working papers

UNIVERSIDAD CARLOS III DE MADRID

Working Paper 06-61
Statistic and Econometric Series 17
October 2006

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

UNCERTAINTY UNDER A MULTIVARIATE NESTED-ERROR REGRESSION MODEL WITH LOGARITHMIC TRANSFORMATION *

Isabel Molina[1]

Abstract

Assuming a multivariate linear regression model with one random factor, we consider the parameters defined as exponentials of mixed effects, i.e., linear combinations of fixed and random effects. Such parameters are of particular interest in prediction problems where the dependent variable is the logarithm of the variable that is the object of inference. We derive bias-corrected empirical predictors of such parameters. A second order approximation for the mean crossed product error of the predictors of two of these parameters is obtained, and an estimator is derived from it. The mean squared error is obtained as a particular case.

Keywords: multivariate nested-error regression model, random effects, empirical predictor, mean squared error, small areas.

[1] Departamento de Estadística, Universidad Carlos III de Madrid

* Supported by the grant MTM2006-05693.

October 11, 2006

Uncertainty under a multivariate nested-error regression model with logarithmic transformation[1]

Isabel Molina

Departamento de Estadística, Universidad Carlos III de Madrid

Abstract

Assuming a multivariate linear regression model with one random factor, we consider the parameters defined as exponentials of mixed effects, i.e., linear combinations of fixed and random effects. Such parameters are of particular interest in prediction problems where the dependent variable is the logarithm of the variable that is the object of inference. We derive bias-corrected empirical predictors of such parameters. A second order approximation for the mean crossed product error of the predictors of two of these parameters is obtained, and an estimator is derived from it. The mean squared error is obtained as a particular case.

Key Words: multivariate nested-error regression model, random effects, empirical predictor, mean squared error, small areas.

1 Introduction

Linear mixed models are nowadays a common tool in many statistical applications, like biostatistics, engineering, econometrics and social sciences. When heteroscedasticity and/or lack of normality is detected in a linear (mixed or not) regression model, a common approach is to transform the dependent variable to the logarithmic scale. However, often the object of inferential interest is a characteristic of the variable in the original scale, which is then the exponential of the transformed variable. This is the case for instance in prediction problems. In such situations, it can be of interest to predict the value of the exponential of a mixed effect; that is, a linear combination of fixed and random effects.

In this work we assume that the logarithm of the target variables follow a multivariate linear regression model with one random factor, also called nested-error regression model, and that the target parameters are exponentials of mixed effects. We derive bias-corrected empirical predictors of these parameters, and obtain a second-order approximation for the mean crossed product error (MCPE) of the predictors of two parameters. The mean squared error (MSE) can be obtained as a particular case.

The results described in this work are relevant e.g. for small-area estimation, where the parameters are usually linear combinations of the values of the target variable in the units of the population (typically means or totals). For illustration, we will introduce the problem focussing on this application.

Consider a (large) population partitioned into D (small) subpopulations, also called areas or domains. The "small-area" problem arises when the sample has been extracted from the whole population, but estimates are required for the small areas, and the sample data coming from some of these small areas are not enough for deriving direct estimates with acceptable precision. Here, a direct estimator of a small area characteristic is an estimator calculated using only the sample data from that small area.

Typically, the way of dealing with this problem is to obtain some kind of indirect estimates, calculated using the sample data from outside the target area, in order to "borrow strength"

[1]Suported by the grant MTM2006-05693

from other areas. When auxiliary information is available, successful indirect estimates are the so-called model-based estimates, obtained under the assumption of a regression model. Among these are the Empirical Best Linear Unbiased Predictors (EBLUPs), derived from a linear regression model including random area-specific effects that represent the between-area variation not explained by the auxiliary variables. See the monograph of Rao (2003) for a collection of the main advances in small-area estimation till the publication date and Jiang and Lahiri (2006) for a review of the last developments in this field.

It is common to accompany the point estimates with an accuracy measure like the mean squared error (MSE). Prasad and Rao (1990) obtained a second order approximation of the MSE of the EBLUP under three different types of models; namely nested-error, random coefficient and Fay-Herriot regression models. They estimate the dispersion parameters by a method of moments. The analogous result for maximum likelihood (ML) or restricted ML (RML) was given by Das et al. (2004). Under a univariate Fay-Herriot model with logarithmic transformation of the dependent variable, recently Slud and Maiti (2006) have proposed a bias-corrected empirical predictor, and obtained an approximation of the mean squared error of their proposed predictor.

Suppose now that we are interested in estimating a vector $\boldsymbol{\delta} = (\delta_1, \dots, \delta_K)$ of linear characteristics of K dependent random variables $W_1, \dots, W_K$; for instance, the means of the K variables in the small areas. Then a multivariate model, that takes into account the dependency structure among the variables, is likely to improve the precision over the univariate modelling (see Fay, 1987 and Datta *et al.*, 1991). The problem of estimation of linear parameters like means on small areas can be reduced to a prediction problem, which in turn deals with predicting exponentials of mixed effects (see Section 3).

Here we extend the bias-corrected empirical predictor of Slud and Maiti (2006) to the setup of a multivariate nested-error regression model with logarithmic transformation of the K responses. The univariate model is the particular case $K = 1$.

Further, if we intend to estimate a continuously differentiable function $f(\boldsymbol{\delta}) \in I\!R$ of the K-dimensional characteristic $\boldsymbol{\delta}$ and we estimate it with $f(\hat{\boldsymbol{\delta}})$, where $\hat{\boldsymbol{\delta}}$ is an estimator of $\boldsymbol{\delta}$, then application of Taylor linerization to calculate the mean squared error of $f(\hat{\boldsymbol{\delta}})$ requires the specification of the mean crossed product errors $E[(\hat{\delta}_i - \delta_i)(\hat{\delta}_j - \delta_j)]$, $i, j = 1, \dots, K$. Here we obtain an approximation of the MCPE of the predictors of two parameters defined as exponentials of mixed effects. The mean squared error of a predictor is a particular case.

The paper is organized as follows. The multivariate nested-error regression model is introduced in Section 2, and the bias-corrected empirical predictors are introduced in Section 3. In Section 4 we give asymptotic representations for the maximum likelihood estimators of the model dispersion parameters and of the predictors proposed in Section 3. In Section 5 we list the results leading to the approximation of the MCPE of two predictors. In Section 6 we propose an estimator of the MCPE and finally, the Appendix contains the proofs of the theoretical results.

2 Description of the model

Consider a population partitioned into D small areas of sizes $N_1, \dots, N_D$. Let $\mathbf{W} = (W_1, \dots, W_K)'$ be the random vector of interest, whose values in the units of the population are $\mathbf{w}_{di} = (w_{di1}, \dots, w_{diK})'$, $i = 1, \dots, N_d$, $d = 1, \dots, D$. In small area estimation a common target is to estimate the small area means $\bar{\mathbf{W}}_d = (\bar{W}_{d1}, \dots, \bar{W}_{dK})'$, for $d = 1, \dots, D$, where $\bar{W}_{dr} = N_d^{-1} \sum_{i=1}^{N_d} w_{dir}$, $r = 1, \dots, K$. Now consider the vector $\mathbf{Y} = (Y_1, \dots, Y_K)'$ of transformed variables $Y_r = \log(W_r)$, $r = 1, \dots, K$. Let $\mathbf{y}_{di} = (y_{di1}, \dots, y_{diK})'$ be the value of $\mathbf{Y}$ in the i-th unit of the d-th small area, and accordingly, let $X_{di} = (\mathbf{x}_{di1}, \dots, \mathbf{x}_{diK})'$ be the $K \times p$

matrix with the values of the auxiliary variables for the same unit, $\mathbf{e}_{di} = (e_{di1}, \ldots, e_{diK})'$ the vector of random errors and u_d the random effect of area d. We assume that the population units satisfy the model

$$\mathbf{y}_{di} = X_{di}\boldsymbol{\beta} + u_d \mathbf{1}_K + \mathbf{e}_{di}, \tag{1}$$

where $\mathbf{1}_r$ denotes an r-vector of ones, $\boldsymbol{\beta}$ is the p-vector containing the coefficients of the auxiliary variables, and u_d and $\mathbf{e}_{di}$ are independent with distributions

$$\mathbf{e}_{di} \stackrel{iid}{\sim} N(\mathbf{0}_K, \Sigma), \quad u_d \stackrel{iid}{\sim} N(0, \sigma^2), \quad i = 1, \ldots, N_d, \; d = 1, \ldots, D. \tag{2}$$

The following structure is assumed for the covariance matrix Σ, where the restrictions $\sigma^2 > 0$ and $-\sigma^2 < \phi < \sigma^2$ ensure the positive definiteness,

$$\Sigma = \begin{pmatrix} \sigma^2 & \phi & \cdots & \phi \\ \phi & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \phi \\ \phi & \cdots & \phi & \sigma^2 \end{pmatrix}.$$

This structure of the covariance matrix Σ, although very simple, is commonly used in multivariate problems. See for instance the works related to familial data (Srivastava, 1984, Srivastava and Katapa, 1986, Bhandary and Alam, 2000 or Hobza *et al.*, 2002) or to longitudinal data (Diggle *et al.*, 2002).

In the inference process, a sample of size n is extracted from the whole population. Let s_d be the set of units sampled from the d-th small area, with size n_d, and s_d^c the complementary of s_d, that is, the set of units of the same area that have not been sampled, $d = 1, \ldots, D$, where $n = \sum_{d=1}^D n_d$. Let us construct the following column vectors and matrices containing sample elements

$$\begin{array}{llll} \mathbf{y}_d = \text{col}_{i \in s_d}(\mathbf{y}_{di}), & \mathbf{X}_d = \text{col}_{i \in s_d}(\mathbf{X}_{di}), & \mathbf{e}_d = \text{col}_{i \in s_d}(\mathbf{e}_{di}), & \mathbf{R}_d = \text{diag}_{i \in s_d}(\Sigma), \\ \mathbf{y} = \text{col}_{1 \le d \le D}(\mathbf{y}_d), & \mathbf{X} = \text{col}_{1 \le d \le D}(\mathbf{X}_d), & \mathbf{e} = \text{col}_{1 \le d \le D}(\mathbf{e}_d), & \mathbf{R} = \text{diag}_{1 \le d \le D}(\mathbf{R}_d), \end{array}$$

and additionally $\mathbf{u} = \text{col}_{1 \le d \le D}(u_d)$ and $\mathbf{Z} = \text{diag}_{1 \le d \le D}(\mathbf{1}_{Kn_d})$. The notation $\text{col}_{i \in A}(B_i)$ indicates stacking the elements B_i, $i \in A$ into a column, and $\text{diag}_{i \in A}(B_i)$ denotes the block-diagonal matrix with blocks B_i, $i \in A$. In this notation, the model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}_{Kn}, \mathbf{R}), \quad \mathbf{u} \sim N(\mathbf{0}_D, \sigma_u^2 I_D),$$

where I_r denotes the $r \times r$ identity matrix. The variance-covariance matrix of $\mathbf{y}$ is given by $\mathbf{V} = \sigma_u^2 \mathbf{Z}\mathbf{Z}' + \mathbf{R}$. This matrix is block-diagonal; more explicitly, $\mathbf{V} = \text{diag}_{1 \le d \le D}(\mathbf{V}_d)$, where $\mathbf{V}_d = \sigma_u^2 \mathbf{1}_{Kn_d} \mathbf{1}'_{Kn_d} + \mathbf{R}_d$, $d = 1, \ldots, D$.

The dispersion parameter space is

$$\Theta = \{\boldsymbol{\theta} = (\sigma_u^2, \sigma^2, \phi)' : \sigma_u^2, \sigma^2 \in (0, M), \; -\sigma^2 < \phi < \sigma^2\},$$

where $M < \infty$ is a constant. We denote by $\boldsymbol{\theta}_0 = (\sigma_{u0}^2, \sigma_0^2, \phi_0)'$ the true, unknown value of the parameter $\boldsymbol{\theta}$. Sometimes we will also use the notation $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$ and $\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02}, \theta_{03})'$. Hereafter, for a quantity A that is function of $\boldsymbol{\theta}$, we will omit the argument $\boldsymbol{\theta}$ when A is evaluated at $\boldsymbol{\theta}_0$; that is, we will denote $A = A(\boldsymbol{\theta}_0)$, and we will maintain it when it is evaluated at $\boldsymbol{\theta} \in \Theta$; that is, $A(\boldsymbol{\theta})$. Similarly, we will use $\partial A / \partial \boldsymbol{\theta}$ when the derivative is evaluated at $\boldsymbol{\theta}_0$, and $\partial A(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ when is evaluated at $\boldsymbol{\theta} \in \Theta$.

3 Predictors of exponentials of mixed effects

Prediction of the original variables w_{dir} based on model (1) involves prediction of the quantities $\exp(\mathbf{x}'_{dir}\boldsymbol{\beta} + u_d)$. This work deals with predicting exponentials of mixed effects; that is, of the parameters

$$\tau_k = \exp(\mu_k), \quad \text{for} \quad \mu_k = \boldsymbol{\lambda}'_k\boldsymbol{\beta} + \mathbf{m}'_k\mathbf{u},$$

where $\boldsymbol{\lambda}_k$ and $\mathbf{m}_k$ are constant known vectors. When $\boldsymbol{\lambda}_k = \mathbf{x}_{dir}$ and $\mathbf{m}_k$ is a vector of zeros except for a one in position d we obtain the particular case $\tau_k = \exp(\mathbf{x}'_{dir}\boldsymbol{\beta} + u_d)$.

In small area estimation, typical parameters are the small area means $\bar{W}_{dr}$, that can be decomposed as

$$\bar{W}_{dr} = N_d^{-1}\left(\sum_{i \in s_d} w_{dir} + \sum_{i \in s_d^c} w_{dir}\right), \quad r = 1, \ldots, K.$$

An estimator of $\bar{W}_{dr}$ can be obtained through prediction of nonsampled units w_{dir}, $i \in s_d^c$.

If $\boldsymbol{\theta}$ is known, then the Best Linear Unbiased Predictor (BLUP) of $\mu_k = \boldsymbol{\lambda}'_k\boldsymbol{\beta} + \mathbf{m}'_k\mathbf{u}$ is given by $\hat{\mu}_k = \boldsymbol{\lambda}'_k\hat{\boldsymbol{\beta}} + \mathbf{m}'_k\hat{\mathbf{u}}$, where

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad \text{and} \quad \hat{\mathbf{u}} = \sigma_u^2\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{3}$$

(Henderson, 1975). Following the ideas of Slud and Maiti (2006), an approximately bias-corrected predictor of τ_k is $\hat{\rho}_k \exp(\hat{\mu}_k)$, where $\hat{\rho}_k$ is a bias-correction factor given by

$$\hat{\rho}_k = \hat{E}\left[\exp(\mu_k)\right] / \hat{E}\left[\exp(\hat{\mu}_k)\right].$$

Here $\hat{E}$ is an estimate of the expectation obtained by replacing the unknown parameters by their estimates. An asymptotically ($D \to \infty$) correct expression for this predictor under model (1)-(2) is

$$\hat{\tau}_k = \exp(\hat{\mu}_k + \alpha_k), \quad \text{where} \quad \alpha_k = \sigma_u^2\mathbf{m}'_k(I_D - \sigma_u^2\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})\mathbf{m}_k/2.$$

If $\boldsymbol{\theta}$ is unknown, then $\hat{\tau}_k$ depends on $\boldsymbol{\theta}$ through $\hat{\mu}_k = \hat{\mu}_k(\boldsymbol{\theta})$ and $\alpha_k = \alpha_k(\boldsymbol{\theta})$; that is, $\hat{\tau}_k = \hat{\tau}_k(\boldsymbol{\theta})$. Substituting an estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ into $\hat{\tau}_k$ we obtain what is generally called an empirical predictor. The maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ can be obtained by maximization of the profile loglikelihood

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

where c denotes a constant and

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

Let $\hat{\boldsymbol{\theta}}$ be the MLE of $\boldsymbol{\theta}$ and $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\theta}})$ the covariance matrix $\mathbf{V}$ evaluated at the MLE. Then, an empirical BLUP of μ_k is $\hat{\mu}_k^E = \hat{\mu}_k(\hat{\boldsymbol{\theta}}) = \boldsymbol{\lambda}'_k\hat{\boldsymbol{\beta}}^E + \mathbf{m}'_k\hat{\mathbf{u}}^E$, where

$$\hat{\boldsymbol{\beta}}^E = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y} \quad \text{and} \quad \hat{\mathbf{u}}^E = \hat{\sigma}_u^2\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^E).$$

A bias-corrected empirical predictor of τ_k is then $\hat{\tau}_k^E = \exp(\hat{\mu}_k^E + \hat{\alpha}_k)$, where $\hat{\alpha}_k = \alpha_k(\hat{\boldsymbol{\theta}})$.

4 Asymptotic representations

In Theorem 2.1, Das *et al.* (2004) gave an asy,ptotic representation of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$, listing the conditions ensuring that result. Baillo and Molina (2005) verified these conditions for model (1)-(2). Following a similar approach, in this section we provide a more precise asymptotic representation of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ under model (1)-(2). As a consequence, we obtain an asymptotic formula for $\hat{\tau}_k^E - \hat{\tau}_k$, result which will lead to an approximated formula for the MCPE in Section 5.

The following notation is used throughout the paper:

$$\mathbf{s} = \partial l_P/\partial\boldsymbol{\theta}, \quad H = \partial^2 l_P/\partial\boldsymbol{\theta}^2, \quad \mathcal{I} = E(-H), \tag{4}$$

$$D_j = \partial H/\partial\theta_j, \quad d_j = \mathbf{s}'\mathcal{I}^{-1}D_j\mathcal{I}^{-1}\mathbf{s}, \; j = 1,2,3, \quad \mathbf{d} = (d_1, d_2, d_3)', \tag{5}$$
$$\mathbf{h}_k = \partial\hat{\tau}_k/\partial\boldsymbol{\theta} \quad \text{and} \quad \boldsymbol{\gamma}_k = \sigma_u^2\mathbf{V}^{-1}\mathbf{Z}\mathbf{m}_k.$$

Further, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of a square matrix A respectively.

The results of this paper require the following conditions:

(H1) The elements of the vector $\boldsymbol{\lambda}_k$ are uniformly bounded as $D \to \infty$. The vector $\mathbf{m}_k$ contains only zeros except for one element, and this element is bounded as $D \to \infty$;

(H2) $p < \infty$, $\limsup_{D\to\infty}(\max_{1\le d\le D} n_d) < \infty$ and $\liminf_{D\to\infty}(\min_{1\le d\le D} n_d) > 0$;

(H3) The elements of the matrix $\mathbf{X}$ are uniformly bounded as $D \to \infty$;

(H4) $\liminf_{D\to\infty} D^{-1}\lambda_{\min}(\mathbf{X}'\mathbf{X}) > 0$;

(H5) $\liminf_{D\to\infty} D^{-1}\lambda_{\min}(\mathcal{I}) > 0$;

For the sake of clarity and completeness, we include below Lemma 1 of Baillo and Molina (2005), which is an adaptation of Theorem 2.1 of Das *et al.* (2004) to model (1)–(2).

Lemma 1 *(Das et al. 2004, Baillo and Molina 2005)*
Let model (1)–(2) satisfy conditions (H2)–(H5). Then, for every $\eta \in (0,1)$*, there exists a subset of the sample space* $\mathcal{B}$ *on which, for large D, it holds that* $|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < D^{-\eta/2}$ *and*

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \mathcal{I}^{-1}\mathbf{s} + \mathbf{r}_*, \tag{6}$$

where $|\mathbf{r}_*| < D^{-\eta}\, v_*$*, for a random variable* v_* *with* $E(v_*^b)$ *bounded for every* $b > 0$*. Furthermore,* $P(\mathcal{B}^c) = O(D^{-\zeta b/2})$*, where* $\zeta = min(1/4, 1-\eta)$*.*

The set $\mathcal{B}$ mentioned in Lemma 1 is defined in the proof of Theorem 2.1 of Das *et al.* (2004), and is the set where the existence of a solution of the likelihood equations $\partial l_P(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$ can be ensured.

Das *et al.* (2004), based on the representation (6), approximated the mean squared error of the EBLUP of a mixed effect, were the neglected terms were $o(D^{-1})$. However, if we follow their approach for approximating the MSE of the empirical predictor $\hat{\tau}_k^E$ of a parameter τ_k, the exponential function appearing in our parameter τ_k makes the order of the neglected terms in the MSE to be slower; thus, a rate $o(D^{-1})$ cannot be ensured. For this reason, a more exact asymptotic representation than (6) is needed. Lemma 2 gives this more precise asymptotic formula.

Lemma 2 *Let model (1)–(2) satisfy conditions (H2)–(H5). Then, for any $\eta \in (0,1)$, on the same set $\mathcal{B}$ as in Lemma 1 and for large D, it holds that*

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \mathcal{I}^{-1}\mathbf{s} + \mathcal{I}^{-1}(H+\mathcal{I})\mathcal{I}^{-1}\mathbf{s} + \mathcal{I}^{-1}\mathbf{d}/2 + \mathbf{r}, \tag{7}$$

where $|\mathbf{r}| < D^{-3\eta/2}v$, for a random variable v with $E(v^b)$ bounded for every $b > 0$.

A sketch of the proof is given in the appendix. As a consequence of Lemma 2, the following result provides an asymptotic representation for $\hat{\tau}_k^E - \hat{\tau}_k$. This result leads to the approximation of the MCPE obtained in Section 5.

Lemma 3 *Under assumptions (H1)–(H5), on the same set $\mathcal{B}$ as in Lemmas 1 and 2, and for large D, it holds that*

$$\hat{\tau}_k^E - \hat{\tau}_k = \mathbf{h}_k'\mathcal{I}^{-1}\mathbf{s} + \mathbf{h}_k'\mathcal{I}^{-1}(H+\mathcal{I})\mathcal{I}^{-1}\mathbf{s} + \mathbf{h}_k'\mathcal{I}^{-1}\mathbf{d}/2 + \mathbf{s}'\mathcal{I}^{-1}(\partial^2\hat{\tau}_k/\partial\boldsymbol{\theta}^2)\mathcal{I}^{-1}\mathbf{s}/2 + r_k, \tag{8}$$

where $|r_k| < D^{-3\eta/2}v_k$, for a random variable v_k with bounded first and second moments.

5 Mean crossed product error

Now let us consider two parameters $\tau_k = \exp(\mu_k)$, where $\mu_k = \boldsymbol{\lambda}_k'\boldsymbol{\beta} + \mathbf{m}_k'\mathbf{u}$ for $k = 1,2$, and their bias-corrected predictors $\hat{\tau}_k = \exp(\hat{\mu}_k + \alpha_k)$, where $\hat{\mu}_k$ is the BLUP of μ_k, $k = 1,2$. Proposition 1 gives the exact expression of the mean crossed product error of $\hat{\tau}_1$ and $\hat{\tau}_2$ under the assumption that $\boldsymbol{\theta}$ is known. The mean squared error of the predictor $\hat{\tau}$ of a parameter τ is obtained by setting $\tau_1 = \tau_2 = \tau$.

Let us define $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$, $\mathbf{m} = \mathbf{m}_1 + \mathbf{m}_2$, $\alpha = \alpha_1 + \alpha_2$ and $\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$, where

$$\mathbf{b}_k = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\boldsymbol{\lambda}_k + \sigma_u^2\mathbf{P}\mathbf{Z}\mathbf{m}_k, \quad k = 1,2. \tag{9}$$

Proposition 1 *The mean crossed product error of the predictors $\hat{\tau}_1 = exp(\hat{\mu}_1 + \alpha_1)$ and $\hat{\tau}_2 = exp(\hat{\mu}_2 + \alpha_2)$ is given by*

$$\begin{aligned} MCPE(\hat{\tau}_1,\hat{\tau}_2) \;=\;& exp(\boldsymbol{\lambda}'\boldsymbol{\beta})\left\{exp(\alpha + \mathbf{b}'\mathbf{V}\mathbf{b}/2) + exp(\sigma_u^2\mathbf{m}'\mathbf{m}/2)\right.\\ & -exp\left[\alpha_1 + \left(\mathbf{b}_1'\mathbf{V}\mathbf{b}_1 + \sigma_u^2\mathbf{m}_2'\mathbf{m}_2 + 2\sigma_u^2\mathbf{m}_2'\mathbf{Z}'\mathbf{b}_1\right)/2\right]\\ & \left.-exp\left[\alpha_2 + \left(\mathbf{b}_2'\mathbf{V}\mathbf{b}_2 + \sigma_u^2\mathbf{m}_1'\mathbf{m}_1 + 2\sigma_u^2\mathbf{m}_1'\mathbf{Z}'\mathbf{b}_2\right)/2\right]\right\}. \end{aligned} \tag{10}$$

In the rest of this section regard the parameter $\boldsymbol{\theta}$ as unknown. The following results provide an approximation correct up to order $o(D^{-1})$ for the MCPE of the empirical predictors $\hat{\tau}_1^E = \exp(\hat{\mu}_1^E + \hat{\alpha}_1)$ and $\hat{\tau}_2^E = \exp(\hat{\mu}_2^E + \hat{\alpha}_2)$. The mentioned approximation is obtained by analyzing each term in the decomposition

$$\begin{aligned} \text{MCPE}(\hat{\tau}_1^E,\hat{\tau}_2^E) \;=\;& \text{MCPE}(\hat{\tau}_1,\hat{\tau}_2) + E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2^E - \hat{\tau}_2)\right]\\ & + E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2 - \tau_2)\right] + E\left[(\hat{\tau}_1 - \tau_1)(\hat{\tau}_2^E - \hat{\tau}_2)\right]. \end{aligned} \tag{11}$$

The first term on the right-hand side is already given in (10). A second order approximation for the second term is given in Theorems 1 and 2 below. Finally, Theorem 3 states that the remaining terms in (11) are $o(D^{-1})$. The proofs of these results appear in the appendix.

Theorem 1 *Suppose that (H1)–(H5) hold. Then, the predictors $\hat{\tau}_k = exp(\hat{\mu}_k + \alpha_k)$ and $\hat{\tau}_k^E = exp(\hat{\mu}_k^E + \hat{\alpha}_k)$ of the parameter $\tau_k = exp(\mu_k)$, $k = 1,2$, satisfy*

$$E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2^E - \hat{\tau}_2)\right] = E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\right] + o(D^{-1}). \tag{12}$$

Theorem 2 provides a simpler formula for practical calculation of the MCPE. This formula is a second order approximation of the right-hand side of (12).

Theorem 2 *Under (H1)–(H5), it holds*

$$E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\right] = exp(\alpha + \boldsymbol{\lambda}'\boldsymbol{\beta} + \mathbf{b}'\mathbf{V}\mathbf{b}/2) \times \\ \left[tr\left(\frac{\partial\boldsymbol{\gamma}_1'}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_2}{\partial\boldsymbol{\theta}}\right) + \left(\mathbf{b}'\mathbf{V}\frac{\partial\boldsymbol{\gamma}_1}{\partial\boldsymbol{\theta}} + \left(\frac{\partial\alpha_1}{\partial\boldsymbol{\theta}}\right)'\right)\mathcal{I}^{-1}\left(\frac{\partial\boldsymbol{\gamma}_2'}{\partial\boldsymbol{\theta}}\mathbf{V}\mathbf{b} + \frac{\partial\alpha_2}{\partial\boldsymbol{\theta}}\right)\right] + o(D^{-1}).$$

Theorem 3 states that in the approximation of the MCPE, the last two terms in (11) can be neglected.

Theorem 3 *Under (H1)–(H5), it holds*

$$E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2 - \tau_2)\right] = o(D^{-1}).$$

From the decomposition given in (11) and Theorems 1–3, we obtain the following formula of the MCPE of the predictors $\hat{\tau}_1^E$ and $\hat{\tau}_2^E$,

$$\text{MCPE}(\hat{\tau}_1^E, \hat{\tau}_2^E) = \text{MCPE}(\hat{\tau}_1, \hat{\tau}_2) + \exp(\alpha + \boldsymbol{\lambda}'\boldsymbol{\beta} + \mathbf{b}'\mathbf{V}\mathbf{b}/2) \times \\ \left[\text{tr}\left(\frac{\partial\boldsymbol{\gamma}_1'}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_2}{\partial\boldsymbol{\theta}}\right) + \left(\mathbf{b}'\mathbf{V}\frac{\partial\boldsymbol{\gamma}_1}{\partial\boldsymbol{\theta}} + \left(\frac{\partial\alpha_1}{\partial\boldsymbol{\theta}}\right)'\right)\mathcal{I}^{-1}\left(\frac{\partial\boldsymbol{\gamma}_2'}{\partial\boldsymbol{\theta}}\mathbf{V}\mathbf{b} + \frac{\partial\alpha_2}{\partial\boldsymbol{\theta}}\right)\right] + o(D^{-1}). \quad (13)$$

6 Estimation of the mean crossed product error

Estimation of (13) can be done with a bias of order $o(D^{-1})$ by plugging ML estimators $\hat{\boldsymbol{\beta}}^E$ and $\hat{\boldsymbol{\theta}}$ instead of the unknown values $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ except for the first term, MCPE$(\hat{\tau}_1, \hat{\tau}_2)$, where a bias appears due to the plug-in procedure. In this section we approximate this bias and, based on this result, we propose a bias-corrected estimator for MCPE$(\hat{\tau}_1, \hat{\tau}_2)$.

Let us denote $g(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \text{MCPE}(\hat{\tau}_1, \hat{\tau}_2)$. The following theorem provides an approximation up to $o(D^{-1})$ for the bias of the estimator $g(\hat{\boldsymbol{\beta}}^E, \hat{\boldsymbol{\theta}})$ of $g(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$.

Theorem 4 *Under (H1)-(H5), it holds*

$$E[g(\hat{\boldsymbol{\beta}}^E, \hat{\boldsymbol{\theta}})] = g(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + \sum_{j=1}^{5}\Lambda_j(\boldsymbol{\theta}_0) + o(D^{-1}),$$

where

$$\Lambda_1(\boldsymbol{\theta}_0) = (\partial g/\partial\boldsymbol{\theta})'\,E(\mathcal{I}^{-1}\mathbf{s}), \qquad \Lambda_2(\boldsymbol{\theta}_0) = (\partial g/\partial\boldsymbol{\theta})'\,E\left[\mathcal{I}^{-1}(H+\mathcal{I})\mathcal{I}^{-1}\right],$$
$$\Lambda_3(\boldsymbol{\theta}_0) = (\partial g/\partial\boldsymbol{\theta})'\,E(\mathcal{I}^{-1}\mathbf{d})/2, \qquad \Lambda_4(\boldsymbol{\theta}_0) = E\left[\mathbf{s}\mathcal{I}^{-1}\left(\partial^2 g/\partial\boldsymbol{\theta}^2\right)\mathcal{I}^{-1}\mathbf{s}\right]/2,$$
$$\Lambda_5(\boldsymbol{\theta}_0) = g(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)\,\boldsymbol{\lambda}'(\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}\boldsymbol{\lambda}.$$

Straightforward algebra yields the following formulas for the $\Lambda_j(\boldsymbol{\theta}_0)$:

$$\begin{aligned}
\Lambda_1(\boldsymbol{\theta}_0) &= (\partial g/\partial\boldsymbol{\theta})'\mathcal{I}^{-1}\boldsymbol{\nu}, \\
\Lambda_2(\boldsymbol{\theta}_0) &= (\partial g/\partial\boldsymbol{\theta})'\mathcal{I}^{-1}\text{col}_{1\le i\le 3}\left(\text{tr}(\Phi_i\mathcal{I}^{-1})\right) + o(D^{-1}), \\
\Lambda_3(\boldsymbol{\theta}_0) &= (1/4)\,(\partial g/\partial\boldsymbol{\theta})'\mathcal{I}^{-1}\text{col}_{1\le i\le 3}\left(\text{tr}\left[(3\Phi_i - B_i)\mathcal{I}^{-1}\Phi\mathcal{I}^{-1}\right]\right) + o(D^{-1}), \\
\Lambda_4(\boldsymbol{\theta}_0) &= (1/4)\text{tr}\left[\left(\partial^2 g/\partial\boldsymbol{\theta}^2\right)\mathcal{I}^{-1}\Phi\mathcal{I}^{-1}\right] + o(D^{-1}).
\end{aligned}$$

Here, $\boldsymbol{\nu} = (\nu_1, \nu_2, \nu_3)'$, $\Phi = (\phi_{ij})_{i,j=1,2,3}$, $\Phi_i = (\phi_{ijk})_{j,k=1,2,3}$, and $B_i = (a_{ijk})_{j,k=1,2,3}$, where

$$\nu_j = \mathrm{tr}[(\mathbf{P} - \mathbf{V}^{-1})\boldsymbol{\Delta}_j], \qquad \phi_{ij} = \mathrm{tr}(\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_j),$$
$$\phi_{ijk} = \mathrm{tr}(\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_k), \quad a_{ijk} = \mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_i\mathbf{V}^{-1}\boldsymbol{\Delta}_j\mathbf{V}^{-1}\boldsymbol{\Delta}_k).$$

From Theorem 4, an unbiased estimator of $g(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ up to $o(D^{-1})$ terms is

$$g^* = g(\hat{\boldsymbol{\beta}}^E, \hat{\boldsymbol{\theta}}) - \sum_{j=1}^{5} \Lambda_j(\hat{\boldsymbol{\theta}}).$$

Appendix

In this appendix we outline the proofs of the results of Sections 4–6. Hereafter, the norm of a vector $\mathbf{v}$ is denoted by $|\mathbf{v}| = (\mathbf{v}'\mathbf{v})^{1/2}$, and for a matrix A, the norms $\|A\| = \lambda_{\max}^{1/2}(A'A)$ and $\|A\|_2 = \mathrm{tr}^{1/2}(A'A)$ are used. Further, let us denote $\boldsymbol{\Delta}_j = \partial\mathbf{V}/\partial\theta_j$, $j = 1, 2, 3$, and $\delta_k = \hat{\mu}_k + \alpha_k$, so that $\hat{\tau}_k = \exp(\delta_k)$, $k = 1, 2$. Finally, for $\eta \in (0, 1)$, let us introduce the following neighborhood of $\boldsymbol{\theta}_0$,

$$N(\boldsymbol{\theta}_0) = \left\{\boldsymbol{\theta} \in \Theta : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| < D^{-\eta/2}\right\}.$$

PROOF OF LEMMA 2 Second-order Taylor expansions about $\boldsymbol{\theta}_0$ of the functions $\partial l_P(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_i$, $i = 1, 2, 3$, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, lead to

$$\mathbf{0} = \mathbf{s} + H(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \tilde{\mathbf{d}}/2 + \tilde{\mathbf{r}},$$

where $\tilde{\mathbf{d}} = (\tilde{d}_1, \tilde{d}_2, \tilde{d}_3)'$ with $\tilde{d}_i = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'D_i(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, and $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \tilde{r}_3)'$ with

$$\tilde{r}_i = \frac{1}{6}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{\ell=1}^{3} \frac{\partial^4 l_P(\boldsymbol{\theta}_*)}{\partial\theta_i\partial\theta_j\partial\theta_k\partial\theta_\ell}(\hat{\theta}_j - \theta_{0j})(\hat{\theta}_k - \theta_{0k})(\hat{\theta}_\ell - \theta_{0\ell}),\ i = 1, 2, 3.$$

Adding and subtracting $\mathcal{I}$ to H, multiplying by $\mathcal{I}^{-1}$ and solving for $\hat{\boldsymbol{\theta}}$ we obtain

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \mathcal{I}^{-1}\mathbf{s} + \mathcal{I}^{-1}(H + \mathcal{I})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathcal{I}^{-1}\tilde{\mathbf{d}}/2 + \mathbf{r}_1, \tag{14}$$

where $\mathbf{r}_1 = \mathcal{I}^{-1}\,\tilde{\mathbf{r}}$. It is not difficult to prove that on $\mathcal{B}$ and for large D, $|\mathbf{r}_1| < D^{-3\eta/2}v_1$, where $E(v_1^b)$ is bounded for $b > 0$. The proof of this result involves showing that

$$E\left(\sup_{N(\boldsymbol{\theta}_0)}\left|\frac{1}{D}\frac{\partial^4 l_P(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j\partial\theta_k\partial\theta_\ell}\right|\right)^b = O(1),\ b > 0.$$

This fact is shown in the proof of Theorem 3 of Baillo and Molina (2005).

Now let us replace (6) in the third and fourth terms on the right of (14). On the one hand,

$$\mathcal{I}^{-1}(H + \mathcal{I})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathcal{I}^{-1}(H + \mathcal{I})\mathcal{I}^{-1}\mathbf{s} + \mathbf{r}_2, \tag{15}$$

where $|\mathbf{r}_2| < D^{-1/2-\eta}v_2$ with $E(v_2^b)$ bounded, for $b > 0$. On the other hand,

$$\mathcal{I}^{-1}\tilde{\mathbf{d}} = \mathbf{d} + \mathbf{r}_3 + \mathbf{r}_4, \tag{16}$$

for $\mathbf{d}$ as defined in (5), and where $|\mathbf{r}_3| < D^{-1/2-\eta}v_3$ and $|\mathbf{r}_4| < D^{-2\eta}v_4$, where $E(v_3^b)$ and $E(v_4^b)$ are bounded, for $b > 0$. The proof of the boundedness of these expectations involves checking the condition

$$E\left(\left|\frac{1}{D}\frac{\partial^3 l_P}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|^b\right) = O(1), \quad b > 0.$$

This is shown in the proof of Lemma 1 of Baillo and Molina (2005). Replacing (15) and (16) in (14) and denoting $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2 + (\mathbf{r}_3 + \mathbf{r}_4)/2$, we obtain the result. □

PROOF OF LEMMA 3 A second-order Taylor expansion of $\hat{\tau}_k(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$, evaluated at the point $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, leads to

$$\hat{\tau}_k^E - \hat{\tau}_k = \mathbf{h}_k'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + (1/2)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\left(\partial^2\hat{\tau}_k/\partial\boldsymbol{\theta}^2\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + r_{k1}, \tag{17}$$

where $|r_{k1}| < D^{-3\eta/2}v_{k1}$ and the first and second moments of v_{k1} are bounded. Now let us insert (7) into the first term on the right of (17), and (6) into the second term. This leads to

$$\hat{\tau}_k^E - \hat{\tau}_k = \mathbf{h}_k'\mathcal{I}^{-1}\mathbf{s} + \mathbf{h}_k'\mathcal{I}^{-1}(H + \mathcal{I})\mathcal{I}^{-1}\mathbf{s} + \frac{1}{2}\,\mathbf{h}_k'\mathcal{I}^{-1}\mathbf{d} + \frac{1}{2}\,\mathbf{s}'\mathcal{I}^{-1}\frac{\partial^2\hat{\tau}_k}{\partial\boldsymbol{\theta}^2}\mathcal{I}^{-1}\mathbf{s} + r_{k1} + r_{k2} + r_{k3},$$

for $r_{k2} = \mathbf{r}_*'(\partial^2\hat{\tau}_k/\partial\boldsymbol{\theta}^2)\mathcal{I}^{-1}\mathbf{s}$ and $r_{k3} = \mathbf{r}_*'(\partial^2\hat{\tau}_k/\partial\boldsymbol{\theta}^2)\mathbf{r}_*$. By Lemma 1, for large D and on $\mathcal{B}$, it holds that $|r_{k2}| < D^{-1/2-\eta}v_{k2}$ and $|r_{k3}| < D^{-2\eta}v_{k3}$, where both v_{k2} and v_{k3} have the first two moments bounded, and further, the expectations $E(v_{ki}v_{kj})$ for $i \neq j$ are also bounded. Then the result follows by calling $r_k = r_{k1} + r_{k2} + r_{k3}$. □

PROOF OF PROPOSITION 1 The mean crossed product error of $\hat{\tau}_1$ and $\hat{\tau}_2$ is

$$\begin{aligned}\text{MCPE}(\hat{\tau}_1, \hat{\tau}_2) &= E\left\{[\exp(\hat{\mu}_1 + \alpha_1) - \exp(\mu_1)]\,[\exp(\hat{\mu}_2 + \alpha_2) - \exp(\mu_2)]\right\} \\ &= E\left[\exp(\hat{\mu}_1 + \hat{\mu}_2 + \alpha)\right] + E\left[\exp(\mu_1 + \mu_2)\right] \\ &\quad - E\left[\exp(\hat{\mu}_1 + \mu_2 + \alpha_1)\right] - E\left[\exp(\mu_1 + \hat{\mu}_2 + \alpha_2)\right].\end{aligned}$$

The exponents involved can be written as linear combinations of multivariate normal vectors as

$$\begin{aligned}\mu_1 + \mu_2 &= \boldsymbol{\lambda}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{u}; \\ \hat{\mu}_1 + \mu_2 &= \boldsymbol{\lambda}'\boldsymbol{\beta} + (\mathbf{b}_1'\mathbf{Z} + \mathbf{m}_2')\mathbf{u} + \mathbf{b}_1'\mathbf{e}; \\ \hat{\mu}_1 + \hat{\mu}_2 &= \boldsymbol{\lambda}'\boldsymbol{\beta} + \mathbf{b}'\mathbf{v}, \quad \text{for} \quad \mathbf{v} = \mathbf{Z}\mathbf{u} + \mathbf{e} \sim N_n(\mathbf{0}_n, \mathbf{V}).\end{aligned} \tag{18}$$

Then (10) follows by using the moment generating function of linear combinations of the multidimensional normal variables $\mathbf{u}$, $\mathbf{e}$ and $\mathbf{v}$. □

PROOF OF THEOREM 1 The proof is based on the following chain results:

A) For large D and on $\mathcal{B}$, it holds

$$\hat{\tau}_k^E - \hat{\tau}_k = \mathbf{h}_k'\mathcal{I}^{-1}\mathbf{s} + t_k, \quad k = 1, 2, \tag{19}$$

where $|t_k| \leq D^{-\eta}\,w_k$ and the first and second moments of the random variable w_k are bounded.
B) $E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2^E - \hat{\tau}_2)1_{\mathcal{B}}\right] = E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})1_{\mathcal{B}}\right] + o(D^{-1})$, where $1_{\mathcal{B}}$ denotes the indicator function of the set $\mathcal{B}$ mentioned in Lemma 1.

C) $E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2^E - \hat{\tau}_2)1_{\mathcal{B}^c}\right] = o(D^{-1})$.

D) $E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})1_{\mathcal{B}}\right] = E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\right] + o(D^{-1})$.

The result then follows directly from the decomposition

$$E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2^E - \hat{\tau}_2)\right] = E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2^E - \hat{\tau}_2)1_{\mathcal{B}}\right] + E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2^E - \hat{\tau}_2)1_{\mathcal{B}^c}\right],$$

and using B), C) and D).

Now we detail the proofs of A)–D). Result A) is obtained by a first-order Taylor expansion of the function $\hat{\tau}_k^E = \hat{\tau}_k(\hat{\boldsymbol{\theta}})$ around $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$,

$$\hat{\tau}_k^E - \hat{\tau}_k = \mathbf{h}_k'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + t_k, \quad \text{for} \quad t_k = (1/2)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \left(\partial^2 \hat{\tau}_k(\boldsymbol{\theta}^*)/\partial \boldsymbol{\theta}^2\right) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \tag{20}$$

where $\boldsymbol{\theta}^*$ is on the line joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Using Lemma 1 and the formula of the second derivative of $\hat{\tau}_k(\boldsymbol{\theta}) = \exp(\delta_k(\boldsymbol{\theta}))$,

$$\partial^2 \hat{\tau}_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^2 = \exp(\delta_k(\boldsymbol{\theta}) \left[(\partial \delta_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta}) (\partial \delta_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta})' + \partial^2 \delta_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^2\right],$$

we obtain that $|t_k| \leq D^{-\eta} w_k$, for $w_k = w_{k1} + w_{k2}$, where

$$\begin{aligned} w_{k1} &= (1/2) \exp(\sup_{N(\boldsymbol{\theta}_0)} \delta_k(\boldsymbol{\theta})) \sup_{N(\boldsymbol{\theta}_0)} |\partial \delta_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|^2 ; \\ w_{k2} &= (1/2) \exp(\sup_{N(\boldsymbol{\theta}_0)} \delta_k(\boldsymbol{\theta})) \sup_{N(\boldsymbol{\theta}_0)} \left\|\partial^2 \delta_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^2\right\| . \end{aligned}$$

It remains to prove that $E(w_k)$ and $E(w_k^2)$ are bounded. From the Hölder inequality,

$$E(w_{k1}) \leq (1/2)\, E^{1/2}[\exp(2 \sup_{N(\boldsymbol{\theta}_0)} \delta_k(\boldsymbol{\theta}))]\, E^{1/2}(\sup_{N(\boldsymbol{\theta}_0)} |\partial \delta_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|^4). \tag{21}$$

We are going to show that both expectations on the right of the inequality above are bounded. We start with the second one. It holds that

$$\hat{\mu}_k(\boldsymbol{\theta}) = \mathbf{b}_k'(\boldsymbol{\theta})\, \mathbf{y}, \quad \text{where} \quad \mathbf{b}_k(\boldsymbol{\theta}) = \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{X} \mathbf{Q}(\boldsymbol{\theta}) \boldsymbol{\lambda}_k + \sigma_u^2\, \mathbf{P}(\boldsymbol{\theta}) \mathbf{Z} \mathbf{m}_k, \quad k = 1, 2. \tag{22}$$

Let $b_{kj}(\boldsymbol{\theta})$ be the j-th element of $\mathbf{b}_k(\boldsymbol{\theta})$. After some algebra, it can be seen that

$$\sup_{N(\boldsymbol{\theta}_0)} |\mathbf{b}_k(\boldsymbol{\theta})| = O(1); \quad \sum_{j=1}^{n} \sup_{N(\boldsymbol{\theta}_0)} \|\partial b_{kj}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\| = O(1). \tag{23}$$

Then, using the Hölder inequality, (23) and taking into account and that the vector $\mathbf{y}$ is normally distributed, we obtain

$$E\left(\sup_{N(\boldsymbol{\theta}_0)} |\partial \hat{\mu}_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|^4 \right) \leq \left(\max_{1 \leq j \leq n} E|y_j|^4 \right) \sum_{j=1}^{n} \sup_{N(\boldsymbol{\theta}_0)} \|\partial b_{kj}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\| = O(1). \tag{24}$$

Furthermore, it holds

$$\sup_{N(\boldsymbol{\theta}_0)} |\alpha_k(\boldsymbol{\theta})| = O(1); \quad \sup_{N(\boldsymbol{\theta}_0)} |\partial \alpha_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta}| = O(1). \tag{25}$$

From the Minkowski inequality, (24) and (25), we obtain

$$E\left(\sup_{N(\boldsymbol{\theta}_0)} \left| \frac{\partial \delta_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|^4 \right) \leq \left[E^{1/4} \left(\sup_{N(\boldsymbol{\theta}_0)} \left| \frac{\partial \hat{\mu}_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|^4 \right) + E^{1/4} \left(\sup_{N(\boldsymbol{\theta}_0)} \left| \frac{\partial \alpha_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|^4 \right) \right]^4 = O(1). \tag{26}$$

Concerning the first expectation on the right of (21), by the equality $\delta_k(\boldsymbol{\theta}) = \alpha_k(\boldsymbol{\theta}) + \hat{\mu}_k(\boldsymbol{\theta})$, (22) and using the moment generating function of a multivariate normal distribution, we obtain

$$E[\exp(2 \sup_{N(\boldsymbol{\theta}_0)} \delta_k(\boldsymbol{\theta}))] = \exp[2 \boldsymbol{\lambda}_k' \boldsymbol{\beta} + 2(\sup_{N(\boldsymbol{\theta}_0)} \alpha_k(\boldsymbol{\theta})) + 2(\sup_{N(\boldsymbol{\theta}_0)} \mathbf{b}_k(\boldsymbol{\theta}))' \mathbf{V} (\sup_{N(\boldsymbol{\theta}_0)} \mathbf{b}_k(\boldsymbol{\theta}))].$$

On the one hand, (H1) states that $|\boldsymbol{\lambda}_k| = O(1)$, and on the other hand, (H2) implies that $\|\mathbf{V}\| = O(1)$. These results, together with the left counterparts of (23) and (25), imply that

$$E[\exp(2 \sup_{N(\boldsymbol{\theta}_0)} \delta_k(\boldsymbol{\theta}))] = O(1), \quad k = 1, 2. \tag{27}$$

By (21), (26) and (27), we have that $E(w_{k1}) = O(1)$. In a similar fashion it can be seen that the quantities $E(w_{k2})$, $E(w_{k1}^2)$, $E(w_{k2}^2)$ and $E(w_{k1}\, w_{k2})$ are also bounded. These facts imply that the first two moments of $w_k = w_{k1} + w_{k2}$ are bounded, $k = 1, 2$.

Finally, by replacing (6) in (20) we obtain

$$\hat{\tau}_k^E - \hat{\tau}_k = \mathbf{h}_k' \mathcal{I}^{-1}\mathbf{s} + t_k^*, \quad \text{where} \quad t_k^* = \mathbf{h}_k' \mathbf{r}_* + t_k, \quad k = 1, 2. \tag{28}$$

Here the error term satisfies $|t_k^*| \le D^{-\eta}\, w_k^*$, where $w_k^* = |\mathbf{h}_k|\, v_* + w_k$ and it is not difficult to see that $E(w_k^*) = O(1)$ and $E(w_k^{*2}) = O(1)$.

Now we prove B). By (28), we can write

$$E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2^E - \hat{\tau}_2)1_{\mathcal{B}}\right] = E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})1_{\mathcal{B}}\right]$$
$$+E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})\, t_2^* 1_{\mathcal{B}}\right] + E\left[(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\, t_1^* 1_{\mathcal{B}}\right] + E(t_1^*\, t_2^* 1_{\mathcal{B}}).$$

We are going to show that for any $\eta \in (0, 1)$,

$$E\left|(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})\, t_2^*\, 1_{\mathcal{B}}\right| = O(D^{-1/2-\eta}) \quad \text{and} \quad E|t_1^*\, t_2^*\, 1_{\mathcal{B}}| = O(D^{-2\eta}). \tag{29}$$

The second part of (29) follows easily from the definition of t_k^* given in (28) and Lemma 1, using the Hölder inequality. As to the first part, since $\mathbf{h}_1 = \exp(\delta_1)(\partial\delta_1/\partial\boldsymbol{\theta})$, by the Hölder inequality,

$$E\left|(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})\, t_2^*\, 1_{\mathcal{B}}\right| \le E^{1/4}\left[\exp(4\delta_1)\right]\, E^{1/4}\left|(\partial\delta_1/\partial\boldsymbol{\theta})'\, \mathcal{I}^{-1}\mathbf{s}\right|^4\, E^{1/2}[(t_2^*)^2 1_{\mathcal{B}}].$$

The first expectation on the right-hand side of the inequality above is bounded due to the same reasons as (27). Further, we have seen that

$$E^{1/2}[(t_2^*)^2\, 1_{\mathcal{B}}] \le D^{-\eta}\, E^{1/2}(w_k^*)^2 = O(D^{-\eta}). \tag{30}$$

Using the Hölder inequality again, we have

$$E^{1/4}\left|(\partial\delta_1/\partial\boldsymbol{\theta})'\, \mathcal{I}^{-1}\mathbf{s}\right|^4 \le E^{1/8}\left|\partial\delta_1/\partial\boldsymbol{\theta}\right|^8 E^{1/8}|\mathcal{I}^{-1}\mathbf{s}|^8. \tag{31}$$

Proceeding as in (26), it can be seen that $E^{1/8}|\partial\delta_1/\partial\boldsymbol{\theta}|^8 = O(1)$. Furthermore it is not difficult to prove that $E|D^{-1/2}\mathbf{s}|^8 = O(1)$. Since (H5) implies that $\|\mathcal{I}^{-1}\| = O(D^{-1})$, then

$$E^{1/8}|\mathcal{I}^{-1}\mathbf{s}|^8 \le D^{1/2}\, \|\mathcal{I}^{-1}\|\, E^{1/8}|D^{-1/2}\mathbf{s}|^8 = O(D^{-1/2}).$$

Therefore,

$$E^{1/4}\left|(\partial\delta_1/\partial\boldsymbol{\theta})'\, \mathcal{I}^{-1}\mathbf{s}\right|^4 = O(D^{-1/2}). \tag{32}$$

Thus, (30) and (32) lead to the first statement of (29); analogously, it holds $E|(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\, t_1^*\, 1_{\mathcal{B}}| = O(D^{-1/2-\eta})$. The result then follows by taking $\eta \in (1/2, 1)$.

Concerning C), by the Hölder inequality, (22), (23), (25) and Lemma 1 with $\eta \in (1/2, 1)$, we can write

$$E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2^E - \hat{\tau}_2)\, 1_{\mathcal{B}^c}\right] \le E\left[\exp\left(\hat{\alpha} + \hat{\mu}_1^E + \hat{\mu}_2^E\right) 1_{\mathcal{B}^c}\right] + E\left[\exp\left(\alpha + \hat{\mu}_1 + \hat{\mu}_2\right) 1_{\mathcal{B}^c}\right]$$
$$\le 2\exp(\sup_{N(\boldsymbol{\theta}_0)} \alpha(\boldsymbol{\theta}))\, E^{1/2}[\exp(2 \sup_{N(\boldsymbol{\theta}_0)} \mathbf{b}'(\boldsymbol{\theta})\mathbf{y})]\, P^{1/2}(\mathcal{B}^c) = O(D^{-b/16}).$$

and this is $o(D^{-1})$ for b large enough ($b > 16$).

Finally, D) yields from the equality

$$\begin{aligned} E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\,1_{\mathcal{B}}\right] &= E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s} - \mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s}\,1_{\mathcal{B}^c})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s} - \mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s}\,1_{\mathcal{B}^c})\right] \\ &= E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\right] - E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\,1_{\mathcal{B}^c}\right], \end{aligned}$$

and showing that the last term is $o(D^{-1})$. Indeed, by the Hölder inequality,

$$\begin{aligned} E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\,1_{\mathcal{B}^c}\right] \le E^{1/2}\left[\exp(2\delta_1 + 2\delta_2)\right] \times \\ E^{1/4}\left|(\partial\delta_1/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\,1_{\mathcal{B}^c}\right|^4 E^{1/4}\left|(\partial\delta_2/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\,1_{\mathcal{B}^c}\right|^4 \end{aligned}$$

The first expectation on the right-hand side of the inequality is bounded by the same arguments used for (27). Further, applying again the Hölder inequality,

$$E^{1/4}\left|(\partial\delta_k/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\,1_{\mathcal{B}^c}\right|^4 \le E^{1/8}\left|(\partial\delta_k/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\right|^8 P^{1/8}(\mathcal{B}^c).$$

But proceeding as in (31), we obtain $E^{1/8}\left|(\partial\delta_k/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\right|^8 = O(D^{-1/2})$, and taking $\eta > 1/4$ in Lemma 1, we get $P^{1/8}(\mathcal{B}^c) = O(D^{-b/64})$. Therefore, for any $b > 0$ and for $\eta > 1/4$, we have

$$E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\,1_{\mathcal{B}^c}\right] = \left[O(D^{-1/2-b/64})\right]^2 = o(D^{-1}).$$

□

The following two results are technical lemmas needed for the proof of Theorem 2.

Lemma 4 *Let A_i, $i = 1, 2, 3$ be $n \times n$ nonstochastic symmetric matrices and $\mathbf{v} \sim N_n(\mathbf{0}_n, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is positive definite. Then,*

(i) $E\left[\mathbf{v}(\mathbf{v}'A_1\mathbf{v})\mathbf{v}'\right] = tr(A_1\boldsymbol{\Sigma})\boldsymbol{\Sigma} + 2\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}$;

(ii) $E\left\{\mathbf{v}[\mathbf{v}'A_1\mathbf{v} - E(\mathbf{v}'A_1\mathbf{v})]\mathbf{v}'\right\} = 2\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}$;

(iii) $E\left\{\prod_{i=1}^{2}[\mathbf{v}'A_i\mathbf{v} - E(\mathbf{v}'A_i\mathbf{v})]\right\} = 2tr(A_1\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma})tr(\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma})tr(A_2\boldsymbol{\Sigma})$;

(iv) $E\left\{\mathbf{v}\prod_{i=1}^{2}[\mathbf{v}'A_i\mathbf{v} - E(\mathbf{v}'A_i\mathbf{v})]\mathbf{v}'\right\} = 2tr(A_1\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma})\boldsymbol{\Sigma} + 4\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma} + 4\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}$;

(v) $E\left\{\prod_{i=1}^{3}[\mathbf{v}'A_i\mathbf{v} - E(\mathbf{v}'A_i\mathbf{v})]\right\} = 4tr(A_1\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma}) + 4tr(A_1\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma})$;

(vi)
$$\begin{aligned} E\left\{\mathbf{v}\prod_{i=1}^{3}[\mathbf{v}'A_i\mathbf{v} - E(\mathbf{v}'A_i\mathbf{v})]\mathbf{v}'\right\} &= 4tr(A_1\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma})\boldsymbol{\Sigma} + 4tr(A_1\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma})\boldsymbol{\Sigma} \\ &+4tr(A_1\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma})\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma} + 4tr(A_1\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma})\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma} + 4tr(A_2\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma})\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma} \\ &+2tr(A_2\boldsymbol{\Sigma})\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma} + 2tr(A_3\boldsymbol{\Sigma})\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma} - 2tr(A_2\boldsymbol{\Sigma})\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma} \\ &-2tr(A_3\boldsymbol{\Sigma})\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma} + 12\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma} + 8\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma} + 12\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma} \\ &+8\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma} + 4\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma} + 4\boldsymbol{\Sigma}A_3\boldsymbol{\Sigma}A_2\boldsymbol{\Sigma}A_1\boldsymbol{\Sigma}. \end{aligned}$$

PROOF OF LEMMA 4 (i) appears in Lemma A.1 of Prasad and Rao (1990), and (iii) is a direct consequence of the same lemma. (ii) is easily obtained from (i). (iv) follows by straightforward algebra and application of (i). Finally, (v) and (vi) are obtained by application of the recurrence formula of Srivastava and Tiwari (1976) and straightforward algebra. □

In the following lemma, we use the notation $\boldsymbol{\Sigma}_{\mathbf{b}} = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{b}\mathbf{b}'\boldsymbol{\Sigma}$.

Lemma 5 *Let A_1 and A_2 be two $n \times n$ nonstochastic symmetric matrices, $\mathbf{b}$ and $\mathbf{c}$ two nonstochastic vectors of size n and $\mathbf{v} \sim N_n(\mathbf{0}_n, \mathbf{\Sigma})$. Then,*

(i) $E\left[exp(\mathbf{b}'\mathbf{v})\,\mathbf{v}\right] = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\,\mathbf{\Sigma}\,\mathbf{b}$;

(ii) $E\left[exp(\mathbf{b}'\mathbf{v})\,\mathbf{v}\mathbf{v}'\right] = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\,\mathbf{\Sigma}_{\mathbf{b}}$;

(iii) $E\left[exp(\mathbf{b}'\mathbf{v})\,\mathbf{v}\mathbf{v}'(\mathbf{c}'\mathbf{v})\right] = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)(\mathbf{\Sigma}\mathbf{b}\mathbf{c}'\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{c}\mathbf{b}'\mathbf{\Sigma})$;

(iv) $E\left\{exp(\mathbf{b}'\mathbf{v})\,[\mathbf{v}'A_1\mathbf{v} - E(\mathbf{v}'A_1\mathbf{v})]\right\} = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\,\mathbf{b}'\mathbf{V}A_1\mathbf{\Sigma}\mathbf{b}$;

(v) $E\left\{exp(\mathbf{b}'\mathbf{v})\,[\mathbf{v}'A_1\mathbf{v} - E(\mathbf{v}'A_1\mathbf{v})]\mathbf{v}'\right\} = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\;[2\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma} + (\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b})\mathbf{b}'\mathbf{\Sigma}]$;

(vi) $E\left\{exp(\mathbf{b}'\mathbf{v})\,\mathbf{v}[\mathbf{v}'A_1\mathbf{v} - E(\mathbf{v}'A_1\mathbf{v})]\mathbf{v}'\right\} = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\times$
$[2\mathbf{\Sigma}_{\mathbf{b}}A_1\mathbf{\Sigma} + (\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b})\mathbf{\Sigma}_{\mathbf{b}} + 2\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b}\mathbf{b}'\mathbf{\Sigma}]$;

(vii) $E\{exp(\mathbf{b}'\mathbf{v})\,\mathbf{v}[\mathbf{v}'A_1\mathbf{v} - E(\mathbf{v}'A_1\mathbf{v})]\mathbf{v}'(\mathbf{c}'\mathbf{v})\} = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\times$
$[2(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\,\mathbf{c})\mathbf{\Sigma}_{\mathbf{b}}) + 4\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b}\mathbf{c}'\mathbf{\Sigma} + 2\mathbf{\Sigma}\mathbf{b}\mathbf{c}'\mathbf{\Sigma}A_i\mathbf{\Sigma} + 2\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{c}\mathbf{b}'\mathbf{\Sigma}$
$+(\mathbf{b}'\mathbf{\Sigma}A_i\mathbf{\Sigma}\mathbf{b})\mathbf{\Sigma}\mathbf{b}\mathbf{c}'\mathbf{\Sigma} + (\mathbf{b}'\mathbf{\Sigma}A_i\mathbf{\Sigma}\mathbf{b})\mathbf{\Sigma}\mathbf{c}\mathbf{b}'\mathbf{\Sigma} + 2(\mathbf{b}'\mathbf{\Sigma}\mathbf{c})\mathbf{\Sigma}_{\mathbf{b}}A_1\mathbf{\Sigma}$
$+(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\,\mathbf{b})(\mathbf{b}'\mathbf{\Sigma}\mathbf{c})\mathbf{\Sigma}_{\mathbf{b}} + 2(\mathbf{b}'\mathbf{\Sigma}\mathbf{c})\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b}\mathbf{b}'\mathbf{\Sigma}]$;

(viii) $E\left\{exp(\mathbf{b}'\mathbf{v})\,\prod_{i=1}^{2}[\mathbf{v}'A_i\mathbf{v} - E(\mathbf{v}'A_i\mathbf{v})]\right\} = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\times$
$[2tr(A_1\mathbf{\Sigma}A_2\mathbf{\Sigma}) + 4\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b} + 2(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b})(\mathbf{b}'\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b})]$;

(ix) $E\left\{exp(\mathbf{b}'\mathbf{v})\,\prod_{i=1}^{2}[\mathbf{v}'A_i\mathbf{v} - E(\mathbf{v}'A_i\mathbf{v})]\mathbf{v}'\right\} = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\times$
$[2tr(A_1\mathbf{\Sigma}A_2\mathbf{\Sigma})\mathbf{b}'\mathbf{\Sigma} + 4(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b})\mathbf{b}'\mathbf{\Sigma} + 2(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b})(\mathbf{b}'\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b})\mathbf{b}'\mathbf{\Sigma}$
$+4\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}A_2\mathbf{\Sigma} + 4\mathbf{b}'\mathbf{\Sigma}A_2\mathbf{\Sigma}A_1\mathbf{\Sigma} + 2(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b})\mathbf{b}'\mathbf{\Sigma}A_2\mathbf{\Sigma}$
$+2(\mathbf{b}'\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b})\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}]$;

(x) $E\left\{exp(\mathbf{b}'\mathbf{v})\,\mathbf{v}\prod_{i=1}^{2}[\mathbf{v}'A_i\mathbf{v} - E(\mathbf{v}'A_i\mathbf{v})]\mathbf{v}'\right\} = exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\times$
$[2tr(A_1\mathbf{\Sigma}A_2\mathbf{\Sigma})\mathbf{\Sigma}_{\mathbf{b}} + 2(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b})\mathbf{\Sigma}_{\mathbf{b}}A_2\mathbf{\Sigma} + 2(\mathbf{b}'\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b})\mathbf{\Sigma}_{\mathbf{b}}A_1\mathbf{\Sigma}$
$+(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b})(\mathbf{b}'\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b})\mathbf{\Sigma}_{\mathbf{b}} + 4(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b})\mathbf{\Sigma}_{\mathbf{b}} + 4\mathbf{\Sigma}_{\mathbf{b}}A_1\mathbf{\Sigma}A_2\mathbf{\Sigma}$
$+4\mathbf{\Sigma}_{\mathbf{b}}A_2\mathbf{\Sigma}A_1\mathbf{\Sigma} + 8\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b}\mathbf{b}'\mathbf{\Sigma}A_2\mathbf{\Sigma} + 4\mathbf{\Sigma}A_2\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b}\mathbf{b}'\mathbf{\Sigma} + 4\mathbf{\Sigma}A_1\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b}\mathbf{b}'\mathbf{\Sigma}$
$+2(\mathbf{b}'\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b})\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b}\mathbf{b}'\mathbf{\Sigma} + 2(\mathbf{b}'\mathbf{\Sigma}A_2\mathbf{\Sigma}\mathbf{b})\mathbf{\Sigma}A_1\mathbf{\Sigma}\mathbf{b}\mathbf{b}'\mathbf{\Sigma}]$.

PROOF OF LEMMA 5 (i) Observe that the expectation can be written as

$$E\left[\exp(\mathbf{b}'\mathbf{v})\mathbf{v}\right] = \exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)\int \mathbf{v}\,(2\pi)^{-n/2}|\mathbf{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\,(\mathbf{v}-\mathbf{\Sigma}\mathbf{b})'\mathbf{\Sigma}^{-1}(\mathbf{v}-\mathbf{\Sigma}\mathbf{b})\right\}d\mathbf{v},$$

where the integral is the expectation of a random vector $\mathbf{z} \sim N(\mathbf{\Sigma}\mathbf{b}, \mathbf{\Sigma})$; that is,

$$E\left[\exp(\mathbf{b}'\mathbf{v})\mathbf{v}\right] = \exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)E(\mathbf{z}) = \exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b})\,\mathbf{\Sigma}\mathbf{b}.$$

(ii) is obtained similarly as (i). As to (iii), rearranging the integral as in (i) and making the change of variable $\mathbf{w} = \mathbf{v} - \mathbf{\Sigma}\mathbf{b}$, we obtain

$$E\{\exp(\mathbf{b}'\mathbf{v})\mathbf{v}\mathbf{v}'(\mathbf{c}'\mathbf{v})\} = \exp(\mathbf{b}'\mathbf{\Sigma}\mathbf{b}/2)E\left[(\mathbf{w}+\mathbf{\Sigma}\mathbf{b})(\mathbf{w}+\mathbf{\Sigma}\mathbf{b})'\mathbf{c}'(\mathbf{w}+\mathbf{\Sigma}\mathbf{b})\right],$$

for $\mathbf{w} \sim N(\mathbf{0}_n, \mathbf{\Sigma})$. Then the result follows by straightforward algebra, and taking into account that the expectations where $\mathbf{w}$ appears an odd number of times are zero. Results (iv)–(x) are obtained using similar arguments as in (iii) and application of Lemma 4. □

PROOF OF THEOREM 2 Since $\mathbf{h}_k = \exp(\delta_k)\,(\partial\delta_k/\partial\boldsymbol{\theta}),\ k=1,2$, we can write

$$E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\right] = E\left\{\exp(\delta_1+\delta_2)\left[(\partial\delta_1/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\right]\left[(\partial\delta_2/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\right]\right\} \tag{33}$$

On the one hand, the derivatives can be written as

$$\partial\delta_k/\partial\boldsymbol{\theta} = \mathbf{d}_k + D_k\mathbf{v}, \quad \text{for} \quad \mathbf{d}_k = \partial\alpha_k/\partial\boldsymbol{\theta} \quad \text{and} \quad D_k = F_k + \partial\boldsymbol{\gamma}_k'/\partial\boldsymbol{\theta},$$

where $\partial\boldsymbol{\gamma}_k'/\partial\boldsymbol{\theta} = \mathrm{col}_{1\le j\le 3}(\partial\boldsymbol{\gamma}_k'/\partial\theta_j)$ and $F_k = \mathrm{col}_{1\le j\le 3}(\mathbf{f}_{kj})$ with

$$\mathbf{f}_{kj} = -\mathbf{P}\boldsymbol{\Delta}_j\mathbf{V}^{-1}\mathbf{XQ}(\boldsymbol{\lambda}_k - \mathbf{X}'\gamma_k) - \mathbf{V}^{-1}\mathbf{XQX}'\left(\partial\boldsymbol{\gamma}_k'/\partial\boldsymbol{\theta}\right).$$

On the other hand, the score vector $\mathbf{s}$ is equal to

$$\mathbf{s} = (\mathbf{q} - E\mathbf{q})/2 + \boldsymbol{\nu}, \tag{34}$$

where $\mathbf{q} = (q_1, q_2, q_3)'$ and $\boldsymbol{\nu} = (\nu_1,\nu_2,\nu_3)'$, with

$$q_j = \mathbf{v}'\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\mathbf{v} \quad \text{and} \quad \nu_j = \left[\mathrm{tr}(\mathbf{P}\boldsymbol{\Delta}_j) - \mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_j)\right],\ j=1,2,3.$$

Let us denote $\boldsymbol{\omega}_k = \mathcal{I}^{-1}\partial\delta_k/\partial\boldsymbol{\theta} = \mathbf{g}_k + C_k\mathbf{v}$, where $\mathbf{g}_k = \mathcal{I}^{-1}\mathbf{d}_k = (g_{k1}, g_{k2}, g_{k3})'$ and $C_k = \mathcal{I}^{-1}D_k = \mathrm{col}_{1\le j\le 3}(\mathbf{c}_{kj}'),\ k=1,2$. With this notation and (34),

$$(\partial\delta_k/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s} = \boldsymbol{\omega}_k'(\mathbf{q}-E\mathbf{q})/2 + \boldsymbol{\omega}_k'\boldsymbol{\nu},\ k=1,2. \tag{35}$$

Further, by (18), for $\mathbf{v} = \mathbf{Zu}+\mathbf{e}$ it holds that $\delta_1+\delta_2 = \alpha + \boldsymbol{\lambda}'\boldsymbol{\beta} + \mathbf{b}'\mathbf{v}$. Inserting this and (35) into (33) we get

$$\begin{aligned}
&E\left[(\mathbf{h}_1'\mathcal{I}^{-1}\mathbf{s})(\mathbf{h}_2'\mathcal{I}^{-1}\mathbf{s})\right] = \frac{1}{4}E\left[\exp(\alpha+\boldsymbol{\lambda}'\boldsymbol{\beta}+\mathbf{b}'\mathbf{v})\,\boldsymbol{\omega}_1'(\mathbf{q}-E\mathbf{q})\,\boldsymbol{\omega}_2'(\mathbf{q}-E\mathbf{q})\right]\\
&+\frac{1}{2}E\left[\exp(\alpha+\boldsymbol{\lambda}'\boldsymbol{\beta}+\mathbf{b}'\mathbf{v})\,\boldsymbol{\omega}_1'(\mathbf{q}-E\mathbf{q})\,\boldsymbol{\omega}_2'\boldsymbol{\nu}\right] + \frac{1}{2}E\left[\exp(\alpha+\boldsymbol{\lambda}'\boldsymbol{\beta}+\mathbf{b}'\mathbf{v})\,\boldsymbol{\omega}_2'(\mathbf{q}-E\mathbf{q})\,\boldsymbol{\omega}_1'\boldsymbol{\nu}\right]\\
&+E\left[\exp(\alpha+\boldsymbol{\lambda}'\boldsymbol{\beta}+\mathbf{b}'\mathbf{v})\,\boldsymbol{\omega}_1'\boldsymbol{\nu}\,\boldsymbol{\omega}_2'\boldsymbol{\nu}\right] = B_1+B_2+B_3+B_4.
\end{aligned} \tag{36}$$

We start calculating B_1. For this, first let us denote $\mathbf{A}_j = \mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}$, so that $q_j = \mathbf{v}'\mathbf{A}_j\mathbf{v}$, $j=1,2,3$. Using the expressions $\boldsymbol{\omega}_k = \mathbf{g}_k + C_k\mathbf{v},\ k=1,2$, we can write

$$\begin{aligned}
B_1 &= (1/4)\exp(\alpha+\boldsymbol{\lambda}'\boldsymbol{\beta})\sum_{i=1}^{3}\sum_{j=1}^{3}\left\{\mathbf{c}_{1i}'E\left[\exp(\mathbf{b}'\mathbf{v})\mathbf{v}(q_i-Eq_i)(q_j-Eq_j)\mathbf{v}'\right]\mathbf{c}_{2j}\right.\\
&\quad+g_{1i}E\left[\exp(\mathbf{b}'\mathbf{v})(q_i-Eq_i)(q_j-Eq_j)\mathbf{v}'\right]\mathbf{c}_{2j} + \mathbf{c}_{1i}'E\left[\exp(\mathbf{b}'\mathbf{v})\mathbf{v}(q_i-Eq_i)(q_j-Eq_j)\right]g_{2j}\\
&\quad\left.+g_{1i}E\left[\exp(\mathbf{b}'\mathbf{v})(q_i-Eq_i)(q_j-Eq_j)\right]g_{2j}\right\}\\
&= \exp(\alpha+\boldsymbol{\lambda}'\boldsymbol{\beta})(B_{11}+B_{12}+B_{13}+B_{14}).
\end{aligned}$$

Applying Lemma 5 (x) we obtain

$$\begin{aligned}
B_{11} \;=\;& (1/4)\exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\times\\
&\sum_{i=1}^{3}\sum_{j=1}^{3}\left[2\,\mathrm{tr}(\mathbf{A}_i\mathbf{V}\mathbf{A}_j\mathbf{V})\,\mathbf{c}_{1i}'\mathbf{V}\mathbf{c}_{2j} + 2\,\mathrm{tr}(\mathbf{A}_i\mathbf{V}\mathbf{A}_j\mathbf{V})\,\mathbf{c}_{1i}'\mathbf{V}\mathbf{b}\mathbf{b}'\mathbf{V}\mathbf{c}_{2j}\right.\\
&+2\,(\mathbf{b}'\mathbf{V}\mathbf{A}_i\mathbf{V}\mathbf{b})\,\mathbf{c}_{1i}'\mathbf{V}_{\mathbf{b}}\mathbf{A}_j\mathbf{V}\mathbf{c}_{2j} + 2\,(\mathbf{b}'\mathbf{V}\mathbf{A}_j\mathbf{V}\mathbf{b})\,\mathbf{c}_{1i}'\mathbf{V}_{\mathbf{b}}\mathbf{A}_i\mathbf{V}\mathbf{c}_{2j}\\
&+(\mathbf{b}'\mathbf{V}\mathbf{A}_i\mathbf{V}\mathbf{b})(\mathbf{b}'\mathbf{V}\mathbf{A}_j\mathbf{V}\mathbf{b})\,\mathbf{c}_{1i}'\mathbf{V}_{\mathbf{b}}\mathbf{c}_{2j} + 4\,(\mathbf{b}'\mathbf{V}\mathbf{A}_i\mathbf{V}\mathbf{A}_j\mathbf{V}\mathbf{b})\,\mathbf{c}_{1i}'\mathbf{V}_{\mathbf{b}}\mathbf{c}_{2j}\\
&+4\,\mathbf{c}_{1i}'\mathbf{V}_{\mathbf{b}}\mathbf{A}_i\mathbf{V}\mathbf{A}_j\mathbf{V}\mathbf{c}_{2j} + 4\,\mathbf{c}_{1i}'\mathbf{V}_{\mathbf{b}}\mathbf{A}_j\mathbf{V}\mathbf{A}_i\mathbf{V}\mathbf{c}_{2j}\\
&+8\,\mathbf{c}_{1i}'\mathbf{V}\mathbf{A}_i\mathbf{V}\mathbf{b}\mathbf{b}'\mathbf{V}\mathbf{A}_j\mathbf{V}\mathbf{c}_{2j} + 4\,\mathbf{c}_{1i}'\mathbf{V}\mathbf{A}_j\mathbf{V}\mathbf{A}_i\mathbf{V}\mathbf{b}\mathbf{b}'\mathbf{V}\mathbf{c}_{2j}\\
&+4\,\mathbf{c}_{1i}'\mathbf{V}\mathbf{A}_i\mathbf{V}\mathbf{A}_j\mathbf{V}\mathbf{b}\mathbf{b}'\mathbf{V}\mathbf{c}_{2j} + 2\,(\mathbf{b}'\mathbf{V}\mathbf{A}_i\mathbf{V}\mathbf{b})\,\mathbf{c}_{1i}'\mathbf{V}\mathbf{A}_j\mathbf{V}\mathbf{b}\mathbf{b}'\mathbf{V}\mathbf{c}_{2j}\\
&\left.+\,2\,(\mathbf{b}'\mathbf{V}\mathbf{A}_j\mathbf{V}\mathbf{b})\,\mathbf{c}_{1i}'\mathbf{V}\mathbf{A}_i\mathbf{V}\mathbf{b}\mathbf{b}'\mathbf{V}\mathbf{c}_{2j}\right].
\end{aligned} \tag{37}$$

We study each term in (37). The first one is equal to

$$\sum_{i=1}^{3}\sum_{j=1}^{3} 2\,\mathrm{tr}(\mathbf{A}_i\mathbf{V}\mathbf{A}_j\mathbf{V})\,\mathbf{c}_{1i}'\mathbf{V}\mathbf{c}_{2j} = \mathrm{tr}\left[Cov(C_1\mathbf{v}, C_2\mathbf{v})Var(\mathbf{q})\right]. \tag{38}$$

It holds that $Cov(C_1\mathbf{v}, C_2\mathbf{v}) = \mathcal{I}^{-1}D_1\mathbf{V}D_2'\mathcal{I}^{-1}$, and it is not difficult to see that $Var(\mathbf{q}) = 4\mathcal{I} + \mathcal{K}$, where $\mathcal{K}$ is a matrix whose element (i,j) is

$$k_{ij} = 2\left[\mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_i\mathbf{V}^{-1}\boldsymbol{\Delta}_j) - \mathrm{tr}(\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_j)\right] = O(1), \quad i,j = 1,2,3.$$

Using these results and the facts that $\|\mathcal{I}^{-1}\| = O(D^{-1})$, $\|D_k\| = O(1)$, $k = 1,2$ and $\|\mathcal{K}\| = O(1)$, we obtain

$$(1/4)\,\mathrm{tr}\left[Cov(C_1\mathbf{v}, C_2\mathbf{v})Var(\mathbf{q})\right] = \mathrm{tr}(\mathcal{I}^{-1}D_1\mathbf{V}D_2') + O(D^{-1}),$$

where

$$\begin{aligned}\mathrm{tr}(\mathcal{I}^{-1}D_1\mathbf{V}D_2') &= \mathrm{tr}\left(\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}_1'}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_2}{\partial\boldsymbol{\theta}}\right) + \mathrm{tr}\left(\mathcal{I}^{-1}F_1\mathbf{V}F_2'\right)\\ &+ \mathrm{tr}\left(\mathcal{I}^{-1}F_1\mathbf{V}\frac{\partial\boldsymbol{\gamma}_2}{\partial\boldsymbol{\theta}}\right) + \mathrm{tr}\left(\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}_1'}{\partial\boldsymbol{\theta}}\mathbf{V}F_2'\right) + O(D^{-2}).\end{aligned} \tag{39}$$

Assumptions (H2) and (H5) imply respectively that $\|\mathbf{V}\| = O(1)$ and $\|\mathcal{I}^{-1}\| = O(D^{-1})$. Further, using (H1)–(H4), we obtain $\|F_k\| = O(D^{-1/2})$ and $\|\partial\boldsymbol{\gamma}_k/\partial\boldsymbol{\theta}\| = O(1)$, $k = 1,2$. Therefore,

$$\mathrm{tr}\left(\mathcal{I}^{-1}F_1\mathbf{V}F_2'\right) \le 3\,\|\mathcal{I}^{-1}F_1\mathbf{V}F_2'\| \le 3\,\|\mathbf{V}\|\|F_1\|\|F_2'\|\|\mathcal{I}^{-1}\| = O(D^{-2}); \tag{40}$$

$$\mathrm{tr}\left[\mathcal{I}^{-1}F_1\mathbf{V}\left(\partial\boldsymbol{\gamma}_2/\partial\boldsymbol{\theta}\right)\right] \le 3\,\|\mathcal{I}^{-1}\|\|F_1\|\|\mathbf{V}\|\,\|\partial\boldsymbol{\gamma}_2/\partial\boldsymbol{\theta}\| = O(D^{-3/2}), \tag{41}$$

and $\mathrm{tr}\left[\mathcal{I}^{-1}\left(\partial\boldsymbol{\gamma}_1'/\partial\boldsymbol{\theta}\right)\mathbf{V}F_2'\right] = O(D^{-3/2})$ analogously to (41). From (39)–(41) we obtain

$$(1/4)\mathrm{tr}\left[Cov(C_1\mathbf{v}, C_2\mathbf{v})Var(\mathbf{q})\right] = \mathrm{tr}\left(\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}_1'}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_2}{\partial\boldsymbol{\theta}}\right) + o(D^{-1}). \tag{42}$$

The second term in the sum of equation (37) is equal to

$$\sum_{i=1}^{3}\sum_{j=1}^{3} 2\,\mathrm{tr}(\mathbf{A}_i\mathbf{V}\mathbf{A}_j\mathbf{V})\,\mathbf{c}_{1i}'\mathbf{V}\mathbf{b}\mathbf{b}'\mathbf{V}\mathbf{c}_{2j} = Cov(C_1\mathbf{v}, \mathbf{b}'\mathbf{v})'Var(\mathbf{q})\,Cov(C_2\mathbf{v}, \mathbf{b}'\mathbf{v}) \tag{43}$$

where $Cov(C_k\mathbf{v}, \mathbf{b}'\mathbf{v}) = \mathcal{I}^{-1}D_k\mathbf{V}\mathbf{b}$, $k = 1,2$, and $Var(\mathbf{q}) = 4\mathcal{I} + \mathcal{K}$. Since $|\mathbf{b}| = O(1)$, it can be easily seen that

$$(1/4)Cov(C_1\mathbf{v}, \mathbf{b}'\mathbf{v})'Var(\mathbf{q})\,Cov(C_2\mathbf{v}, \mathbf{b}'\mathbf{v}) = \mathbf{b}'\mathbf{V}D_1'\mathcal{I}^{-1}D_2\mathbf{V}\mathbf{b} + o(D^{-1}).$$

Now inserting $D_k = F_k + \partial\boldsymbol{\gamma}_k'/\partial\boldsymbol{\theta}$, $k = 1,2$, we get

$$\begin{aligned}\mathbf{b}'\mathbf{V}D_1'\mathcal{I}^{-1}D_2\mathbf{V}\mathbf{b} &= \mathbf{b}'\mathbf{V}\frac{\partial\boldsymbol{\gamma}_1}{\partial\boldsymbol{\theta}}\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}_2'}{\partial\boldsymbol{\theta}}\mathbf{V}\mathbf{b} + \mathbf{b}'\mathbf{V}F_1'\mathcal{I}^{-1}F_2\mathbf{V}\mathbf{b}\\ &+\mathbf{b}'\mathbf{V}F_1'\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}_2'}{\partial\boldsymbol{\theta}}\mathbf{V}\mathbf{b} + \mathbf{b}'\mathbf{V}\frac{\partial\boldsymbol{\gamma}_1}{\partial\boldsymbol{\theta}}\mathcal{I}^{-1}F_2\mathbf{V}\mathbf{b},\end{aligned} \tag{44}$$

where

$$\mathbf{b}'\mathbf{V}F_1'\mathcal{I}^{-1}F_2\mathbf{V}\mathbf{b} \le 3\,\|\mathcal{I}^{-1}\|\|\mathbf{b}'\mathbf{V}|^2\|F_1\|\|F_2\| = O(D^{-2}).$$

In a similar way it can be proved that the third and fourth terms on the right of (44) are $o(D^{-1})$. Therefore,

$$(1/4)Cov(C_1\mathbf{v},\mathbf{b}'\mathbf{v})'Var(\mathbf{q})\,Cov(C_2\mathbf{v},\mathbf{b}'\mathbf{v}) = \mathbf{b}'\mathbf{V}\frac{\partial\boldsymbol{\gamma}_1}{\partial\boldsymbol{\theta}}\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}_2'}{\partial\boldsymbol{\theta}}\mathbf{V}\mathbf{b} + o(D^{-1}). \quad (45)$$

Since $|\mathbf{b}| = O(1)$, $\|\mathbf{V}\| = O(1)$, $\|A_i\| = O(1)$ and $|\mathbf{c}_{1i}| = O(D^{-1})$, $i = 1,2,3$, all remaining terms on the right of (37) are $O(D^{-2})$. Further, we know that $\exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2) = O(1)$. Therefore, from (38), (42), (43) and (45) we obtain

$$B_{11} = \exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\left[\mathrm{tr}\left(\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}_1'}{\partial\boldsymbol{\theta}}\mathbf{V}\frac{\partial\boldsymbol{\gamma}_2}{\partial\boldsymbol{\theta}}\right) + \mathbf{b}'\mathbf{V}\frac{\partial\boldsymbol{\gamma}_1}{\partial\boldsymbol{\theta}}\mathcal{I}^{-1}\frac{\partial\boldsymbol{\gamma}_2'}{\partial\boldsymbol{\theta}}\mathbf{V}\mathbf{b}\right] + o(D^{-1}). \quad (46)$$

As to B_{12}, applying Lemma 5 (ix) and using the facts that $|\mathbf{b}| = O(1)$, $\|\mathbf{V}\| = O(1)$ and $\|A_i\| = O(1)$, $i = 1,2,3$, we obtain

$$\begin{aligned} B_{12} &= (1/4)\exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\sum_{i=1}^{3}\sum_{j=1}^{3} g_{1i}\,2\mathrm{tr}(A_i\mathbf{V}A_j\mathbf{V})\,\mathbf{b}'\mathbf{V}\mathbf{c}_{2j} + O(D^{-2}) \\ &= (1/4)\exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\;\mathbf{g}_1 Var(\mathbf{q})Cov(C_2\mathbf{v},\mathbf{b}'\mathbf{v}) + O(D^{-2}). \end{aligned}$$

Now inserting $Var(\mathbf{q}) = 4\mathcal{I} + \mathcal{K}$, $D_2 = F_2 + \partial\boldsymbol{\gamma}_2'/\partial\boldsymbol{\theta}$ and $\mathbf{g}_1 = \mathcal{I}^{-1}\partial\alpha_1/\partial\boldsymbol{\theta}$, and taking into account that $\|\mathcal{I}^{-1}\| = O(D^{-1})$, $|\mathbf{g}_1| = O(D^{-1})$ and $\|F_2\| = O(D^{-1/2})$, and that the rest of the matrices and vectors involved have bounded norm, then

$$B_{12} = \exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\;(\partial\alpha_1/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\left(\partial\boldsymbol{\gamma}_2'/\partial\boldsymbol{\theta}\right)\mathbf{V}\mathbf{b} + O(D^{-3/2}). \quad (47)$$

Similarly, we obtain

$$B_{13} = \exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\;(\partial\boldsymbol{\gamma}_1/\partial\boldsymbol{\theta})\,\mathcal{I}^{-1}\,(\partial\alpha_2/\partial\boldsymbol{\theta})\,\mathbf{V}\mathbf{b} + O(D^{-3/2}). \quad (48)$$

Finally, B_{14} is obtained by application of Lemma 5 (viii),

$$\begin{aligned} B_{14} &= \exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\sum_{i=1}^{3}\sum_{j=1}^{3} g_{1i}g_{2j}\,2\mathrm{tr}(A_i\mathbf{V}A_j\mathbf{V}) + O(D^{-2}) \\ &= \exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\,\mathbf{g}_1 Var(\mathbf{q})\mathbf{g}_2 + O(D^{-2}) \\ &= \exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\;(\partial\alpha_1/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\,(\partial\alpha_2/\partial\boldsymbol{\theta}) + O(D^{-2}). \quad (49) \end{aligned}$$

From results (46)–(49), we obtain

$$\begin{aligned} &B_1 = \exp(\alpha + \boldsymbol{\lambda}'\boldsymbol{\beta} + \mathbf{b}'\mathbf{V}\mathbf{b}/2)\times \\ &\left[\mathrm{tr}\left(\frac{\partial\boldsymbol{\gamma}_1'}{\partial\boldsymbol{\theta}}\,\mathbf{V}\,\frac{\partial\boldsymbol{\gamma}_2}{\partial\boldsymbol{\theta}}\right) + \left(\mathbf{b}'\mathbf{V}\frac{\partial\boldsymbol{\gamma}_1}{\partial\boldsymbol{\theta}} + \left(\frac{\partial\alpha_1}{\partial\boldsymbol{\theta}}\right)'\right)\mathcal{I}^{-1}\left(\frac{\partial\boldsymbol{\gamma}_2'}{\partial\boldsymbol{\theta}}\mathbf{V}\mathbf{b} + \frac{\partial\alpha_2}{\partial\boldsymbol{\theta}}\right)\right] + o(D^{-1}). \quad (50) \end{aligned}$$

Concerning B_2, using $\boldsymbol{\omega}_k = \mathbf{g}_k + C_k\mathbf{v}$, $k = 1,2$, we get

$$\begin{aligned} B_2 = &(1/2)\exp(\alpha + \boldsymbol{\lambda}'\boldsymbol{\beta})\sum_{i=1}^{3}\sum_{j=1}^{3}\nu_j\left\{\mathbf{c}_{1i}'\,E\left[e^{\mathbf{b}'\mathbf{v}}\mathbf{v}(q_i - Eq_i)\mathbf{v}'\right]\mathbf{c}_{2j}\right. \\ &\left. + g_{1i}\mathbf{c}_{2j}'E\left[e^{\mathbf{b}'\mathbf{v}}\mathbf{v}(q_i - Eq_i)\right] + g_{2j}E\left[e^{\mathbf{b}'\mathbf{v}}(q_i - Eq_i)\mathbf{v}'\right]\mathbf{c}_{1i} + g_{1i}g_{2j}E\left[e^{\mathbf{b}'\mathbf{v}}(q_i - Eq_i)\right]\right\} \\ = &\exp(\alpha + \boldsymbol{\lambda}'\boldsymbol{\beta})(B_{21} + B_{22} + B_{23} + B_{24}). \end{aligned}$$

By applying Lemma 5 (vi), we obtain

$$\begin{aligned} B_{21} &= (1/2)\exp(\mathbf{b}'\mathbf{V}\mathbf{b}/2)\sum_{i=1}^{3}\sum_{j=1}^{3}\nu_j\left[\mathbf{c}'_{1i}\mathbf{V}A_i\mathbf{V}\mathbf{c}_{2j} + (\mathbf{b}'\mathbf{V}A_i\mathbf{V}\mathbf{b})\,\mathbf{c}'_{1i}\mathbf{V}_{\mathbf{b}}\mathbf{c}_{2j}\right. \\ &\quad \left. + 2\mathbf{c}'_{1i}\mathbf{V}\mathbf{b}\mathbf{b}'\mathbf{V}A_i\mathbf{V}\mathbf{c}_{2j} + 2\mathbf{c}'_{1i}\mathbf{V}A'_i\mathbf{V}\mathbf{b}\mathbf{b}'\mathbf{V}\mathbf{c}_{2j}\right]. \end{aligned} \tag{51}$$

By the facts that $|\mathbf{c}_{1i}| = O(D^{-1})$ and $|\nu_i| = O(1)$, $i = 1, 2, 3$, all terms involved in this sum are $O(D^{-2})$. Similarly, after using Lemma 5 (iv) and (v) we obtain that all B_{22}, B_{23} and B_{24} are also $O(D^{-2})$.

The proof that $B_3 = O(D^{-2})$ is analogous to B_2. Finally, using Lemma 5 (i) and (ii) we get that $B_4 = O(D^{-2})$. We have seen that B_2–B_4 are all $O(D^{-2})$. Then, the result follows from (50). □

The following lemma is required in the proof of Theorem 3.

Lemma 6 *Let $f(\mathbf{v})$ be a function of $\mathbf{v} = \mathbf{Z} + \mathbf{e}$ such that $\mathbf{v}$ is the only stochastic element in $f(\mathbf{v})$. Then under assumptions (H1)–(H5), for the vector $\boldsymbol{\ell} = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\boldsymbol{\lambda}_2 - \mathbf{X}'\boldsymbol{\gamma}_2)$ it holds*

$$\begin{aligned} E\left[f(\mathbf{v})(\hat{\tau}_2 - \tau_2)\right] = exp(\alpha_2 + \boldsymbol{\lambda}_2\boldsymbol{\beta})\left\{E\left[exp(\boldsymbol{\gamma}'_2\mathbf{v})(\boldsymbol{\ell}'\mathbf{v})f(\mathbf{v})\right]\right. \\ \left. +(1/2)E\left[exp(\boldsymbol{\gamma}'_2\mathbf{v})(\boldsymbol{\ell}'\mathbf{v})^2 f(\mathbf{v})\right]\right\} + o(D^{-1}), \end{aligned}$$

PROOF OF LEMMA 6 Let us define the random variable $\xi_2 = \boldsymbol{\gamma}'_2\mathbf{v} - \mathbf{m}'_2\mathbf{u}$. Then $\xi_2 \sim N(0, 2\alpha_2)$ and is independent of $\mathbf{v}$. By symmetry, $-\xi_2$ has the same distribution, and therefore, the following relation holds

$$E\left[1 - \exp(-\xi_2 - \alpha_2)\right] = 0. \tag{52}$$

Further, from the definition of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ in (3), it follows that $\hat{\mu}_2 = \mu_2 + \xi_2 + \boldsymbol{\ell}'\mathbf{v}$. This last relation allows us to write

$$\begin{aligned} \hat{\tau}_2 - \tau_2 &= \exp(\hat{\mu}_2 + \alpha_2) - \exp(\mu_2) \\ &= \exp(\alpha_2 + \boldsymbol{\lambda}_2\boldsymbol{\beta} + \boldsymbol{\gamma}'_2\mathbf{v})\left[\exp(\boldsymbol{\ell}'\mathbf{v}) - \exp(-\xi_2 - \alpha_2)\right]. \end{aligned} \tag{53}$$

Now let us use Taylor formula for $\exp(\boldsymbol{\ell}'\mathbf{v})$,

$$\exp(\boldsymbol{\ell}'\mathbf{v}) = 1 + \boldsymbol{\ell}'\mathbf{v} + (\boldsymbol{\ell}'\mathbf{v})^2/2 + \epsilon. \tag{54}$$

Here, $\epsilon = x^3/6$, where x satisfies $|x| < |\boldsymbol{\ell}'\mathbf{v}|$. Since $|\boldsymbol{\ell}| = O(D^{-1/2})$, then

$$E(\epsilon^2) \leq (1/36)\,|\boldsymbol{\ell}|^6\,E|\mathbf{v}|^6 = O(D^{-3}).$$

Substituting (54) in (53) and taking into account the independence between ξ_2 and $\mathbf{v}$, we obtain

$$\begin{aligned} E\left[f(\mathbf{v})(\hat{\tau}_2 - \tau_2)\right] = \exp(\alpha_2 + \boldsymbol{\lambda}_2\boldsymbol{\beta})\left\{E\left[\exp(\boldsymbol{\gamma}'_2\mathbf{v})f(\mathbf{v})\right]E\left[1 - \exp(-\xi_2 - \alpha_2)\right]\right. \\ \left. +E\left[\exp(\boldsymbol{\gamma}'_2\mathbf{v})(\boldsymbol{\ell}'\mathbf{v})f(\mathbf{v})\right] + (1/2)E\left[\exp(\boldsymbol{\gamma}'_2\mathbf{v})(\boldsymbol{\ell}'\mathbf{v})^2 f(\mathbf{v})\right]\right\} + o(D^{-1}), \end{aligned}$$

and finally the relation (52) leads to the statement. □

PROOF OF THEOREM 3 By the asymptotic representation (8) and the inequality

$$\begin{aligned} E\left[r_1(\hat{\tau}_2 - \tau_2)\right] &\leq E^{1/2}(r_1^2)E^{1/2}\left[(\hat{\tau}_2 - \tau_2)^2\right] \\ &\leq D^{-3\eta/2}\,E(v_k^2)\left[MSE(\hat{\tau}_2)\right]^{1/2} = O(D^{-3\eta/2}), \end{aligned}$$

taking $\eta \in (2/3, 1)$, we get

$$E\left[(\hat{\tau}_1^E - \hat{\tau}_1)(\hat{\tau}_2 - \tau_2)\right] = E\left[\mathbf{h}_1' \mathcal{I}^{-1}\mathbf{s}\,(\hat{\tau}_2 - \tau_2)\right] + E\left[\mathbf{h}_1' \mathcal{I}^{-1}(H + \mathcal{I})\mathcal{I}^{-1}\mathbf{s}\,(\hat{\tau}_2 - \tau_2)\right]$$
$$+(1/2)E\left[\mathbf{h}_k' \mathcal{I}^{-1}\mathbf{d}\,(\hat{\tau}_2 - \tau_2)\right] + (1/2)E\left[\mathbf{s}'\mathcal{I}^{-1}\frac{\partial^2 \hat{\tau}_k}{\partial \boldsymbol{\theta}^2}\mathcal{I}^{-1}\mathbf{s}\,(\hat{\tau}_2 - \tau_2)\right] + o(D^{-1}). \qquad (55)$$

We are going to prove that all terms on the right-hand side of (55) are also $o(D^{-1})$. As to the first one, since $\mathbf{h}_1 = \exp(\delta_1)(\partial\delta_1/\partial\boldsymbol{\theta})$ and $\delta_1 = \alpha_1 + \boldsymbol{\lambda}_1\boldsymbol{\beta} + \mathbf{b}_1'\mathbf{v}$, then

$$E\left[\mathbf{h}_1' \mathcal{I}^{-1}\mathbf{s}\,(\hat{\tau}_2 - \tau_2)\right] = \exp(\alpha_1 + \boldsymbol{\lambda}_1'\boldsymbol{\beta})\left[\exp(\mathbf{b}_1'\mathbf{v})\,(\partial\delta_1/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\,(\hat{\tau}_2 - \tau_2)\right].$$

Observe that $(\partial\delta_1/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}$ is a function of $\mathbf{v}$, so that Lemma 6 can be applied. By this lemma,

$$E\left[\mathbf{h}_1' \mathcal{I}^{-1}\mathbf{s}\,(\hat{\tau}_2 - \tau_2)\right] = \exp(\alpha + \boldsymbol{\lambda}'\boldsymbol{\beta})\left\{E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})(\boldsymbol{\ell}'\mathbf{v})\,(\partial\delta_1/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\right]\right.$$
$$\left. + E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})(\boldsymbol{\ell}'\mathbf{v})^2\,(\partial\delta_1/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\right]\right\}, \qquad (56)$$

where $\boldsymbol{\kappa} = \mathbf{b}_1 + \boldsymbol{\gamma}_2$. It remains to show that both expectations on the right are $o(D^{-1})$. Using the relations $\mathbf{s} = (\mathbf{q} - E\mathbf{q})/2 + \boldsymbol{\nu}$ and $\boldsymbol{\omega}_1 = \mathcal{I}^{-1}\,(\partial\delta_1/\partial\boldsymbol{\theta}) = \mathbf{g}_1 + C_1\mathbf{v}$, we obtain

$$E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})(\boldsymbol{\ell}'\mathbf{v})\,(\partial\delta_1/\partial\boldsymbol{\theta})'\,\mathcal{I}^{-1}\mathbf{s}\right] = \frac{1}{2}\sum_{i=1}^{3} g_{1i}\,\boldsymbol{\ell}'E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})\mathbf{v}(q_i - Eq_i)\right]$$
$$+\frac{1}{2}\sum_{i=1}^{3}\boldsymbol{\ell}'E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})\mathbf{v}(q_i - Eq_i)\mathbf{v}'\right]\mathbf{c}_{1i} + \boldsymbol{\ell}'E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})\mathbf{v}\right]\mathbf{g}_1'\boldsymbol{\nu} + \boldsymbol{\ell}'E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})\mathbf{v}\mathbf{v}'\right]\mathbf{c}_{1i}'\boldsymbol{\nu}.$$

Then Lemma 5 (i), (ii), (v) and (vi), and the facts that $|g_{1i}| = O(D^{-1})$ and $|\boldsymbol{\ell}| = O(D^{-1/2})$ imply that the first expectation on the right of (56) is $O(D^{-3/2})$. Similarly, after straightforward algebra and the use of Lemma 5 we obtain that the second expectation on the right of (56) is $O(D^{-2})$.

Now we study the second term on the right of (55). As above, by Lemma 6, it suffices to prove that

$$E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})(\boldsymbol{\ell}'\mathbf{v})^j\,\mathbf{h}_1'\mathcal{I}^{-1}(H + \mathcal{I})\mathcal{I}^{-1}\mathbf{s}\right] = o(D^{-1}),\; j = 1, 2. \qquad (57)$$

For this, let us denote by H_{ij} and $\mathcal{I}_{ij}$ respectively the elements (i, j) of the matrices H and $\mathcal{I}$. Observe that

$$H_{ij} = \mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_i\mathbf{V}^{-1}\boldsymbol{\Delta}_j)/2 + \mathbf{v}'\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\mathbf{v},$$

and then, for $A_{ij} = \mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}$, it holds

$$H_{ij} + \mathcal{I}_{ij} = -\left[\mathbf{v}'A_{ij}\mathbf{v} - E(\mathbf{v}'A_{ij}\mathbf{v})\right].$$

Then, the expectation in (57) with $j = 1$ can be written as

$$E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})(\boldsymbol{\ell}'\mathbf{v})\,\mathbf{h}_1'\mathcal{I}^{-1}(H + \mathcal{I})\mathcal{I}^{-1}\mathbf{s}\right] = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}(\mathcal{I}^{-1})_{jk} \times$$
$$\left\{(1/2)\,g_{1i}E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})\left(\mathbf{v}'A_{ij}\mathbf{v} - E(\mathbf{v}'A_{ij}\mathbf{v})\right)(q_k - Eq_k)\mathbf{v}'\right]\boldsymbol{\ell}\right.$$
$$+ g_{1i}\nu_k E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})\left(\mathbf{v}'A_{ij}\mathbf{v} - E(\mathbf{v}'A_{ij}\mathbf{v})\right)\mathbf{v}'\right]\boldsymbol{\ell}$$
$$+ (1/2)\,\mathbf{c}_{1i}'E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})\,\mathbf{v}\left(\mathbf{v}'A_{ij}\mathbf{v} - E(\mathbf{v}'A_{ij}\mathbf{v})\right)(q_k - Eq_k)\mathbf{v}'\right]\boldsymbol{\ell}$$
$$\left. + \nu_k\mathbf{c}_{1i}'E\left[\exp(\boldsymbol{\kappa}'\mathbf{v})\,\mathbf{v}\left(\mathbf{v}'A_{ij}\mathbf{v} - E(\mathbf{v}'A_{ij}\mathbf{v})\right)\mathbf{v}'\right]\boldsymbol{\ell}\right\},$$

where $(\mathcal{I}^{-1})_{jk}$ denotes the element (j,k) of $\mathcal{I}^{-1}$. Then Lemma 5 (v), (vi), (ix) and (x), and the facts $\|A_{ij}\| = O(1)$, $|\nu_k| = O(1)$, $|(\mathcal{I}^{-1})_{ij}| = O(D^{-1})$, $|\mathbf{c}_{1i}| = O(D^{-1})$ and $|\boldsymbol{\ell}| = O(D^{-1/2})$, for $i,j,k = 1,2,3$, imply (57) with $j=1$. The result for $j=2$ can be proven similarly after straightforward but tedious algebra. Using similar arguments it can be seen that the remaining two terms in (55) are also $o(D^{-1})$. For the third term on the right of (55), just remind that $\mathbf{d} = (d_1, d_2, d_3)'$, where $d_i = \mathbf{s}'\mathcal{I}^{-1}D_i\mathcal{I}^{-1}\mathbf{s}$, for $D_i = \partial H/\partial\theta_i$, $i = 1,2,3$. Further, note that the element (j,k) of matrix D_i can be written as a linear combination of $\mathbf{v}$; concretely, $D_{ijk} = a_{ijk} + \mathbf{v}'A_{ijk}\mathbf{v}$, where

$$\begin{aligned} a_{ijk} &= -\mathrm{tr}(\mathbf{V}^{-1}\boldsymbol{\Delta}_i\mathbf{V}^{-1}\boldsymbol{\Delta}_j\mathbf{V}^{-1}\boldsymbol{\Delta}_k); \\ A_{ijk} &= \mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P} + \mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P} + \mathbf{P}\boldsymbol{\Delta}_j\mathbf{P}\boldsymbol{\Delta}_i\mathbf{P}\boldsymbol{\Delta}_k\mathbf{P}. \end{aligned}$$

Finally, concerning the last term in (55), just observe that

$$\partial^2\hat{\tau}_1/\partial\boldsymbol{\theta}^2 = \exp(\alpha_1 + \boldsymbol{\lambda}_1'\boldsymbol{\beta} + \mathbf{b}_1'\mathbf{v})\left[(\partial\delta_1/\partial\boldsymbol{\theta})(\partial\delta_1/\partial\boldsymbol{\theta})' + \partial^2\delta_1/\partial\boldsymbol{\theta}^2\right],$$

and that the elements of the first and second order derivative of δ_1 are also linear functions of $\mathbf{v}$; more concretely,

$$\begin{aligned} \partial\delta_1/\partial\theta_i &= \partial\alpha_1/\partial\theta_i + (\partial\mathbf{b}_1/\partial\theta_i)'\,\mathbf{v}; \\ \partial^2\delta_1/\partial\theta_i\partial\theta_j &= \partial^2\alpha_1/\partial\theta_i\partial\theta_j + \left(\partial^2\mathbf{b}_1/\partial\theta_i\partial\theta_j\right)'\mathbf{v}, \end{aligned}$$

for $i,j = 1,2,3$. $\square$

PROOF OF THEOREM 4 A third order Taylor expansion of $g(\boldsymbol{\beta},\boldsymbol{\theta})$ around $(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ evaluated at $(\boldsymbol{\beta},\boldsymbol{\theta}) = (\hat{\boldsymbol{\beta}}^E,\hat{\boldsymbol{\theta}})$ yields

$$\begin{aligned} g(\hat{\boldsymbol{\beta}}^E,\hat{\boldsymbol{\theta}}) = g(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) &+ \left(\frac{\partial g}{\partial\boldsymbol{\theta}}\right)'(\hat{\boldsymbol{\theta}}^E - \boldsymbol{\theta}_0) + \left(\frac{\partial g}{\partial\boldsymbol{\beta}}\right)'(\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0) + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\frac{\partial^2 g}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &+ \frac{1}{2}(\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0)'\frac{\partial^2 g}{\partial\boldsymbol{\beta}^2}(\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0) + \frac{1}{2}(\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0)'\frac{\partial^2 g}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + r_g, \end{aligned}$$

where, using Lemma 1, it can be seen that $E(r_g) = o(D^{-1})$. Now from Lemma 2, the following relations hold

$$\begin{aligned} E\left[(\partial g/\partial\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right] &= (\partial g/\partial\boldsymbol{\theta})'\left\{E(\mathcal{I}^{-1}\mathbf{s}) + E\left[\mathcal{I}^{-1}(H+\mathcal{I})\mathcal{I}^{-1}\mathbf{s}\right]\right. \\ &\quad \left. + (1/2)\,E(\mathcal{I}^{-1}\mathbf{d})\right\} + o(D^{-1}); \qquad (58) \\ E\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\left(\partial^2 g/\partial\boldsymbol{\theta}^2\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right] &= E\left[\mathbf{s}'\mathcal{I}^{-1}\left(\partial^2 g/\partial\boldsymbol{\theta}^2\right)\mathcal{I}^{-1}\mathbf{s}\right] + o(D^{-1}). \qquad (59) \end{aligned}$$

We are going to show that

$$E\left[(\partial g/\partial\boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] = o(D^{-1}). \qquad (60)$$

For this, observe that

$$(\partial g/\partial\boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0) = g(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\,\boldsymbol{\lambda}'(\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0), \qquad (61)$$

where $\hat{\boldsymbol{\beta}}^E = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$. Let us denote $f(\boldsymbol{\theta}) = \boldsymbol{\lambda}'(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{\beta}_0) = \boldsymbol{\lambda}'\mathbf{Q}(\boldsymbol{\theta})\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{v}$, and perform a second order Taylor expansion of $f(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$. At the point $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, the expansion is

$$f(\hat{\boldsymbol{\theta}}) = f(\boldsymbol{\theta}_0) + (\partial f/\partial\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + r_f, \qquad (62)$$

where $r_f = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \left(\partial^2 f(\boldsymbol{\theta}^*)/\partial \boldsymbol{\theta}^2\right) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. By the Hölder inequality,

$$E(r_f) \leq E^{1/2} \left(\sup_{N(\boldsymbol{\theta}_0)} \left\| \partial^2 f(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \right\| \right)^2 E^{1/2} |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|^4.$$

It can be seen that the first expectation on the right of this inequality is bounded. Furthermore, from the Minkowski inequality and Lemma 1 with $\eta > 1/2$, it follows that $E^{1/2}|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|^4 = o(D^{-1})$, so that $E(r_f) = o(D^{-1})$. Then if we replace $f(\hat{\boldsymbol{\theta}}) = \boldsymbol{\lambda}'(\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0)$ by (62) in (61) and take expectation, we obtain

$$E\left[(\partial g/\partial \boldsymbol{\beta})' (\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0) \right] = g(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)\, E\left[(\partial f/\partial \boldsymbol{\theta})' (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right] + o(D^{-1}),$$

since $E[f(\boldsymbol{\theta}_0)] = 0$. But inserting again $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathcal{I}^{-1}\mathbf{s} + \mathbf{r}_*$, we obtain

$$E\left[(\partial f/\partial \boldsymbol{\theta})' (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right] = E\left[(\partial f/\partial \boldsymbol{\theta})' \mathcal{I}^{-1}\mathbf{s} \right] + E\left[(\partial f/\partial \boldsymbol{\theta})' \mathbf{r}_* \right],$$

where, by the Hölder inequality and Lemma 1, since $E^{1/2} \left|\partial f/\partial \boldsymbol{\theta}\right|^2 = O(D^{-1})$,

$$E\left[(\partial f/\partial \boldsymbol{\theta})' \mathbf{r}_* \right] \leq E^{1/2} \left|\partial f/\partial \boldsymbol{\theta}\right|^2 E^{1/2}(r_*^2) = O(D^{-1/2-\eta}).$$

However,

$$E\left[(\partial f/\partial \boldsymbol{\theta})' \mathcal{I}^{-1}\mathbf{s} \right] = \sum_{i=1}^{3} \sum_{j=1}^{3} (\mathcal{I}^{-1})_{ij}\, \boldsymbol{\lambda}'\mathbf{Q}\, \mathbf{X}'\mathbf{V}^{-1} \boldsymbol{\Delta}_i \mathbf{P}\, E(\mathbf{v}s_j) = 0.$$

Therefore, (60) holds for $\eta > 1/2$. Following similar arguments, it is easy to see that

$$E\left[(\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0)' \frac{\partial^2 g}{\partial \boldsymbol{\beta}^2} (\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0) \right] = g(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \boldsymbol{\lambda}'\mathbf{Q}\boldsymbol{\lambda} + o(D^{-1}). \quad (63)$$

$$E\left[(\hat{\boldsymbol{\beta}}^E - \boldsymbol{\beta}_0)' \frac{\partial^2 g}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}'} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right] = o(D^{-1}). \quad (64)$$

Thus, (58)–(60) and (63)–(64) lead to the desired result. □

References

Baillo, A. and Molina, I. (2005). Mean squared errors of small area estimators under a unit-level multivariate model. *Working Paper, Statistics and Econometrics Series. Universidad Carlos III de Madrid*, **05-40 (07)**.

Bhandary, M. and Alam, M. K. (2000). Test for the equality of intraclass correlation coefficients under unequal family sizes for several populations. *Communications in Statistics–Theory and Methods*, **29**, 755–768.

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., 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*, 63–79, U. S. Bureau of the Census, Washington, DC.

Diggle, P. J., Heagerty, P., Liang, K. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford University Press.

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

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

Hobza, T., Molina, I. and Morales, D. (2002). Likelihood divergence statistics for testing hypotheses in familial data. *Communications in Statistics–Theory and Methods*, **312**, 415–434.

Jiang, J. and Lahiri, P. (2006). Mixed Model Prediction and Small Area Estimation. *Test*, **15**, 1–84.

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.

Slud, E. and Maiti, T. (2006). Mean-squared error estimation in transformed Fay-Herriot models. *Journal of the Royal Statistical Society B*. **68**, 239–257.

Srivastava, M. S. (1984). Estimations of interclass correlations in familial data. *Biometrika*, **71**, 177–85.

Srivastava, M. S. and Katapa, R. S. (1986). Comparison of estimators of interclass and intraclass correlations from familial data. *The Canadian Journal of Statistics*, **14**, 29–42.

Srivastava, V. K. and Tiwari, R. (1976). Evaluation of Expectations of Products of Stochastic Matrices. *Scandinavian Journal of Statistics*, **3**, 135–138.