Bootstrapping Mean Squared Errors of Robust Small Area Estimators*

Valery D. Jiongo ‡ Pierre Nguimkeu ‡

January 2016

Abstract

Robust small-area estimation has received considerable attention in recent years, and the mean-squared error has been the main way in which the estimators performance is measured. This paper proposes a new bootstrap procedure for mean squared errors of robust small area estimators. We formally prove the asymptotic validity of the proposed bootstrap method and examine its finite sample performance through Monte Carlo simulations. The results show that our procedure performs reasonably well and outperforms existing ones. We also provide a real data example to illustrate the usefulness of the proposed bootstrap method in practice.

Keywords: Small-area estimation; Bootstrap; Mean squared error; Linear mixed model

1 Introduction

The growing demand for reliable statistics of small geographical areas and subgroups of populations has generated a lot of interest for small-area estimation (SAE) during the last decades (see Rao 2003, 2005, Pfeffermann 2013, for a recent review of methods used for SAE). Many theoretically sound and yet practical estimation procedures have been proposed of which the empirical bayes (EB) and the empirical best linear unbiased predictor (EBLUP) are used. In particular, the empirical best linear unbiased predictor is asymptotically unbiased and efficient under correct model specification and distributional assumptions. However, it becomes highly sensitive in the presence of outliers or departures from the assumed normality of the random effects in the underlying model (see, e.g., Fellner 1986, Stahel and Welsh 1997, Sinha & Rao 2009).

*The authors would like to thank Jon Rao, Mike Hidiroglou, Jean-François Beaumont, Franois Verret, Wesley Benjamin, Chris Mohl and Wesley Yung for useful comments and suggestions. Valery D. Jiongo is particularly grateful to David Haziza and Pierre Duchesne for their availability and extensive advice. We thank Edouard Djeutem for valuable research assistance.

†Corresponding author. Statistics Canada; Business Survey Methods Division; 100 Tunney’s Pasture Driveway, Ottawa, Ontario K1A 0T6, Canada; Email: Valery.DongmoJiongo@statcan.gc.ca

‡Department of Economics, Georgia State University, PO Box 3992 Atlanta, GA 30302-3992, USA, Email: mnguimkeu@gsu.edu.
Robustified versions of the classical estimators have therefore been recently investigated by several authors to downweight influential observations in the data when estimating the model (Sinha & Rao 2009, Chambers et al. 2013, Jiongo et al. 2013). The robust estimation of these small-area quantities also requires the estimation of the Mean Square Error (MSE) which provides a measure of the precision of the point estimators. Sinha & Rao (2009) proposed a parametric bootstrap procedure based on the robust EBLUP estimators of the underlying linear mixed model to estimate the MSE. But Jiongo et al. (2013) point out that the use of robust variance estimates to generate bootstrap replicas leads to bootstrap samples whose variability are significantly smaller than the variability in the original data. Other analytical and bootstrap procedures for the MSE of the robust empirical best linear unbiased predictors (REBLUPs) have been proposed in Chambers et al. (2013) and Jiongo et al. (2013), respectively. However, their theoretical validity has not been formally established and their empirical performance are not fully satisfactory (as evidenced by the simulations results in Section 4 below).

In this paper, we propose a new semiparametric bootstrap procedure for estimating the mean-squared error of robust small area estimators. We focus on small area methods based on unit-level models that linearly relate the small area quantities of inferential interest to some unit level auxiliary covariates and includes random effects associated with the small-areas. Since robust estimates of the variance components are typically smaller than their nonrobust counterparts and could yield bootstrap data on a smaller scale than the original data (Field et al. 2010), our bootstrap procedure uses (non-robust) maximum likelihood estimators to generate bootstrap samples, and robust bootstrap predictors to estimate the MSE. This produces bootstrap samples whose variability is similar to the one associated with the original data, and the resulting MSE estimator therefore has improved coverage rates. We formally prove the theoretical validity of our bootstrap procedure, examine its empirical performance through simulations and illustrate its use via a real data application. Our new estimator is attractive for several reasons: First, it does not require specific distributional assumptions about error distributions. It is therefore unaffected by misspecification biases that could arise in case of the not so uncommon non-normality. Second, since our bootstrap samples mimic the original data, the use of the same estimation procedure for robust bootstrap predictors leads to a consistent MSE estimator. Third, the proposed procedure is easy to implement and our simulations results show that it performs quite well in finite samples and outperforms existing ones.

The validity of our bootstrap procedure uses an approach similar to the one used by Bickel & Friedman (1981) and Freedman (1981). Our procedure is an extension of the Freedman (1981) methodology to the MSE estimation of robust small-area estimators in the linear mixed model framework. To our knowledge, this is the first study that provides sufficient conditions and a rigorous proof of the convergence of a MSE estimator of robust small-area estimators. Although our proofs and procedure are based on the Sinha & Rao (2009) robust estimator, the derivation can be easily adapted to other existing robust predictors. A Monte Carlo simulation study computes the relative biases and relative root mean squared error rates of the proposed bootstrap MSE estimator and compares it to several analytical and bootstrap existing alternatives, including the bootstrap MSE estimator of Sinha & Rao (2009), the analytical pseudolinearization MSE estimator and linearization-based MSE estimators of Chambers et al. (2013), the bootstrap MSE of Jiongo et al. (2013) and the MSE estimator of Prasad & Rao (1990). This comparison
is provided for different robust small-area point estimators and various modes of data contamination.

The paper is organized as follows. Section 2 presents the model, notation and reviews some existing results. In Section 3, we present our proposed bootstrap procedure. Asymptotic properties and validity of the proposed method are also discussed. The validity proof of our bootstrap MSE proceeds in two main steps. Lemma 1 provides the asymptotic properties of the robust estimators of the main model, while Lemma 3 provides the requirements for the asymptotic validity of our bootstrap procedure. Our main result is given in Theorem 2. Section 4 presents Monte Carlo simulation results showing that our procedure has satisfactory finite sample properties, and its performance is compared with the above-mentioned alternative estimators. A real data example on county crop areas is provided in Section 5 to show the usefulness of our method in practice. Concluding remarks are given in Section 6, followed by a technical appendix in Section 7.

2 Preliminaries

This section presents the basic linear mixed model that linearly relates the small area quantities of inferential interest to some unit level auxiliary covariates and includes random effects associated to the areas. We then briefly discuss the Sinha & Rao (2009) robust estimator that is used to construct our bootstrap MSE estimation.

2.1 Underlying Model

Consider a population $\mathcal{U}$ of size $N$, partitioned into $k$ domains (areas) $\mathcal{U}_1, \ldots, \mathcal{U}_k$, of known sizes $N_1, \ldots, N_k$, respectively; that is, $\mathcal{U} = \bigcup_{i=1}^{k} \mathcal{U}_i$ such that $\mathcal{U}_i \cap \mathcal{U}_l = \emptyset$, $i \neq l$, and $N = \sum_{i=1}^{k} N_i$. Let $y$ define the variable of interest, and denote by $y_{ij}$ the response value for unit $j$ belonging to area $i$, $i = 1, \ldots, k$, $j = 1, \ldots, N_i$. The area mean associated with $\mathcal{U}_i$ is given by

$$\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}, \quad (i = 1, \ldots, k),$$

Let $s$ be the sample of size $n$, selected from the population $\mathcal{U}$ according to a given sampling plan $\mathcal{P}(s)$. The overall sample $s$ can be partitioned as $s = \bigcup_{i=1}^{k} s_i$, where $s_i = s \cap \mathcal{U}_i$, of size $n_i$ is the sample observed for sampled area $i$, $n = \sum_{i=1}^{k} n_i$. Note that $n_i$ is random unless a planned sample of fixed size is taken in that area.

Traditional area-specific direct estimation methods (design-based or model-based) are not suitable in the small area context because of small (or even zero) area-specific sample sizes $n_i$. As a result, indirect estimation methods that borrow information across related areas through explicit models and auxiliary information, such as census and administrative data, are used for small area estimation. Denote by $x_{ij} = (x_{1ij}, \ldots, x_{p_{ij}})^\top$ a $p$-dimensional deterministic vector of covariate values available for unit $(i, j)$ and by $\bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$ the column vector of sample means of these covariates for area $i$. The corresponding vector of true area means is given by $\bar{X}_i = N_i^{-1} \sum_{j=1}^{N_i} x_{ij}$, $i = 1, \ldots, k$, and is assumed to be available as well.
The nested error unit-level model considered can be expressed as

\[ y_{ij} = x_{ij}^T \beta + v_i + e_{ij}, \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, n_i, \tag{1} \]

where \( \beta \) is an unknown \( p \)-dimensional fixed-effects regression parameter vector. The random-effects \( v_i \) are independently and identically distributed with a known distribution \( F_v \) of mean 0 and variance \( \sigma_v^2 \). The error terms \( e_{ij} \) are assumed to be independent with a distribution \( F_e \) of mean 0 and variance \( \sigma_e^2 \), and independent to the \( v_i \). We suppose that \( F_v \) and \( F_e \) both belong to the same family of stable distributions \( F \). The regressor \( x_{ij} \) is a \( p \)-dimensional vector of observed responses. Model (1) can be rewritten as

\[ y_i = X_i \beta + v_i 1_{n_i} + e_i, \quad i = 1, \ldots, k, \tag{2} \]

where \( i \) is the area index, \( y_i \) is an \( n_i \)-dimensional vector of observed responses, \( X_i \) is a known \( n_i \times p \) full-rank design matrix, and \( 1_{n_i} \) is an \( n_i \)-vector of ones. Denote by \( \theta = (\beta^T, \delta^T)^T \), the vector of model parameters, where \( \delta \) is the vector of variances parameters \( \delta = (\sigma_v^2, \sigma_e^2)^T \). The variance-covariance matrix of \( y_i \) is given by \( V_i = \sigma_e^2 I_{n_i} + \sigma_v^2 1_{n_i} 1_{n_i}^T \), where \( I_{n_i} \) is the identity matrix of order \( n_i \). The random-effect component, \( v_i \), accounts for the between-area variations that are not explained by the available auxiliary information \( X_i \).

### 2.2 Robust Estimation

For our bootstrap MSE procedure, we consider the class of robust estimators \( \hat{\theta}_R \) that are solutions to the following estimating equation:

\[ S(y, X, \theta) = \sum_{i=1}^{k} \Psi(y_i, X_i, \theta) = 0, \tag{3} \]

where \( \Psi(y_i, X_i, \theta) = (\Psi_1(y_i, X_i, \theta)^T, \Psi_2(y_i, X_i, \theta)^T)^T \); \( \Psi_1 \) is a \( p \)-dimensional vector of estimating functions associated to the regression parameter \( \beta \) and \( \Psi_2 = (\Psi_{21}^T, \Psi_{22}^T)^T \) is a \( 2 \)-dimensional vector of estimating functions associated to the variance parameters \( \delta = (\sigma_v^2, \sigma_e^2) \). This class of estimators includes robust maximum likelihood estimators developed by Sinha & Rao (2009) for which

\[
\begin{align*}
\Psi_1(y_i, X_i, \theta) & = X_i^T V_i^{-1} U_i^{1/2} \Psi_b(r_i) \\
\Psi_2(y_i, X_i, \theta) & = \Psi_b(r_i) U_i^{1/2} V_i^{-1} \frac{\partial V_i}{\partial \delta_l} V_i^{-1/2} \Psi_b(r_i) - \text{tr} \left( K_i V_i^{-1} \frac{\partial V_i}{\partial \delta_l} \right),
\end{align*}
\tag{4,5}
\]

where \( l = 1, 2 \), \( r_i = U_i^{-1/2} (y_i - X_i \beta) \); \( \Psi_b(r_i) = (\psi_b(r_{i1}), \ldots, \psi_b(r_{in}))^T \) is an \( n_i \)-vector of bounded functions, \( U_i = \text{diag}(V_i) \) is a diagonal matrix whose elements are the diagonal elements of the matrix \( V_i \), and \( K_i = E \{ \psi_b^2(r) \} I_{n_i} \) where \( r \) has the standard normal distribution \( r \sim \mathcal{N}(0, 1) \). An example of function \( \psi_b \) is the Huber-type function defined by\footnote{That is, the sum of two independent copies has the same distribution, up to location and scale parameters. Examples include normal, Cauchy, Levy distributions, etc.}

\[ \psi_b(u) = \min\{|b|, \max(-|b|, u)\}, \tag{6} \]
where $b$ is a user-chosen positive constant. In a classical robust estimation framework under normality, a popular choice of the tuning constant $b$ dictated by efficiency considerations is $b = 1.345$. Smoother versions of these functions can also be used as desired. Note that the case where $b \to \infty$, or, equivalently, $\psi_b(u) = u$ corresponds to the classical (nonrobust) maximum likelihood estimation.

Newton-Raphson algorithms to solve for these robust estimators numerically can be found in Sinha & Rao (2009). From the robustly estimated parameters, $\hat{\theta}_R = (\hat{\beta}_R^T, \hat{\delta}_R^T)^T$, obtained from (3), the Sinha-Rao robust empirical best linear unbiased predictor (REBLUP) for the area mean $\hat{Y}_i$, denoted $\hat{Y}_{isR}$, is of a plug-in type given by

$$
\hat{Y}_{isR} = N_i^{-1} \sum_{j \in s_i} y_{ij} + (1 - n_i N_i^{-1}) \bar{x}_{ic} \hat{\beta}_R + (1 - n_i N_i^{-1}) \hat{v}_{iR}.
$$

(7)

where $\bar{x}_{ic} = \frac{1}{N_i - n_i} \sum_{j \in U_i - s_i} x_{ij}$, and the robust predictors of the random effects, $\hat{v}_{iR} \equiv \hat{v}_{iR}(\hat{\delta}_R)$, are obtained by solving the following Fellner (1986) system of estimating equations, conditionally on $\hat{\theta}_R = (\hat{\beta}_R^T, \hat{\delta}_R^T)^T$:

$$
\sigma_e^{-1} X_i^T \Psi \{ \sigma_e^{-1} (y_i - X_i \beta - v_i 1_{n_i}) \} = 0,
$$

(8)

$$
\sigma_e^{-1} \sum_{i=1}^k \Psi \{ \sigma_e^{-1} (y_i - X_i \beta - v_i 1_{n_i}) \} - \sigma_v^{-1} \psi_b(\sigma_v^{-1} v_i) = 0, \quad (i = 1, \ldots, k).
$$

(9)

An alternative expression for the Sinha-Rao estimator is given by

$$
\hat{Y}_{isR} = N_i^{-1} \sum_{j \in s_i} y_{ij} + (1 - n_i N_i^{-1}) \bar{x}_{ic} \hat{\beta}_R + (1 - n_i N_i^{-1}) \hat{\rho}_R \sum_{j \in s_i} \hat{f}_{ijR} \left( y_{ij} - \bar{x}_{ij} \hat{\beta}_R \right)
$$

where

$$
\hat{\rho}_R = \frac{\sigma_v^2 \sum_{j=1}^{n_i} \hat{a}_{ijR}}{\sigma_v^2 \sum_{j=1}^{n_i} \hat{a}_{ijR} + \sigma_{eR}^2 \hat{b}_{iR}} \quad \text{and} \quad \hat{f}_{ijR} = \frac{\hat{a}_{ijR}}{\sum_{j=1}^{n_i} \hat{a}_{ijR}},
$$

with

$$
\hat{a}_{ijR} = \frac{\psi_b \{ \hat{\sigma}_{eR}^{-1} (y_{ij} - \bar{x}_{ij} \hat{\beta}_R - \hat{v}_{iR}) \}}{\hat{\sigma}_{eR}^{-1} (y_{ij} - \bar{x}_{ij} \hat{\beta}_R - \hat{v}_{iR})} \quad \text{and} \quad \hat{b}_{iR} = \frac{\psi_b (\hat{\sigma}_{vR}^{-1} \hat{v}_{iR})}{\hat{\sigma}_{vR}^{-1} \hat{v}_{iR}}.
$$

Denote by $\theta_R = (\beta_R^T, \delta_R^T)^T$ the probability limit of $\hat{\theta}_R = (\hat{\beta}_R^T, \hat{\delta}_R^T)^T$ (also usually referred to as the robust target parameter). An expression for the prediction error, i.e. the difference between the predictor and the true area mean, can be obtained as follows:

$$
(1 - n_i N_i^{-1})^{-1} \left( \hat{Y}_{isR} - \bar{Y}_i \right) = (\bar{x}_{ic} - \hat{\rho}_R \bar{x}_{icR})^T \left( \hat{\beta}_R - \beta_R \right) - (1 - \hat{\rho}_R) v_i + \hat{\rho}_R c_{iR} - \bar{c}_{ic}
$$

$$
+ (\bar{x}_{ic} - \hat{\rho}_R \bar{x}_{icR})^T (\beta_R - \beta),
$$

(10)

where

$$
\bar{x}_{ic} = \frac{1}{N_i - n_i} \sum_{j \in U_i - s_i} x_{ij}, \quad \bar{c}_{iR} = \sum_{j \in U_i - s_i} \hat{f}_{ijR} c_{ij} \quad \text{and} \quad \bar{c}_{ic} = \frac{1}{N_i - n_i} \sum_{j \in U_i - s_i} e_{ij}.
$$

As it will become clearer later, the expression for the prediction error provides a useful means to establish the convergence requirements for the validity of the bootstrapped MSE
developed in this paper. Specifically, we will show that sufficient conditions to establish
the convergence of our bootstrap using the Bickel & Freedman (1981) approach is to es-
tablish the convergence of the random effects \( v_i \), the average error of the units of the area
of interest, \( \bar{e}_{iR} \), the average error of nonsampled units of the area of interest, \( \bar{e}_{ic} \), and the
robust ML estimator of the fixed effects \( \hat{\beta}_R \). For the latter, asymptotic properties of the
whole parameter vector \( \hat{\theta}_R \) are needed. We state these properties in what follows.

### 2.3 Asymptotic Properties of the Robust Parameter Estimator

Denote by \( E_m[\cdot] \) the expectation using Model (2). The asymptotic properties of the robust
estimator \( \hat{\theta}_R \) are based on the following assumptions, which can be found in other related
studies. These assumptions allow to rule out cases for which the limiting distributions of
the estimated parameters either degenerate or blow-up.

**Assumption A0.** The function \( \psi_b(\cdot) \) is continuously differentiable and bounded, and
its derivative is bounded.

**Assumption A1.** \( \lim_{k \to \infty} \frac{k}{n} = c \in [0, 1] \)

**Assumption A2.** The covariates \( X_i \) are distributed over a bounded support.

**Assumption A3.** The \( p \times p \) matrix \( J_1 \) defined by
\[
J_1(\theta) = \lim_{k \to \infty} \sum_{i=1}^{k} I_{1k}^{-1/2} X_i^\top V_i^{-1/2} E_m \{ \Psi_b(r_i) \Psi_b(r_i)^\top \} U_i^{1/2} V_i^{-1} X_i I_{1k}^{-1/2}
\]
exists, is positive definite and continuous in \( \theta \); where \( I_{1k} \) is the \( p \times p \) diagonal matrix
defined by
\[
I_{1k} = \text{diag}(k, n, \ldots, n) = \begin{pmatrix}
k & 0 & \ldots & 0 \\
0 & n & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \ldots & 0 & n
\end{pmatrix}
\]

**Assumption A4.** The \( 2 \times 2 \) matrix \( J_2 \) defined by
\[
J_2(\theta) = \lim_{k \to \infty} \sum_{i=1}^{k} I_{2k}^{-1/2} E_m \{ \Psi_2(y_i, X_i, \theta) \Psi_2^\top(y_i, X_i, \theta) \} I_{2k}^{-1/2}
\]
exists, is positive definite and continuous in \( \theta \); where
\[
I_{2k} = \text{diag}(k, n) = \begin{pmatrix}
k & 0 \\
0 & n
\end{pmatrix}
\]

**Assumption A5.** The \( (p + 2) \times (p + 2) \) matrix \( G \) defined by
\[
G(\theta) = \text{Plim } G_k(\theta)
\]
exists, is finite, positive definite and continuous in \( \theta \); where

\[
G_k(\theta) = - \left[ \sum_{i=1}^k I_{1k}^{-1/2} \frac{\partial \Psi_1(y_i, X_i, \theta)}{\partial \beta} I_{1k}^{-1/2} \begin{array}{cc} 0 \\ \sum_{i=1}^k I_{2k}^{-1/2} \frac{\partial \Psi_2(y_i, X_i, \theta)}{\partial \delta} I_{2k}^{-1/2} \end{array} \right] ;
\]

and the above convergence in probability is uniform on compacts of \( \theta \).

**Assumption A6.**

\[
I_k^{-1} S(y, X, \theta_R) \xrightarrow{p} 0, \quad \text{where} \quad I_k = \begin{pmatrix} I_{1k} & 0 \\ 0 & I_{2k} \end{pmatrix} , \ I_{1k}, I_{2k} \text{ are defined in A3 and A4.}
\]

**Assumption A7.**

\[
I_k^{-1/2} S(y, X, \theta_R) \xrightarrow{d} \mathcal{N}_{p+2}(0, \Sigma_R) , \quad \text{where} \quad \Sigma_R = \Sigma(\theta_R) = \begin{pmatrix} J_1(\theta_R) & 0 \\ 0 & J_2(\theta_R) \end{pmatrix} .
\]

Assumption A1 states that the ratio of the number of areas over the total number of observations is asymptotically a constant fraction. This condition is weaker than the one required by Field *et al.* (2008) to establish the validity of the random-effect bootstrap (for linear mixed models). Field *et al.* (2008) require that each of the area’s sample size converges to infinity as the number of areas increases. In contrast, in our framework, all the areas could remain small as the number of areas increases. This condition is therefore similar to Assumption 3.2 of Miller (1977) which is a direct application of those that Weiss (1971, 1973, 1975) used to establish the asymptotic properties of maximum likelihood estimators in some nonstandard cases. As pointed out by Miller (1977), such an assumption is reasonable and easily holds in most practical situations.

Assumptions A3 and A4 are similar to Assumptions 3.4 and 3.5 of Miller (1977). The matrices \( J_1 \) and \( J_2 \) defined within these assumptions determine the asymptotic covariance matrices of the fixed and random effects estimates respectively. Assumptions A3 and A4 ensure the existence and positive definiteness of these matrices. It should be noted that if either \( J_1 \) or \( J_2 \) does not exist or is not positive definite, then the associated estimates do not converge to a nondegenerate distribution. As explained by Miller (1977), any design or set of designs that might be used in practice would naturally satisfy these two assumptions.

Assumptions A5-A7 are equivalent to conditions A.1-A.4 of Huggins (1993). Assumption A5 is usually checked in an ad-hoc manner. For example, the existence of bounded derivatives or the Hölder or Lipschitz continuity of \( G_k(\cdot) \) on compacts of \( \theta \) would suffice for these conditions to hold. Assumptions A6 and A7 readily follow from the law of large numbers, the central limit theorem and the appropriate standard regularity conditions.

The above assumptions guarantee that conditions A.1-A.4 of Huggins (1993) are satisfied. The following result is therefore a corollary of Theorem A of Huggins (1993), and is thus given without proof.

**Lemma 1.** Under Assumptions A0-A7,

\[
I_k^{1/2} (\hat{\theta}_R - \theta_R) \xrightarrow{d} \mathcal{N}(0, G_R^{-1} \Sigma_R G_R^{-1}), \quad (11)
\]
where $\hat{\theta}_R = (\hat{\beta}_R^\top, \hat{\delta}_R^\top)^\top$ is the unique solution to \(3\), and $\theta_R$ is its probability limit.

Likewise, if we take $\psi_\delta(t) = t$ in all of the functions given above, we obtain:

$$I_k^{1/2}(\hat{\theta} - \theta_0) \overset{d}{\to} \mathcal{N}(0, G_0^{-1}\Sigma_0 G_0^{-1}),$$

where $\hat{\theta} = (\hat{\beta}^\top, \hat{\delta}^\top)^\top$ is solution to \(3\), and $\hat{\theta}$ is the maximum likelihood estimator of the true parameter vector $\theta_0$.


This result gives the asymptotic normality of both the robust estimator and the maximum likelihood estimator under correct specification. Note that the probability limit of the robust estimator, $\theta_R$, is possibly different from the true parameter vector $\theta_0$. However, since the classic MLE $\hat{\theta}$ is based on all the original data including the outliers, the estimator $\hat{\theta}_R$ would usually be preferred in practice because of its lesser sensitivity to influential observations.

3 The MSE Bootstrap Estimator

This section proposes a bootstrap procedure designed for estimating the MSE of the robust empirical best linear unbiased predictors described in the previous section. We show that if the bootstrap samples are generated similarly to the process that generated the original data, then the bootstrap procedure for the MSE will be valid. As explained earlier, generating bootstrap samples using the robust estimators of the model parameters as in Sinha & Rao (2009), leads to bootstrap replicas whose variability is lower than that of the original data, yielding poor coverage rates (see Jiongo et al. 2013). To overcome this issue, we propose a bootstrap procedure that uses the nonrobust maximum likelihood estimators which are asymptotically unbiased. This allows to obtain a bootstrap sample whose variability is similar to that of the original data. Moreover, we generalize the above procedures by relaxing the usual normality distributional assumption. Our bootstrap is semi-parametric and therefore avoids the possible bias due to incorrect specification of the random effects or the error distribution (Opsomer et al. 2008).

3.1 Description of the Bootstrap Method

In the following, we present the method of generating the bootstrap samples and estimating the MSE of the robust estimators. The method is described for the Sinha-Rao robust predictor and it can be easily adapted for other predictors. The bootstrap procedure works as follows.

**Step 1:** Generate $k$ random variables $u^*_i$, $i = 1, \ldots, k$, by drawing independently with replacement among $\hat{u}_i - \frac{1}{k} \sum_{i=1}^k \hat{u}_i$, $i = 1, \ldots, k$; and generate $N$ random variables $e^*_ij$, $i = 1, \ldots, k$, $j = 1, \ldots, N_i$, by drawing independently with replacement among $\hat{e}_{ijg} - \frac{1}{n} \sum_{l=1}^k \sum_{g=1}^{n_l} \hat{e}_{lg}$, $i = 1, \ldots, k$, $g = 1, \ldots, n_l$, respectively; where $\hat{u}_i$ and $\hat{e}_{lg}$ are defined as
where \( e \) component of follows.

...to non-normality of the random components of the model. are in the fixed effects, or in the random effects or in the error term, and should be robust expected to work reasonably well regardless of the nature of the outlier, i.e., whether they do not substantially affect the estimated values. The proposed bootstrap method is generated, it is recommended to choose a number sufficiently large such that further increases

\[ \hat{\tau}_i = 1 - \sqrt{1 - \hat{\rho}_i}, \quad \hat{\rho}_i = \frac{n_i \hat{\sigma}_v^2}{\hat{\sigma}_v^2 + n_i \hat{\sigma}_v^2}, \]

and \( \hat{\nu} \) is the empirical best linear unbiased predictor of the random effect, given by

\[ \hat{\nu}_i = \hat{\rho}_i (\hat{y}_i - \bar{x}_i \hat{\beta}), \quad (13) \]

Note that the estimates \( \hat{\theta} = (\hat{\beta}^\top, \hat{\delta}^\top)^\top \), where \( \hat{\delta} = (\hat{\sigma}_v^2, \hat{\sigma}_v^2) \), used at this step are the nonrobust estimators of \( \theta_0 \).

**Step 2:** Compute the mean of the bootstrap population:

\[ \bar{Y}_i^* = N_i^{-1} \sum_{j=1}^{N_i} x_{ij} \hat{\beta} + u_i^* + N_i^{-1} \sum_{j=1}^{N_i} e_{ij}^*. \quad (14) \]

**Step 3:** Generate a bootstrap sample \((X_i, y_i^*)\), \( i = 1, \ldots, k \), from the model

\[ y_{ij}^* = x_{ij} \hat{\beta} + u_i^* + e_{ij}^*, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \quad (15) \]

where \( \{e_{ij}^*; \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i\} \) is a sample of size \( n \) drawn from the population of bootstrapped errors using the same sampling plan \( P(s) \) that was used to draw the original sample. Equation \((15)\) can be rewritten in the form

\[ y_{ij}^* = X_i \hat{\beta} + u_i^* 1_{ni} + e_{ij}^*, \quad i = 1, \ldots, k. \quad (16) \]

**Step 4:** Robust bootstrap estimators \( \hat{\beta}_{SR}^*, \hat{\delta}_{SR}^* \), and \( \hat{\nu}_{SR}^* \) are computed from the bootstrap samples. The Sinha-Rao robust bootstrap estimator for small-area means, \( \hat{Y}_{iSR}^* \), is obtained as

\[ \hat{Y}_{iSR}^* = N_i^{-1} \sum_{j \in s_i} y_{ij}^* + \left( 1 - n_i N_i^{-1} \right) \bar{x}_{ic} \hat{\beta}_R^* + \left( 1 - n_i N_i^{-1} \right) \hat{\nu}_{SR}^*. \quad (17) \]

**Step 5:** Repeat the above process a large number of times, say \( B \) times, to obtain \( B \) bootstrap samples and compute the estimator of the mean squared error of \( \hat{Y}_{iSR}^* \) by

\[ \text{MSE} \left( \hat{Y}_{iSR}^* \right) = B^{-1} \sum_{b=1}^{B} \left( \hat{Y}_{iSR}^{* (b)} - \bar{Y}_{iSR}^{* (b)} \right)^2, \]

where \( \hat{Y}_{iSR}^{* (b)} \) and \( \bar{Y}_{iSR}^{* (b)} \) correspond to Expressions \((17)\) and \((14)\), respectively, for the \( b \)th bootstrap sample.

Although the procedure does not specify the number of bootstrap samples to be generated, it is recommended to choose a number sufficiently large such that further increases do not substantially affect the estimated values. The proposed bootstrap method is expected to work reasonably well regardless of the nature of the outlier, i.e., whether they are in the fixed effects, or in the random effects or in the error term, and should be robust to non-normality of the random components of the model.

\[ \text{The use of different subscripts notations here is made to emphasize the fact that the randomly selected component of } e^* \text{ for which the coordinate } (i, j) \text{ is assigned, } e_{ij}^*, \text{ is independent to the corresponding area and units from the original residual } e_{ij}. \]
3.2 Validity of the Bootstrap Estimator

Denote by \(d_t, t = 1, 2, \ldots\), the Mallows (1972) metric for probabilities in \(\mathbb{R}^{p+2}\), relative to the Euclidean norm \(||\cdot||\). If \(\mu\) and \(\nu\) are two probabilities in \(\mathbb{R}^{p+2}\), then \(d_t(\mu, \nu)\) is the infimum of \([E(||U - V||^t)]]^{1/t}\) over all pairs of random vectors \(U\) and \(V\) whose distributions are \(\mu\) and \(\nu\), respectively. Also, for two random variables \(U\) and \(V\), write \(d_t(U, V)\) for the \(d_t\)-distance between the distributions of \(U\) and \(V\). Only the cases \(t = 1, 2, 3\) or 4 are of interest in this paper.

Let \(\hat{F}_{uk}\) be the empirical distribution of \(\hat{u}_i, i = 1, \ldots, k\), centered at their mean, and let \(F_{uk}\) be the empirical distribution of \(u_i, i = 1, \ldots, k\). Likewise, let \(\hat{F}_{ek}\) be the empirical distribution of \(\hat{e}_{ij}, i = 1, \ldots, j, j = 1, \ldots, n_i\), centered at their mean, and let \(F_{ek}\) be the empirical distribution of the \(e_{ij}, i = 1, \ldots, j, j = 1, \ldots, n_i\). Define by \(\Phi_k(F_{u,e})\) the distribution of \(I_{k}^{1/2}(\hat{\theta}_R - \theta_R)\), and by \(\Phi_k(\hat{F}_{u,e})\) the distribution of \(I_{k}^{1/2}(\hat{\theta}_R - \theta_R)\), where \(\hat{\theta}_R\) is the robust estimate of \(\theta\) obtained from the bootstrap sample \((X_i, y_i), i = 1, \ldots, k\).

Denote by \(E_*[\cdot]\) the bootstrap expectation. To derive the asymptotic properties of the bootstrap estimators we use the following bootstrap analogues of Assumptions A3 to A7 stated in Section 2.3 which given conditionally on the original sample \((X_i, y_i), i = 1, \ldots, k\).

**Assumption B3.** The \(p \times p\) matrix \(J_1^*(\theta)\) defined by

\[
J_1^*(\theta) = \lim_{k \to \infty} \sum_{i=1}^{k} I_{1k}^{-1/2} X_i^T V_i^{-1} U_i^{1/2} E_* \{ \Psi_k(r_i)^T \Psi_k(r_i)^T \} U_i^{1/2} V_i^{-1} X_i I_{1k}^{-1/2}
\]

exists, is positive definite and is a continuous function of \(\theta\).

**Assumption B4.** The \(2 \times 2\) matrix \(J_2^*(\theta)\) defined by

\[
J_2^*(\theta) = \lim_{k \to \infty} \sum_{i=1}^{k} I_{2k}^{-1/2} E_* \{ \Psi_2(y_i^*, X_i, \theta) \Psi_2(y_i^*, X_i, \theta) \} I_{2k}^{-1/2}
\]

exists, is positive definite and is a continuous function of \(\theta\).

**Assumption B5.** The \((p + 2) \times (p + 2)\) matrix \(G^*\) defined by

\[
G^*(\theta) = \text{Plim} G_k^*(\theta)
\]

exists, is positive definite and continuous in \(\theta\); where

\[
G_k^*(\theta) = - \begin{bmatrix}
\sum_{i=1}^{k} I_{1k}^{-1/2} \frac{\partial \Psi_1(y_i^*, X_i, \theta)}{\partial \hat{\theta}} I_{1k}^{-1/2} & 0 \\
0 & \sum_{i=1}^{k} I_{2k}^{-1/2} \frac{\partial \Psi_2(y_i^*, X_i, \theta)}{\partial \hat{\theta}} I_{2k}^{-1/2}
\end{bmatrix}
\]

The above convergence of \(G_k^*(\theta)\) in probability is uniform over compacts of \(\theta\).

**Assumption B6.**

\[
I_{k}^{-1} S(y^*, X, \hat{\theta}_R) \overset{p}{\to} 0
\]
Assumption B7.

\[ I_k^{-1/2} S(y^*, X, \hat{\theta}_R) \xrightarrow{d} N_{p+2}(0, \Sigma_R), \quad \text{where} \quad \Sigma_R = \Sigma^*(\theta_R) = \begin{pmatrix} J_1^*(\theta_R) & 0 \\ 0 & J_2^*(\theta_R) \end{pmatrix}. \]

With these conditions, a bootstrap version of Lemma 1 is given by the following result.

**Lemma 2.** Under Assumptions A0-A2, B3-B7, and conditionally on the sample,

\[ I_k^{1/2}(\hat{\theta}_R^*-\hat{\theta}_R) \xrightarrow{d} N(0, G_{R}^{-1} \Sigma_R G_{R}^{-1}); \quad (18) \]

Likewise, when we take \( \psi_b(t) = t \), we get:

\[ I_k^{1/2}(\hat{\theta}_R^*-\hat{\theta}_R) \xrightarrow{d} N(0, G_{R}^{-1} \Sigma_R G_{R}^{-1}), \quad (19) \]

To show the validity of our bootstrap, the first step is to show that the bootstrap samples, as well as the bootstrap matrices given in the above conditions, converge in distribution to the original sample and in probability to the original matrices, respectively. These results are gathered in the following lemma.

**Lemma 3.** Let Assumptions A0-A2, B3- B7 hold. Then, for \( k \to \infty \), and uniformly over \( \theta \),

\[ d_4(F_v, \hat{F}_{ku}) \xrightarrow{p} 0 \quad \text{and} \quad d_4(F_e, \hat{F}_{ke}) \xrightarrow{p} 0 \quad (20) \]

\[ J_1^*(\theta) \xrightarrow{p} J_1(\theta) \quad (21) \]

\[ J_2^*(\theta) \xrightarrow{p} J_2(\theta) \quad (22) \]

\[ G^*(\theta) \xrightarrow{p} G(\theta) \quad (23) \]

**Proof.** It uses the Mallows (1972) metric for \( t = 4 \), and results from Bickel & Freeman (1981). See the Appendix at Section 7.

We next show that the asymptotic distribution of the robust bootstrap estimator is asymptotically equivalent to the asymptotic distribution of the robust initial estimator, conditionally on the sample.

**Theorem 1.** Under Assumptions A0-A7 and B3-B7, and conditionally on the sample,

\[ d_{p+2}^p \left\{ \Phi_k(F_{v,e}), \Phi_k(\hat{F}_{a,e}) \right\} \xrightarrow{p} 0 \quad \text{as} \quad k \to \infty. \]

**Proof.** The proof follows immediately from the above results and Lemma 8.3 of Bickel and Freeman (1981). Denote \( \xi_R = I_k^{1/2}(\hat{\theta}_R^*-\theta_R) \) and \( \xi_R^* = I_k^{1/2}(\hat{\theta}_R^*-\hat{\theta}_R) \). Recall that their finite sample distribution are defined by \( \Phi_k(F_{v,e}) \) and \( \Phi_k(\hat{F}_{a,e}) \), respectively. By Lemmas 3 and 2 their asymptotic distributions are given by \( N(0, G_{R}^{-1} \Sigma_R G_{R}^{-1}) \) and \( N(0, G_{R}^{-1} \Sigma_R G_{R}^{-1}) \), respectively. It then follows by Lemma 3 and the Levy’s Continuity Theorem that conditionally on the sample,

\[ \xi_R^* \xrightarrow{d} \xi_R \]
It also easily follows that conditionally on the sample
\[ E_* \left[ \| \xi \|^2 \right] \longrightarrow \text{tr} \left( G_*^{-1} \Sigma_R G_*^{-1} \right), \quad \text{and} \quad E_m \left[ \| \xi \|^2 \right] \longrightarrow \text{tr} \left( G_R^{-1} \Sigma_R G_R^{-1} \right) \]
which, by Lemma 3 and the continuous mapping theorem, implies that
\[ E_* \left[ \| \xi \|^2 \right] \xrightarrow{p} E_m \left[ \| \xi \|^2 \right] \]
It then follows by Lemma 8.3 a) of Bickel and Freeman (1981) that
\[ d_2^{\| \xi \|^2} \left\{ \Phi_k(F_{v,e}), \Phi_k(\hat{F}_{u,e}) \right\} \xrightarrow{p} 0 \quad \text{as} \quad k \to \infty. \]

The following theorem is the main result of this paper. It states that under conditions given above, the proposed bootstrap MSE estimator of the Sinha & Rao (2009) robust empirical best linear unbiased predictor is a consistent estimator of the MSE.

**Theorem 2.** Under Assumptions A0 to A7 and B3 to B7, and conditionally on the sample,
\[ \left| E_* \left( \hat{Y}_{iSR}^* - \bar{Y}_i \right)^2 - E_m \left( \hat{Y}_{iSR} - \bar{Y}_i \right)^2 \right| \xrightarrow{p} 0 \quad \text{as} \quad k \to \infty. \]

**Proof.** By Lemma 8.3 a) of Bickel and Freedman (1981), it is sufficient to show that
\[ d_2 \left( \hat{Y}_{iSR}^* - \bar{Y}_i, \hat{Y}_{iSR} - \bar{Y}_i \right) \xrightarrow{p} 0. \]
Denote \( \hat{\gamma}_R = I_{1k}^{1/2} \left( \hat{\beta}_R - \beta_R \right) \) and \( \hat{\gamma}_R^* = I_{1k}^{1/2} \left( \hat{\beta}_R^* - \beta_R \right) \).
Then, from Equation (10) above, we can write \( (1 - n_i N_i)^{-1} \left( \hat{Y}_{iSR} - \bar{Y}_i \right) \) as an affine function of \( \hat{\gamma}_R, v_i, \bar{e}_R, \bar{e}_{ic}. \) That is, \( (1 - n_i N_i)^{-1} \left( \hat{Y}_{iSR} - \bar{Y}_i \right) = \Lambda_i \left( \hat{\gamma}_R, v_i, \bar{e}_R, \bar{e}_{ic} \right). \) It then follows from Assumptions A0 and A2 that there exists a positive constant, \( M > 0 \) such that
\[ \| \Lambda_i \left( \hat{\gamma}_R, v_i, \bar{e}_R, \bar{e}_{ic} \right) \|^2 \leq M \left[ 1 + \| (\hat{\gamma}_R, v_i, \bar{e}_R, \bar{e}_{ic})^\top \|^2 \right]. \]
Given that, by Theorem 1 and Condition 20 of Lemma 3 above, we must have
\[ d_2 \left( \left( \hat{\gamma}_R, v_i^*, \bar{e}_{R}, \bar{e}_{ic}^* \right), \left( \hat{\gamma}_R, v_i, \bar{e}_R, \bar{e}_{ic} \right) \right) \xrightarrow{p} 0. \]
It then follows by Lemma 8.5 of Bickel and Freedman (1981) that
\[ d_2 \left( \hat{Y}_{iSR}^* - \hat{Y}_i^*, \hat{Y}_{iSR} - \hat{Y}_i \right) \xrightarrow{p} 0. \]

**4 Monte Carlo Simulations**

In this section we carry out Monte Carlo simulations to explore the finite sample performance of the proposed bootstrap procedure for estimating the MSE. For this purpose we consider four small-area estimators. The empirical best linear unbiased estimator, EBLUP, the robust estimator of Sinha & Rao (2009), SR, the robust estimator of Chambers et al. (2013), CCST3, and the robust estimator of Jiongo et al. (2013) based on the conditional bias concept of Beaumont et al. (2013), CB.
For each of these small-area estimators, the performance of the proposed bootstrap MSE procedure, denoted JNBOOT, is assessed and compared with several other alternative MSE estimators. For the small-area estimators CB, we compare our results with the bootstrap MSE estimators of Sinha & Rao (2009), denoted SRBOOT, and Jiongo et al. (2013), denoted JHDBOOT. For the robust estimators SR and CCST3, we also compare our results with the analytical linearization MSE and linearization-based MSE estimators developed by Chambers et al. (2013), denoted CCT and CCST, respectively. Finally, for the EBLUP, we compare our results with all of the above including the estimator of Prasad and Rao (1990), denoted PR.

4.1 Simulation Design

We consider the same type of contamination design as in Jiongo et al. (2013). Outliers are generated from a mixture model \( \zeta_n \) satisfying \( y_{ij} = (1 - A_{ij})y_{0ij} + A_{ij}y_{1ij} \), where the \( A_{ij} \) are independently generated according to a Bernoulli distribution with parameter \( p = 0.1 \), and \( y_{0ij} \) and \( y_{1ij} \) are given by two mixed linear models defined by

\[
\begin{align*}
\zeta_0 : y_{0ij} & = \beta_0 + \beta_{01}x_{ij} + v_{0i} + e_{0ij}, \quad (j = 1, \ldots, N_i; i = 1, \ldots, k), \\
\zeta_1 : y_{1ij} & = \beta_{10} + \beta_{11}x_{ij} + v_{1i} + e_{1ij}, \quad (j = 1, \ldots, N_i; i = 1, \ldots, k).
\end{align*}
\]

We take \( k = 40 \) and \( N_1 = \ldots = N_{40} = 50 \). The error terms and random effects are assumed to be normally distributed and are given by \( v_{0i} \sim \mathcal{N}(0, \sigma^2_{v0}), v_{1i} \sim \mathcal{N}(0, \sigma^2_{v1}), e_{0ij} \sim \mathcal{N}(0, \sigma^2_{e0}) \) and \( e_{1ij} \sim \mathcal{N}(0, \sigma^2_{e1}) \), \( (k = 1, \ldots, 40; j = 1, \ldots, 50) \).

The values of the auxiliary variable are generated from a normal distribution with mean \( E(X) = 2 \) and standard deviation \( \text{var}(X) = 0.35 \). In each area of the population, random samples of size \( n_1 = \ldots = n_{40} = 5 \) have been selected by simple random sampling without replacement. Three contamination scenarios are considered and Table 1 provides a description of each scenario. Scenario \((0, 0, 0)\) corresponds to the absence of contamination, while scenario \((e, v, 0)\) corresponds to having both the random errors and the area random effects contaminated. Finally, scenario \((e, v, b)\) corresponds to the situation where the contamination comes from the random effects, the random errors and the fixed effects. These scenarios are similar to those given in Sinha & Rao (2009), but an additional scenario has been added here to allow for \( \beta_0 = (\beta_{00}, \beta_{01}) \) to be different from \( \beta_1 = (\beta_{10}, \beta_{11}) \). For the parameters given in Table 1 the correlation between the units in a given area is equal to \( \rho_0 = 0.5 \) in the absence of contamination, where \( \rho_0 \) satisfies the relation \( \sigma^2_{e0} = \rho_0\sigma^2_{v0}/(1 - \rho_0) \). Figure 1 provides a picture of the simulated data for various modes of contamination.

For each scenario, we generate \( T = 500 \) populations and \( B = 200 \) bootstrap replications. The tuning constant for the small-area estimators CB is set as defined in Beaumont et al. (2013). The tuning constant of the robust predictor CCST3 is set at \( b = 3 \) as in the simulation experiments of Chambers et al. (2013). Although the robust estimation of the small area means is not the subject of this paper, we present the results of the relative absolute bias and root relative mean squared errors (RRMSE) of each of the small-area estimators considered, for completeness.

\footnote{Note that for the EBLUP, the bootstrap procedures SRBOOT and JHDBOOT are equivalent. Hence, only JHDBOOT is reported for this case.}
Table 1: Description of the contamination scenarios.

The populations are generated according to \( y_{ij} = (1 - A_{ij}) \bar{y}_{0ij} + A_{ij} \bar{y}_{1ij}, A_{ij} \sim Bernoulli(0.1) \), using the unit-level models (24) and (25), assuming normality for the random effects and error terms in \( \zeta_0 \) and \( \zeta_1 \). Under the scenario \((0, 0, 0)\), the correlation between the units of the same domain equals 0.5.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Sources of the contamination</th>
</tr>
</thead>
</table>
| \((0, 0, 0)\) | \begin{align*}
\text{Variances of} & \quad \text{Error terms} \\
& \quad (\sigma_0^2, \sigma_1^2) = (6, 6) \\
\text{Variances of} & \quad \text{Random effects} \\
& \quad (\sigma_0^2, \sigma_1^2) = (6, 6) \\
\text{Intercepts and} & \quad \beta_0 = \beta_1 = (100, 3)^\top \\
\text{slopes} & \quad (e, v, b) \\
\end{align*} |
| \((e, v, 0)\) | \begin{align*}
\text{Variances of} & \quad \text{Error terms} \\
& \quad (\sigma_0^2, \sigma_1^2) = (6, 150) \\
\text{Variances of} & \quad \text{Random effects} \\
& \quad (\sigma_0^2, \sigma_1^2) = (6, 150) \\
\text{Intercepts and} & \quad \beta_0 = \beta_1 = (100, 3)^\top \\
\text{slopes} & \quad (e, v, b) \\
\end{align*} |
| \((e, v, b)\) | \begin{align*}
\text{Variances of} & \quad \text{Error terms} \\
& \quad (\sigma_0^2, \sigma_1^2) = (6, 150) \\
\text{Variances of} & \quad \text{Random effects} \\
& \quad (\sigma_0^2, \sigma_1^2) = (6, 150) \\
\text{Intercepts and} & \quad \beta_0 = (100, 3)^\top, \beta_1 = (150, 1)^\top \\
\text{slopes} & \quad (e, v, b) \\
\end{align*} |

Figure 1: Scatter plots of the populations generated from the mixture model (24) and (25). The model parameters are given in Table 1.

Let \( \hat{Y}_i \) denotes an arbitrary estimator of the small-area mean \( \bar{Y}_i \). Then the absolute relative bias for the area mean \( \bar{Y}_i \) associated to \( \hat{Y}_i \) is given by

\[
\text{ARB}(\hat{Y}_i) = 100 \times \left| T^{-1} \sum_{t=1}^{T} \frac{\hat{Y}_{it}^{(t)} - \bar{Y}_i^{(t)}}{\bar{Y}_i^{(t)}} \right|, \quad (i = 1, \ldots, k),
\]

and the root relative mean squared error is given by

\[
\text{RRMSE}(\hat{Y}_i) = 100 \times \sqrt{T^{-1} \sum_{t=1}^{T} \left( \frac{\hat{Y}_{it}^{(t)} - \bar{Y}_i^{(t)}}{\bar{Y}_i^{(t)}} \right)^2}, \quad (i = 1, \ldots, k).
\]

For the estimation of the MSE, we also compute the empirical values of the relative bias (RB) and the root relative mean squared error. Denote by \( \text{MSE}(\hat{Y}_i) \) the estimator of the mean squared error of \( \hat{Y}_i \). The relative bias associate to \( \text{MSE}(\hat{Y}_i) \) is given by

\[
\text{RB}(\text{MSE}(\hat{Y}_i)) = 100 \times T^{-1} \sum_{t=1}^{T} \frac{\text{MSE}(\hat{Y}_{it}^{(t)}) - \text{MSE}(\hat{Y}_i)}{\text{MSE}(\hat{Y}_i)}, \quad (i = 1, \ldots, k),
\]

and the root relative mean squared error of \( \text{MSE}(\hat{Y}_i) \) is calculated as

\[
\text{RRMSE}(\text{MSE}(\hat{Y}_i)) = 100 \times \sqrt{T^{-1} \sum_{t=1}^{T} \left( \frac{\text{MSE}(\hat{Y}_{it}^{(t)}) - \text{MSE}(\hat{Y}_i)}{\text{MSE}(\hat{Y}_i)} \right)^2}, \quad (i = 1, \ldots, k).
\]
Section 4.2 presents simulation results based on all the domains. We use boxplots and measures of central tendency such as the median of all the areas. Simulation results are obtained under scenarios \((0,0,0)\), \((e,v,0)\) and \((e,v,b)\) described in Table 1.

### 4.2 Simulation Results

Table 2: Monte Carlo absolute relative biases (\%) and relative root mean squared error (\%) for the predictors of the small area means (median of the areas).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Absolute relative bias</th>
<th>Root relative mean squared error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EBLUP</td>
<td>CB</td>
</tr>
<tr>
<td>((0,0,0))</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>((e,v,0))</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>((e,v,b))</td>
<td>0.10</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 3: Monte Carlo relative biases (RB \%) and root relative mean squared error (RRMSE \%) for the mean squared error estimator of the predictors of small area means (at the median of the areas).

<table>
<thead>
<tr>
<th>SAE</th>
<th>MSE</th>
<th>RB</th>
<th>RRMSE</th>
<th>RB</th>
<th>RRMSE</th>
<th>RB</th>
<th>RRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBLUP</td>
<td>PR</td>
<td>-3.39</td>
<td>11.74</td>
<td>3.33</td>
<td>31.28</td>
<td>34.87</td>
<td>57.09</td>
</tr>
<tr>
<td>CCT</td>
<td>-2.31</td>
<td>43.41</td>
<td>49.13</td>
<td>195.70</td>
<td>238.50</td>
<td>549.90</td>
<td></td>
</tr>
<tr>
<td>CCST</td>
<td>0.21</td>
<td>44.96</td>
<td>53.48</td>
<td>221.90</td>
<td>243.50</td>
<td>587.80</td>
<td></td>
</tr>
<tr>
<td>JHDBOOT</td>
<td>-1.28</td>
<td>15.19</td>
<td>-3.48</td>
<td>32.63</td>
<td>12.22</td>
<td>49.27</td>
<td></td>
</tr>
<tr>
<td>JNBOOT</td>
<td>-1.02</td>
<td>15.60</td>
<td>-5.36</td>
<td>33.02</td>
<td>10.93</td>
<td>48.61</td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>CCT</td>
<td>-3.34</td>
<td>56.76</td>
<td>-23.90</td>
<td>48.96</td>
<td>-70.03</td>
<td>71.71</td>
</tr>
<tr>
<td>CCST</td>
<td>-0.88</td>
<td>57.43</td>
<td>-18.10</td>
<td>89.58</td>
<td>-65.56</td>
<td>81.95</td>
<td></td>
</tr>
<tr>
<td>SRBOOT</td>
<td>2.77</td>
<td>17.67</td>
<td>-42.14</td>
<td>43.46</td>
<td>-91.44</td>
<td>91.46</td>
<td></td>
</tr>
<tr>
<td>JHDBOOT</td>
<td>1.66</td>
<td>15.59</td>
<td>41.42</td>
<td>62.58</td>
<td>-37.65</td>
<td>46.82</td>
<td></td>
</tr>
<tr>
<td>JNBOOT</td>
<td>0.97</td>
<td>15.90</td>
<td>-9.21</td>
<td>25.36</td>
<td>-9.33</td>
<td>38.36</td>
<td></td>
</tr>
<tr>
<td>CCST3</td>
<td>CCT</td>
<td>40.75</td>
<td>97.16</td>
<td>4.09</td>
<td>150.40</td>
<td>-46.72</td>
<td>122.00</td>
</tr>
<tr>
<td>CCST</td>
<td>42.81</td>
<td>104.80</td>
<td>7.50</td>
<td>175.80</td>
<td>-43.84</td>
<td>138.80</td>
<td></td>
</tr>
<tr>
<td>SRBOOT</td>
<td>1.11</td>
<td>18.04</td>
<td>-54.39</td>
<td>55.08</td>
<td>-91.44</td>
<td>91.46</td>
<td></td>
</tr>
<tr>
<td>JHDBOOT</td>
<td>-0.25</td>
<td>16.61</td>
<td>24.51</td>
<td>85.56</td>
<td>65.38</td>
<td>81.61</td>
<td></td>
</tr>
<tr>
<td>JNBOOT</td>
<td>-0.56</td>
<td>16.65</td>
<td>6.32</td>
<td>33.08</td>
<td>1.00</td>
<td>44.48</td>
<td></td>
</tr>
<tr>
<td>CB</td>
<td>SRBOOT</td>
<td>0.74</td>
<td>17.19</td>
<td>-48.25</td>
<td>49.20</td>
<td>-82.22</td>
<td>82.32</td>
</tr>
<tr>
<td>JHDBOOT</td>
<td>0.05</td>
<td>15.30</td>
<td>9.41</td>
<td>50.85</td>
<td>40.03</td>
<td>70.02</td>
<td></td>
</tr>
<tr>
<td>JNBOOT</td>
<td>-0.51</td>
<td>15.68</td>
<td>-3.75</td>
<td>28.32</td>
<td>3.95</td>
<td>41.12</td>
<td></td>
</tr>
</tbody>
</table>

The results reported in Table 2 present the percent Monte Carlo absolute relative biases (ARB \%) and the percent relative root mean squared error (RRMSE \%) for the EBLUP and the robust predictors of the small area means, where the computation is given for the median of all the areas. The results show that the estimator CB proposed by Jiongo et al. (2013) perform well with the given value of the tuning constant regardless of the mode of contamination. That is, whether the contamination occurs at the errors level, the random effect level or the fixed effects level. On the other hand, the biases of the Sinha & Rao (2009) and the Chambers et al. (2013) predictors are quite similar for
the case \((e,v,0)\), while the former tend to yield smaller mean squared error than the latter. For the case \((e,v,b)\), the Chambers et al. (2013) predictor has less bias than the Sinha & Rao (2009), but is more variable.

The results reported in Table 3 present the percent Monte Carlo relative biases (RB %) and the percent root relative mean squared error (RRMSE %) of the mean squared error estimator of the predictors of the small area means, obtained at the median of the areas. In the absence of outliers (scenario \((0,0,0)\) in Table 3), only the analytical pseudolinearization MSE estimators (CCT) and linearization-based MSE estimators (CCST) of the MSE are biased when the Chambers et al. (2013) robust small area predictor CCST3 is used. All the other MSE estimators are equivalent in terms of bias and display negligible biases, regardless of the small-area estimator considered. Likewise, for the MSE estimators it can be noted that only analytical pseudolinearization MSE estimators (CCT) and linearization-based MSE estimator (CCST) are unstable throughout. In contrast, all the bootstrap estimators are stable and equivalent to each other regardless of the small-area estimator considered. For the particular case of the empirical best linear unbiased predictor (EBLUP), the Prasad & Rao (1990) MSE estimator (PR) is, as one would expect, more stable than the bootstrap estimators.

Figure 2: Boxplots of the relative biases of the MSE estimators. Scenario \((e,v,0)\)

Consider the case where outliers are present in the errors and the random effects (scenario \((e,v,0)\) in Table 3). We also note that, except for the Chambers et al. (2013) robust estimator (CCST), the analytical pseudolinearization MSE estimator (CCT) and linearization-based MSE estimator (CCST) of the MSE are biased for all other small area
estimators considered. In general, it can be seen that for all the robust predictors of the small area mean under consideration, the bootstrap MSE of Sinha & Rao (2009), SRBOOT, is consistently negatively biased while the bootstrap MSE of Jiongo et al. (2013), JHDBOOT, is consistently positively biased. Moreover, these biases are generally higher than 20% in absolute value. In sharp contrast, the proposed bootstrap MSE, JNBOOT, performs the best and exhibits smaller biases, always less than 12% in absolute value. This clearly shows the striking difference in terms of performance between our estimator and the others. As for the variability of these MSE estimators, the proposed bootstrap is again the most efficient. Indeed, regardless of the robust predictor considered the relative root mean square error of the competitors CCT, CCST, SRBOOT or JHDBOOT is generally higher than the RRMSE of the proposed bootstrap, JNBOOT. Figure 2 and 3 depict the boxplots of the relative biases and the root relative mean squared error of the MSE estimators, respectively, which further confirms the superiority of the proposed bootstrap method over existing alternatives for all the robust small-area predictors.

Figure 3: Boxplots of the root relative mean squared error of the MSE estimators. Scenario \((e,v,0)\)

Finally, if we look at the case where outliers are simultaneously present in the error terms, the random effects and the fixed effects (that is, scenario \((e,v,b)\) in Table 3), the estimators, CCT, CCST, SRBOOT or JHDBOOT, are highly biased, regardless of the robust small-area predictor considered. In contrast, the proposed bootstrap, JNBOOT, is unbiased with relative bias as low as 1.00% for the robust predictors CCST3 of Chambers.
et al. (2013), and 3.95% for the robust predictors CB of Jiongo et al. (2013). The relative bias of the proposed estimator is moderate (−9.21%) for SR of Sinha & Rao (2009), yet by far still lower than that of the other predictors. This shows once again that the JNBOOT outperforms its competitors in terms of bias. In terms of efficiency, the proposed bootstrap also has the best performance. The relative root mean squared error of the alternative MSE estimators are mostly at least three to four times as high as that of the proposed MSE estimator. Even for the particular case of EBLUP where the CCT and the CCST display extremely large biases and error rates, the proposed JNBOOT still performs the best and is within reasonable ranges. Unreported boxplots (available from the authors) further confirm the superiority of the proposed bootstrap method, JNBOOT, over all the existing alternatives considered, as in the previous scenario.

5 Application: County crop areas

The data used in this application are taken from Battese, Harter and Fuller (1988). They estimate the acreage of corn and soybeans of \( k = 12 \) counties (small areas) of North-Central Iowa from LANDSAT satellite images and observations from \( n = 37 \) segments obtained from a farmers survey. The data include (a) the sample size for each area, (b) the number of acres of corn and soybeans for each unit of the sample (as collected in the survey), (c) the number of image pixels classified as corn or soybeans for each unit in the sample, and (d) the population mean of each area of pixels classified as corn or soybeans. For a detailed description of these data, see Battese, Harter and Fuller (1988).

Battese et al. (1988) identified an outlier and deleted it from their study. Sinha & Rao (2009) incorporated this outlier in their estimation procedure to investigate the influence of this observation on the EBLUP, and also to assess the ability of their robust method to identify and reduce the influence of this unit on the estimation. These data are of interest to us because they provide a good example in which the proposed bootstrap MSE estimator for outlier-robust predictors of small-area means can be applied.

The model is given by:

\[
y_{ij} = \beta_0 + x_{1ij} \beta_1 + x_{2ij} \beta_2 + v_i + e_{ij}, \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, n_i,
\]

where the random effects \( v_i \) are independently and identically distributed as \( \mathcal{N}(0, \sigma^2_v) \); the error terms \( e_{ij} \) are independently and identically distributed as \( \mathcal{N}(0, \sigma^2_e) \); the random effects \( v_i \) and the error terms \( e_{ij} \) are independent for all \( i = 1, \ldots, k \) and \( j = 1, \ldots, n_i \); \( x_{1ij} \) and \( x_{2ij} \) correspond to the number of corn pixels and soybeans pixels, respectively; and finally \( y_{ij} \) is the number of acres of corn. The parameters estimation results of Model (27) can be found in Sinha & Rao (2009).

The small-area population mean is given by \( \bar{Y}_i = \frac{N_i^{-1}}{N_i} \sum_{j=1}^{N_i} y_{ij} \). Tables 4 and 5 present the bootstrap MSE estimates for the small-area predictors EBLUP and SR, respectively, based on \( B = 1000 \) bootstrap replications. The results for the proposed bootstrap MSE, JNBOOT, and the Jiongo et al. (2013) bootstrap estimator, JHDBOOT, are similar for all the crop areas, regardless of the predictor used. For the (non-robust) EBLUP, the results obtained for the analytical MSE estimators, PR, CCT and CCST, are quite different and usually higher than those of the bootstrap MSE estimators, JHDBOOT.
Table 4: EBLUP Predicted hectares of corn with estimated standard errors.

<table>
<thead>
<tr>
<th>County</th>
<th>Sample Segments</th>
<th>EBLUP Predictor</th>
<th>PR</th>
<th>CCT</th>
<th>CCST</th>
<th>JHDBOOT</th>
<th>JNBOOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerro Gordo</td>
<td>1</td>
<td>122.2</td>
<td>8.4</td>
<td>11.5</td>
<td>n.a.</td>
<td>7.7</td>
<td>7.8</td>
</tr>
<tr>
<td>Hamilton</td>
<td>1</td>
<td>123.2</td>
<td>8.4</td>
<td>8.8</td>
<td>n.a.</td>
<td>7.8</td>
<td>7.4</td>
</tr>
<tr>
<td>Worth</td>
<td>2</td>
<td>113.8</td>
<td>8.4</td>
<td>23.6</td>
<td>n.a.</td>
<td>7.7</td>
<td>8.0</td>
</tr>
<tr>
<td>Humboldt</td>
<td>3</td>
<td>115.4</td>
<td>8.5</td>
<td>8.6</td>
<td>8.7</td>
<td>7.6</td>
<td>7.4</td>
</tr>
<tr>
<td>Franklin</td>
<td>3</td>
<td>136.1</td>
<td>8.0</td>
<td>15.4</td>
<td>15.4</td>
<td>6.4</td>
<td>6.8</td>
</tr>
<tr>
<td>Pocanhontas</td>
<td>3</td>
<td>108.4</td>
<td>8.1</td>
<td>9.4</td>
<td>9.4</td>
<td>6.9</td>
<td>6.7</td>
</tr>
<tr>
<td>Winnibago</td>
<td>3</td>
<td>116.8</td>
<td>8.0</td>
<td>7.2</td>
<td>7.2</td>
<td>7.1</td>
<td>6.8</td>
</tr>
<tr>
<td>Wright</td>
<td>3</td>
<td>122.6</td>
<td>8.1</td>
<td>5.9</td>
<td>5.9</td>
<td>6.9</td>
<td>7.0</td>
</tr>
<tr>
<td>Webster</td>
<td>3</td>
<td>111.0</td>
<td>7.8</td>
<td>8.7</td>
<td>8.7</td>
<td>6.5</td>
<td>6.5</td>
</tr>
<tr>
<td>Hancock</td>
<td>5</td>
<td>124.4</td>
<td>7.4</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.1</td>
</tr>
<tr>
<td>Kossuth</td>
<td>5</td>
<td>113.4</td>
<td>7.4</td>
<td>10.7</td>
<td>10.7</td>
<td>6.1</td>
<td>5.9</td>
</tr>
<tr>
<td>Hardin</td>
<td>6</td>
<td>131.3</td>
<td>7.2</td>
<td>5.3</td>
<td>5.4</td>
<td>6.3</td>
<td>5.9</td>
</tr>
</tbody>
</table>

Note that it is not possible to calculate the analytical linearization-based MSE estimator CCST of Chambers et al. (2013) for the first three domains (Cerro Gordo, Hamilton, and Worth) because their sample sizes are equal to \( n_i = 1 \).

As for the robust predictor SR, it can be noted that the standard errors of the bootstrap of Sinha & Rao (2009), SRBOOT are significantly higher than the proposed bootstrap, JNBOOT, for the first three areas for which the sample sizes are equal to \( n_i = 1 \). These differences decrease when the sample size of the area increases. The standard errors of the two bootstrap procedures tend to be similar when the sample size of the area is 4 and above. On the other hand, the analytical estimators CCT and CCST are quite different from the bootstrap MSE estimators and from one another, except for the case of the robust predictor SR where they are similar (with the exception of the first three areas), but still different from each of the three bootstraps.

Table 5: SR Predicted hectares of corn with estimated standard errors.

<table>
<thead>
<tr>
<th>County</th>
<th>Sample Segments</th>
<th>SR Predictor</th>
<th>CCT</th>
<th>CCST</th>
<th>SRBOOT</th>
<th>JHDBOOT</th>
<th>JNBOOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerro Gordo</td>
<td>1</td>
<td>123.7</td>
<td>6.1</td>
<td>n.a.</td>
<td>9.8</td>
<td>7.6</td>
<td>7.7</td>
</tr>
<tr>
<td>Hamilton</td>
<td>1</td>
<td>125.3</td>
<td>7.8</td>
<td>n.a.</td>
<td>9.6</td>
<td>7.7</td>
<td>7.3</td>
</tr>
<tr>
<td>Worth</td>
<td>1</td>
<td>110.2</td>
<td>19.0</td>
<td>n.a.</td>
<td>9.6</td>
<td>7.7</td>
<td>7.8</td>
</tr>
<tr>
<td>Humboldt</td>
<td>2</td>
<td>114.1</td>
<td>7.8</td>
<td>7.8</td>
<td>8.7</td>
<td>7.6</td>
<td>7.2</td>
</tr>
<tr>
<td>Franklin</td>
<td>3</td>
<td>140.8</td>
<td>10.6</td>
<td>10.6</td>
<td>7.4</td>
<td>6.5</td>
<td>6.8</td>
</tr>
<tr>
<td>Pocanhontas</td>
<td>3</td>
<td>110.8</td>
<td>8.3</td>
<td>8.3</td>
<td>7.5</td>
<td>6.9</td>
<td>6.7</td>
</tr>
<tr>
<td>Winnibago</td>
<td>3</td>
<td>115.2</td>
<td>7.1</td>
<td>7.1</td>
<td>7.4</td>
<td>7.2</td>
<td>6.8</td>
</tr>
<tr>
<td>Wright</td>
<td>3</td>
<td>122.7</td>
<td>6.2</td>
<td>6.2</td>
<td>7.6</td>
<td>6.9</td>
<td>6.9</td>
</tr>
<tr>
<td>Webster</td>
<td>4</td>
<td>113.5</td>
<td>7.5</td>
<td>7.4</td>
<td>6.9</td>
<td>6.5</td>
<td>6.4</td>
</tr>
<tr>
<td>Hancock</td>
<td>5</td>
<td>124.1</td>
<td>5.9</td>
<td>5.9</td>
<td>6.4</td>
<td>6.1</td>
<td>6.3</td>
</tr>
<tr>
<td>Kossuth</td>
<td>5</td>
<td>109.4</td>
<td>8.1</td>
<td>8.1</td>
<td>6.5</td>
<td>6.1</td>
<td>6.0</td>
</tr>
<tr>
<td>Hardin</td>
<td>6</td>
<td>136.9</td>
<td>5.8</td>
<td>5.9</td>
<td>6.3</td>
<td>6.4</td>
<td>6.0</td>
</tr>
</tbody>
</table>

The estimation results from the methods discussed clearly display considerable differ-
ences in their values and standard errors and could lead to different inferences about the small-area means. The analytical MSE estimators give higher standard errors compared to the standard errors produced by bootstrap MSE estimators including the proposed bootstrap JNBOOT. Hence, it is important to compare the accuracy of these methods in a rigorous way. Meanwhile, our results suggest that the latter should be preferred since its theoretical validity has just been established by the main theoretical result of this paper and it is empirically more accurate as per the simulations outcomes obtained in Section 4.

6 Concluding Remarks

In this paper, we considered the problem of bootstrapping the mean squared error of robust small-area estimators. The underlying model is the unit-level model where error variance, random effects and fixed effects can be estimated using existing approaches. Given that robust estimates of the variance components are typically smaller than their nonrobust counterparts it is difficult to construct bootstrap data on the same scale as the original data (Field et al. 2010). We overcome this difficulty by using the nonrobust maximum likelihood estimators for generating the bootstrap samples and apply the robust estimation technique on this sample to obtain outlier-robust bootstrap predictors. It is from this starting point that our proposed MSE estimator is built.

Existing bootstrap MSE procedures that have been proposed in this literature are not justified theoretically, whereas we formally prove the theoretical validity of our proposed bootstrap. This is the first time, to our knowledge, that the asymptotic validity of a bootstrap method for MSE has been formally established for robust small-area estimation. Moreover, the semi-parametric nature of the proposed method makes it particularly attractive, as it does not rely on the often misleading normality assumption. Our theoretical results are derived using an approach similar to Bickel & Freedman (1981) and Freedman (1981) and convergence results established by Huggins (1993). The proofs of the proposed bootstrap MSE estimator are provided for the robust estimator of Sinha & Rao (2009), but the argument can be easily extended to accommodate other robust predictors.

We examined the behaviour of the proposed method through Monte Carlo simulations and compare its performance with five other methods: the bootstrap MSE estimator of Sinha & Rao (2009), the analytical pseudolinearization MSE estimator and linearization-based MSE estimator of Chambers et al. (2013), the bootstrap MSE of Jiongo et al. (2013) and the MSE estimator of Prasad & Rao (1990). The results showed that for all the different robust small-area estimators and all the various modes of contamination considered, the proposed bootstrap MSE performs the best, both in terms of bias and efficiency. Finally, an empirical application using county crops area data from North-Central Iowa farmers and LANDSAT satellite images illustrates the usefulness of the proposed method in practice.

Finally, we note that, although our bootstrap MSE estimator was developed under the linear mixed model, it should be possible to develop a version of this MSE estimator under a semiparametric mixed model that allow for possibly nonparametric effects in the dependency. This presents an avenue for further research.
7 Appendix: Proofs

This section provides the proofs of Conditions (20) - (23) stated in Lemma 3.

Proof of Lemma 3

Proof of (20): \( d_4 \left( F_v, \hat{F}_{uk} \right) \xrightarrow{p} 0 \) as \( k \rightarrow \infty \) and \( d_4 \left( F_e, \hat{F}_{ek} \right) \xrightarrow{p} 0 \) as \( k \rightarrow \infty \).

Using the triangular inequality and a binomial expansion, it can be shown that

\[
\frac{1}{8} d_4 \left( F_v, \hat{F}_{uk} \right)^4 \leq d_4 \left( F_v, F_u \right)^4 + d_4 \left( F_u, \hat{F}_{uk} \right)^4
\]

and

\[
\frac{1}{8} d_4 \left( F_u, \hat{F}_{uk} \right)^4 \leq d_4 \left( F_u, F_k \right)^4 + d_4 \left( F_k, \hat{F}_{uk} \right)^4
\]

and this implies that

\[
d_4 \left( F_v, \hat{F}_{uk} \right)^4 \leq 8 d_4 \left( F_v, F_u \right)^4 + 64 d_4 \left( F_u, F_k \right)^4 + 64 d_4 \left( F_k, \hat{F}_{uk} \right)^4,
\]

Notice that \( \hat{u}_i = u_i + O_p(k^{-1/2}) \) where \( u_i = \sqrt{p_i} (v_i + \bar{e}_i) \). By the stability of the family distribution of \( v_i \) and \( e_i \), it follows that \( u_i \) is distributed as \( F_v \).

We then have \( d_4 \left( F_v, F_u \right)^4 = 0 \), and by Lemma 8.4 of Bickel & Freeman (1981) we also have \( d_4 \left( F_u, F_k \right)^4 \xrightarrow{p} 0 \) as \( k \rightarrow \infty \).

On the other hand, since \( F_{uk} \) and \( \hat{F}_{uk} \) are two empirical distributions, this implies that

\[
d_4 \left( F_{uk}, \hat{F}_{uk} \right)^4 \leq \frac{1}{k} \sum_{i=1}^{k} (\hat{u}_i - u_i)^4 = O_p(k^{-2}),
\]

so that, finally, \( d_4 \left( F_v, \hat{F}_{uk} \right) \xrightarrow{p} 0 \) as \( k \rightarrow \infty \).

Likewise, we have

\[
d_4 \left( F_e, \hat{F}_{ek} \right)^4 \leq 8 d_4 \left( F_e, F_s \right)^4 + 64 d_4 \left( F_s, F_k \right)^4 + 64 d_4 \left( F_k, \hat{F}_{ek} \right)^4,
\]

where \( \epsilon_{ij} = (1 - \tau_i) v_i + e_{ij} - \tau_i \bar{e}_i, i = 1, \ldots, k \quad j = 1, \ldots, n_i \).

Note that the sampling residuals are \( \hat{e}_{ij} = \hat{e}_{ij} - \frac{1}{n} \sum_{g=1}^{k} \sum_{l \in s_g} \hat{e}_{gl} \) and that the \( \epsilon_{ij} \) are independent and identically distributed with the same distribution \( F_s \). Then \( d_4 \left( F_e, F_s \right) = 0 \), and by lemma 8.4 of Bickel & Freeman (1981) \( d_4 \left( F_e, F_k \right)^4 \xrightarrow{p} 0 \).

On the other hand, we can write \( \hat{e}_{ij} = \epsilon_{ij} + O_p(k^{-1/2}) \),

which implies that \( \hat{e}_{ij} - \epsilon_{ij} = O_p(k^{-1/2}) \). It follows that

\[
d_4 \left( F_{ek}, \hat{F}_{ek} \right)^4 \leq \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\hat{e}_{ij} - \epsilon_{ij})^4
\]

\[
= O_p(k^{-2}) \xrightarrow{p} 0 \quad \text{as} \quad k \rightarrow \infty.
\]
Hence, \( d_4 \left( F_c, \hat{F}_{ck} \right) \xrightarrow{p} 0 \) as \( k \to \infty \). \( \square \)

**Proof of (21):** \( J_1 \xrightarrow{p} J_1 \), where

\[
J_1^* = \lim_{k \to \infty} \sum_{j=1}^{k} I_{1k}^{-1/2} X_i A_k^{-1} B_k = \lim_{k \to \infty} \sum_{j=1}^{k} I_{1k}^{-1/2} X_i A_k^{-1} B_k \]

Denote: \( X_i = \left( n_i, \tilde{X}_i \right) \), where \( X_i \) is a \( n_i \times (p - 1) \) matrix of auxiliary variables.

\[
J_{1k} = \sum_{i=1}^{k} I_{1k}^{-1/2} X_i A_k^{-1} B_k = \lim_{k \to \infty} \sum_{i=1}^{k} \left[ J_{111k} J_{112k} J_{122k} \right],
\]

where

\[
J_{111k} = \frac{1}{k} \sum_{i=1}^{k} \left\{ \frac{\sigma^2}{\sigma^2} \rho^2 a_{11} + \frac{\sigma^2}{\sigma^2} (\rho_i - \rho^2) (a_2 - a_{11}) \right\},
\]

\[
J_{112k} = \sqrt{\frac{k}{n}} \left[ \frac{1}{k} \sum_{i=1}^{k} \left\{ \frac{\sigma^2}{\sigma^2} \rho^2 a_{11} + \frac{\sigma^2}{\sigma^2} (\rho_i - \rho^2) (a_2 - a_{11}) \right\} \right] \hat{X}_i.
\]

\[
J_{122k} = \frac{k}{n} \left[ \frac{1}{k} \sum_{i=1}^{k} \left\{ \frac{\sigma^2}{\sigma^2} \rho^2 a_{11} + \frac{\sigma^2}{\sigma^2} (\rho_i - \rho^2) (a_2 - a_{11}) \right\} \right] \tilde{X}_i^\top \tilde{X}_i + \sum_{i=1}^{k} \frac{n_i}{n} \left\{ \frac{1}{n_i} \tilde{X}_i^\top \tilde{X}_i - \tilde{X}_i^\top \tilde{X}_i \right\}.
\]

and \( \rho_i = \frac{n_i \sigma^2}{\sigma^2 + n_i \sigma^2} \).

Taking the expression to the limit as \( k \to \infty \), yields

\[
J_1 = \left( \begin{array}{ccc}
J_{111} & J_{112} & J_{122}
\end{array} \right),
\]

where

\[
J_{111} = \frac{\sigma^2}{\sigma^2} \nu a_{11} + \frac{\sigma^2}{\sigma^2} (\nu_1 - \nu_2) (a_2 - a_{11}),
\]

\[
J_{112} = \sqrt{c} \lim_{k \to \infty} \left[ \frac{1}{k} \sum_{i=1}^{k} \left\{ \frac{\sigma^2}{\sigma^2} \rho^2 a_{11} + \frac{\sigma^2}{\sigma^2} (\rho_i - \rho^2) (a_2 - a_{11}) \right\} \right] \hat{X}_i,
\]

\[
J_{122} = \lim_{k \to \infty} \left[ \frac{1}{k} \sum_{i=1}^{k} \left\{ \frac{\sigma^2}{\sigma^2} \rho^2 a_{11} + \frac{\sigma^2}{\sigma^2} (\rho_i - \rho^2) (a_2 - a_{11}) \right\} \right] \tilde{X}_i^\top \tilde{X}_i + \sum_{i=1}^{k} \frac{n_i}{n} \left\{ \frac{1}{n_i} \tilde{X}_i^\top \tilde{X}_i - \tilde{X}_i^\top \tilde{X}_i \right\}.
\]

with \( \nu_1 = \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} \rho_i \), and \( \nu_2 = \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} \rho_i^2 \).
Using the same reasoning we can obtain the bootstrap version $J_1^*$ of $J_1$ defined by

$$J_1^* = \left( \begin{array}{ccc} J_{111}^* & J_{112}^* & J_{122}^* \end{array} \right),$$

where

$$J_{111}^* = \frac{\sigma_e^4}{\sigma_v^4} \hat{\nu}_1 a_{11}^* + \frac{\hat{\sigma}_e^2}{\sigma_v^2} \left( \hat{\nu}_1 - \hat{\nu}_2 \right) (a_2^* - a_{11}^*),$$

$$J_{112}^* = \sqrt{c} \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} \left\{ \frac{\hat{\sigma}_e^4}{\sigma_v^4} \rho_{11}^2 a_{11}^* + \frac{\hat{\sigma}_e^2}{\sigma_v^2} \left( \hat{\rho}_i - \hat{\rho}_2 \right) (a_2^* - a_{11}^*) \right\} \tilde{X}_i,$$

$$J_{122}^* = \lim_{k \to \infty} \left[ \frac{1}{k} \sum_{i=1}^{k} \left\{ \frac{\hat{\sigma}_e^4}{\sigma_v^4} \rho_{11}^2 a_{11}^* + \frac{\hat{\sigma}_e^2}{\sigma_v^2} \left( \hat{\rho}_i - \hat{\rho}_2 \right) (a_2^* - a_{11}^*) \right\} \tilde{X}_i \tilde{X}_i + \frac{n_i}{n} \left\{ \frac{1}{n_i} \tilde{X}_i \tilde{X}_i - \tilde{\tilde{X}}_i \tilde{\tilde{X}}_i \right\} \right];$$

and

$$\hat{\nu}_1 = \frac{1}{k} \sum_{i=1}^{k} \hat{\rho}_i, \quad \hat{\rho}_2 = \frac{1}{k} \sum_{i=1}^{k} \hat{\rho}_i^2, \quad a_{11}^* = E_* \{ \psi^{(r)}(\eta) \}, \quad a_{11}^* = E_* \{ \psi^{(r)}(\eta) \}, \quad j_1 \neq j_2.$$

By Lemmas 2.1 and 8.5 of Bickel and Freedman (1981), $a_{11}^*$ and $a_{2}^*$ converge in probability to $a_{11}$ and $a_2$ respectively. Lemma 1 obtained above implies that $(\hat{\sigma}_e^2, \hat{\nu}_1, \hat{\nu}_2)$ converge in probability to $(\sigma_e^2, \nu_1, \nu_2)$. Hence, by continuity, $J_1^* \xrightarrow{p} J_1$. □

**Proof of (22):** $J_2 \xrightarrow{p} J_2$, where

$$J_2^* = \lim_{k \to \infty} \sum_{i=1}^{k} I_{2k}^{-1/2} E_* \left\{ \Psi_2(y_i^*, X_i, \theta) \Psi_2^T(y_i^*, X_i, \theta) \right\} I_{2k}^{-1/2}$$

The proof of (22) proceeds exactly as for (21). The derivation is however more cumbersome because it requires to calculate fourth-order moments. Thus after a lengthy algebraic expansion, one could write

$$J_{22k} = \left( \begin{array}{cc} J_{211k} & J_{212k} \\ J_{212k} & J_{222k} \end{array} \right),$$

with

$$J_{211k} = \frac{1}{\sigma_e^2} \left\{ \frac{1 - \eta}{\eta} \right\}^2 \left( \frac{a_{11} - \eta a_2}{1 - \eta} \right)^2 \left( \frac{\rho - 1}{\eta} \right)^2 k + A_{11k},$$

where $\eta = \frac{\sigma_v}{\sigma_e^2 + \sigma_v^2}$, and $A_{11k}$ is a bounded sequence of real numbers given by

$$A_{11k} = \frac{1}{\sigma_e^2 \eta^2} \left( a_4 - 4a_{31} - 4a_{22} - 12a_{211} - 6a_{1111} \right) \left( \rho - 3 \rho^2 + 3 \rho^3 + \rho^4 \right)$$

$$+ \frac{1}{\sigma_e^2 \eta^2} \left\{ 4a_{31} + 4a_{22} - 18a_{211} + 11a_{1111} - (a_2 - a_{11})^2 \right\} \left( \rho^2 - 2 \rho^3 + \rho^4 \right)$$

$$+ \frac{1}{\sigma_e^2 \eta^2} \left\{ 6a_{211} - 6a_{1111} - 2a_{11} (a_2 - a_{11}) \right\} \left( \rho^3 - \rho^4 \right)$$

$$+ \frac{(1 - \eta)^2}{\sigma_e^2 \eta^2} \left( a_{1111} - a_{1111}^2 \right) \rho^4$$

$$= A_{11} \left( \sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \rho, \rho^2, \rho^3, \rho^4 \right).$$
The numbers $a_4, a_{31}, a_{22}, a_{211}, a_{1111}$ are fourth-order moments defined by

$$a_4 = E_m \{ \psi^4_b(r_{ij}) \}, \quad a_{31} = E_m \{ \psi^3_b(r_{ij}) \psi_b(r_{ij}) \}, \quad a_{22} = E_m \{ \psi^2_b(r_{ij}) \psi^2_b(r_{ij}) \},$$

$$a_{211} = E_m \{ \psi^2_b(r_{ij}) \psi_b(r_{ij}) \psi_b(r_{ij}) \}, \quad a_{1111} = E_m \{ \psi_b(r_{ij}) \psi_b(r_{ij}) \psi_b(r_{ij}) \psi_b(r_{ij}) \},$$

where $j_1 \neq j_2, \ j_1 \neq j_3, \ j_1 \neq j_4, \ j_2 \neq j_3, \ j_2 \neq j_4, \ j_3 \neq j_4,$

and the numbers $\tilde{\rho}^l, \ l = 1, 2, 3, 4$ are defined by $\tilde{\rho}^l = \frac{1}{k} \sum_{i=1}^k \rho^l_i, \ l = 1, 2, 3, 4.$

Note that $A_{111}(\cdot)$, as defined above, is a continuous function of its arguments.

Likewise,

$$J_{212k} = \frac{1}{\sigma^4_e} \left( \frac{1 - \eta}{\eta} \right) \left( \frac{a_{11} - \eta a_2}{1 - \eta} \right)^2 \left( 1 - \frac{1}{\eta} \tilde{\rho}^2 \right) \left( 1 - \tilde{\rho} \frac{\sqrt{k}}{n} \right) \sqrt{n k}$$

$$+ \frac{1}{\sigma^4_e \eta^2} \left( a_{211} - a_{1111} + a_{11}(a_2 - a_{11}) \right) \sqrt{n \frac{k}{k}} + A_{12k} \sqrt{n \frac{k}{k}},$$

where, as for the above derivation, $A_{12k}$ is a bounded sequence of real numbers which depends on the fourth moments of $\psi_b(r_{ij})$ and the sample moments of $\rho^l_i, \ l = 1, 2, 3, 4.$ That is, $A_{12k} = A_{12} \left( \sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \tilde{\rho}, \tilde{\rho}^2, \tilde{\rho}^3, \tilde{\rho}^4 \right),$ and $A_{12}(\cdot)$ is a continuous function of its arguments.

The last component $J_{222k}$ of matrix $J_{2k}$ is given by

$$J_{222k} = \frac{1}{\sigma^4_e} \left( \frac{a_{11} - \eta a_2}{1 - \eta} \right)^2 \left\{ 1 - \left( \tilde{\rho} + \frac{1}{\eta} \tilde{\rho} - \frac{1}{\eta} \tilde{\rho}^2 \right) \frac{k}{n} \right\}^2 n$$

$$+ \frac{1}{\sigma^4_e (1 - \eta)^2} \left( a_{22} - 2a_{211} + a_{1111} - (a_2 - a_{11})^2 \right) \left( 1 - \frac{1}{n} \sum_{i=1}^k a_i^2 \right)$$

$$+ \frac{1}{\sigma^4_e (1 - \eta)^2} \left( a_4 - 3a_{22} - 4a_{31} + 12a_{211} - a_{1111} + 2(a_2 - a_{11})^2 \right) + A_{22k} \times \frac{k}{n},$$

where, as above, $A_{22k}$ is a bounded sequence of real numbers and can be written as $A_{22k} = A_{22} \left( \sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \tilde{\rho}, \tilde{\rho}^2, \tilde{\rho}^3, \tilde{\rho}^4 \right),$ where $A_{22}(\cdot)$ is also a continuous function of its arguments.

Since Assumption A1 implies that $\frac{k}{n}$ converges to a possibly zero constant $c$, and the limit of $J_{2k}$ is assumed to always exist and be finite by Assumption A4, then we must have

$$a_{11} - \eta a_2 = 0,$$
$$a_{211} - a_{1111} + a_{11}(a_2 - a_{11}) = 0,$$
$$a_{22} - 2a_{211} + a_{1111} - (a_2 - a_{11})^2 = 0.$$

It follows that

$$J_2 = \lim_{k \to \infty} J_{2k} = \begin{pmatrix} J_{211} & J_{212} \\ J_{212} & J_{222} \end{pmatrix},$$

24
where

\[ J_{211} = \frac{1}{\sigma_e^4 \eta^2} (a_4 - 4a_{31} - 4a_{22} - 12a_{211} - 6a_{1111}) (\nu_1 - 3\nu_2 + 3\nu_3 - \nu_4) \]

\[ + \frac{1}{\sigma_e^4 \eta^2} \{ 4a_{31} + 4a_{22} - 18a_{211} + 11a_{1111} - (a_2 - a_{11})^2 \} (\nu_2 - 2\nu_3 + \nu_4) \]

\[ + \frac{1 - \eta}{\sigma_e^4 \eta^3} \{ 6a_{211} - 6a_{1111} - 2a_{11} (a_2 - a_{11}) \} (\nu_3 - \nu_4) \]

\[ + \frac{(1 - \eta)^2}{\sigma_e^4 \eta^4} (a_{1111} - a_{11}) \nu_4, \]

\[ = A_{11} \left( \sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \nu_1, \nu_2, \nu_3, \nu_4 \right), \]

\[ J_{212} = \sqrt{c} A_{12} \left( \sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \nu_1, \nu_2, \nu_3, \nu_4 \right). \]

and

\[ J_{222} = \frac{1}{\sigma_e^4 (1 - \eta)^2} \{ a_4 - 3a_{22} - 4a_{31} + 12a_{211} - a_{1111} + 2(a_2 - a_{11})^2 \} \]

\[ + c A_{22} \left( \sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \nu_1, \nu_2, \nu_3, \nu_4 \right). \]

Likewise, the bootstrap version \( J^*_2 \) of \( J_2 \) is given by

\[ J^*_2 = \lim_{k \to \infty} J^{*k}_{2} = \left( \frac{J^{*k}_{211}}{J^{*k}_{212}}, \frac{J^{*k}_{212}}{J^{*k}_{222}} \right), \]

where

\[ J^{*k}_{211} = A_{11} \left( \sigma_e^2, \eta, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, \nu_1^*, \nu_2^*, \nu_3^*, \nu_4^* \right), \]

\[ J^{*k}_{212} = \sqrt{c} A_{12} \left( \sigma_e^2, \eta, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, \nu_1^*, \nu_2^*, \nu_3^*, \nu_4^* \right), \]

\[ J^{*k}_{222} = \frac{1}{\sigma_e^4 (1 - \eta)^2} \{ a_4^* - 3a_{22}^* - 4a_{31}^* + 12a_{211}^* - a_{1111}^* + 2(a_2^* - a_{11}^*)^2 \} \]

\[ + c A_{22} \left( \sigma_e^2, \eta, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, \nu_1^*, \nu_2^*, \nu_3^*, \nu_4^* \right). \]

By Lemmas 2.1 and 8.5 of Bickel and Freedman (1981), \( a_{11}^*, a_2^*, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^* \), converge in probability to \( a_{11}, a_{2}, a_4, a_{31}, a_{22}, a_{211}, a_{1111} \), respectively. Since, by Lemma 1 above \( (\hat{\sigma}_e^2, \hat{\sigma}_e^2, \hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\nu}_4) \) converges in probability to \( (\sigma_e^2, \sigma_e^2, \nu_1, \nu_2, \nu_3, \nu_4) \) it then follows by the continuous mapping theorem that \( J^*_2 \xrightarrow{p} J_2 \). \( \square \)

**Proof of (23):** \( G^*(\theta) \xrightarrow{p} G(\theta) \) where

\[ G^*(\hat{\theta}) = \lim_{k \to \infty} \sum_{i=1}^{k} \left[ -I_{1k}^{-1/2} E_s \left\{ \frac{\partial \Psi_i(y_i^*, X, \hat{\theta})}{\partial \hat{\theta}} \right\} I_{1k}^{-1/2} \right. \]

\[ \left. -I_{2k}^{-1/2} E_s \left\{ \frac{\partial \Psi_i(y_i^*, X, \hat{\theta})}{\partial \hat{\delta}} \right\} I_{2k}^{-1/2} \right], \]

and

\[ G(\theta) = \lim_{k \to \infty} \sum_{i=1}^{k} \left[ -I_{1k}^{-1/2} E_m \left\{ \frac{\partial \Psi_i(y_i, X, \theta)}{\partial \theta} \right\} I_{1k}^{-1/2} \right. \]

\[ \left. -I_{2k}^{-1/2} E_m \left\{ \frac{\partial \Psi_i(y_i, X, \theta)}{\partial \delta} \right\} I_{2k}^{-1/2} \right]. \]
We use the same reasoning as for the proofs of (21) and (22). A straightforward expansion of the components of \( G_k(\theta) \) allows to see that we can express it as a sum of two components, one which is a bounded sequence and another which depends on \( a_2, a_{11}, d_2, d_{11} \), where \( d_2 = E_m \{ r_{ij} \psi_b(r_{ij}) \psi_b(r_{ij}) \} \) and \( d_{11} = E_m \{ r_{ij} \psi_b(r_{ij}) \psi_b(r_{ij}) \} \), \( j_1 \neq j_2 \).

By Assumptions A1 and A5 which respectively assume that \( \frac{k}{n} \) converges to a possibly zero constant \( c \in [0, 1] \) and that the limit \( G(\theta) \) of \( G_k(\theta) \) always exists and is finite, we must have

\[
a_2 - a_{11} - d_2 + d_{11} = 0.
\]

It then follows that

\[
G(\theta) = \begin{bmatrix}
G_{111} & G_{112} & 0 & 0 \\
G_{112}^T & G_{122} & 0 & 0 \\
0 & 0 & G_{211} & G_{212} \\
0 & 0 & G_{212} & G_{222}
\end{bmatrix},
\]

where

\[
G_{111} = \frac{d_1}{\sigma^2} \frac{1 - \eta}{\eta} \nu_1,
\]
\[
G_{112} = \frac{d_1}{\sigma^2} \frac{1 - \eta}{\eta} \sqrt{c} \lim_{k \to \infty} \left[ \frac{1}{k} \sum_{i=1}^{k} \rho_i \bar{X}_i \right],
\]
\[
G_{122} = \frac{d_1}{\sigma^2} \lim_{k \to \infty} \left[ \frac{1 - \eta}{\eta} \frac{1}{k} \sum_{i=1}^{k} \rho_i \bar{X}_i^T \bar{X}_i + \sum_{i=1}^{k} \frac{n_i}{n} \left\{ \frac{1}{n_i} \bar{X}_i^T \bar{X}_i - \bar{X}_i^T \bar{X}_i \right\} \right],
\]
\[
G_{211} = \frac{1}{\sigma^4} \left( \frac{1 - \eta}{\eta} \right)^2 \{(1 - \eta) a_2 + d_{11}\} \nu_2,
\]
\[
G_{212} = \frac{1}{\sigma^4} \left( \frac{1 - \eta}{\eta} \right)^2 \sqrt{c} \{(1 - \eta) a_2 + d_{11}\} (\nu_1 - \nu_2),
\]
\[
G_{222} = \frac{a_2}{\sigma^2} + \frac{c}{\sigma^2} \left[ a_2 \{2 \nu_2 - 3 \nu_1 + (\nu_1 - \nu_2)\eta\} + \frac{1 - \eta}{\eta} (\nu_1 - \nu_2) d_{11} \right];
\]

where \( d_1 = E_m \{ \psi_b(r_{ij}) \} \).

Likewise, we have

\[
G^*(\hat{\theta}) = \begin{bmatrix}
G^*_{111} & G^*_{112} & 0 & 0 \\
G^*_{112}^T & G^*_{122} & 0 & 0 \\
0 & 0 & G^*_{211} & G^*_{212} \\
0 & 0 & G^*_{212} & G^*_{222}
\end{bmatrix},
\]

26
where 
\[
\begin{align*}
G_{111}^* &= \frac{d_1^*}{\tilde{\sigma}^2} \frac{1 - \hat{\eta}}{\hat{\eta}} \nu_1, \\
G_{112}^* &= \frac{d_1^*}{\tilde{\sigma}^2} \frac{1 - \hat{\eta}}{\hat{\eta}} \sqrt{c} \lim_{k \to \infty} \left[ \frac{1}{k} \sum_{i=1}^{k} \hat{\rho}_i \tilde{X}_i \right], \\
G_{122}^* &= \frac{d_1^*}{\tilde{\sigma}^2} \lim_{k \to \infty} \left[ \frac{1 - \hat{\eta}}{\hat{\eta}} \frac{1}{c} \sum_{i=1}^{k} \hat{F}_i \tilde{X}_i \tilde{X}_i + \sum_{i=1}^{k} \frac{n_i}{n} \left\{ \frac{1}{n_i} \tilde{X}_i^{\top} \tilde{X}_i - \bar{\tilde{X}}^{\top} \bar{\tilde{X}} \right\} \right], \\
G_{211}^* &= \frac{1}{\tilde{\sigma}^2} \left( \frac{1 - \hat{\eta}}{\hat{\eta}} \right)^2 \{ (1 - \hat{\eta})a_2^* + d_{11}^* \} \nu_2, \\
G_{212}^* &= \frac{1}{\tilde{\sigma}^2} \left( \frac{1 - \hat{\eta}}{\hat{\eta}} \right) \sqrt{c} \{ (1 - \hat{\eta})a_2^* + d_{11}^* \} \{ \nu_1 - \nu_2 \}, \\
G_{222}^* &= \frac{a_2^*}{\tilde{\sigma}^2} + \frac{c}{\tilde{\sigma}^2} \left[ a_2^* \{ 2\nu_2 - 3\nu_1 + (\nu_1 - \nu_2)\hat{\eta} \} + \frac{1 - \hat{\eta}}{\hat{\eta}} (\nu_1 - \nu_2)d_{11}^* \right];
\end{align*}
\]

with 
\[
\begin{align*}
d_1^* &= E_\star \left\{ \psi_{b}(r_{ij}^*) \right\}, \quad d_2^* = E_\star \left\{ \{r_{ij}^* \psi_{b}(r_{ij}^*)\} \right\}, \quad d_{11}^* = E_\star \left\{ \{r_{ij}^* \psi_{b}(r_{ij}^*)\} \right\}, \quad j_1 \neq j_2,
\end{align*}
\]

It follows by the continuous mapping theorem that \( G^*(\hat{\theta}) \) converges in probability to \( G(\theta) \) as \( k \to \infty \). □

References


