gains in all areas to the direct bootstrap estimator for any value of D , and for $D = 50$ it gains in all areas to the bias-corrected estimator $mcpe_{d\ell m}^{*3}$. However, for $D \geq 100$, the latter estimator presents less mean squared error in most of the areas. The best results are for the term-to-term bootstrap estimator $mcpe_{d\ell m}^{*2}$, since even for $D = 50$ it has lower MSE for the most of areas. For D large all estimators become very close, and the given percentages are not as representative. It is important to mention that, as already observed in Table 3, the differences in the mean squared errors of the estimators are small.

D	P_{11}^{*1}	P_{11}^{*2}	P_{11}^{*3}	P_{12}^{*1}	P_{12}^{*2}	P_{12}^{*3}	P_{22}^{*1}	P_{22}^{*2}	P_{22}^{*3}
50	1.00	0.22	1.00	1.00	0.20	1.00	1.00	0.18	1.00
100	1.00	0.02	0.11	1.00	0.01	0.09	1.00	0.01	0.09
200	1.00	0.06	0.07	1.00	0.07	0.07	1.00	0.07	0.08
300	1.00	0.23	0.29	1.00	0.24	0.29	1.00	0.24	0.29
400	1.00	0.20	0.20	1.00	0.20	0.19	1.00	0.21	0.19
500	1.00	0.40	0.42	1.00	0.40	0.42	1.00	0.39	0.42

Table 4. Percentage of areas $P_{\ell m}^{*1}$, $P_{\ell m}^{*2}$ and $P_{\ell m}^{*3}$ where the mean squared errors of $mcpe_{d\ell m}^{*1}$, $mcpe_{d\ell m}^{*2}$ and $mcpe_{d\ell m}^{*3}$ are greater than that of $mcpe_{d\ell m}^{PR}$.

The normality assumption is essential for deriving the Prasad-Rao approximation, and thus for applying the arguments given in Section 6 about the consistency of the bootstrap estimators. However, the bootstrap procedure described in Section 5 does not require normality, and can be used in practice even with unknown probability distribution. In this sense, it can be regarded as a nonparametric bootstrap. In order to study the robustness of the estimators in a nonparametric setting with absence of normality, a second simulation study was carried out. In this simulation, the initial random effects u_d were generated from a Gumbel distribution with zero mean and variance $\sigma_u^2 = 2$, and the vector of random errors e_d from a bivariate Logistic with mean vector equal to zero, and covariance matrix equal to $\Sigma_d = (\sigma_{dij})_{i,j=1,2}$, with $\sigma_{dij} = \sigma_e^2 r_{ij}/w_d$, where $r_{11} = 1$, $r_{22} = 2$, $r_{12} = \rho_e \sqrt{r_{11}r_{22}}$, for $\rho_e = 1/2$ and $\sigma_e^2 = 1$. The bootstrap random effects u_d^* and the random errors e_d^* were generated as in *Step 2-3* of Section 5 with W_1 and W_2 simulated

from a standard normal distribution.

In the obtained results, the direct bootstrap estimator $mcpe_{d11}^{*1}$ presented uniformly greater mean squared errors for all areas. Thus, in Figure 4 we plotted the mean squared errors over simulations of the remaining three estimators $mcpe_{d11}^{PR}$, $mcpe_{d11}^{*2}$ and $mcpe_{d11}^{*3}$. It can be appreciated that in the most of areas, the two bootstrap estimators get very similar results, and both are more accurate than the Prasad-Rao extension. For the components (1,2) and (2,2) of the matrices, the figures show similar results.

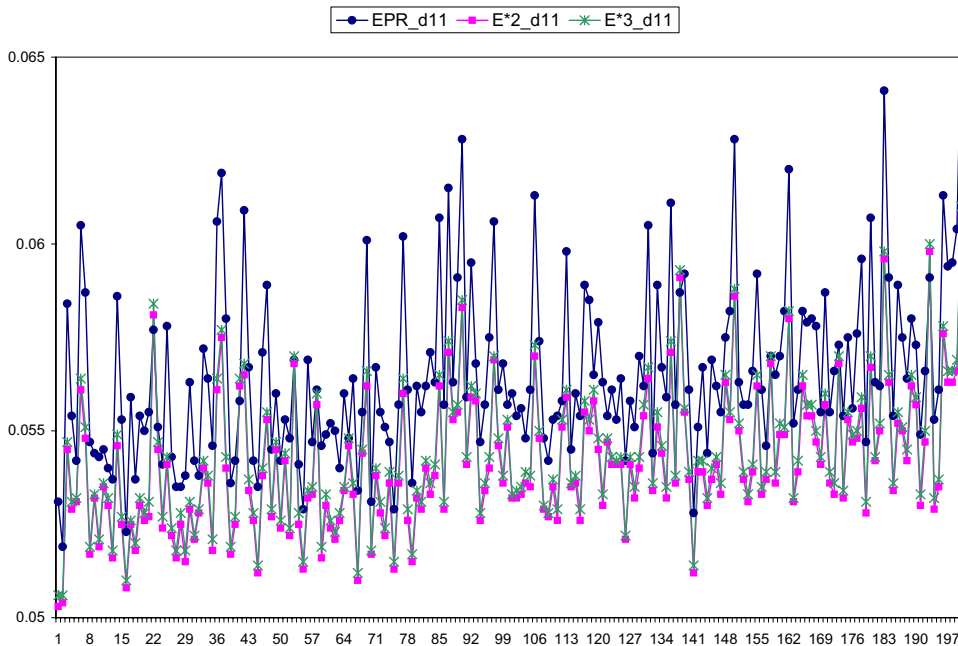


Figure 4. Mean squared errors of $mcpe_{d11}^{PR}$, $mcpe_{d11}^{*2}$ and $mcpe_{d11}^{*3}$ by areas.

Practitioners may have the reasonable interest in knowing which number of bootstrap iterations B is enough in order to get a desired precision of bootstrap estimators. Of course, the answer to this question depends on the particular data set at hand, but in order to put some light to this point, simulations have been repeated for increasing values of B , for $D = 200$ areas, and two arbitrary areas have been selected, concretely areas with $d = 60$ and $d = 180$. The mean squared error over simulations of the three proposed bootstrap estimators of the element (1,1) of $MCPE_d(\hat{\mu}_{Ed})$ are shown in Figure 5 for $d = 60$. In Figure 6, the same results are shown for $d = 180$.

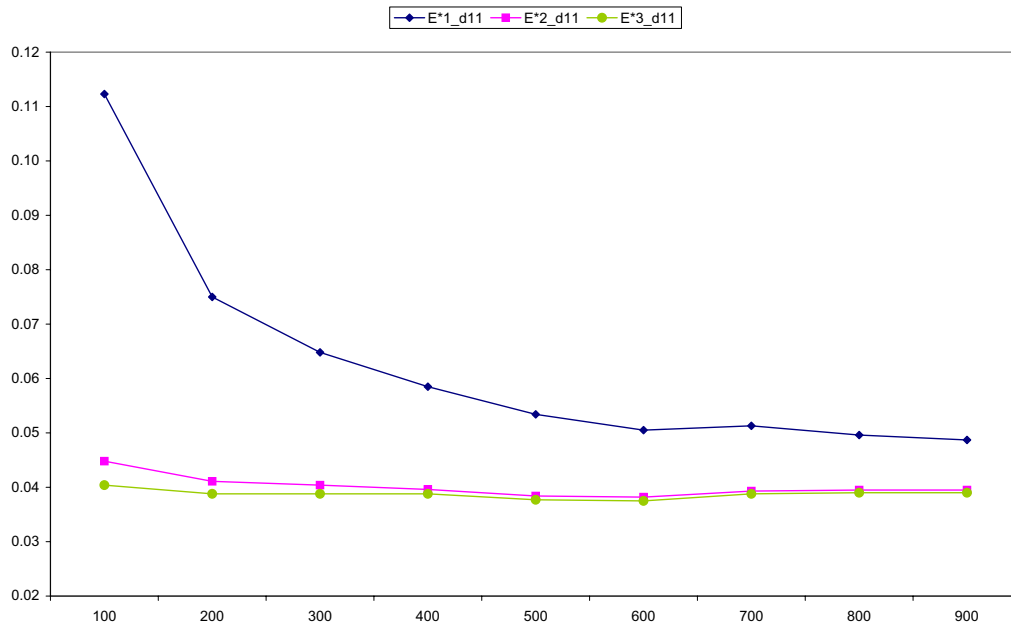


Figure 5. Mean squared errors of mcp_{d11}^{*1} , mcp_{d11}^{*2} and mcp_{d11}^{*3} for $d = 60$ versus B .

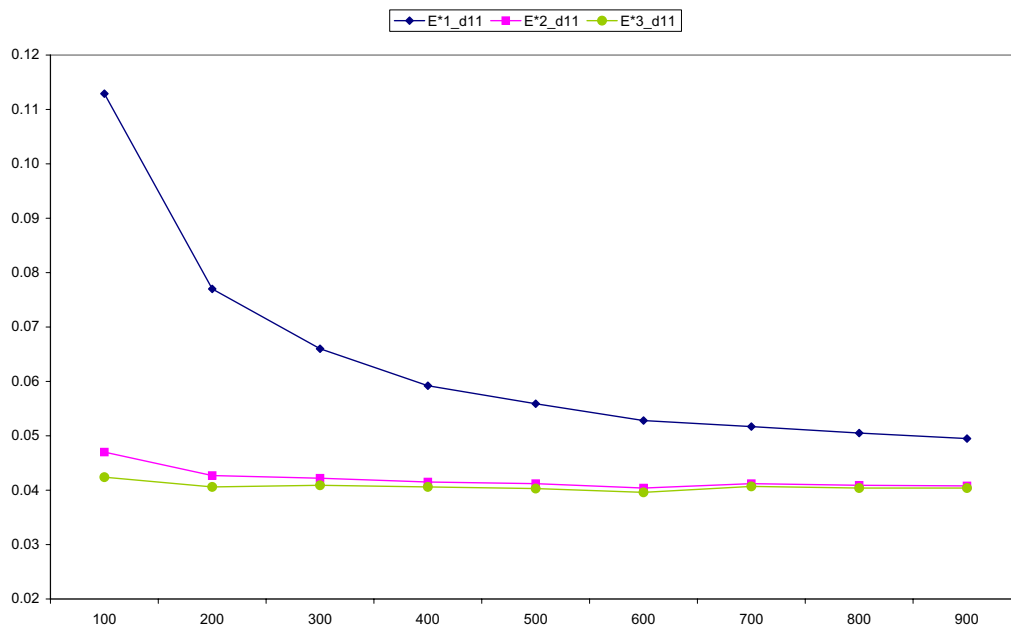


Figure 6. Mean squared errors of mcp_{d11}^{*1} , mcp_{d11}^{*2} and mcp_{d11}^{*3} for $d = 180$ versus B .

We observe in both figures that the mean squared error of the bootstrap estimators $mcpe_{d11}^{*2}$ and $mcpe_{d11}^{*3}$ may be considered acceptably stable for $B = 200$ bootstrap replications, while the estimator $mcpe_{d11}^{*1}$ becomes stable for $B \geq 600$. The graphics for the rest of elements of the matrix estimators present similar shape and not much variation has been observed when analyzing other arbitrary areas.

In conclusion, in these simulations the Prasad-Rao type estimator performs acceptably well even under non normality of the model. The direct bootstrap estimator works a little worse (see Figures 1-3) but is easy to implement, being enough $B = 600$ replications. As alternatives to both estimators, we have the term-to-term and the bias-corrected bootstrap estimators, which work well in all situations, including the absence of normality of the model (see Figure 4), and do not need many bootstrap replications ($B = 200$) to be precise enough.

Appendix

Using similar arguments as those in Prasad and Rao (1990), the following lemmas can be proved.

Lemma 1. *Let us define the vector containing the random part of model $(M_1)-(M_2)$, $\mathbf{v} = \mathbf{Z}\mathbf{u} + \mathbf{e}$. Then*

$$\mathbf{s} = (\mathbf{F} + \mathbf{L})\mathbf{v}, \quad (17)$$

where the matrices \mathbf{F} and \mathbf{L} are such that

$$(i) \quad \mathbf{L} = \text{diag}\{L_1, \dots, L_D\}, \text{ with } L_d = [O(1)]_{r \times r}, \quad d = 1, \dots, D.$$

$$(ii) \quad \mathbf{F} = [O(D^{-1})]_{Dr \times Dr}.$$

Lemma 2. *Let \mathbf{z} be a random vector with $\mathbf{z} \sim N(\mathbf{0}, \mathbf{V})$, let $s_1 = \boldsymbol{\lambda}'_1 \mathbf{z}$ and $s_2 = \boldsymbol{\lambda}'_2 \mathbf{z}$ be two linear combinations of \mathbf{z} , and let $q = \mathbf{z}'\mathbf{C}\mathbf{z}$ be a quadratic form. Then,*

$$E[s_1 s_2 (q - E[q])^2] = \text{Cov}(s_1, s_2) \text{Var}[q] + 8\boldsymbol{\lambda}'_1 \mathbf{V} \mathbf{C} \mathbf{V} \mathbf{C} \mathbf{V} \boldsymbol{\lambda}_2.$$

Lemma 3. *Under the conditions of Theorem 1, the variance of vector \mathbf{s} (cf. 17) satisfies*

$$\text{Var}[\mathbf{s}] = \mathbf{L} \mathbf{V} \mathbf{L}' + [O(D^{-1})]_{Dr \times Dr}.$$

Proof of Theorem 1. By Lemma 1, the components of \mathbf{s} are linear transformations of $\mathbf{v} = \mathbf{Z}\mathbf{u} + \mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, that is, if \mathbf{f}_{di} and \mathbf{L}_{di} are the di -th rows of matrices \mathbf{F} and \mathbf{L} respectively, then

$$s_{di} = (\mathbf{f}_{di} + \mathbf{L}_{di})' \mathbf{v}, \quad i = 1, \dots, r, \quad d = 1, \dots, D.$$

By assumption (H_8) , $\hat{\sigma}_u^2(\mathbf{y}) = \hat{\sigma}_u^2(\mathbf{v}) = k + \mathbf{y}'C\mathbf{y}$ is unbiased and translation invariant, so that $\hat{\sigma}_u^2(\mathbf{y}) = \hat{\sigma}_u^2(\mathbf{v}) = k + \mathbf{v}'C\mathbf{v}$ and $\sigma_u^2 = k + E[\mathbf{v}'C\mathbf{v}]$. Subtracting both equations, we get

$$\hat{\sigma}_u^2 - \sigma_u^2 = \mathbf{v}'C\mathbf{v} - E[\mathbf{v}'C\mathbf{v}].$$

From Lemma 2 with $\boldsymbol{\lambda}_1 = \mathbf{f}_{di} + \mathbf{L}_{di}$, $\boldsymbol{\lambda}_2 = \mathbf{f}_{\ell j} + \mathbf{L}_{\ell j}$, $s_1 = s_{di}$, $s_2 = s_{\ell j}$ and $q = \mathbf{v}'C\mathbf{v}$, we get

$$E[s_{di}s_{\ell j}(\hat{\sigma}_u^2 - \sigma_u^2)^2] = Cov(s_{di}, s_{\ell j})Var[\hat{\sigma}_u^2] + 8(\mathbf{f}_{di} + \mathbf{L}_{di})'VCVVCV(\mathbf{f}_{\ell j} + \mathbf{L}_{\ell j}).$$

Applying the same formula for all pairs $i, j \in \{1, \dots, r\}$ and all pairs $d, \ell \in \{1, \dots, D\}$, and writing the result in matrix form we get

$$E[\mathbf{s}\mathbf{s}'(\hat{\sigma}_u^2 - \sigma_u^2)^2] = Var[\mathbf{s}]Var[\hat{\sigma}_u^2] + 8(F + L)VCVVCV(F + L)'$$

From Lemma A.3(g) of Prasad and Rao (1990) with $r = \mathbf{f}_{di}$, $s = \mathbf{L}_{di}$ and $\Sigma = V$, we get

$$(\mathbf{f}_{di} + \mathbf{L}_{di})'VCVVCV(\mathbf{f}_{di} + \mathbf{L}_{di}) = O(D^{-2}), \quad i = 1, \dots, r, \quad d = 1, \dots, D.$$

But in the mentioned lemma, the particular values of \mathbf{f}_{di} and \mathbf{L}_{di} are not important, just their asymptotic orders. Thus, it holds for any element \mathbf{f}_{di} and \mathbf{L}_{di} with the required asymptotic order, even when on the right of matrix $VCVVCV$ they have different index ℓj . Therefore

$$(F + L)VCVVCV(F + L)' = [O(D^{-2})]_{Dr \times Dr},$$

and then

$$E[\mathbf{s}\mathbf{s}'(\hat{\sigma}_u^2 - \sigma_u^2)^2] = Var[\mathbf{s}]Var[\hat{\sigma}_u^2] + [o(D^{-1})]_{Dr \times Dr}.$$

Furthermore, from Lemma 3, we have

$$Var[\mathbf{s}] = LVL' + [O(D^{-1})]_{Dr \times Dr}.$$

Since $\hat{\sigma}_u^2$ is unbiased and consistent, $Var[\hat{\sigma}_u^2] = o(1)$, then

$$E[\mathbf{s}\mathbf{s}'(\hat{\sigma}_u^2 - \sigma_u^2)^2] = LVL'Var[\hat{\sigma}_u^2] + [o(D^{-1})]_{Dr \times Dr}. \quad \square$$

Acknowledgments. This work has been supported by grants 06/HSE/0181/2004, BMF2003-04820, BFM2002-03213, GV04B-670 and PGIDIT03PXIC20702PN.

References

- Butar, F. B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *Journal of Statistical Planning and Inference*, **112**, 63–76.
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under selection model. *Biometrics*, **31**, 423–427.
- 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.
- Fay, R. E. (1987). Application of Multivariate Regression of Small Domain Estimation. *Small Area Statistics*, (R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh Eds.), Wiley, New York, 91–102.
- Fay, R. E. and Herriot, R. A. (1979). Estimates of income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **74** (366), 269–277.
- Jiang, J., Lahiri, P. and Wan, S. (2002). A unified jackknife theory for empirical best prediction with M -estimation. *The Annals of Statistics*, **30**, 1782–1810.
- Kackar, R. and Harville, D. A. (1984). Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association*, **79**, 853–862.
- Lahiri, P. (2003). On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation. *Statistical Science*, **18**, 199–210.
- Pfefferman, D. and Tiller, R. (2002). Bootstrap Approximation of Prediction MSE for State-Space Models with Estimated Parameters. *S3RI Survey Methodology*, Working Paper M03/05. University of Southampton.
- 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.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag New York, Inc.

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