The Annals of Statistics
2018, Vol. 46, No. 5, 1961–1993
https://doi.org/10.1214/17-AOS1608

EMPIRICAL BEST PREDICTION UNDER A NESTED ERROR MODEL WITH LOG TRANSFORMATION

BY ISABEL MOLINA[1] AND NIRIAN MARTÍN[2]

Universidad Carlos III de Madrid and Universidad Complutense de Madrid

In regression models involving economic variables such as income, log transformation is typically taken to achieve approximate normality and stabilize the variance. However, often the interest is predicting individual values or means of the variable in the original scale. Under a nested error model for the log transformation of the target variable, we show that the usual approach of back transforming the predicted values may introduce a substantial bias. We obtain the optimal (or "best") predictors of individual values of the original variable and of small area means under that model. Empirical best predictors are defined by estimating the unknown model parameters in the best predictors. When estimation is desired for subpopulations with small sample sizes (small areas), nested error models are widely used to "borrow strength" from the other areas and obtain estimators with greater efficiency than direct estimators based on the scarce area-specific data. We show that naive predictors of small area means obtained by back-transformation under the mentioned model may even underperform direct estimators. Moreover, assessing the uncertainty of the considered predictor is not straightforward. Exact mean squared errors of the best predictors and second-order approximations to the mean squared errors of the empirical best predictors are derived. Estimators of the mean squared errors that are second-order correct are also obtained. Simulation studies and an example with Mexican data on living conditions illustrate the procedures.

1. Introduction. In econometric regression models, variables such as income or expenditure are often transformed with a logarithm to achieve homoscedastic errors with approximately normal distribution. However, the variable of interest remains to be the untransformed one. Target characteristics of the study variable such as the values for out-of-sample individuals or the means for specific subpopulations become then functions of the exponentials of the dependent variable in the model. We show that, under a nested error linear regression model [1], the predictors obtained by transforming back the individual predicted values may be severely biased and derive optimal predictors under that model. The nested-error model is widely used in small area estimation, because it solves the problem of lack of data

Received October 2016; revised March 2017.

[1]Supported by the Spanish Grants SEJ-2007-64500 and MTM2012-37077-C02-01.

[2]Supported by the Spanish Grants MTM-2012-33740 and ECO-2011-25706.

MSC2010 subject classifications. Primary 62D05; secondary 62G09.

Key words and phrases. Empirical best estimator, mean squared error, parametric bootstrap.

in some of the areas of interest by linking all areas through the common regression parameters. At the same time, these models include random area effects that represent the unexplained between-area variation. The common parameters are estimated using the sample observations from all the areas together, and this often leads to great efficiency gains with respect to estimators that use only the area-specific sample data (direct estimators). In econometric applications, this model has been used to estimate poverty indicators in small areas; see, for example, [4] or [16]. For further details on small area estimation methods, see the monograph by [19] and the recent review by [17].

In many applications, predicting the individual values of the original target variable in the model rather than (or beyond) the area means may be of interest. For example, the approach of Elbers, Lanjouw and Lanjouw [4] widely extended in World Bank applications, is based on obtaining predicted censuses of the variable of interest at the individual level. Important advantages of prediction at the individual level are that, once a predicted census is obtained, it may be used to estimate whatever desired population characteristic and at whatever level of disaggregation; see, for example, [10] for prediction at different levels (including the individual level) in the context of forest inventories.

Assessing the reliability, or uncertainty, of the obtained predictors is crucial in practical applications. A popular uncertainty measure is the mean squared error (MSE), also called mean squared prediction error. Second-order correct approximations to the MSEs of the optimal predictors of small area parameters have been obtained under certain models but only for simple parameters; see, for example, [3]. The MSE of an individual prediction under a nested-error model with log-transformation that is second-order correct has not been obtained yet. Moreover, when predicting the mean of the original variable in a given area, the optimal predictor depends on the predicted values for the out-of-sample individuals from that area. Since the individuals belong to the same area, due to the presence of the area effects, individual predictors are not independent. Then mean crossed product errors (MCPEs) between pairs of individual predictions are needed to derive the MSE of the predictor of the mean in that area.

Here, we obtain optimal predictors for individual values of the target variable in out-of-sample units and also for small area means. Additionally, second-order asymptotic approximations for the MCPEs of pairs of individual predictions are derived, which lead to good approximations for the MSEs of predicted area means. In the small area estimation literature, this was done previously only under area-level models by [21]. Under a unit-level model, [12] dealt with estimation of exponentials of mixed effects, that is, exponentials of linear functions of the fixed and the random effects in the model; the individual values of the original variable cannot be expressed as special cases of these parameters. Thus, the target parameters are not the same, and consequently results are also different. In particular, certain crossed-product terms appearing in the MCPE that are of lower order in [12], are not negligible when predicting individual observations. In fact, those crossed-product terms are typically neglected in small area estimation applications. Here, we show that these terms cannot be neglected and give their analytical expression up to $o(D^{-1})$ terms, where D is the number of areas.

Analytical approximations for the uncertainty measures have a complex shape and users might prefer to use resampling procedures such as bootstrap methods. González-Manteiga et al. [6] proposed a parametric bootstrap method designed for finite populations under a nested error model that is suitable in this paper. Here, we use the technique of imitation used in that paper to obtain the consistency of the bootstrap MSE estimates in our setup.

The paper is organized as follows. The considered model and the target quantities are introduced in Section 2. That section also gives the best predictor and first- and second-stage empirical best predictors of the target quantities. Section 3 describes usual likelihood-based fitting methods. MCPEs and MSEs of first-stage empirical best predictors are obtained in Section 4, and for second-stage empirical best predictors, second-order approximations to the analogous uncertainty measures are given in Section 5. Second-order unbiased estimators of these uncertainty measures are provided in Section 6. Section 7 describes a parametric bootstrap procedure for estimation of the uncertainty. Section 8 describes the result of a simulation experiment comparing the proposed predictor with existing ones. Section 9 illustrates the procedures through the estimation of mean income in municipalities from Mexico. The proofs of theorems are included in the Appendix, and finally, Supplement A [14] studies the bias of the proposed predictors compared to existing ones and includes additional simulation and application results.

2. Model, target quantities and predictors. When estimating characteristics of subpopulations that have varying sizes, it seems convenient to work under a finite population setup. Here, we consider that the population U is finite and contains N units. This population is partitioned into D subpopulations $U_1, \dots, U_D$, called areas or domains, of sizes $N_1, \dots, N_D$. The data is obtained from a sample s of size n drawn from the population U. We denote by s_d the subsample from domain d, of (fixed) size n_d and by $\bar{s}_d = U_d - s_d$ the sample complement from area d, of size $N_d - n_d$, $d = 1, \dots, D$, with $\sum_{d=1}^{D} n_d = n$.

The goal is to predict the value w_{di} of the variable of interest for an out-of-sample individual i within area d, or the area mean $N_d^{-1} \sum_{i=1}^{N_d} w_{di}$, based on a regression model for w_{di}. If w_{di} represents a measurement of an economic variable such as income or expenditure, it is customary to consider $\log(w_{di})$ as dependent variable in the model. Moreover, in many applications such as in small area estimation, the available auxiliary variables do not explain sufficiently well all the between-area heterogeneity that data exhibit. Then random area effects representing this unexplained heterogeneity are included in the model. Thus, here we assume the following superpopulation model, known as the nested-error model, for the log-transformed variables $y_{di} = \log(w_{di})$,

$$
(2.1)\qquad \begin{aligned} &y_{di} = \mathbf{x}_{di}'\boldsymbol{\beta} + u_d + e_{di},\\ &\quad u_d \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2),\, e_{di} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_e^2),\, i = 1, \dots, N_d,\, d = 1, \dots, D. \end{aligned}
$$

Here, $\mathbf{x}_{di}$ is a vector containing the values of p explanatory variables for the ith individual in the dth area, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of unknown regression coefficients, e_{di} is the individual error, u_d is the random effect of area d, with random effects $\{u_d\}$ and errors $\{e_{di}\}$ assumed to be mutually independent, and finally σ_u^2 and σ_e^2 are respectively the unknown random effects and individual error variances, called variance components.

In model (2.1), the area effects u_d induce a constant correlation among the units belonging to the same area, whereas observations from different areas are kept uncorrelated. Moreover, the log-transformation implies a log-normal distribution for the original variables w_{di}. If residuals from the fitted model to the log-transformed responses y_{di} indicate severe departure from the normal distribution, other transformations of the original variables w_{di} such as those from the Box–Cox or power families can be considered. Corresponding EB predictors can be obtained and the Taylor-linearization approach followed in this paper to obtain uncertainty measures may also be applied. However, log is the most commonly used transformation especially for economic variables such as income, probably due to the nice interpretation of the resulting estimated model coefficients.

Let $\boldsymbol{\theta} = (\sigma_u^2, \sigma_e^2)'$ be the vector of variance components and $\Theta = \{(\sigma_u^2, \sigma_e^2)';\ \sigma_u^2 \geq 0, \sigma_e^2 > 0\}$ the space where these parameters lie. We will denote by $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ to generic elements from $\mathbb{R}^p$ and Θ, whereas $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ will be their respective true values, where $\boldsymbol{\theta}_0$ is supposed to be in the interior of Θ. A quantity $A(\boldsymbol{\beta}, \boldsymbol{\theta})$ depending on $\boldsymbol{\beta}$ and/or $\boldsymbol{\theta}$ will be sometimes denoted simply by A, omitting the explicit dependence on $\boldsymbol{\beta}$ and/or $\boldsymbol{\theta}$.

If we intend to estimate the mean of an area with a poor sample size n_d, the estimators that use only the n_d area-specific observations, called direct estimators, are highly inefficient. Model (2.1) links all the areas through the common parameters $\boldsymbol{\beta}$, σ_u^2 and σ_e^2, which allows us to "borrow strength" from all the areas when estimating a particular area mean. However, even though the model is assumed for $y_{di} = \log(w_{di})$, the target parameter remains to be the area mean of the untransformed variables, which can be expressed in terms of the dependent variables in the model as

$$
\tau_d = \frac{1}{N_d}\sum_{i=1}^{N_d} w_{di} = \frac{1}{N_d}\sum_{i=1}^{N_d} \exp(y_{di}), \qquad d = 1, \dots, D.
$$

Here, we intend to estimate single values $w_{di} = \exp(y_{di})$ of the target variable in out-of-sample units $i \in \bar{s}_d$ and area means $\tau_d = N_d^{-1}\sum_{i=1}^{N_d} \exp(y_{di})$, when the variables y_{di} in the population units follow model (2.1). These target quantities are special cases of a general parameter of the form $h(\mathbf{y}_d)$, where $h(\cdot)$ is a measurable function and $\mathbf{y}_d = (y_{d1}, \dots, y_{dN_d})'$ is the vector of outcomes for domain d. Defining also $\mathbf{X}_d = (\mathbf{x}_{d1}, \dots, \mathbf{x}_{dN_d})'$ and $\mathbf{e}_d = (e_{d1}, \dots, e_{dN_d})'$, the model reads

$$\begin{aligned} &\mathbf{y}_d = \mathbf{X}_d\boldsymbol{\beta} + u_d\mathbf{1}_{N_d} + \mathbf{e}_d, \\ &u_d \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2), \mathbf{e}_d \overset{\text{ind}}{\sim} \mathcal{N}_{N_d}(\mathbf{0}_{N_d}, \sigma_e^2\mathbf{I}_{N_d}), d = 1, \dots, D, \end{aligned} \tag{2.2}$$

where $\mathbf{0}_k$ is a k-vector of zeros, $\mathbf{1}_k$ is a k-vector of ones and $\mathbf{I}_k$ is the $k \times k$ identity matrix. The covariance matrix of $\mathbf{y}_d$ is equal to $\mathbf{V}_d = \sigma_u^2\mathbf{1}_{N_d}\mathbf{1}'_{N_d} + \sigma_e^2\mathbf{I}_{N_d} = \mathbf{V}_d(\boldsymbol{\theta})$. Let us arrange the elements from domain d into sample and out-of-sample elements, as

$$\mathbf{y}_d = \begin{pmatrix} \mathbf{y}_{ds} \\ \mathbf{y}_{dr} \end{pmatrix}, \qquad \mathbf{X}_d = \begin{pmatrix} \mathbf{X}_{ds} \\ \mathbf{X}_{dr} \end{pmatrix}, \qquad \mathbf{V}_d = \begin{pmatrix} \mathbf{V}_{ds} & \mathbf{V}_{dsr} \\ \mathbf{V}_{drs} & \mathbf{V}_{dr} \end{pmatrix}.$$

The "best predictor" $\tilde{\delta}_d$ of a general parameter $\delta_d = h(\mathbf{y}_d)$ is the function of the sample data $\mathbf{y}_{ds}$ with minimum mean squared error $\mathrm{MSE}(\tilde{\delta}_d) = E(\tilde{\delta}_d - \delta_d)^2$ and is given by $\tilde{\delta}_d = E_{\mathbf{y}_{dr}}\{h(\mathbf{y}_d)|\mathbf{y}_{ds}\}$, where the expectation is taken with respect to the distribution of $\mathbf{y}_{dr}|\mathbf{y}_{ds}$. The best predictor is exactly unbiased in the sense $E_{\mathbf{y}_{ds}}(\tilde{\delta}_d) = E_{\mathbf{y}_d}(\delta_d)$. Since by (2.2) we have $\mathbf{y}_d \sim \mathcal{N}(\mathbf{X}_d\boldsymbol{\beta}, \mathbf{V}_d)$, the desired conditional distribution is

$$\mathbf{y}_{dr}|\mathbf{y}_{ds} \overset{\text{ind}}{\sim} \mathcal{N}_{N_d-n_d}(\boldsymbol{\mu}_{dr|s}, \mathbf{V}_{dr|s}), \qquad d = 1, \dots, D, \tag{2.3}$$

with mean vector and covariance matrix given by

$$\boldsymbol{\mu}_{dr|s} = \mathbf{X}_{dr}\boldsymbol{\beta} + \mathbf{V}_{drs}\mathbf{V}_{ds}^{-1}(\mathbf{y}_{ds} - \mathbf{X}_{ds}\boldsymbol{\beta}), \qquad \mathbf{V}_{dr|s} = \mathbf{V}_{dr} - \mathbf{V}_{drs}\mathbf{V}_{ds}^{-1}\mathbf{V}_{dsr}.$$

Under the nested-error model (2.1), they reduce to

$$\boldsymbol{\mu}_{dr|s} = \mathbf{X}_{dr}\boldsymbol{\beta} + \mathbf{1}_{N_d-n_d}\gamma_d(\bar{y}_{ds} - \bar{\mathbf{x}}'_{ds}\boldsymbol{\beta}), \tag{2.4}$$

$$\mathbf{V}_{dr|s} = \sigma_u^2(1 - \gamma_d)\mathbf{1}_{N_d-n_d}\mathbf{1}'_{N_d-n_d} + \sigma_e^2\mathbf{I}_{N_d-n_d}, \tag{2.5}$$

where $\gamma_d = \sigma_u^2/(\sigma_u^2 + \sigma_e^2/n_d)$, $\bar{y}_{ds} = n_d^{-1}\sum_{i\in s_d} y_{di}$ and $\bar{\mathbf{x}}_{ds} = n_d^{-1}\sum_{i\in s_d}\mathbf{x}_{di}$.

Based on the conditional distribution (2.3) with mean vector given in (2.4) and covariance matrix (2.5), the next theorem gives closed-form expressions for the best predictors of $w_{di} = \exp(y_{di})$ and $\tau_d = N_d^{-1}\sum_{i=1}^{N_d}\exp(y_{di})$.

THEOREM 2.1. *Under the nested-error model with log-transformation* (2.1), *it holds*:

(i) *The best predictor of* $w_{di} = \exp(y_{di})$, *for* $i \in \bar{s}_d$, *is given by*

$$\tilde{w}_{di} = \tilde{w}_{di}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \exp(\tilde{y}_{di} + \alpha_d), \tag{2.6}$$

where $\tilde{y}_{di} = \mathbf{x}'_{di}\boldsymbol{\beta} + \gamma_d(\bar{y}_{ds} - \bar{\mathbf{x}}'_{ds}\boldsymbol{\beta})$ *and* $\alpha_d = \{\sigma_u^2(1 - \gamma_d) + \sigma_e^2\}/2$.

(ii) *The best predictor of* $\tau_d = N_d^{-1}\sum_{i=1}^{N_d}\exp(y_{di})$ *is given by*

$$\tilde{\tau}_d = \tilde{\tau}_d(\boldsymbol{\beta},\boldsymbol{\theta}) = \frac{1}{N_d}\biggl(\sum_{i\in s_d} w_{di} + \sum_{i\in\bar{s}_d}\tilde{w}_{di}\biggr). \tag{2.7}$$

REMARK 2.1. In contrast with the case of estimation of a small area mean under a nested error model without log-transformation, the best predictor of the small area mean τ_d given in (2.7) requires to predict the values of the target variables w_{di} for each out-of-sample unit $i \in \bar{s}_d$ from area d. In applications to official statistics, the unit-level values of the auxiliary variables $\mathbf{x}_{di}$ required for prediction of w_{di} might be obtained from a census or an administrative register with microdata for all of the population. These kinds of microdata are available in practically all European countries and many other, although their use is typically subject to the signature of proper confidentiality contracts. In other areas such as agriculture or forestry (see, e.g., [1] or [10]), satellite or laser scanner images play the role of a census with pixels acting as population units, which may be more easily available. Other useful data sets for small area estimation are included in the R package *sae* [13].

In fact, writing the best predictor as $\tilde{\tau}_d = N_d^{-1}\{E_{ds} + \exp(\tilde{u}_d + \alpha_d)E_d\}$, where $\tilde{u}_d = \gamma_d(\bar{y}_{ds} - \bar{\mathbf{x}}'_{ds}\boldsymbol{\beta})$, $E_{ds} = \sum_{i\in s_d}\exp(\mathbf{x}'_{di}\boldsymbol{\beta})$ and $E_d = \sum_{i=1}^{N_d}\exp(\mathbf{x}'_{di}\boldsymbol{\beta})$, we only need the area totals E_d rather than all individual values $\mathbf{x}_{di}$. These area totals E_d might be estimated from larger surveys that contain the same auxiliary variables and adequately cover the areas. In the case of area-level covariates $\mathbf{x}_{di} = \mathbf{x}_d$ for all i, we get $E_d = N_d\exp(\mathbf{x}'_d\boldsymbol{\beta})$ and only the aggregates $\mathbf{x}_d$ are needed. If the covariates $\mathbf{x}_{di}$ take only a finite number of values $\mathbf{x}_{di} \in \{\mathbf{z}_1,\dots,\mathbf{z}_K\}$, as it occurs when using only categorical variables (the most common in household surveys), then for $\mathbf{x}_{di} = \mathbf{z}_k$ we get $E_d = N_{dk}\exp(\mathbf{z}'_k\boldsymbol{\beta})$, and only the counts N_{dk} of individuals in the area with $\mathbf{x}_{di} = \mathbf{z}_k$ are needed, $k = 1,\dots,K$. These counts can be obtained from aggregated data sources or may be estimated from a larger survey; see, for example, [16] or [15].

The bias-corrected predictor $\tilde{w}_{di}^M = \exp(\tilde{y}_{di} + \alpha_d^M)$, where $\alpha_d^M = \sigma_u^2(1-\gamma_d)/2$, which is similar to the best predictor $\tilde{w}_{di}$ given in (2.6), was proposed in [12]. However, these predictors are not exactly the same because the target parameters in [12] are of the type $\exp(\mathbf{x}'_{di}\boldsymbol{\beta} + u_d)$, which differ from our target parameters here given by the individual observations $w_{di} = \exp(y_{di}) = \exp(\mathbf{x}'_{di}\boldsymbol{\beta} + u_d + e_{di})$. Nevertheless, it is interesting to study how Molina's predictor $\tilde{w}_{di}^M$ performs for $w_{di} = \exp(y_{di})$. In Supplement A, we study the bias of $\tilde{w}_{di}^M$ and of the naive predictor obtained by back-transforming the predicted model responses, $\tilde{w}_{di}^N = \exp(\tilde{y}_{di})$.

The best predictors $\tilde{w}_{di}(\boldsymbol{\beta},\boldsymbol{\theta})$ and $\tilde{\tau}_d(\boldsymbol{\beta},\boldsymbol{\theta})$ depend on the true values of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, which are unknown in practice. Next, we define first- and second-stage empirical best (EB) predictors obtained by estimating these unknown parameters in two stages. First, define the following vectors and matrices containing the sample elements from all the areas:

$$\mathbf{y}_s = (\mathbf{y}'_{1s}, \dots, \mathbf{y}'_{Ds})', \qquad \mathbf{X}_s = (\mathbf{X}'_{1s}, \dots, \mathbf{X}'_{Ds})', \qquad \mathbf{e}_s = (\mathbf{e}'_{1s}, \dots, \mathbf{e}'_{Ds})',$$
$$\mathbf{Z}_s = \operatorname{diag}_{1\le d\le D}(\mathbf{1}_{n_d}), \qquad \mathbf{u} = (u_1, \dots, u_D)'.$$

Then the model for the sample units can be written as

$$\mathbf{y}_s = \mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{u} + \mathbf{e}_s, \qquad \mathbf{u} \sim \mathcal{N}_D(\mathbf{0}_D, \sigma_u^2\mathbf{I}_D), \qquad \mathbf{e}_s \sim \mathcal{N}_n(\mathbf{0}_n, \sigma_e^2\mathbf{I}_n),$$

and the covariance matrix of $\mathbf{y}_s$ is $\mathbf{V}_s = \operatorname{diag}_{1\le d\le D}(\mathbf{V}_{ds})$, where $\mathbf{V}_{ds} = \sigma_u^2\mathbf{1}_{n_d}\mathbf{1}'_{n_d} + \sigma_e^2\mathbf{I}_{n_d}$.

The first-stage EB predictor is obtained under the assumption that $\boldsymbol{\theta}$ is known but $\boldsymbol{\beta}$ is unknown. The maximum likelihood (ML) estimator of $\boldsymbol{\beta}$ under normality, which is also the weighted least squares (WLS) estimator of $\boldsymbol{\beta}$ without normality reads

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left(\mathbf{X}'_s\mathbf{V}_s^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}'_s\mathbf{V}_s^{-1}\mathbf{y}_s. \tag{2.8}$$

The first-stage EB predictors of w_{di} and τ_d are then

$$\hat{w}_{di} = \hat{w}_{di}(\boldsymbol{\theta}) = \tilde{w}_{di}\big(\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}\big), \qquad \hat{\tau}_d = \hat{\tau}_d(\boldsymbol{\theta}) = \tilde{\tau}_d\big(\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}\big). \tag{2.9}$$

Finally, the second-stage EB predictors of w_{di} and τ_d are obtained by replacing the unknown $\boldsymbol{\theta}$ in (2.9) by a consistent estimator $\hat{\boldsymbol{\theta}}$, that is,

$$\hat{w}_{di}^E = \hat{w}_{di}(\hat{\boldsymbol{\theta}}) = \tilde{w}_{di}\big(\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\big), \qquad \hat{\tau}_d^E = \hat{\tau}_d(\hat{\boldsymbol{\theta}}) = \tilde{\tau}_d\big(\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\big). \tag{2.10}$$

Section 3 describes typical methods for consistent estimation of $\boldsymbol{\theta}$ under model (2.1).

3. Fitting methods. A typical estimation method is maximum likelihood (ML), which delivers consistent and asymptotically efficient estimators of the variance components, provided that the following conditions hold [11]:

$$n \ge p + 2, \qquad \operatorname{range}(\mathbf{X}_s) = p, \qquad \operatorname{range}(\mathbf{X}_s|\mathbf{Z}_s) > p. \tag{3.1}$$

The ML estimator $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2)'$ of $\boldsymbol{\theta} = (\sigma_u^2, \sigma_e^2)'$ maximizes the penalized log-likelihood,

$$l_P(\boldsymbol{\theta}) = c - \frac{1}{2}\big(\log|\mathbf{V}_s| + \mathbf{y}'_s\mathbf{P}_s\mathbf{y}_s\big), \qquad \mathbf{P}_s = \mathbf{V}_s^{-1} - \mathbf{V}_s^{-1}\mathbf{X}_s\mathbf{Q}_s\mathbf{X}'_s\mathbf{V}_s^{-1}, \tag{3.2}$$

where c denotes a generic constant. The score vector is defined as $\mathbf{s}(\boldsymbol{\theta}) = \partial l_P(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = (s_1(\boldsymbol{\theta}), s_2(\boldsymbol{\theta}))'$. In terms of $\mathbf{v}_s = \mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta} = \mathbf{Z}_s\mathbf{u} + \mathbf{e}_s$, the elements of the score vector are

$$s_h(\boldsymbol{\theta}) = -\frac{1}{2}\operatorname{tr}\big(\mathbf{V}_s^{-1}\boldsymbol{\Delta}_h\big) + \frac{1}{2}\mathbf{v}'_s\mathbf{P}_s\boldsymbol{\Delta}_h\mathbf{P}_s\mathbf{v}_s, \qquad h = 1, 2, \tag{3.3}$$

where $\boldsymbol{\Delta}_h = \partial \mathbf{V}_s/\partial\theta_h$, that is, $\boldsymbol{\Delta}_1 = \mathbf{Z}_s\mathbf{Z}_s'$ and $\boldsymbol{\Delta}_2 = \mathbf{I}_n$. The ML estimator of $\boldsymbol{\theta}$ is then obtained solving the equation system $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{0}_2$ together with equation (2.8) for $\boldsymbol{\beta}$. Since equations are nonlinear, numerical algorithms such as Newton–Raphson or Fisher scoring are typically applied. These algorithms require respectively the elements of Hessian matrix or the Fisher information matrix. The Hessian matrix is defined as $H(\boldsymbol{\theta}) = \partial^2 l_P(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^2 = (H_{h\ell}(\boldsymbol{\theta}))$, where

$$H_{h\ell}(\boldsymbol{\theta}) = \frac{1}{2}\operatorname{tr}\bigl(\mathbf{V}_s^{-1}\boldsymbol{\Delta}_h\mathbf{V}_s^{-1}\boldsymbol{\Delta}_\ell\bigr) - \mathbf{v}_s'\mathbf{P}_s\boldsymbol{\Delta}_h\mathbf{P}_s\boldsymbol{\Delta}_\ell\mathbf{P}_s\mathbf{v}_s, \qquad h,\ell = 1,2.$$

Finally, the Fisher information matrix is $\mathcal{F}(\boldsymbol{\theta}) = E\{-H(\boldsymbol{\theta})\} = (\mathcal{F}_{h\ell}(\boldsymbol{\theta}))$, where

$$\mathcal{F}_{h\ell}(\boldsymbol{\theta}) = -\frac{1}{2}\operatorname{tr}\bigl(\mathbf{V}_s^{-1}\boldsymbol{\Delta}_h\mathbf{V}_s^{-1}\boldsymbol{\Delta}_\ell\bigr) + \operatorname{tr}(\mathbf{P}_s\boldsymbol{\Delta}_h\mathbf{P}_s\boldsymbol{\Delta}_\ell), \qquad h,\ell = 1,2.$$

A drawback of the ML estimator of $\boldsymbol{\theta}$ is that it does not account for the degrees of freedom due to estimation of $\boldsymbol{\beta}$. Restricted ML (REML) corrects for this problem, providing estimators with bias of lower order. This is achieved by transforming the data $\mathbf{y}$ as $\mathbf{F}'\mathbf{y}$, where $\mathbf{F}$ is any $n \times (n-p)$ matrix with rank $n-p$ and satisfying $\mathbf{F}'\mathbf{X} = \mathbf{0}_{n-p}$. The REML estimator is the value of $\boldsymbol{\theta}$ maximizing the so-called restricted log-likelihood l_R, which is the logarithm of the joint density function of the transformed data $\mathbf{F}'\mathbf{y}$. Noting that $\mathbf{F}(\mathbf{F}'\mathbf{V}_s\mathbf{F})^{-1}\mathbf{F}' = \mathbf{P}_s$ (see [20], p. 451), this function can be written as

$$l_R(\boldsymbol{\theta}) = c - \frac{1}{2}\bigl(\log\bigl|\mathbf{F}'\mathbf{V}_s\mathbf{F}\bigr| + \mathbf{y}_s'\mathbf{P}_s\mathbf{y}_s\bigr). \tag{3.4}$$

The score vector obtained from l_R is $\mathbf{s}_R(\boldsymbol{\theta}) = \partial l_R(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = (s_{R,1}(\boldsymbol{\theta}), s_{R,2}(\boldsymbol{\theta}))'$. Using again the relation $\mathbf{F}(\mathbf{F}'\mathbf{V}_s\mathbf{F})^{-1}\mathbf{F}' = \mathbf{P}_s$, the elements of $\mathbf{s}_R$ can be expressed as

$$s_{R,h}(\boldsymbol{\theta}) = -\frac{1}{2}\operatorname{tr}\bigl(\mathbf{P}_s^{-1}\boldsymbol{\Delta}_h\bigr) + \frac{1}{2}\mathbf{v}_s'\mathbf{P}_s\boldsymbol{\Delta}_h\mathbf{P}_s\mathbf{v}_s, \qquad h = 1,2. \tag{3.5}$$

The Hessian matrix obtained from l_R is $H_R(\boldsymbol{\theta}) = \partial^2 l_R(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^2 = (H_{R,h\ell}(\boldsymbol{\theta}))$, where

$$H_{R,h\ell}(\boldsymbol{\theta}) = \frac{1}{2}\operatorname{tr}\bigl(\mathbf{P}_s^{-1}\boldsymbol{\Delta}_h\mathbf{P}_s^{-1}\boldsymbol{\Delta}_\ell\bigr) - \mathbf{v}_s'\mathbf{P}_s\boldsymbol{\Delta}_h\mathbf{P}_s\boldsymbol{\Delta}_\ell\mathbf{P}_s\mathbf{v}_s, \qquad h,\ell = 1,2.$$

Finally, the corresponding Fisher information matrix is in this case given by $\mathcal{F}_R(\boldsymbol{\theta}) = E\{-H_R(\boldsymbol{\theta})\} = (\mathcal{F}_{R,h\ell}(\boldsymbol{\theta}))$, with elements

$$\mathcal{F}_{R,h\ell}(\boldsymbol{\theta}) = \frac{1}{2}\operatorname{tr}(\mathbf{P}_s\boldsymbol{\Delta}_h\mathbf{P}_s\boldsymbol{\Delta}_\ell), \qquad h,\ell = 1,2.$$

4. Uncertainty of first-stage EB predictors. The reliability of a point predictor is typically assessed by its MSE. When estimating a small area mean τ_d, in virtue of (2.7), the MSE of a predictor $\tilde{\tau}_d$ can be directly obtained as a function of the MCPEs of pairs of predictors $\hat{w}_{di}$ and $\hat{w}_{dj}$ for out-of-sample units $i, j \in \bar{s}_d$. For this reason, in the following we focus on giving the expressions for the MCPEs of pairs of individual predictors.

Theorem 4.1 spells out the MCPE of the best predictors $\tilde{w}_{di}$ and $\tilde{w}_{dj}$ for out-of-sample units $i, j \in \bar{s}_d$, defined by $\mathrm{MCPE}(\tilde{w}_{di}, \tilde{w}_{dj}) = E\{(\tilde{w}_{di} - w_{di})(\tilde{w}_{dj} - w_{dj})\}$. The mean squared error (MSE) of the best predictor of a single out-of-sample observation $\mathrm{MSE}(\tilde{w}_{di}) = E(\tilde{w}_{di} - w_{di})^2$, $i \in \bar{s}_d$ is then obtained taking $i = j$. For the area mean τ_d, the MSE of the best predictor $\mathrm{MSE}(\tilde{\tau}_d) = E(\tilde{\tau}_d - \tau_d)^2$ is given in Corollary 4.1. In these results, $1_{\{i=j\}}$ is equal to 1 if $i = j$ and 0 otherwise.

THEOREM 4.1. *Under the nested-error model with log-transformation* (2.1), *the mean crossed product error of the best predictors $\tilde{w}_{di}$ and $\tilde{w}_{dj}$ of w_{di} and w_{dj}, for $i, j \in \bar{s}_d$, is given by*

$$\begin{aligned}
&\mathrm{MCPE}(\tilde{w}_{di}, \tilde{w}_{dj}) \\
&\quad = \exp\{2\sigma_u^2 + \sigma_e^2 + (\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta}\} \\
&\qquad \times \left[1 + \{\exp(\sigma_e^2) - 1\}1_{\{i=j\}} - \exp\{-\sigma_u^2(1-\gamma_d)\}\right].
\end{aligned}$$

COROLLARY 4.1. *The mean squared error of the best predictor $\tilde{\tau}_d$ of τ_d is given by*

$$\begin{aligned}
\mathrm{MSE}(\tilde{\tau}_d) = N_d^{-2}\exp(2\sigma_u^2 + \sigma_e^2)\Bigg(&2\left[1 - \exp\{-\sigma_u^2(1-\gamma_d)\}\right] \\
&\times \sum_{i\in\bar{s}_d}\sum_{j\in\bar{s}_d, j>i} \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta}\} \\
&+ \left[\exp(\sigma_e^2) - \exp\{-\sigma_u^2(1-\gamma_d)\}\right]\sum_{i\in\bar{s}_d}\exp\{2\mathbf{x}_{di}'\boldsymbol{\beta}\}\Bigg).
\end{aligned}$$

For a pair of first-stage EB predictors obtained by estimating $\boldsymbol{\beta}$ using the WLS estimator given in (2.8) but assuming that $\boldsymbol{\theta}$ is known, Theorem 4.2 gives the MCPE. The MSE of a single first-stage EB predictor is obtained setting $j = i$. The following notation is required:

$$\begin{aligned}
\mathbf{Q}_s &= (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}, \qquad h_{d,ij} = \mathbf{x}_{di}'\mathbf{Q}_s\mathbf{x}_{dj}, \\
h_{d,i} &= \mathbf{x}_{di}'\mathbf{Q}_s\bar{\mathbf{x}}_{ds}, \qquad h_d = \bar{\mathbf{x}}_{ds}'\mathbf{Q}_s\bar{\mathbf{x}}_{ds}.
\end{aligned}$$

THEOREM 4.2. *Under the nested-error model with log-transformation* (2.1), *the mean crossed product error of the first-stage EB predictors* $\hat{w}_{di}$ *and* $\hat{w}_{dj}$, *for* $i, j \in \bar{s}_d$, *is given by*

$$\begin{aligned}\mathrm{MCPE}(\hat{w}_{di}, \hat{w}_{dj}) = {} & \exp\{2\sigma_u^2 + \sigma_e^2 + (\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta}\} \\ & \times \big[1 + \{\exp(\sigma_e^2) - 1\}1_{\{i=j\}} \\ & + \exp\{(h_{d,ii} + h_{d,jj})/2 + h_{d,ij} - 2\gamma_d^2 h_d - \sigma_u^2(1-\gamma_d)\} \\ & - \exp\{(h_{d,jj} - \gamma_d^2 h_d)/2 + \gamma_d(h_{d,j} - \gamma_d h_d) - \sigma_u^2(1-\gamma_d)\} \\ & - \exp\{(h_{d,ii} - \gamma_d^2 h_d)/2 + \gamma_d(h_{d,i} - \gamma_d h_d) - \sigma_u^2(1-\gamma_d)\}\big] \\ =: {} & M_{1d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}).\end{aligned} \tag{4.1}$$

5. Uncertainty of second-stage EB predictors. In practice, the vector of variance components $\boldsymbol{\theta} = (\sigma_u^2, \sigma_e^2)'$ is also unknown. Estimation of $\boldsymbol{\theta}$ to obtain second-stage EB predictors entails an increase in uncertainty and this increase should be accounted for in the MCPE. The additional uncertainty depends on the estimation method used for $\boldsymbol{\theta}$. This section gives an approximation up to $o(D^{-1})$ terms for the MCPE of pairs of individual second-stage EB predictors when model parameters are estimated by ML or REML.

For the second-stage EB predictors $\hat{w}_{di}^E = \hat{w}_{di}(\hat{\boldsymbol{\theta}})$ and $\hat{w}_{dj}^E = \hat{w}_{dj}(\hat{\boldsymbol{\theta}})$ of w_{di} and w_{dj}, for $i, j \in \bar{s}_d$, the MCPE can be decomposed as

$$\begin{aligned}\mathrm{MCPE}(\hat{w}_{di}^E, \hat{w}_{dj}^E) = {} & \mathrm{MCPE}(\hat{w}_{di}, \hat{w}_{dj}) + E\{(\hat{w}_{di}^E - \hat{w}_{di})(\hat{w}_{dj}^E - \hat{w}_{dj})\} \\ & + E\{(\hat{w}_{di}^E - \hat{w}_{di})(\hat{w}_{dj} - w_{dj})\} \\ & + E\{(\hat{w}_{di} - w_{di})(\hat{w}_{dj}^E - \hat{w}_{dj})\}.\end{aligned} \tag{5.1}$$

The first term on the right-hand side of (5.1) is already given in Theorem 4.2 above. The remaining terms will be approximated up to $o(D^{-1})$ terms under the following assumptions, where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of A:

(H1) $p < \infty$, $\limsup_{D\to\infty} \max_{1\le d\le D} n_d < \infty$ and $\liminf_{D\to\infty} \times \min_{1\le d\le D} n_d > 1$;
(H2) the elements of the matrix $\mathbf{X}$ are uniformly bounded as $D \to \infty$;
(H3) $\liminf_{D\to\infty} D^{-1}\lambda_{\min}(\mathbf{X}_s'\mathbf{X}_s) > 0$;
(H4) $\liminf_{D\to\infty} D^{-1}\lambda_{\min}(\mathcal{F}) > 0$.

Theorem 5.1 gives an approximation for the second term on the right-hand side of (5.1). This result uses the additional notation

$$\mathbf{x}_{dij} = \mathbf{x}_{di} + \mathbf{x}_{dj}, \qquad \mathbf{m}_d = (\mathbf{0}_{d-1}', 1, \mathbf{0}_{D-d}')', \qquad \boldsymbol{\eta}_d = \sigma_u^2 \mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{m}_d,$$

$$E_{dij} = \exp\left\{2\alpha_d + \mathbf{x}_{dij}'\boldsymbol{\beta} + \frac{1}{2}\mathbf{x}_{dij}'\mathbf{Q}_s\mathbf{x}_{dij} + 2\gamma_d(\sigma_u^2 - \gamma_d\bar{\mathbf{x}}_{ds}'\mathbf{Q}_s\bar{\mathbf{x}}_{ds})\right\},$$

$$K_d = \mathrm{tr}\left(\mathcal{F}^{-1}\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s\frac{\partial \boldsymbol{\eta}_d}{\partial \boldsymbol{\theta}}\right) + \left(\frac{\partial \alpha_d}{\partial \boldsymbol{\theta}} + 2\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s\boldsymbol{\eta}_d\right)'\mathcal{F}^{-1}\left(\frac{\partial \alpha_d}{\partial \boldsymbol{\theta}} + 2\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s\boldsymbol{\eta}_d\right),$$

$$M_{2d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) = E_{dij}K_d.$$

THEOREM 5.1. *Let $\hat{w}_{di}^E = \hat{w}_{di}(\hat{\boldsymbol{\theta}})$ be the second-stage EB predictor of w_{di}, with $\hat{\boldsymbol{\theta}}$ denoting either ML or REML estimator of $\boldsymbol{\theta}$ under the nested-error model with log-transformation* (2.1). *If assumptions* (H1)–(H4) *hold, then*

$$E\{(\hat{w}_{di}^E - \hat{w}_{di})(\hat{w}_{dj}^E - \hat{w}_{dj})\} = M_{2d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) + o(D^{-1}),$$

where the $o(D^{-1})$ is uniform over i and j in $\bar{s}_d$.

Theorem 5.2 gives a second-order unbiased approximation for the first of the crossed product terms in (5.1); the last term is analogous. For this theorem, we need to introduce additional notation. We define

$$E_{dij}^* = \exp\{\alpha_d + \mathbf{x}_{dij}'\boldsymbol{\beta} + \sigma_e^2 + \sigma_u^2(3 + \gamma_d) + h_{d,ii} + 2h_{d,ij} - 2\gamma_d h_{d,j} - \gamma_d^2 h_d\}. \tag{5.2}$$

We also define $\mathbf{E}_d = 2(\boldsymbol{\Delta}_1\boldsymbol{\eta}_d, \boldsymbol{\Delta}_2\boldsymbol{\eta}_d)$, $\mathbf{A}_d = (\alpha_{d,ht})$, with $\alpha_{d,ht} = \partial^2\alpha_d/\partial\theta_h\partial\theta_t$, $\mathbf{B}_d = (b_{d,ht})$ with $b_{d,ht} = 2\boldsymbol{\eta}_d'\mathbf{V}_s(\partial^2\boldsymbol{\eta}_d/\partial\theta_h\partial\theta_t)$,

$$\mathbf{G}_d = \operatorname*{col}_{1\le k\le 2}\left\{\left(\frac{\partial \alpha_d}{\partial \boldsymbol{\theta}} + 2\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s\boldsymbol{\eta}_d\right)'\mathcal{F}^{-1}\Phi_k\right\},$$

for $\Phi_k = (\phi_{hk\ell})_{h,\ell}$ with $\phi_{hk\ell} = \mathrm{tr}(\mathbf{V}_s^{-1}\boldsymbol{\Delta}_h\mathbf{V}_s^{-1}\boldsymbol{\Delta}_t\mathbf{V}_s^{-1}\boldsymbol{\Delta}_k)$, $\boldsymbol{\varepsilon}_d = \operatorname*{col}_{1\le h\le 2}(4\boldsymbol{\eta}_d'\boldsymbol{\Delta}_h\boldsymbol{\eta}_d)$, $\boldsymbol{\varsigma} = (\varsigma_1, \varsigma_2)'$, with $\varsigma_h = 2\,\mathrm{tr}(\mathcal{F}^{-1}\Phi_h)$, $h = 1, 2$, and $\boldsymbol{\nu} = (\nu_1, \nu_2)'$, with $\nu_h = \mathrm{tr}(\mathbf{P}_s\boldsymbol{\Delta}_h) - \mathrm{tr}(\mathbf{V}_s^{-1}\boldsymbol{\Delta}_h)$, $h = 1, 2$, and

$$C_d = \mathrm{tr}\left[\mathcal{F}^{-1}\left(\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{E}_d + \frac{\mathbf{A}_d + \mathbf{B}_d}{2} - \mathbf{G}_d\right)\right] + \left(\frac{\partial \alpha_d}{\partial \boldsymbol{\theta}} + 2\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s\boldsymbol{\eta}_d\right)'\mathcal{F}^{-1}\left(\boldsymbol{\nu} + \frac{\boldsymbol{\varepsilon}_d + \boldsymbol{\varsigma}}{2}\right).$$

Finally, we define

$$\begin{aligned} M_{2d,ij}^*(\boldsymbol{\beta}, \boldsymbol{\theta}) &= E_{dij}^* K_d, \qquad T_{d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) = E_{dij}C_d, \\ T_{d,ij}^*(\boldsymbol{\beta}, \boldsymbol{\theta}) &= E_{dij}^* C_d, \\ M_{3d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \frac{1}{2}M_{2d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) + T_{d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ &\quad - \frac{1}{2}M_{2d,ij}^*(\boldsymbol{\beta}, \boldsymbol{\theta}) - T_{d,ij}^*(\boldsymbol{\beta}, \boldsymbol{\theta}). \end{aligned} \tag{5.3}$$

THEOREM 5.2. *Let $\hat{w}_{di}^E = \hat{w}_{di}(\hat{\boldsymbol{\theta}})$ be the second-stage EB predictor of w_{di} under the nested-error model with log-transformation* (2.1), *with $\hat{\boldsymbol{\theta}}$ denoting either ML or REML estimator of $\boldsymbol{\theta}$. If assumptions* (H1)–(H4) *hold, then for $i, j \in \bar{s}_d$, we have*

$$E\{(\hat{w}_{di}^E - \hat{w}_{di})(\hat{w}_{dj} - w_{dj})\} = M_{3d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) + o(D^{-1}),$$

where the $o(D^{-1})$ is uniform over i and j in $\bar{s}_d$. If $\hat{\boldsymbol{\theta}}$ is the REML estimator, set $\boldsymbol{\nu} = \mathbf{0}_2$ in $M_{3d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta})$.

Finally, Theorem 5.3 gives a second-order approximation to the MCPE of $\hat{w}_{di}^E$ and $\hat{w}_{dj}^E$, as a direct consequence of decomposition (5.1) and Theorems 4.2, 5.1 and 5.2.

THEOREM 5.3. *Let $\hat{w}_{di}^E = \hat{w}_{di}(\hat{\boldsymbol{\theta}})$ be the second-stage EB predictor of w_{di} under the nested-error model with log-transformation* (2.1), *with $\hat{\boldsymbol{\theta}}$ denoting either ML or REML estimator of $\boldsymbol{\theta}$. Under assumptions* (H1)–(H4), *it holds*

$$\begin{aligned}\mathrm{MCPE}(\hat{w}_{di}^E, \hat{w}_{dj}^E) = {} & M_{1d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) + M_{2d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) + M_{3d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ & + M_{3d,ji}(\boldsymbol{\beta}, \boldsymbol{\theta}) + o(D^{-1}),\end{aligned}$$

where the $o(D^{-1})$ is uniform over i and j in $\bar{s}_d$.

The following corollary gives a second-order approximation to the MSE of the second-stage EB predictor $\hat{\tau}_d^E$ of the area mean τ_d.

COROLLARY 5.1. *A second-order approximation to the MSE of $\hat{\tau}_d^E$ is obtained writing*

$$\begin{aligned}\mathrm{MSE}(\hat{\tau}_d^E) &= \frac{1}{N_d^2}\Biggl\{\sum_{i \in \bar{s}_d} \mathrm{MSE}(\hat{w}_{di}^E) + 2 \sum_{i \in \bar{s}_d} \sum_{j \in \bar{s}_d, j > i} \mathrm{MCPE}(\hat{w}_{di}^E, \hat{w}_{dj}^E)\Biggr\} \\ &= M_{1d}(\boldsymbol{\beta}, \boldsymbol{\theta}) + M_{2d}(\boldsymbol{\beta}, \boldsymbol{\theta}) + 2M_{3d}(\boldsymbol{\beta}, \boldsymbol{\theta}) + o(D^{-1}),\end{aligned} \tag{5.4}$$

for $M_{kd}(\boldsymbol{\beta}, \boldsymbol{\theta}) = N_d^{-2}(\sum_{i \in \bar{s}_d} M_{kd,ii}(\boldsymbol{\beta}, \boldsymbol{\theta}) + 2 \sum_{i \in \bar{s}_d} \sum_{j \in \bar{s}_d, j > i} M_{kd,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}))$, $k = 1, 2, 3$, where we have applied Theorem 5.3 with $i = j$ for $\mathrm{MSE}(\hat{w}_{di}^E)$ *and with $i \neq j$.*

6. Estimation of the uncertainty. Theorem 6.1 states that replacing the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ by their corresponding ML estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ in $M_{1d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta})$ leads to a $O(D^{-1})$ bias. It also gives a second-order approximation for that bias, which can then be corrected. The proof follows closely that of Theorem 4 in [12].

THEOREM 6.1. *Let $\hat{\boldsymbol{\theta}}$ denote either ML or REML estimator of $\boldsymbol{\theta}$ under the nested-error model with log-transformation* (2.1) *and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$. If assumptions* (H1)–(H4) *hold, then*

$$E\{M_{1d,ij}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})\} = M_{1d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) + \sum_{k=1}^{3} \Lambda_{d,ij,k}(\boldsymbol{\beta}, \boldsymbol{\theta}) + o(D^{-1}),$$

where $\Lambda_{d,ij,1}(\boldsymbol{\beta}, \boldsymbol{\theta}) = 2(\partial M_{1d,ij}/\partial \boldsymbol{\theta})' \mathcal{F}^{-1} \boldsymbol{\nu}$, $\Lambda_{d,ij,2}(\boldsymbol{\beta}, \boldsymbol{\theta}) = (1/2) \operatorname{tr}[(\partial^2 M_{1d,ij}/\partial \boldsymbol{\theta}^2) \mathcal{F}^{-1}]$ *and* $\Lambda_{d,ij,3}(\boldsymbol{\beta}, \boldsymbol{\theta}) = M_{1d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) \mathbf{x}'_{dij} \mathbf{Q}_s \mathbf{x}_{dij}$. *If $\hat{\boldsymbol{\theta}}$ is the REML estimator,* $\Lambda_{d,ij,1}(\boldsymbol{\beta}, \boldsymbol{\theta}) = 0$ *because* $\boldsymbol{\nu} = \mathbf{0}_2$.

It is not difficult to see that plugging the ML estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ for the true values $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ in the above bias correction terms leads to negligible bias in the sense

$$E\{\Lambda_{d,ij,k}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})\} = \Lambda_{d,ij,k}(\boldsymbol{\beta}, \boldsymbol{\theta}) + o(D^{-1}), \qquad k = 1, 2, 3. \tag{6.1}$$

The same occurs for REML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. According to Theorem 6.1 and equation (6.1), an unbiased estimator of $\text{MCPE}(\tilde{w}_{di}, \tilde{w}_{dj})$ up to $o(D^{-1})$ terms is given by

$$\text{mcpe}(\tilde{w}_{di}, \tilde{w}_{dj}) = M_{1d,ij}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \sum_{k=1}^{3} \Lambda_{d,ij,k}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}). \tag{6.2}$$

Moreover, by [12], it holds that

$$E\{M_{2d,ij}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})\} = M_{2d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) + o(D^{-1}). \tag{6.3}$$

So far we have obtained unbiased estimators up to $o(D^{-1})$ terms of the first two terms on the right-hand side of (5.1). Thus, in order to have an unbiased estimator of (5.1) of the same order, it only remains to estimate unbiasedly $M_{3d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta})$. The next theorem states that plugging the ML estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ in $M_{3d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta})$ yields an unbiased estimator of the desired order.

THEOREM 6.2. *Let $\hat{\boldsymbol{\theta}}$ denote either ML or REML estimator of $\boldsymbol{\theta}$ under the nested-error model with log-transformation* (2.1) *and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$. If assumptions* (H1)–(H4) *hold, then*

$$E\{M_{3d,ij}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})\} = M_{3d,ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) + o(D^{-1}).$$

For the proof of this result, the reader is addressed to the preprint in arXiv: 1404.5465.

The analogous result holds for $M_{3d,ji}(\boldsymbol{\beta},\boldsymbol{\theta}) = E\{(\hat{w}_{di} - \hat{w}_{di})(\hat{w}_{dj}^E - w_{dj})\} + o(D^{-1})$. Finally, from (6.2), (6.1) and Theorem 6.2, the estimator

$$\begin{aligned}\operatorname{mcpe}(\hat{w}_{di}^E, \hat{w}_{dj}^E) = M_{1d,ij}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}) - \sum_{k=1}^{3} \Lambda_{d,ij,k}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}) + M_{2d,ij}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}) \\ + M_{3d,ij}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}) + M_{3d,ji}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}})\end{aligned}$$

satisfies

$$E\{\operatorname{mcpe}(\hat{w}_{di}^E, \hat{w}_{dj}^E)\} = \operatorname{MCPE}(\hat{w}_{di}^E, \hat{w}_{dj}^E) + o(D^{-1}).$$

7. Bootstrap estimation of the uncertainty. Resampling methods are very popular among practitioners due to their conceptual simplicity, which also makes them less prone to coding errors. Under the setup of this paper, the naive bootstrap procedure for finite populations proposed by [6] can be applied for the estimation of the MSE of either an individual predictor $\hat{w}_{di}^E$ or for the predicted area mean $\hat{\tau}_d^E$. It can also be applied to estimate the MCPE of two individual predictors $\hat{w}_{di}^E$ and $\hat{w}_{dj}^E$, with $j \neq i$. Here, we describe only the steps of the bootstrap procedure for estimation of the MSE of $\hat{\tau}_d^E$, because for the other cases is analogous:

(1) With the available data $(\mathbf{y}_s, \mathbf{X}_s)$ coming from the sample s, calculate the ML estimators of the model parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2)'$.

(2) Generate bootstrap random effects $u_d^* \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \hat{\sigma}_u^2)$, $d = 1, \ldots, D$.

(3) Generate bootstrap errors $e_{di}^* \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \hat{\sigma}_e^2)$, $i = 1, \ldots, N_d$, $d = 1, \ldots, D$.

(4) Generate a bootstrap population of response variables from the fitted model

$$y_{di}^* = \mathbf{x}_{di}'\hat{\boldsymbol{\beta}} + u_d^* + e_{di}^*, \qquad i = 1, \ldots, N_d, d = 1, \ldots, D. \tag{7.1}$$

Let $\tau_d^* = N_d^{-1} \sum_{i=1}^{N_d} \exp(y_{di}^*)$ be the true mean of area d in this bootstrap population.

(5) Let $\mathbf{y}_s^*$ be the vector with the bootstrap elements whose subscripts are in the original sample s, $\{y_{di}^*; i \in s_d, d = 1, \ldots, D\}$. Using the bootstrap sample data $\mathbf{y}_s^*$ and $\mathbf{X}_d$, fit the bootstrap model (7.1), obtaining new model parameter estimators $\hat{\boldsymbol{\beta}}^*$ and $\hat{\boldsymbol{\theta}}^* = (\hat{\sigma}_u^{2*}, \hat{\sigma}_e^{2*})'$. Calculate the bootstrap second-stage EB predictor

$$\hat{\tau}_d^{E*} = \tilde{\tau}_d^*(\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\theta}}^*) = \frac{1}{N_d}\left\{\sum_{i \in s_d} \exp(y_{di}^*) + \sum_{i \in \bar{s}_d} \exp(\tilde{y}_{di}^* + \hat{\alpha}_d^*)\right\},$$

for $\hat{\alpha}_d^* = \{\hat{\sigma}_u^{2*}(1 - \hat{\gamma}_d^*) + \hat{\sigma}_e^{2*}\}/2$ and $\tilde{y}_{di}^* = \mathbf{x}_{di}'\hat{\boldsymbol{\beta}}^* + \hat{\gamma}_d^*(\bar{y}_{ds}^* - \bar{\mathbf{x}}_{ds}'\hat{\boldsymbol{\beta}}^*)$, where $\bar{y}_{ds}^* = n_d^{-1} \sum_{i \in s_d} y_{di}^*$ and $\hat{\gamma}_d^* = \hat{\sigma}_u^{2*}/(\hat{\sigma}_u^{2*} + \hat{\sigma}_e^{2*}/n_d)$.

(6) The bootstrap MSE of $\hat{\tau}_d^{E*}$ is then

$$\operatorname{MSE}_*(\hat{\tau}_d^{E*}) = E_*(\hat{\tau}_d^{E*} - \tau_d^*)^2, \tag{7.2}$$

where E_* indicates expectation with respect to the probability distribution induced by model (7.1) given the original sample data $\{y_{di}; i \in s_d, d = 1, \dots, D\}$.

In practice, (7.2) is approximated by Monte Carlo, by repeating Steps 2–5 a large number of times B, and then averaging over the B replicates. Let $\tau_d^{*(b)}$ be the true parameter in bth replicate and $\hat{\tau}_d^{E*(b)}$ be the corresponding second-stage EB predictor. The Monte Carlo approximation of (7.2), used here as an estimator of $\mathrm{MSE}(\hat{\tau}_d^E)$, is given by

$$\mathrm{mse}_*(\hat{\tau}_d^E) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\tau}_d^{E*(b)} - \tau_d^{*(b)})^2. \tag{7.3}$$

For a a linear parameter, the consistency of the bootstrap MSE of the second-stage EB predictor was proved by [6] using the technique of imitation. With the available analytical formula for the MCPE given in Theorem 5.3, here the result is analogous.

THEOREM 7.1. *If assumptions* (H1)–(H4) *hold, then under the bootstrap model* (7.1) *given the sample data* $(\mathbf{y}_s, \mathbf{X}_s)$, *it holds*

$$\mathrm{MCPE}_*(\hat{w}_{di}^{E*}, \hat{w}_{dj}^{E*}) = \mathrm{MCPE}_{N*}(\hat{w}_{di}^{E*}, \hat{w}_{dj}^{E*}) + o(D^{-1}), \tag{7.4}$$

where $\mathrm{MCPE}_{N*}(\hat{w}_{di}^{E*}, \hat{w}_{dj}^{E*}) = M_{1d,ij}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) + M_{2d,ij}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) + M_{3d,ij}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) + M_{3,j,i}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$.

The result follows by imitating the proofs of Theorems 4.2, 5.1 and 5.2 under the bootstrap model (7.1) given the sample data $(\mathbf{y}_s, \mathbf{X}_s)$. The following theorem yields the consistency of the bootstrap MCPE to the real value.

THEOREM 7.2. *Under the model* (2.1) *with assumptions* (3.1) *and* (H1)–(H4), *it holds*

$$|\mathrm{MCPE}_{N*}(\hat{w}_{di}^{E*}, \hat{w}_{dj}^{E*}) - \mathrm{MCPE}_N(\hat{w}_{di}^{E}, \hat{w}_{dj}^{E})| = O_p(D^{-1/2}).$$

The result follows by noting that under assumptions (3.1) and (H1)–(H4), $|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| = O_p(D^{-1/2})$ and $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| = O_p(D^{-1/2})$, and that $\mathrm{MCPE}_{N*}(\hat{w}_{di}^{E*}, \hat{w}_{dj}^{E*})$ is a continuous function of the elements of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$.

The above result implies that the bootstrap MCPE is only first-order unbiased, that is, $E\{\mathrm{MCPE}_{N*}(\hat{w}_{di}^{E*}, \hat{w}_{dj}^{E*})\} = \mathrm{MCPE}(\hat{w}_{di}^{E}, \hat{w}_{dj}^{E}) + O(D^{-1})$, where the $O(D^{-1})$ term comes from the bias terms in Theorem 6.1. The analogous result is obtained for $\hat{\tau}_d^E$ using (5.4).

For bias corrections of the naive bootstrap estimator (7.3) to achieve a $o(D^{-1})$ bias in the case of linear parameters see, for example, [2] and [18]. For a bias correction based on double bootstrap, see [7]. These corrections can be directly extended to estimate our specific nonlinear parameters w_{di} or τ_d, but they might yield negative MSE estimates. Hall and Maiti [8] proposed a positive bias-corrected MSE estimate through double bootstrap, but the second-order unbiasedness property is lost. Thus, ensuring a positive bootstrap MSE estimate and second-order unbiased is still a challenge. Recently, [9] have found a second-order unbiased and positive estimator of the log-MSE based on jackknife.

8. Simulation experiment. We carried out a simulation experiment to compare, in terms of bias and MSE under the simple mean model $y_{di} = \mu + u_d + e_{di}$, the following estimators of the area means τ_d: (i) second-stage EB predictor $\hat{\tau}_d^E$; (ii) naive predictor $\hat{\tau}_d^N = N_d^{-1}(\sum_{i\in s_d} w_{di} + \sum_{i\in \bar{s}_d} \hat{w}_{di}^N)$, where $\hat{w}_{di}^N = \exp(\hat{y}_{di})$; Molina's predictor $\hat{\tau}_d^M = N_d^{-1}(\sum_{i\in s_d} w_{di} + \sum_{i\in \bar{s}_d} \hat{w}_{di}^M)$, for $\hat{w}_{di}^M = \exp(\hat{y}_{di} + \hat{\alpha}_d^M)$, with $\hat{\alpha}_d^M = \hat{\sigma}_u^2(1-\hat{\gamma}_d)/2$; (iii) direct estimator $\hat{\tau}_d^D = n_d^{-1}\sum_{i\in s_d} w_{di}$ and (iv) the estimator obtained assuming the area-level model of [5], $\hat{\tau}_d^D = \mu + v_d + \varepsilon_d$, where v_d are assumed i.i.d. with $E(v_d) = 0$, $\text{var}(v_d) = \sigma_v^2$, and ε_d are independent with $E(\varepsilon_d) = 0$ and $\text{var}(\varepsilon_d) = \psi_d$, with ψ_d assumed to be known and fixed to the sampling variance of the direct estimator $\hat{\tau}_d^D$, $d = 1, \ldots, D$. We will also analyze the contribution of each term of $\text{MSE}(\hat{\tau}_d)$.

We consider a limited number of areas, $D = 12$, in order to analyze the small sample properties of the estimators. Population sizes of the areas are taken as $N_d = 150, 200, 250$, each value repeated for four consecutive areas, which gives a total population size of $N = 2400$. Model parameters are taken as $\mu = 1$, $\sigma_e^2 = 1$ and $\sigma_u^2 = 0.3$. A total of $K = 10{,}000$ Monte Carlo (MC) populations were generated from the mentioned mean model. In each MC simulation replicate, simple random samples s_d without replacement of sizes $n_d = 5, 10, 20$ were drawn from the areas with $N_d = 150, 200, 250$, respectively, independently from each area d, making a total sample size of $n = 140$. With these sample sizes, the variance fractions are respectively $\gamma_d = \sigma_u^2/(\sigma_u^2 + \sigma_e^2/n_d) = 0.6, 0.75, 0.86$. In this case, by Proposition 1.1 in Supplement A, the naive predictors $\tilde{w}_{di}^N$ have a relative bias that amounts to $RB(\tilde{w}_{di}^N) = 100 \times B(\tilde{w}_{di}^N)/E(w_{di}) = -42.9\%, -41.6\%, -40.6\%$ for the areas with sample sizes $n_d = 5, 10, 20$, respectively. For Molina's predictor, it is constantly equal to $RB(\tilde{w}_{di}^M) = -39.3\%$.

Concerning the empirical biases and MSEs of the estimators of the area means τ_d, Figure 1 (left) plots the MC means of the true values τ_d and of the estimators (i)–(iv) and the corresponding MSEs (right). This figure illustrates how the naive and Molina's predictors are both considerably biased low and also how the EB predictor $\hat{\tau}_d^E$ proposed in this paper has a negligible bias together with a substantially smaller MSE than all other estimators. The direct estimator $\hat{\tau}_d^D$ is unbiased as long as $n_d^{-1}\sum_{i\in s_d}\exp(\mathbf{x}_{di}'\boldsymbol{\beta}) = N_d^{-1}\sum_{i=1}^{N_d}\exp(\mathbf{x}_{di}'\boldsymbol{\beta})$, which holds in this case since $\mathbf{x}_{di} = 1\ \forall i$. However, we can see that for very small sample sizes, it has large MSE. Still, see that for $n_d \geq 10$, it overperforms the naive and Molina's estimators, and also the estimator based on the FH model. In fact, for the areas with $n_d = 20$

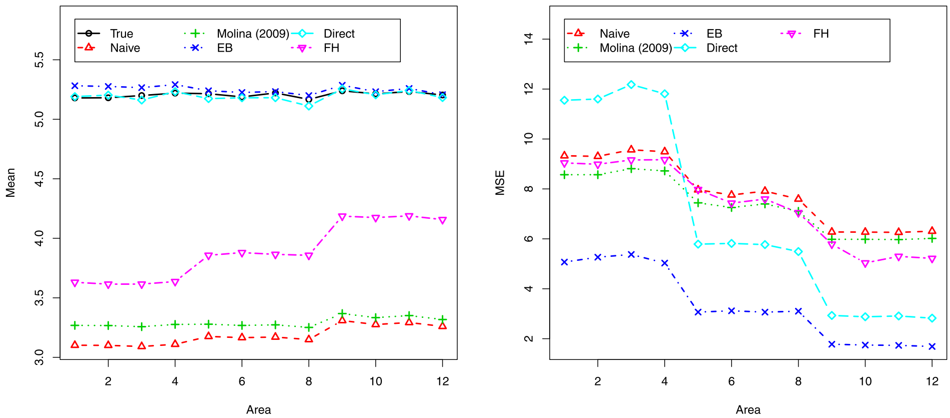

FIG. 1. *Monte Carlo means of true values, naive, Molina's and second-stage EB predictors, direct estimators and estimators based on FH model (left). Monte Carlo MSEs of all the estimators (right).*

(last four), $\mathrm{MSE}(\hat{\tau}_d^D) = 2.9$ whereas the squared biases of $\tilde{\tau}_d^N$ and $\tilde{\tau}_d^M$ amount to 5 and 3.5, respectively; see Supplement A, Section 2. Hence, in this case the biases of the naive and Molina estimators ruin the efficiency provided (or "strength borrowed") by the model. For results on individual prediction, see Figure 1 of Supplement A, which yields similar conclusions.

Next, we analyze the contribution of each MSE term to the total $\mathrm{MSE}(\hat{\tau}_d^E)$. Figure 2 displays the MC approximation to $\mathrm{MSE}(\hat{\tau}_d^E)$ labelled "MC MSE(EB)", $\mathrm{MSE}(\tilde{\tau}_d)$ given in Corollary 4.1 labelled "MSE(B)", $\mathrm{MSE}(\hat{\tau}_d)$ given in Theorem 4.2 labelled "MSE(EB1)", the same but adding the crossed-product terms $M_{2d,ij}$ given in Theorem 5.1, and finally the analytical approximation to $\mathrm{MSE}(\hat{\tau}_d^E)$ obtained from Theorem 5.3 and Corollary 5.1 that includes the terms $M_{3d,ij} + M_{3d,ji}$. It is clear that the naive estimators $\mathrm{MSE}(\tilde{\tau}_d)$ and $\mathrm{MSE}(\hat{\tau}_d)$ underestimate seriously the true MSE and the additional MSE terms of Theorems 5.1 and 5.3 seem to be needed to avoid undesired underestimation of the MSE. In fact, the average relative difference of $\mathrm{MSE}(\tilde{\tau}_d)$ and $\mathrm{MSE}(\hat{\tau}_d)$ to the MC MSE values is -14.1% and -11.7%, respectively, while for $\mathrm{MSE}(\hat{\tau}_d^E)$ it is 1%. The differences that we observe in the MC MSE(EB) for the areas with the same sample and population sizes are due to MC error since, under a mean model with $\mathbf{x}_{di} = 1\ \forall i$, the true MSE is constant for those areas.

Figure 2 in Supplement A shows the MC means, obtained with $K = 1000$ simulations, of the analytical estimate of $\mathrm{MSE}(\hat{\tau}_d^E)$ given below Theorem 6.2, and of the bootstrap estimate of Section 7, against the true MSE values obtained previously with $K = 10{,}000$. In this plot, the analytical MSE estimates are clearly closer to the true values than the bootstrap estimates. Averages across areas of percent relative biases are 20.1% and 9.4% for the bootstrap and analytical MSE estimates, respectively. These results illustrate the slower convergence rate of the naive bootstrap MSE estimate to the true MSE.

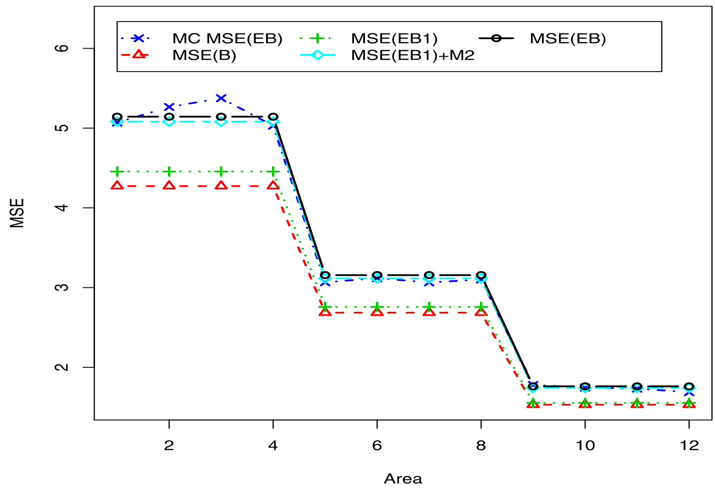

FIG. 2. *MC MSE of second-stage EB predictor* $\hat{\tau}_d^E$ *labelled "MC MSE(EB)", MSE of best predictor* $\tilde{\tau}_d$ *labelled "MSE(B)", MSE of first-stage EB predictor* $\hat{\tau}_d$ *labelled "MSE(EB1)", the same but adding the crossed-product terms* $M_{2d,ij}$, *and total MSE of second-stage EB predictor* $\hat{\tau}_d^E$ *labelled "MSE(EB)".*

9. Estimation of mean income in municipalities from Mexico. In this section, we apply the obtained results to the prediction of incomes for individuals and of mean income in municipalities from the State of Mexico. An advantage of EB prediction based on a unit level model is that estimates can be obtained at whatever level of disaggregation that is desired, since a predicted census can be constructed. Data comes from two different sources. One is the Module of Socio-economic Conditions (MCS in Spanish) from the 2010 Mexican National Survey on Income and Expense of Households (ENIGH in Spanish). The MCS collects microdata on income, health, nutrition, education, social security, quality of household, basic equipment and social cohesion in Mexico. We also have available microdata from the census of the same year. The census contains several of the variables also contained in the MCS, but the income variable used officially (monthly total per capita income) is collected only in the mentioned survey. Based on both data sources, we estimate the mean income, as well as the individual incomes, in each municipality that appears in the MCS survey data (many of them are not sampled by the MCS), except for one which, after a preliminary study of the considered variables, turned out to be very different from the other municipalities (outlier). This makes a total of $D = 57$ municipalities. From these, the minimum sample size is 8 and the maximum is 2037, with a median of 96 and an average of 185.

After a preliminary check of the relationships between income and the available variables in the MCS, we selected as auxiliary variables age, age^2, age^3, the indicators of gender, indigenous population, activity sectors (including unemployed and inactive), composition of household, quality of dwelling, indicator of receiving social benefits, classification according to the available equipment, years of schooling, indicator of rural/urban area and the interactions between quality of dwelling with rural/urban area and of composition of household with gender. Since income distribution in Mexico is heavily skewed, the model was fitted to $\log(\text{income} + k)$ where $k = 171$ was selected to achieve an approximately symmetric distribution of model residuals. Histograms of income before and after the transformation are shown in Supplement A (Figure 3), as well as the resulting fitted regression parameters (Table 1). The fitted variance components are $\sigma_u^2 = 0.0160$ and $\sigma_e^2 = 0.3245$, which lead to an average contribution of σ_u^2 to the total variance of $D^{-1}\sum_{d=1}^{D}\gamma_d = 0.789$.

We computed also direct Horvitz–Thompson estimators of mean income w_{di} together with their sampling variances, obtained as

$$\hat{\tau}_d^{\mathrm{DIR}} = N_d^{-1}\sum_{i\in s_d}\pi_{di}^{-1}w_{di}, \qquad \operatorname{var}(\hat{\tau}_d^{\mathrm{DIR}}) = N_d^{-2}\sum_{i\in s_d}\pi_{di}^{-2}(1-\pi_{di})w_{di}^2,$$

where π_{di} is the inclusion probability of ith unit in the sample from municipality d. The sampling variance is obtained using the following approximation for the second-order inclusion probabilities $\pi_{d,ij} \approx \pi_{di}\pi_{dj}$, $j \neq i$, and noting that $\pi_{d,ii} = \pi_{di}$ for all i. Figure 3 left shows EB, Molina, naive and direct estimators of mean income for the $D = 57$ municipalities. This figure illustrates that direct estimators are somewhat unstable. According to Proposition 1.1 in Supplement A, Molina and naive estimators have an average estimated relative bias of -14.8% and -14.9%, respectively. In this application, these two estimators take very similar values (superposed in the plot) and are lower than EB estimates, which could be due to the mentioned theoretical bias.

Boxplots of the estimated coefficients of variation (CVs), defined for any estimator $\hat{\tau}_d$ as $\mathrm{cv}(\hat{\tau}_d) = 100 \times \sqrt{\mathrm{mse}(\hat{\tau}_d)}/\hat{\tau}_d$, are shown in Figure 3, where $\mathrm{mse}(\hat{\tau}_d^E)$

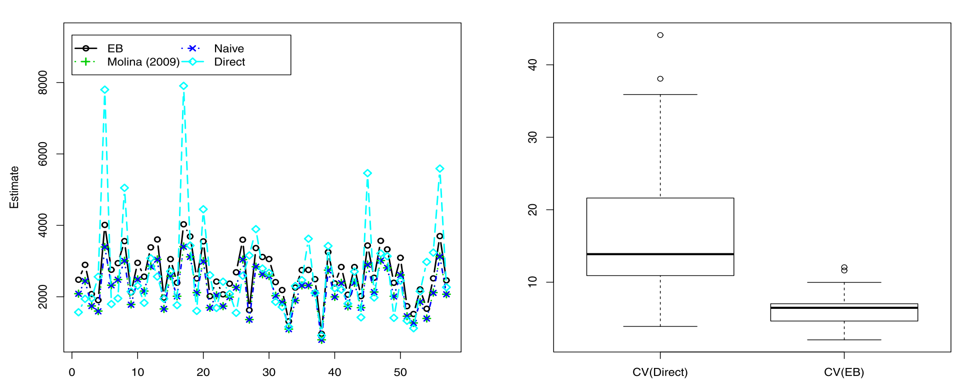

FIG. 3. *EB, Molina, naive and direct estimates of mean income for each municipality (left) and boxplots of estimated CVs of direct and EB estimates of mean income (right).*

is obtained using the analytical estimator below Theorem 6.2. These boxplots show the significant reduction in CV obtained when using EB estimators rather than the default direct ones.

When looking at results for prediction at the individual level, included in Figure 4 of Supplement A, we can see that Molina's predictors are lower than EB predictors (16% lower on average), indicating again the mentioned negative bias of Molina's predictors. Naive and Molina's predictors take practically the same values in this application; see Figure 4 right. Figure 5 in Supplement A shows a substantially smaller CV for individual EB predictors, where this CV has been estimated by the bootstrap method. Finally, Figure 6, showing the bootstrap CV estimates for the predictors of mean income (including also the direct estimator), confirms the obtained conclusions.

APPENDIX: PROOFS

We denote by $|\mathbf{a}| = (\mathbf{a}'\mathbf{a})^{1/2}$ the Euclidean norm of a vector $\mathbf{a}$. For a matrix A, we consider the norms $\|A\| = \lambda_{\max}^{1/2}(A'A)$ and $\|A\|_2 = \mathrm{tr}^{1/2}(A'A)$, where $\lambda_{\max}(A)$ denotes the maximum eigenvalue of A. Asymptotic orders refer to $D \to \infty$.

The next lemma is required in the proofs of several of the remaining results.

LEMMA A.1. *Let $\mathbf{V}_s$ be the covariance matrix of $\mathbf{y}_s$, $\mathcal{F}$ and $\mathcal{F}_R$ the ML and REML Fisher-information matrices, respectively, and $\mathbf{Q}_s = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}$. It holds*:

(i) *Condition* (H1) *implies* $\|\mathbf{V}_s\| = O(1)$.
(ii) $\|\mathbf{V}_s^{-1}\| = O(1)$.
(iii) *Conditions* (H1) *and* (H3) *imply* $\|\mathbf{Q}_s\| = O(D^{-1})$.
(iv) *Condition* (H4) *implies* $\|\mathcal{F}^{-1}\| = O(D^{-1})$ *and* $\|\mathcal{F}_R^{-1}\| = O(D^{-1})$.

PROOF. (i) Since $\mathbf{V}_s$ is symmetric and block-diagonal with blocks equal to $\mathbf{V}_{ds}$, $d = 1, \ldots, D$, we have

$$\|\mathbf{V}_s\| = \lambda_{\max}^{1/2}(\mathbf{V}_s^2) = \lambda_{\max}(\mathbf{V}_s) = \max_{1\le d\le D}\{\lambda_{\max}(\mathbf{V}_{ds})\}.$$

Now since $\mathbf{V}_{ds} = \sigma_u^2\mathbf{1}_{n_d}\mathbf{1}_{n_d}' + \sigma_e^2\mathbf{I}_{n_d}$, we have

$$\lambda_{\max}(\mathbf{V}_d) \le \sigma_u^2\lambda_{\max}(\mathbf{1}_{n_d}\mathbf{1}_{n_d}') + \sigma_e^2\lambda_{\max}(\mathbf{I}_{n_d}) = \sigma_u^2 n_d + \sigma_e^2.$$

Then, by assumption (H1), we obtain (i) from

$$\|\mathbf{V}_s\| = \max_{1\le d\le D}\{\lambda_{\max}(\mathbf{V}_{ds})\} \le \sigma_u^2 \max_{1\le d\le D} n_d + \sigma_e^2 = O(1).$$

(ii) Similarly, as before, we have

$$\|\mathbf{V}_s^{-1}\| = \lambda_{\max}(\mathbf{V}_s^{-1}) = \lambda_{\min}^{-1}(\mathbf{V}_s) = \Big\{\min_{1\le d\le D}\lambda_{\min}(\mathbf{V}_{ds})\Big\}^{-1}.$$

But again, using the expression of $\mathbf{V}_{ds} = \sigma_u^2 \mathbf{1}_{n_d} \mathbf{1}'_{n_d} + \sigma_e^2 \mathbf{I}_{n_d}$, we have

$$\lambda_{\min}(\mathbf{V}_d) \geq \sigma_u^2 \lambda_{\min}\big(\mathbf{1}_{n_d} \mathbf{1}'_{n_d}\big) + \sigma_e^2 \lambda_{\min}(\mathbf{I}_{n_d}) = \sigma_e^2 > 0,$$

which is true for all $d \in \{1, \dots, D\}$ and for all D. Therefore, $\|\mathbf{V}_s^{-1}\| = O(1)$.

(iii) By the definition of $\mathbf{Q}_s = (\mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1}$, we obtain

$$\|\mathbf{Q}_s\| = \lambda_{\max}(\mathbf{Q}_s) = \lambda_{\min}^{-1}(\mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{X}_s).$$

But by the definition of eigenvalue, we have

$$\begin{aligned}
\lambda_{\min}(\mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{X}_s) &= \min_v \frac{v' \mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{X}_s v}{v'v} \\
&= \min_v \left(\frac{v' \mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{X}_s v}{v' \mathbf{X}'_s \mathbf{X}_s v} \frac{v' \mathbf{X}'_s \mathbf{X}_s v}{v'v} \right) \\
&\geq \left(\min_w \frac{w' \mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{X}_s w}{w' \mathbf{X}'_s \mathbf{X}_s w} \right) \left(\min_v \frac{v' \mathbf{X}'_s \mathbf{X}_s v}{v'v} \right) \\
&= \lambda_{\min}(\mathbf{V}_s^{-1}) \lambda_{\min}(\mathbf{X}'_s \mathbf{X}_s) \\
&= \lambda_{\max}^{-1}(\mathbf{V}_s) \lambda_{\min}(\mathbf{X}'_s \mathbf{X}_s).
\end{aligned}$$

Using (i) and assumption (H3), we finally get

$$D\|\mathbf{Q}_s\| = D\lambda_{\min}^{-1}(\mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{X}_s) \leq \frac{\lambda_{\max}(\mathbf{V}_s)}{D^{-1} \lambda_{\min}(\mathbf{X}'_s \mathbf{X}_s)} = O(1).$$

(iv) Condition (H4) implies

$$D\|\mathcal{F}^{-1}\| = D\lambda_{\max}(\mathcal{F}^{-1}) = \{D^{-1} \lambda_{\min}(\mathcal{F})\}^{-1} = O(1).$$

Moreover, note that $\mathcal{F} = -A/2 + B$, where $B = (b_{h\ell})_{h,\ell=1,2}$ and $A = (a_{h\ell})_{h,\ell=1,2}$, for $a_{h\ell} = \operatorname{tr}(\mathbf{V}_s^{-1} \boldsymbol{\Delta}_h \mathbf{V}_s^{-1} \boldsymbol{\Delta}_\ell)$ and $b_{h\ell} = \operatorname{tr}(\mathbf{P}_s \boldsymbol{\Delta}_h \mathbf{P}_s \boldsymbol{\Delta}_\ell)$, whereas $\mathcal{F}_R = B/2$. Then

$$\begin{aligned}
\lambda_{\min}(\mathcal{F}_R) &= \lambda_{\min}(B - B/2) \\
&= \lambda_{\min}(B - B/2 + A/2 - A/2) \\
&= \lambda_{\min}\{\mathcal{F} + (A - B)/2\} \\
&\geq \lambda_{\min}(\mathcal{F}) + \frac{1}{2} \lambda_{\min}(A - B).
\end{aligned}$$

But the diagonal elements of $D^{-1}(A - B)$ tend to zero. Indeed

$$\begin{aligned}
b_{hh} - a_{hh} &= \operatorname{tr}(\mathbf{P}_s \boldsymbol{\Delta}_h \mathbf{P}_s \boldsymbol{\Delta}_h) - \operatorname{tr}(\mathbf{V}_s^{-1} \boldsymbol{\Delta}_h \mathbf{V}_s^{-1} \boldsymbol{\Delta}_h) \\
&= \operatorname{tr}(\mathbf{P}_s \boldsymbol{\Delta}_h \mathbf{W}_s \boldsymbol{\Delta}_h) + \operatorname{tr}(\mathbf{W}_s \boldsymbol{\Delta}_h \mathbf{V}_s^{-1} \boldsymbol{\Delta}_h),
\end{aligned}$$

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

$$|b_{hh} - a_{hh}| \leq |\operatorname{tr}(\mathbf{P}_s \boldsymbol{\Delta}_h \mathbf{W}_s \boldsymbol{\Delta}_h)| + |\operatorname{tr}(\mathbf{W}_s \boldsymbol{\Delta}_h \mathbf{V}_s^{-1} \boldsymbol{\Delta}_h)|.$$

Now, for the second term on the right-hand side, we have

$$
\begin{aligned}
&|\mathrm{tr}(\mathbf{W}_s \mathbf{\Delta}_h \mathbf{V}_s^{-1} \mathbf{\Delta}_h)| \\
&\quad = |\mathrm{tr}(\mathbf{V}_s^{-1}\mathbf{X}_s(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s\mathbf{V}_s^{-1}\mathbf{\Delta}_h\mathbf{V}_s^{-1}\mathbf{\Delta}_h)| \\
&\quad = |\mathrm{tr}\{(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1/2}\mathbf{X}_s\mathbf{V}_s^{-1}\mathbf{\Delta}_h\mathbf{V}_s^{-1}\mathbf{\Delta}_h\mathbf{V}_s^{-1}\mathbf{X}_s(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1/2}\}| \\
&\quad \le p\lambda_{\max}\{(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1/2}\mathbf{X}_s\mathbf{V}_s^{-1}\mathbf{\Delta}_h\mathbf{V}_s^{-1}\mathbf{\Delta}_h\mathbf{V}_s^{-1}\mathbf{X}_s(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1/2}\} \\
&\quad = p\|\mathbf{V}_s^{-1/2}\mathbf{\Delta}_h\mathbf{V}_s^{-1}\mathbf{X}_s(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1/2}\|^2 \\
&\quad \le p\lambda_{\min}^{-4}(\mathbf{V}_s)\|\mathbf{\Delta}_h\|^2\|\mathbf{V}_s^{-1/2}\mathbf{X}_s(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1/2}\|^2 \\
&\quad = p\lambda_{\min}^{-4}(\mathbf{V}_s)\|\mathbf{\Delta}_h\|^2 = O(1).
\end{aligned}
$$

Similarly, it is easy to see that $|\mathrm{tr}(\mathbf{P}_s\mathbf{\Delta}_h\mathbf{W}_s\mathbf{\Delta}_h)| = O(1)$. Therefore, for $h = 1, 2$, it holds $D^{-1}(a_{hh} - b_{hh}) \to 0$ as $D \to \infty$, leading to $\liminf \lambda_{\min}\{D^{-1}(A - B)\} = 0$, which implies

$$\liminf D^{-1}\lambda_{\min}(\mathcal{F}_R) \ge \liminf D^{-1}\lambda_{\min}(\mathcal{F}) + \liminf \lambda_{\min}\{D^{-1}(A - B)\} > 0.$$

Then, similarly as we did for $\mathcal{F}$ above, we obtain $\|\mathcal{F}_R^{-1}\| = O(D^{-1})$. □

PROOF OF THEOREM 2.1. (i) The best predictor of $w_{di} = \exp(y_{di})$ is equal to $\tilde{w}_{di} = E_{\mathbf{y}_{dr}}\{\exp(y_{di})|\mathbf{y}_{ds}\}$. For a nonstochastic vector $\mathbf{b}_d$ of size $N_d - n_d$, using the conditional distribution given in (2.3), we get

$$E_{\mathbf{y}_{dr}}[\exp(\mathbf{b}_d'\mathbf{y}_{dr})|\mathbf{y}_{ds}] = \exp(\boldsymbol{\mu}_{dr|s}'\mathbf{b}_d + \mathbf{b}_d'\mathbf{V}_{dr|s}\mathbf{b}_d/2). \tag{9.1}$$

Now (i) follows from the expressions for $\boldsymbol{\mu}_{dr|s}$ and $\mathbf{V}_{dr|s}$ given in (2.4) and (2.5), and taking $\mathbf{b}_d$ as a vector with 1 in position i and 0 in the rest of elements.

(ii) The best predictor of τ_d is given by

$$
\begin{aligned}
\tilde{\tau}_d &= \tilde{\tau}_d(\boldsymbol{\beta}, \boldsymbol{\theta}) \\
&= E_{\mathbf{y}_{dr}}(\tau_d|\mathbf{y}_{ds}) \\
&= \frac{1}{N_d}\Bigg[\sum_{i\in s_d}\exp(y_{di}) + \sum_{i\in\bar{s}_d}E_{\mathbf{y}_{dr}}\{\exp(y_{di})|\mathbf{y}_{ds}\}\Bigg].
\end{aligned} \tag{9.2}
$$

The result then follows by straightforward application of (i). □

PROOF OF THEOREM 4.1. For $i, j \in \bar{s}_d$, we need to calculate

$$\mathrm{MCPE}(\tilde{w}_{di}, \tilde{w}_{dj}) = E(\tilde{w}_{di}\tilde{w}_{dj}) - E(\tilde{w}_{di}w_{dj}) - E(w_{di}\tilde{w}_{dj}) + E(w_{di}w_{dj}). \tag{9.3}$$

Since u_d and e_{di} are independent for all i, the last term on the right-hand side of (9.3) for $i \neq j$ is given by

$$E(w_{di}w_{dj}) = \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta}\}E\{\exp(2u_d)\}E\{\exp(e_{di})\}E\{\exp(e_{dj})\}. \tag{9.4}$$

In contrast, for $i = j$ we have

$$E(w_{di}^2) = \exp(2\mathbf{x}_{di}'\boldsymbol{\beta})E\{\exp(2u_d)\}E\{\exp(2e_{di})\}. \tag{9.5}$$

Observe that the expectations appearing on the right-hand side of (9.4) and (9.5) are respectively the moment generating function (m.g.f.) of the independent random variables $2u_d$, e_{di}, e_{dj} and $2e_{di}$, evaluated at $t = 1$. Since the m.g.f. of a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is given by $M_X(t) = \exp(\mu t + \sigma^2 t^2/2)$, using this expression we get

$$E(w_{di}w_{dj}) = \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta} + 2\sigma_u^2 + \sigma_e^2(1 + 1_{\{i=j\}})\}. \tag{9.6}$$

Now we obtain $E(\tilde{w}_{di}w_{dj}) = E\{\exp(\tilde{y}_{di} + \alpha_d + y_{dj})\}$. But by model (2.1) we know

$$y_{dj} = \mathbf{x}_{dj}'\boldsymbol{\beta} + u_d + e_{dj},$$
$$\tilde{y}_{di} = \mathbf{x}_{di}'\boldsymbol{\beta} + \gamma_d(\bar{y}_{ds} - \bar{\mathbf{x}}_{ds}'\boldsymbol{\beta}) = \mathbf{x}_{di}'\boldsymbol{\beta} + \gamma_d(u_d + \bar{e}_{ds}).$$

Noting that u_d, e_{dj} for $j \in \bar{s}_d$ and $\bar{e}_{ds}$ are independent, we have

$$\begin{aligned} E(\tilde{w}_{di}w_{dj}) &= E\{\exp(\tilde{y}_{di} + \alpha_d + y_{dj})\} \\ &= \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta}\}\exp(\alpha_d)E[\exp\{(1 + \gamma_d)u_d\}] \\ &\quad \times E\{\exp(e_{dj})\}E[\exp\{\gamma_d\bar{e}_{ds}\}]. \end{aligned} \tag{9.7}$$

Using the m.g.f.'s evaluated at $t = 1$ of the random variables involved in (9.7), using the expression of $\alpha_d = \frac{1}{2}\{\sigma_u^2(1 - \gamma_d) + \sigma_e^2\}$ and the fact that $\gamma_d(\sigma_u^2 + \sigma_e^2/n_d) = \sigma_u^2$, we get

$$\begin{aligned} E(\tilde{w}_{di}w_{dj}) &= \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta} + 2\sigma_u^2 + \sigma_e^2 - \sigma_u^2(1 - \gamma_d)\} \\ &= E(w_{di}\tilde{w}_{dj}). \end{aligned} \tag{9.8}$$

Finally, we calculate $E(\tilde{w}_{di}\tilde{w}_{dj}) = E\{\exp(\tilde{y}_{di} + \tilde{y}_{dj} + 2\alpha_d)\}$. Again, by model (2.1), it holds

$$\begin{aligned} \tilde{y}_{di} + \tilde{y}_{dj} &= (\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta} + 2\gamma_d(\bar{y}_{ds} - \bar{\mathbf{x}}_{ds}'\boldsymbol{\beta}) \\ &= (\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta} + 2\gamma_d(u_d + \bar{e}_{ds}). \end{aligned}$$

Now since

$$2\gamma_d(u_d + \bar{e}_{ds}) \sim \mathcal{N}\left\{0, 4\gamma_d^2\left(\sigma_u^2 + \frac{\sigma_e^2}{n_d}\right)\right\} \equiv \mathcal{N}(0, 4\gamma_d\sigma_u^2),$$

then using again the m.g.f. of $\gamma_d(u_d + \bar{e}_{ds})$ evaluated at $t = 1$, we get

$$E[\exp\{2\gamma_d(u_d + \bar{e}_{ds})\}] = \exp(2\gamma_d\sigma_u^2).$$

Finally, using the expression of $\alpha_d = \{\sigma_u^2(1-\gamma_d) + \sigma_e^2\}/2$, we get

$$E(\tilde{w}_{di}\tilde{w}_{dj}) = \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta}\}\exp\{2\sigma_u^2 + \sigma_e^2 - \sigma_u^2(1-\gamma_d)\} \\ = E(\tilde{w}_{di} w_{dj}). \tag{9.9}$$

The result follows by replacing (9.6), (9.8) and (9.9) in (9.3). □

PROOF OF COROLLARY 4.1. Subtracting $\tau_d = N_d^{-1}\sum_{i=1}^{N_d} w_{di}$ to the best predictor $\tilde{\tau}_d = N_d^{-1}(\sum_{i\in s_d} w_{di} + \sum_{i\in\bar{s}_d}\tilde{w}_{di})$, the sum in the sample elements cancels out and we get $\tilde{\tau}_d - \tau_d = N_d^{-1}\sum_{i\in\bar{s}_d}(\tilde{w}_{di} - w_{di})$. Therefore, the MSE of $\tilde{\tau}_d$ is given by

$$\begin{aligned}\mathrm{MSE}(\tilde{\tau}_d) &= E\{(\tilde{\tau}_d - \tau_d)^2\} \\ &= N_d^{-2}\Biggl\{\sum_{i\in\bar{s}_d}\mathrm{MSE}(\tilde{w}_{di}) + 2\sum_{i\in\bar{s}_d}\sum_{j\in\bar{s}_d, j>i}\mathrm{MCPE}(\tilde{w}_{di}, \tilde{w}_{dj})\Biggr\}.\end{aligned}$$

The result follows by applying Theorem 4.1 separately for $i = j$ and for $i \neq j$. □

PROOF OF THEOREM 4.2. The mean crossed product error of a pair of individual first-stage predictors $\hat{w}_{di}$ and $\hat{w}_{dj}$, for $i, j \in \bar{s}_d$, is given by

$$\mathrm{MCPE}(\hat{w}_{di}, \hat{w}_{dj}) = E(\hat{w}_{di}\hat{w}_{dj}) + E(w_{di}w_{dj}) \\ - E(\hat{w}_{di}w_{dj}) - E(w_{di}\hat{w}_{dj}). \tag{9.10}$$

The second term on the right-hand side of (9.10) is given in (9.6). Concerning the first term on the right-hand side of (9.10), see that for all $i \in \bar{s}_d$, using (9.12), we get

$$E(\hat{w}_{di}\hat{w}_{dj}) = \exp\{2\alpha_d + (\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta}\}E[\exp\{(\mathbf{b}_{di} + \mathbf{b}_{dj})'\mathbf{v}_s\}],$$

where the expectation on the right-hand side is the m.g.f. of the normal random vector $(\mathbf{b}_{di} + \mathbf{b}_{dj})'\mathbf{v}_s$ evaluated at 1, that is,

$$E(\hat{w}_{di}\hat{w}_{dj}) = \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta} + (\mathbf{b}_{di} + \mathbf{b}_{dj})'\mathbf{V}_s(\mathbf{b}_{di} + \mathbf{b}_{dj})/2 + 2\alpha_d\}. \tag{9.11}$$

Concerning the remaining expectations in (9.10), first note that $w_{di} = \exp(y_{di})$ for $y_{di} = \mathbf{x}_{di}'\boldsymbol{\beta} + u_d + e_{di}$. Moreover, the first-stage EB predictor of w_{di} can be expressed as

$$\hat{w}_{di} = \exp(\hat{y}_{di} + \alpha_d), \qquad \hat{y}_{di} = \mathbf{x}_{di}'\boldsymbol{\beta} + \mathbf{b}_{di}'\mathbf{v}_s, \tag{9.12}$$

for the vector

$$\mathbf{b}_{di} = \mathbf{V}_s^{-1}\mathbf{X}_s\mathbf{Q}_s\mathbf{x}_{di} + \sigma_u^2\mathbf{P}_s\mathbf{Z}_s\mathbf{m}_d, \tag{9.13}$$

where $\mathbf{m}_d = (\mathbf{0}_{d-1}', 1, \mathbf{0}_{D-d}')'$. Then we have

$$y_{di} + \hat{y}_{dj} = (\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta} + \mathbf{b}_{dj}'\mathbf{v}_s + u_d + e_{di} + \alpha_d.$$

Replacing now $\mathbf{v}_s = \mathbf{Z}_s\mathbf{u} + \mathbf{e}_s$ and writing $u_d = \mathbf{m}_d'\mathbf{u}$, we obtain

$$E(w_{di}\hat{w}_{dj}) = \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta} + \alpha_d\}E\{\exp(e_{di})\} \\ \times E[\exp\{(\mathbf{m}_d' + \mathbf{b}_{dj}'\mathbf{Z}_s)\mathbf{u}\}]E\{\exp(\mathbf{b}_{dj}'\mathbf{e}_s)\}.$$

Similarly as before, using the m.g.f. of the normal random vectors involved in the previous expression and rearranging the terms, we obtain

$$E(w_{di}\hat{w}_{dj}) = \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta} + \alpha_d + (\sigma_e^2 + \sigma_u^2)/2 \\ + \mathbf{b}_{dj}'\mathbf{V}_s\mathbf{b}_{dj}/2 + \sigma_u^2\mathbf{m}_d'\mathbf{Z}_s\mathbf{b}_{dj}\}. \tag{9.14}$$

Replacing (9.6), (9.11) and (9.14) in (9.10), we get

$$\begin{aligned} \mathrm{MCPE}(\hat{w}_{di}, \hat{w}_{dj}) = \exp\{(\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta}\}[&\exp\{2\sigma_u^2 + \sigma_e^2(1 + 1_{\{i=j\}})\} \\ &+ \exp\{(\mathbf{b}_{di} + \mathbf{b}_{dj})'\mathbf{V}_s(\mathbf{b}_{di} + \mathbf{b}_{dj})/2 + 2\alpha_d\} \\ &- \exp\{(\sigma_e^2 + \sigma_u^2)/2 + \mathbf{b}_{di}'\mathbf{V}_s\mathbf{b}_{di}/2 + \sigma_u^2\mathbf{m}_d'\mathbf{Z}_s\mathbf{b}_{di} + \alpha_d\} \\ &- \exp\{(\sigma_e^2 + \sigma_u^2)/2 + \mathbf{b}_{dj}'\mathbf{V}_s\mathbf{b}_{dj}/2 + \sigma_u^2\mathbf{m}_d'\mathbf{Z}_s\mathbf{b}_{dj}\} + \alpha_d]. \end{aligned} \tag{9.15}$$

Let us calculate the expression of each term in (9.15). Now using the definition of $\mathbf{b}_{di}$ given in (9.13) and $\mathbf{P}_s$ in (3.2), and taking into account that $\mathbf{X}_s'\mathbf{P}_s = \mathbf{0}_{p\times n}$ and $\mathbf{P}_s\mathbf{V}_s\mathbf{P}_s = \mathbf{P}_s$, it is easy to see that

$$\begin{aligned} &(\mathbf{b}_{di} + \mathbf{b}_{dj})'\mathbf{V}_s(\mathbf{b}_{di} + \mathbf{b}_{dj}) \\ &\quad = (\mathbf{x}_{di} + \mathbf{x}_{dj})'\mathbf{Q}_s(\mathbf{x}_{di} + \mathbf{x}_{dj}) + 4(\sigma_u^2)^2\mathbf{m}_d'\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{m}_d \\ &\quad\quad - 4(\sigma_u^2)^2\mathbf{m}_d'\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{X}_s\mathbf{Q}_s\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{m}_d. \end{aligned} \tag{9.16}$$

Since $\mathbf{V}_s = \mathrm{diag}_{1\le d\le D}(\mathbf{V}_{ds})$ with $\mathbf{V}_{ds} = \sigma_u^2\mathbf{1}_{n_d}\mathbf{1}_{n_d}' + \sigma_e^2\mathbf{I}_{n_d}$, $\mathbf{m}_d = (\mathbf{0}_{d-1}', 1, \mathbf{0}_{D-d}')'$, $\mathbf{Z}_{ds} = \mathrm{diag}_{1\le d\le D}(\mathbf{1}_{n_d})$ and $\mathbf{X}_s = (\mathbf{X}_{1s}', \dots, \mathbf{X}_{Ds}')'$, we obtain

$$\mathbf{m}_d'\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{m}_d = \frac{\gamma_d}{\sigma_u^2}, \qquad \mathbf{m}_d'\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{X}_s = \frac{\gamma_d}{\sigma_u^2}\bar{\mathbf{x}}_{ds}'. \tag{9.17}$$

Replacing (9.17) in (9.16), we finally obtain

$$\begin{aligned} (\mathbf{b}_{di} + \mathbf{b}_{dj})'\mathbf{V}_s(\mathbf{b}_{di} + \mathbf{b}_{dj}) &= (\mathbf{x}_{di} + \mathbf{x}_{dj})'\mathbf{Q}_s(\mathbf{x}_{di} + \mathbf{x}_{dj}) \\ &\quad + 4\gamma_d(\sigma_u^2 - \gamma_d\bar{\mathbf{x}}_{ds}'\mathbf{Q}_s\bar{\mathbf{x}}_{ds}). \end{aligned} \tag{9.18}$$

Similarly, we obtain

$$\mathbf{b}_{di}'\mathbf{V}_s\mathbf{b}_{di} = \mathbf{x}_{di}'\mathbf{Q}_s\mathbf{x}_{di} + \gamma_d(\sigma_u^2 - \gamma_d\bar{\mathbf{x}}_{ds}'\mathbf{Q}_s\bar{\mathbf{x}}_{ds}) = \gamma_d\sigma_u^2 + h_{d,ii} - \gamma_d^2 h_d. \tag{9.19}$$

On the other hand, using (9.17), we get

$$\sigma_u^2\mathbf{m}_d'\mathbf{Z}_s\mathbf{b}_{di} = \gamma_d(\sigma_u^2 + \mathbf{x}_{di}'\mathbf{Q}_s\bar{\mathbf{x}}_{ds} - \gamma_d\bar{\mathbf{x}}_{ds}'\mathbf{Q}_s\bar{\mathbf{x}}_{ds}). \tag{9.20}$$

Replacing (9.18), (9.19), (9.20) and the expression for α_d in (9.15), we obtain the desired expression for $\mathrm{MCPE}(\hat{w}_{di}, \hat{w}_{dj})$. □

PROOF OF THEOREM 5.1. We prove it for the case in which $\hat{\boldsymbol{\theta}}$ is the ML estimator of $\boldsymbol{\theta}$. For the REML estimator the proof is analogous, but in fact simpler. Following the same arguments as in the proof of Theorem 1 in [12], we obtain

$$E\{(\hat{w}_{di}^E - \hat{w}_{di})(\hat{w}_{dj}^E - \hat{w}_{dj})\} = E\{(\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{s})(\mathbf{h}_{dj}'\mathcal{F}^{-1}\mathbf{s})\} + o(D^{-1}), \tag{9.21}$$

where $\mathbf{h}_{di} = \partial\hat{w}_{di}/\partial\boldsymbol{\theta}$. Using the same ideas as in Theorem 2 in [12], we get

$$\begin{aligned} &E\{(\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{s})(\mathbf{h}_{dj}'\mathcal{F}^{-1}\mathbf{s})\} \\ &\quad = \exp\left\{2\alpha_d + (\mathbf{x}_{di} + \mathbf{x}_{dj})'\boldsymbol{\beta} + \frac{1}{2}(\mathbf{b}_{di} + \mathbf{b}_{dj})'\mathbf{V}_s(\mathbf{b}_{di} + \mathbf{b}_{dj})\right\} \\ &\qquad \times \left\{\mathrm{tr}\left(\mathcal{F}^{-1}\frac{\partial\boldsymbol{\eta}_d'}{\partial\boldsymbol{\theta}}\mathbf{V}_s\frac{\partial\boldsymbol{\eta}_d}{\partial\boldsymbol{\theta}}\right)\right. \\ &\qquad \left. + \left(\frac{\partial\boldsymbol{\eta}_d'}{\partial\boldsymbol{\theta}}\mathbf{V}_s(\mathbf{b}_{di} + \mathbf{b}_{dj}) + \frac{\partial\alpha_d}{\partial\boldsymbol{\theta}}\right)'\mathcal{F}^{-1}\left(\frac{\partial\boldsymbol{\eta}_d'}{\partial\boldsymbol{\theta}}\mathbf{V}_s(\mathbf{b}_{di} + \mathbf{b}_{dj}) + \frac{\partial\alpha_d}{\partial\boldsymbol{\theta}}\right)\right\} \\ &\qquad + o(D^{-1}). \end{aligned} \tag{9.22}$$

Note that by (9.13), we can express $\mathbf{b}_{di}$ in terms of $\boldsymbol{\eta}_d$ as follows:

$$\mathbf{b}_{di} = \boldsymbol{\eta}_d + \mathbf{V}_s^{-1}\mathbf{X}_s\mathbf{Q}_s(\mathbf{x}_{di} - \mathbf{X}_s'\boldsymbol{\eta}_d), \tag{9.23}$$

but $\|\mathbf{Z}_s\| = O(1)$ by assumption (H1). Moreover, $|\mathbf{m}_d| = 1$. Using Lemma A.1(ii), we get

$$|\boldsymbol{\eta}_d| = \sigma_u^2|\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{m}_d| \le \sigma_u^2\|\mathbf{V}_s^{-1}\|\|\mathbf{Z}_s\||\mathbf{m}_d| = O(1). \tag{9.24}$$

Now observe that by Lemma A.1(iii), we have

$$\|\mathbf{V}_s^{-1/2}\mathbf{X}_s\mathbf{Q}_s\| = \lambda_{\max}^{1/2}(\mathbf{Q}_s\mathbf{X}_s\mathbf{V}_s^{-1}\mathbf{X}_s\mathbf{Q}_s) = \lambda_{\max}^{1/2}(\mathbf{Q}_s) = O(D^{-1/2}).$$

Since $\mathbf{X}_s'\boldsymbol{\eta}_d = \mathbf{X}_{ds}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}$, which has bounded norm, and $|\mathbf{x}_{di} - \mathbf{X}_s'\boldsymbol{\eta}_d| \le |\mathbf{x}_{di}| + |\mathbf{X}_s'\boldsymbol{\eta}_d|$, by assumptions (H1)–(H3), we have

$$\begin{aligned} |\mathbf{V}_s^{-1}\mathbf{X}_s\mathbf{Q}_s(\mathbf{x}_{di} - \mathbf{X}_s'\boldsymbol{\eta}_d)| &\le \|\mathbf{V}_s^{-1/2}\|\|\mathbf{V}_s^{-1/2}\mathbf{X}_s\mathbf{Q}_s\||\mathbf{x}_{di} - \mathbf{X}_s'\boldsymbol{\eta}_d| \\ &= O(D^{-1/2}). \end{aligned} \tag{9.25}$$

From (9.23), (9.24) and (9.25), we have obtained

$$\mathbf{b}_{di} = \boldsymbol{\eta}_d + \mathbf{f}_{di}, \qquad |\boldsymbol{\eta}_d| = O(1), \qquad |\mathbf{f}_{di}| = O(D^{-1/2}). \tag{9.26}$$

Note also that $|\partial\boldsymbol{\eta}_d/\partial\theta_h| = O(1)$, since $\partial\boldsymbol{\eta}_d/\partial\theta_h = \mathbf{V}_s^{-1}\{(\partial\sigma_u^2/\partial\theta_h)\mathbf{I}_n - \boldsymbol{\Delta}_h\mathbf{V}_s^{-1}\}\mathbf{Z}_s\mathbf{m}_d$, $h = 1, 2$. This implies $\|\partial\boldsymbol{\eta}_d/\partial\boldsymbol{\theta}\| = O(1)$, because

$$\left\|\frac{\partial\boldsymbol{\eta}_d}{\partial\boldsymbol{\theta}}\right\| \le \left\|\frac{\partial\boldsymbol{\eta}_d}{\partial\boldsymbol{\theta}}\right\|_2 = \mathrm{tr}^{1/2}\left\{\left(\frac{\partial\boldsymbol{\eta}_d}{\partial\boldsymbol{\theta}}\right)'\frac{\partial\boldsymbol{\eta}_d}{\partial\boldsymbol{\theta}}\right\} = \left(\sum_{h=1}^2\left|\frac{\partial\boldsymbol{\eta}_d}{\partial\theta_h}\right|^2\right)^{1/2} \le 2^{1/2}\max_{h\in\{1,2\}}\left|\frac{\partial\boldsymbol{\eta}_d}{\partial\theta_h}\right|.$$

By (9.23) and (9.25), we get for any i

$$\mathcal{F}^{-1}\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s\mathbf{b}_{di} = \mathcal{F}^{-1}\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s\boldsymbol{\eta}_d + \boldsymbol{\kappa}_{di}, \qquad |\boldsymbol{\kappa}_{di}| = o(D^{-1}). \tag{9.27}$$

Using repeatedly (9.27) we obtain

$$\begin{aligned}&\left\{\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s(\mathbf{b}_{di}+\mathbf{b}_{dj}) + \frac{\partial \alpha_d}{\partial \boldsymbol{\theta}}\right\}'\mathcal{F}^{-1}\left\{\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s(\mathbf{b}_{di}+\mathbf{b}_{dj}) + \frac{\partial \alpha_d}{\partial \boldsymbol{\theta}}\right\}\\ &\quad = \left(2\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s\boldsymbol{\eta}_d + \frac{\partial \alpha_d}{\partial \boldsymbol{\theta}}\right)'\mathcal{F}^{-1}\left(2\frac{\partial \boldsymbol{\eta}_d'}{\partial \boldsymbol{\theta}}\mathbf{V}_s\boldsymbol{\eta}_d + \frac{\partial \alpha_d}{\partial \boldsymbol{\theta}}\right) + o(D^{-1})\end{aligned}$$

and using (9.18), we obtain

$$\exp\{2\alpha_d + (\mathbf{x}_{di}+\mathbf{x}_{dj})'\boldsymbol{\beta} + (\mathbf{b}_{di}+\mathbf{b}_{dj})'\mathbf{V}_s(\mathbf{b}_{di}/2+\mathbf{b}_{dj})\} = E_{dij}. \tag{9.28}$$

Replacing (9.28) in (9.22) and then (9.22) in (9.21), we get the desired result. □

PROOF OF THEOREM 5.2. Again, we show the result for the ML estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, because for REML the proof is analogous but simpler. The proof is based on the following chain of results:

(A) For every $\nu \in (0, 1)$, there exists a subset of the sample space $\mathcal{B}$ on which, for large D, it holds

$$\begin{aligned}\hat{w}_{di}^E - \hat{w}_{di} &= \mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{s} + \mathbf{h}_{di}'\mathcal{F}^{-1}(H+\mathcal{F})\mathcal{F}^{-1}\mathbf{s}\\ &\quad + \frac{1}{2}\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{d} + \frac{1}{2}\mathbf{s}'\mathcal{F}^{-1}S_{di}\mathcal{F}^{-1}\mathbf{s} + \mathbf{r}_{di},\end{aligned}$$

where $\mathbf{h}_{di} = \partial\hat{w}_{di}/\partial\boldsymbol{\theta}$, $S_{di} = \partial^2\hat{w}_{di}/\partial\boldsymbol{\theta}^2$, $\mathbf{d} = (d_1, d_2)'$, with $d_h = \mathbf{s}'\mathcal{F}^{-1}D_h\mathcal{F}^{-1}\mathbf{s}$, $D_h = \partial H/\partial\theta_h$, $h = 1, 2$, and the remainder term $\mathbf{r}_{di}$ satisfies $|\mathbf{r}_{di}| < D^{-3\nu/2}w$, for a random variable w with bounded first and second moments.

(B) If $1_{\mathcal{B}}$ is the indicator function of the set $\mathcal{B}$, it holds that

$$\begin{aligned}&E\{(\hat{w}_{di}^E - \hat{w}_{di})(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}}\}\\ &\quad = E\{\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}}\}\\ &\qquad + E\{\mathbf{h}_{di}'\mathcal{F}^{-1}(H+\mathcal{F})\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}}\}\\ &\qquad + E\left\{\frac{1}{2}\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{d}(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}}\right\}\\ &\qquad + E\left\{\frac{1}{2}\mathbf{s}'\mathcal{F}^{-1}S_{di}\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}}\right\} + o(D^{-1}).\end{aligned} \tag{9.29}$$

(C) $E\{(\hat{w}_{di}^E - \hat{w}_{di})(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}^c}\} = o(D^{-1})$.

(D) It holds that

$$
\begin{aligned}
&E\{\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})\}+E\{\mathbf{h}_{di}'\mathcal{F}^{-1}(H+\mathcal{F})\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})\}\\
&\qquad+E\left\{\frac{1}{2}\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{d}(\hat{w}_{dj}-w_{dj})\right\}\\
&\qquad+E\left\{\frac{1}{2}\mathbf{s}'\mathcal{F}^{-1}S_{di}\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})\right\}\\
&\quad=M_{3d,ij}(\boldsymbol{\beta},\boldsymbol{\theta})+o(D^{-1}).
\end{aligned}
\tag{9.30}
$$

(E) It holds that

$$
\begin{aligned}
E\{\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}^c}\}&=o(D^{-1}),\\
E\{\mathbf{h}_{di}'\mathcal{F}^{-1}(H+\mathcal{F})\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}^c}\}&=o(D^{-1}),\\
E\left\{\frac{1}{2}\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{d}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}^c}\right\}&=o(D^{-1}),\\
E\left\{\frac{1}{2}\mathbf{s}'\mathcal{F}^{-1}S_{di}\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}^c}\right\}&=o(D^{-1}).
\end{aligned}
$$

Applying in turn (C) and (B), we obtain

$$
\begin{aligned}
&E\{(\hat{w}_{di}^E-\hat{w}_{di})(\hat{w}_{dj}-w_{dj})\}\\
&\quad=E\{\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}}\}\\
&\qquad+E\{\mathbf{h}_{di}'\mathcal{F}^{-1}(H+\mathcal{F})\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}}\}\\
&\qquad+E\left\{\frac{1}{2}\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{d}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}}\right\}\\
&\qquad+E\left\{\frac{1}{2}\mathbf{s}'\mathcal{F}^{-1}S_{di}\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}}\right\}+o(D^{-1}).
\end{aligned}
$$

Finally, writing $1_{\mathcal{B}}=1-1_{\mathcal{B}^c}$ and applying (E) and (D), we obtain

$$
E\{(\hat{w}_{di}^E-\hat{w}_{di})(\hat{w}_{dj}-w_{dj})\}=M_{3d,ij}(\boldsymbol{\beta},\boldsymbol{\theta})+o(D^{-1}).
$$

Next, we give the proofs of results (A)–(C).

Proof of (A): It is obtained by applying Lemma 3 of [12] to $\hat{w}_{di}^E=\hat{w}_{di}(\boldsymbol{\theta})$, where $\hat{\boldsymbol{\theta}}$ is the ML estimator of $\boldsymbol{\theta}$.

Proof of (B): Applying (A) we obtain

$$
\begin{aligned}
&E\{(\hat{w}_{di}^E-\hat{w}_{di})(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}}\}\\
&\quad=E\{\mathbf{h}_{di}'\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}}\}\\
&\qquad+E\{\mathbf{h}_{di}'\mathcal{F}^{-1}(H+\mathcal{F})\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj}-w_{dj})1_{\mathcal{B}}\}
\end{aligned}
$$

$$
\begin{aligned}
&+ E\left\{\frac{1}{2}\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{d}(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}}\right\} \\
&+ E\left\{\frac{1}{2}\mathbf{s}'\mathcal{F}^{-1}S_{di}\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}}\right\} + E\{\mathbf{r}_{di}(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}}\}.
\end{aligned}
$$

But by Theorem 4.2, we know that $\mathrm{MSE}(\hat{w}_{dj}) = O(1)$ as D tends to infinity. Then, applying Hölder's inequality and taking $\nu \in (2/3, 1)$, we obtain

$$
\begin{aligned}
E\{\mathbf{r}_{di}(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}}\} &\le E^{1/2}(\mathbf{r}^2_{di}1_{\mathcal{B}})E^{1/2}\{(\hat{w}_{dj} - w_{dj})^2\} \\
&< D^{-3\nu/2}E^{1/2}(w^2)\{\mathrm{MSE}(\hat{w}_{dj})\}^{1/2} \\
&= o(D^{-1}).
\end{aligned}
\tag{9.31}
$$

Proof of (C): Noting that $\hat{w}^E_{di} = \exp(\hat{y}^E_{di} + \hat{\alpha}_d)$, for $\hat{y}^E_{di} = \hat{y}_{di}(\hat{\boldsymbol{\theta}})$ and $\hat{\alpha}_d = \alpha_d(\hat{\boldsymbol{\theta}})$, we have

$$
\begin{aligned}
&E\{(\hat{w}^E_{di} - \hat{w}_{di})(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}^c}\} \\
&\quad = E[\{\exp(\hat{y}^E_{di} + \hat{\alpha}_d) - \exp(\hat{y}_{di} + \alpha_d)\} \\
&\qquad \times \{\exp(\hat{y}_{dj} + \alpha_d) - \exp(y_{dj})\}1_{\mathcal{B}^c}] \\
&\quad \le E[\exp(\hat{y}^E_{di} + \hat{y}_{dj} + \hat{\alpha}_d + \alpha_d)1_{\mathcal{B}^c}] \\
&\qquad + E[\exp(\hat{y}_{di} + y_{dj} + \alpha_d)1_{\mathcal{B}^c}].
\end{aligned}
\tag{9.32}
$$

For $\nu \in (0, 1)$, we define the neighborhood $N(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} \in \Theta : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| < D^{-\nu/2}\}$. Using (9.13) and applying Hölder's inequality, the first expectation on the right-hand side of (9.32) can be bounded as

$$
\begin{aligned}
&E[\exp(\hat{y}^E_{di} + \hat{y}_{dj} + \hat{\alpha}_d + \alpha_d)1_{\mathcal{B}^c}] \\
&\quad \le \exp\Big\{2 \sup_{N(\boldsymbol{\theta}_0)} \alpha_d(\boldsymbol{\theta})\Big\} \\
&\qquad \times E\Big[\exp\Big\{\sup_{N(\boldsymbol{\theta}_0)} (\mathbf{b}_{di}(\boldsymbol{\theta}) + \mathbf{b}_{dj}(\boldsymbol{\theta}))'\mathbf{y}_s\Big\}1_{\mathcal{B}^c}\Big] \\
&\quad \le \exp\Big\{2 \sup_{N(\boldsymbol{\theta}_0)} \alpha_d(\boldsymbol{\theta})\Big\}E^{1/2}\Big[\exp\Big\{2 \sup_{N(\boldsymbol{\theta}_0)} (\mathbf{b}_{di}(\boldsymbol{\theta}) + \mathbf{b}_{dj}(\boldsymbol{\theta}))'\mathbf{y}_s\Big\}\Big]P^{1/2}(\mathcal{B}^c).
\end{aligned}
$$

But the suprema of $|\alpha_d(\boldsymbol{\theta})|$ and $|\mathbf{b}_{di}(\boldsymbol{\theta})|$ over $N(\boldsymbol{\theta}_0)$ are bounded. Moreover, since $\mathbf{y}_s$ is normally distributed, the expected value on the right-hand side of the inequality is bounded. Now by Lemma 1 of [12] with $\nu = \eta \in (0, 3/4)$ and $b > 16$, we get $P^{1/2}(\mathcal{B}^c) = O(D^{-b/16}) = o(D^{-1})$. Therefore,

$$
E[\exp(\hat{y}^E_{di} + \hat{y}_{dj} + \hat{\alpha}_d + \alpha_d)1_{\mathcal{B}^c}] = o(D^{-1}).
\tag{9.33}
$$

Similarly, writing $y_{dj} = \mathbf{a}'_{dj}\mathbf{y}_s$, we have

$$\begin{aligned}
&E\bigl[\exp(\hat{y}_{di} + y_{dj} + \alpha_d)1_{\mathcal{B}^c}\bigr] \\
&\quad \le \exp(\alpha_d)E^{1/2}\Bigl[\exp\Bigl\{\Bigl(\sup_{N(\boldsymbol{\theta}_0)} \mathbf{b}_{di}(\boldsymbol{\theta}) + \mathbf{a}_{dj}\Bigr)'\mathbf{y}_s\Bigr\}\Bigr]P^{1/2}(\mathcal{B}^c) \qquad (9.34)\\
&\quad = o(D^{-1}).
\end{aligned}$$

Replacing (9.33) and (9.34) in (9.32), we obtain $E\{(\hat{w}^E_{di} - \hat{w}_{di})(\hat{w}_{dj} - w_{dj})1_{\mathcal{B}^c}\} = o(D^{-1})$.

Proof of (D): Consider the first term on the left-hand side of (9.30), given by

$$E\{\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj} - w_{dj})\} = E(\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}\hat{w}_{dj}) - E(\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}w_{dj}).$$

Using $\hat{w}_{di} = \exp(\delta_{di})$, for $\delta_{di} = \alpha_d + \mathbf{x}'_{di}\boldsymbol{\beta} + \mathbf{b}'_{di}\mathbf{v}_s$, and taking into account that

$$\mathbf{h}_{di} = \partial\hat{w}_{di}/\partial\boldsymbol{\theta} = \exp(\delta_{di})\partial\delta_{di}/\partial\boldsymbol{\theta}, \qquad (9.35)$$

we obtain

$$E(\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}\hat{w}_{dj}) = \exp(2\alpha_d + \mathbf{x}'_{dij}\boldsymbol{\beta})E\{\exp(\mathbf{b}'_{dij}\mathbf{v}_s)(\partial\delta_{di}/\partial\boldsymbol{\theta})'\mathcal{F}^{-1}\mathbf{s}\}, \qquad (9.36)$$

where $\mathbf{b}_{dij} = \mathbf{b}_{di} + \mathbf{b}_{dj} = 2\boldsymbol{\eta}_d + \mathbf{f}_{di} + \mathbf{f}_{dj}$, with $|\boldsymbol{\eta}_d| = O(1)$ and $|\mathbf{f}_{di}| = O(D^{-1/2})$ by (9.26).

To calculate the expected value in (9.36), note that $\delta_{di} = \alpha_d + \mathbf{x}'_{di}\boldsymbol{\beta} + \mathbf{b}'_{di}\mathbf{v}_s$ and define

$$\mathbf{g}_d = \mathcal{F}^{-1}\frac{\partial\alpha_d}{\partial\boldsymbol{\theta}} = (g_{d1}, g_{d2})', \qquad C_{di} = \mathcal{F}^{-1}\frac{\partial\mathbf{b}'_{di}}{\partial\boldsymbol{\theta}} = (\mathbf{c}_{di1}, \mathbf{c}_{di2})'. \qquad (9.37)$$

Then we can write

$$\mathcal{F}^{-1}\frac{\partial\delta_{di}}{\partial\boldsymbol{\theta}} = \mathcal{F}^{-1}\frac{\partial\alpha_d}{\partial\boldsymbol{\theta}} + \mathcal{F}^{-1}\frac{\partial\mathbf{b}'_{di}}{\partial\boldsymbol{\theta}}\mathbf{v}_s = \mathbf{g}_d + C_{di}\mathbf{v}_s. \qquad (9.38)$$

Moreover, denoting $\mathbf{A}_h = \mathbf{P}_s\boldsymbol{\Delta}_h\mathbf{P}_s$, $q_h = \mathbf{v}'_s\mathbf{A}_h\mathbf{v}_s$, $h = 1, 2$ and $\mathbf{q} = (q_1, q_2)'$, the vector of scores (3.5) can be expressed as

$$\begin{aligned}
\mathbf{s} &= (\mathbf{q} - E\mathbf{q})/2 + \boldsymbol{\nu}, \\
\boldsymbol{\nu} &= (\nu_1, \nu_2)', \qquad (9.39)\\
\nu_h &= \{\mathrm{tr}(\mathbf{P}_s\boldsymbol{\Delta}_h) - \mathrm{tr}(\mathbf{V}_s^{-1}\boldsymbol{\Delta}_h)\}/2.
\end{aligned}$$

Using these expressions, we get

$$\begin{aligned}
&E\biggl\{\exp(\mathbf{b}'_{dij}\mathbf{v}_s)\biggl(\frac{\partial\delta_{di}}{\partial\boldsymbol{\theta}}\biggr)'\mathcal{F}^{-1}\mathbf{s}\biggr\} \\
&\quad = \frac{1}{2}\mathbf{g}'_d E\{\exp(\mathbf{b}'_{dij}\mathbf{v}_s)(\mathbf{q} - E\mathbf{q})\} \\
&\qquad + \mathbf{g}'_d\boldsymbol{\nu}E\{\exp(\mathbf{b}'_{dij}\mathbf{v}_s)\} + \frac{1}{2}E\{\exp(\mathbf{b}'_{dij}\mathbf{v}_s)\mathbf{v}'_s C'_{di}(\mathbf{q} - E\mathbf{q})\} \\
&\qquad + E\{\exp(\mathbf{b}'_{dij}\mathbf{v}_s)\mathbf{v}'_s C'_{di}\}\boldsymbol{\nu}.
\end{aligned}$$

Using repeatedly Lemma 5(iv) of [12], we obtain

$$E(\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}\hat{w}_{dj}) = E_{dij}\biggl\{\operatorname{tr}\biggl(\mathcal{F}^{-1}\frac{\partial\boldsymbol{\eta}'_d}{\partial\boldsymbol{\theta}}\mathbf{E}_{dj}\biggr) + \frac{1}{2}\biggl(\frac{\partial\alpha_d}{\partial\boldsymbol{\theta}} + 2\frac{\partial\boldsymbol{\eta}'_d}{\partial\boldsymbol{\theta}}\mathbf{V}_s\boldsymbol{\eta}_{dj}\biggr)'\mathcal{F}^{-1}(2\boldsymbol{\nu} + \boldsymbol{\varepsilon}_{dj})\biggr\}. \tag{9.40}$$

For the expected value $E(\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}w_{dj})$, note that $w_{dj} = \exp(y_{dj})$, where $y_{dj} = \mathbf{x}'_{dj}\boldsymbol{\beta} + v_{dj}$, for $v_{dj} = u_d + e_{dj}$. However, since $j \in \bar{s}_d$, we cannot express y_{dj} in terms of $\mathbf{v}_s$ as done above. In this case, we construct an extended vector $\mathbf{v}^*_{sj} = (\mathbf{v}'_s, v_{dj})' \sim N(\mathbf{0}_2, \mathbf{V}^*_s)$, for

$$\mathbf{V}^*_s = \begin{pmatrix} \mathbf{V}_s & \sigma_u^2\mathbf{z}_d \\ \sigma_u^2\mathbf{z}'_d & \sigma_u^2 + \sigma_e^2 \end{pmatrix},$$

where $\mathbf{z}_d = \mathbf{Z}_s\mathbf{m}_d$. Defining also $\mathbf{b}^*_{di} = (\mathbf{b}'_{di}, 1)'$, we can express

$$E(\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}w_{dj}) = \exp(\alpha_d + \mathbf{x}'_{dij}\boldsymbol{\beta})E\biggl[\exp\{(\mathbf{b}^*_{di})'\mathbf{v}^*_{sj}\}\biggl(\frac{\partial\delta_{di}}{\partial\boldsymbol{\theta}}\biggr)'\mathcal{F}^{-1}\mathbf{s}\biggr].$$

Writing now $\mathcal{F}^{-1}(\partial\delta_{di}/\partial\boldsymbol{\theta})$ and $\mathbf{s}$ in terms of $\mathbf{v}^*_{sj}$ similarly as in (9.38) and (9.39) by adding zero elements to the vectors and matrices multiplying $\mathbf{v}_s$, we can apply the same results as used for $E(\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}\hat{w}_{dj})$, obtaining (9.40) with E_{dij} replaced by E^*_{dij}.

The rest of terms on the left-hand side of (9.30) are obtained following a similar procedure, by expressing the terms within the expectations as sums of products of quadratic and linear forms in $\mathbf{v}_s$ multiplied by exponentials of linear forms of $\mathbf{v}_s$ and then applying repeatedly Lemma 5 of [12].

Proof of (E): Note that

$$E\{|\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj} - w_{dj})|1_{\mathcal{B}^c}\} \le E\{|\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}\hat{w}_{dj}|1_{\mathcal{B}^c}\} + E\{|\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}w_{dj}|1_{\mathcal{B}^c}\}. \tag{9.41}$$

By the definition of $\mathbf{h}_{di}$ in (9.35) and that of $\hat{w}_{dj}$ in (9.12), we obtain

$$E\{|\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}\hat{w}_{dj}|1_{\mathcal{B}^c}\} = \exp(2\alpha_d + \mathbf{x}'_{dij}\boldsymbol{\beta})E\biggl[\exp\{(\mathbf{b}_{di} + \mathbf{b}_{dj})'\mathbf{v}_s\}\biggl|\biggl(\frac{\partial\delta_{di}}{\partial\boldsymbol{\theta}}\biggr)'\mathcal{F}^{-1}\mathbf{s}\biggr|1_{\mathcal{B}^c}\biggr].$$

Now applying repeatedly Hölder's inequality, we get

$$\begin{aligned} E\{|\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}\hat{w}_{dj}|1_{\mathcal{B}^c}\} &\le \exp(2\alpha_d + \mathbf{x}'_{dij}\boldsymbol{\beta})E^{1/2}[\exp\{2(\mathbf{b}_{di} + \mathbf{b}_{dj})'\mathbf{v}_s\}] \\ &\quad \times E^{1/8}\biggl|\frac{\partial\delta_{di}}{\partial\boldsymbol{\theta}}\biggr|^8 E^{1/8}|\mathcal{F}^{-1}\mathbf{s}|^8 P^{1/4}(\mathcal{B}^c) \\ &= O(D^{-1/2-b/32}) = o(D^{-1}) \end{aligned} \tag{9.42}$$

for $b > 16$, noting that by the proof of Theorem 1 in [12], it holds

$$E^{1/8}\left|\frac{\partial\delta_{di}}{\partial\boldsymbol{\theta}}\right|^8 = O(1), \qquad E^{1/8}\left|\mathcal{F}^{-1}\mathbf{s}\right|^8 = O\left(D^{-1/2}\right), \tag{9.43}$$

that $P^{1/4}(\mathcal{B}^c) = O(D^{-b/32})$, by Lemma 1 in [12] with $\nu = \eta \in (0, 3/4)$, and finally taking into account that $\mathbf{v}_s$ is normally distributed and that $\exp(2\alpha_d + \mathbf{x}'_{dij}\boldsymbol{\beta})$ and $\mathbf{b}_{di}$ are bounded. By a similar reasoning, we obtain

$$\begin{aligned} &E\left\{\left|\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}w_{dj}\right|1_{\mathcal{B}^c}\right\} \\ &\quad \le \exp(\alpha_d + \mathbf{x}'_{dij}\boldsymbol{\beta})E\left[\exp\left\{\left(\mathbf{b}^*_{di}\right)'\mathbf{v}^*_{sj}\right\}\left|\left(\frac{\partial\delta_{di}}{\partial\boldsymbol{\theta}}\right)'\mathcal{F}^{-1}\mathbf{s}\right|1_{\mathcal{B}^c}\right] \\ &\quad = o\left(D^{-1}\right). \end{aligned} \tag{9.44}$$

By (9.44) and (9.42), we obtain $E\{|\mathbf{h}'_{di}\mathcal{F}^{-1}\mathbf{s}(\hat{w}_{dj} - w_{dj})|1_{\mathcal{B}^c}\} = o(D^{-1})$. The remaining results in (E) are proved similarly. $\square$

Acknowledgments. Part of this work was done during a research stay of the second author in the Institute of Statistics of the University of Neuchâtel.

SUPPLEMENTARY MATERIAL

Supplement to "Empirical best prediction under a nested error model with log transformation" (DOI: 10.1214/17-AOS1608SUPP; .pdf). This document contains results on the bias of the proposed and existing predictors, simulation results for prediction at the individual level, on the performance of the bootstrap MSE estimator compared with the analytical estimator, and additional results on the application with Mexican data.

REFERENCES

[1] BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.

[2] BUTAR, F. B. and LAHIRI, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *J. Statist. Plann. Inference* **112** 63–76. MR1961720

[3] DAS, K., JIANG, J. and RAO, J. N. K. (2004). Mean squared error of empirical predictor. *Ann. Statist.* **32** 818–840. MR2060179

[4] ELBERS, C., LANJOUW, J. O. and LANJOUW, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica* **71** 355–364.

[5] FAY, R. E. and HERRIOT, R. A. (1979). Estimation of income from small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277.

[6] GONZÁLEZ-MANTEIGA, W., LOMBARDÍA, M. J., MOLINA, I., MORALES, D. and SANTAMARÍA, L. (2008). Bootstrap mean squared error of a small-area EBLUP. *J. Stat. Comput. Simul.* **78** 443–462. MR2424556

[7] HALL, P. and MAITI, T. (2006). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *Ann. Statist.* **34** 1733–1750. MR2283715

[8] HALL, P. and MAITI, T. (2006). On parametric bootstrap methods for small area prediction. *J. Roy. Statist. Soc. Ser. B* **68** 221–238.

[9] JIANG, J., LAHIRI, P. and NGUYEN, T. (2016). A unified Monte-Carlo jackknife for small area estimation under model selection. *Ann. Math. Sci. Appl.* To appear.

[10] MAURO, F., MOLINA, I., GARCÍA-ABRIL, A., VALBUENA, R. and AYUGA-TÉLLEZ, E. (2016). Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. *Environmetrics* **27** 225–238.

[11] MILLER, J. J. (1973). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746–762. MR0448661

[12] MOLINA, I. (2009). Uncertainty under a multivariate nested-error regression model with logarithmic transformation. *J. Multivariate Anal.* **100** 963–980.

[13] MOLINA, I. and MARHUENDA, Y. (2015). sae: An R package for small area estimation. *R J.* **7** 81–98.

[14] MOLINA, I. and MARTÍN, N. (2018). Supplement to "Empirical best prediction under a nested error model with log transformation." DOI:10.1214/17-AOS1608SUPP.

[15] MOLINA, I., NANDRAM, B. and RAO, J. N. K. (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *Ann. Appl. Stat.* **8** 852–885. MR3262537

[16] MOLINA, I. and RAO, J. N. K. (2010). Small area estimation of poverty indicators. *Canad. J. Statist.* **38** 369–385.

[17] PFEFFERMANN, D. (2013). New important developments in small area estimation. *Statist. Sci.* **28** 40–68. MR3075338

[18] PFEFFERMANN, D. and TILLER, R. (2005). Bootstrap approximation to prediction MSE for state-space models with estimated parameters. *J. Time Series Anal.* **26** 893–916.

[19] RAO, J. N. K. and MOLINA, I. (2015). *Small Area Estimation*, 2nd ed. Wiley, Hoboken, NJ.

[20] SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. Wiley, New York.

[21] SLUD, E. and MAITI, T. (2006). Mean-squared error estimation in transformed Fay–Herriot models. *J. Roy. Statist. Soc. Ser. B* **68** 239–257. MR2188984

DEPARTMENT OF STATISTICS
UNIVERSIDAD CARLOS III DE MADRID
CALLE MADRID 126
28903 MADRID
SPAIN
E-MAIL: isabel.molina@uc3m.es

DEPARTMENT SECTION OF STATISTICS
AND OPERATIONS RESEARCH II
UNIVERSIDAD COMPLUTENSE DE MADRID
AVENIDA DE FILIPINAS 3
28003 MADRID
SPAIN
E-MAIL: nirian@estad.ucm.es