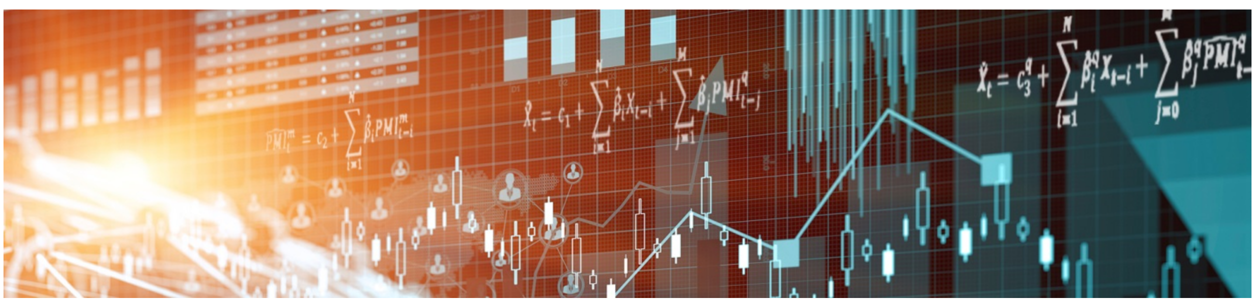
Staff Working Paper/Document de travail du personnel 2018-28

Bootstrapping Mean Squared Errors of Robust Small-Area Estimators: Application to the Method-of-Payments Data



by Valéry D. Jiongo and Pierre Nguimkeu

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2018-28

June 2018

Bootstrapping Mean Squared Errors of Robust Small-Area Estimators: Application to the Method-of-Payments Data

by

Valéry D. Jiongo[1] and Pierre Nguimkeu[2]

[1] Currency Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
vdongmo@cmhc-schl.gc.ca (current address)

2 Department of Economics
Georgia State University
Atlanta, Georgia, USA 30302-3992
nnguimkeu@gsu.edu

ISSN 1701-9397

Acknowledgements

The authors would like to thank Jon Rao, Louis-Paul Rivest and colleagues at Statistics Canada including Mike Hidiroglou, Jean-François Beaumont, François Verret, Wesley Benjamin, Chris Mohl and Wesley Yung for useful comments and suggestions on the earlier draft of this paper. We also thank Kim P. Huynh, Heng Chen, Edouard Tsagué and Geoffrey Dunbar for helpful comments and suggestions. We thank Timo Schmid and Nicola Salvati for providing their code for the robust linear mixed-model estimation and the analytical mean squared error estimation, respectively. Valéry D. Jiongo is particularly grateful to David Haziza and Pierre Duchesne for their availability and extensive advice. We thank Maren Hansen for writing support and Emma McGuire, Chiyoung Ahn and Stan Hatko for valuable research assistance. The views expressed in this paper are solely those of the authors. No responsibility should be attributed to the Bank of Canada.

Abstract

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 well and outperforms existing ones. We also apply our procedure to the estimation of the total volume and value of cash, debit card and credit card transactions in Canada as well as in its provinces and subgroups of households. In particular, we find that there is a significant average annual decline rate of 3.1 percent in the volume of cash transactions, and that this decline is relatively higher among high-income households living in heavily populated provinces. Our bootstrap estimator also provides indicators of quality useful in selecting the best small-area predictors from among several alternatives in practice.

Bank topics: Econometric and statistical methods; Bank notes
JEL codes: C13, C15, C83, E, E41

Résumé

Nous proposons une nouvelle procédure bootstrap pour estimer l'erreur quadratique moyenne associée aux estimateurs sur petits domaines robustes. La validité asymptotique du bootstrap proposé est formellement établie et ses propriétés en échantillons finis sont examinées à l'aide de simulations de Monte Carlo. Les résultats montrent que notre procédure est plus performante que les méthodes existantes. Nous appliquons ensuite celle-ci à l'estimation du volume total et de la valeur totale des transactions monétaires au comptant, par carte de débit et carte de crédit effectuées au Canada, ainsi qu'au niveau des provinces et des sous-groupes de ménages. Les résultats font ressortir une décroissance significative du volume annuel des transactions au comptant, de l'ordre de 3,1 % en moyenne, qui est d'ailleurs plus prononcée dans le cas des ménages à hauts revenus vivant dans les provinces les plus peuplées. Notre estimateur bootstrap permet aussi de construire des indicateurs de qualité permettant de sélectionner les meilleurs prédicteurs sur petits domaines, parmi plusieurs possibilités.

Sujets : Méthodes économétriques et statistiques; Billets de banque
Codes JEL : C13, C15, C83, E, E41

Non-Technical Summary

Traditionally, Bank of Canada surveys have aimed to adequately cover the national population and five regions: Atlantic, Quebec, Ontario, Prairies and British Columbia. However, due to sample size limitations at fine geographic levels, reliable figures are often not available for all provinces. Therefore, this paper overcomes the small sample size issue by using a method called small-area estimation. Applying this method to the Bank of Canada's 2009 and 2013 Methods-of-Payment (MOP) surveys, we compute estimates for all the provinces and for subgroups of the population within each province. Specifically, we estimate the total volume and value of transactions for low-, medium- and high-income household classes within each of the 10 Canadian provinces.

Another perennial problem in samples of economics surveys is the presence of outliers. In recent years, methods have been developed that can robustly estimate in the presence of outliers in a small area. However, assessing the quality of robust small-area estimates remains a challenge.

This paper proposes a new procedure to evaluate their quality, based on a resampling method. We formally prove the theoretical validity of the proposed method. We also assess how well our quality estimator performs on finite samples, using simulations. The results show that our procedure performs well and outperforms its competitors.

We illustrate how to apply the proposed method by analyzing the survey data from the 2009 and 2013 MOP surveys. We identify approximately 1.2 percent of the sample units as outliers. They mostly occur in the province of Quebec and correspond to individuals who are either cash-intensive users or non-cash users. The results suggest that this method is quite applicable to survey data and is a useful tool to select the best small-area predictor.

The results of the best predictor show that, at the national level, the volume share and the value share of cash transactions are similar to those found in Henry et al. (2015) and Fung et al. (2015). We go further and provide estimates at both the national and the household-income-group levels. We note a significant average annual decline rate of 3.1 percent in total volume at the national level. At a finer level, the annual decline in cash transactions is relatively higher among high-income households living in the most populated provinces, such as Ontario and Quebec.

We conclude the paper with some caveats and provide future potential extensions that will better account for the count data modelling and the large proportion of outliers in the resample data.

1 Introduction

Traditionally, sample surveys have been used to produce estimators of totals and means of items of interest for large areas (or domains). Such estimators are "direct" in the sense of using only the domain-specific sample data, and the sample sizes are large enough to provide reliable direct estimators that are design-based and avoid specific modelling assumptions. Reliability of the estimators is typically measured by mean squared error (MSE), and many methods of estimating MSE are available. However, due to cost and operational considerations, it is seldom possible to procure a large enough overall sample size to support direct estimators for all domains of interest, in particular small areas like municipalities, counties, subgroups of populations, etc. Yet the demand for reliable small-area statistics has greatly increased in recent years, and it becomes necessary to use model-based "indirect" estimators that can increase the reliability of estimators by borrowing auxiliary information across related areas through linking models. Rao and Molina (2015) provide a recent review as well as a comprehensive account of theory and methods used for small-area estimation (SAE). Small area models may be broadly classified into two types: (a) area-level models that relate direct area estimates to area-level covariates and (b) unit-level models that relate the unit item values to associated unit-level covariates with known area means. Unit-level models, when unit-level variables are available for the sample, provide more efficient estimators than area-level models. However, unit-level models are sensitive to extreme observations (or outliers) in the unit-level data, and robust methods of estimation that are not sensitive to outliers are therefore needed (Fellner 1986; Stahel and Welsh 1997). This is the case considered in the present paper and applied to Bank of Canada method-of-payment unit-level sample data.

Methods used to study small-area point estimates based on unit-level models are numerous in the literature. The empirical best linear unbiased predictor (EBLUP) is a classical example useful for efficiently estimating the small-area means under normality assumptions. However, it can also be highly influenced by the presence of outliers in the data. Chambers and Tzavidis (2006) suggest regression M-quantiles aiming at overcoming the issue of outliers by avoiding conventional Gaussian assumptions, as well as problems associated with the specification of random effects. Tzavidis et al. (2010) study robust prediction of small-area means and quantiles where the small-area estimator is a functional of a predictor of the small-area cumulative distribution function. Sinha and Rao (2009), Chambers et al. (2014), and Jiongo et al. (2013) study the robustified versions of the classical EBLUP to downweight influential observations in the data. MSE procedures required to estimate the precision of these robust point estimators have also received some attention. Sinha and Rao (2009) propose to estimate the MSE using a parametric bootstrap procedure based on the robust EBLUP estimators of the underlying linear mixed model. But, as pointed out by Jiongo et al. (2013), the use of robust variance estimates to generate bootstrap replicas leads to bootstrap samples whose variability is significantly smaller than the variability in the original data. Other analytical and bootstrap procedures for the MSE of robust empirical best linear unbiased predictors (REBLUPs) have been proposed in Chambers et al. (2014) and Jiongo et al. (2013), respectively. However, their theoretical validity has not been formally established, and their empirical performance is not fully satisfactory (as evidenced by the simulations results in Section 4 below).

This paper proposes a new semi-parametric bootstrap procedure for estimating the MSE within the unit-level model framework. Since robust estimates of the variance components are typically smaller than their non-robust 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 original sample data, and the resulting MSE estimator therefore has improved coverage rates.

The theoretical validity of our bootstrap procedure is proven using an approach similar to the one employed by Bickel and Freedman (1981) and Freedman (1981), and we extend their 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 an MSE estimator of robust small-area estimators. Although our proofs and procedure are based on the Sinha and Rao (2009) robust estimator, the derivation can be easily adapted to other existing robust predictors based on the unit-level linear mixed model. A Monte Carlo simulation study computes the relative biases and relative root mean squared error rates of the proposed bootstrap MSE estimator and compares them favourably to several existing analytical and bootstrap alternatives. These include the bootstrap MSE estimator of Sinha and Rao (2009), the analytical pseudolinearization MSE estimator and linearization-based MSE estimators of Chambers et al. (2014), the bootstrap MSE of Jiongo et al. (2013) and the MSE estimator of Prasad and Rao (1990). This comparison is provided for different robust small-area point estimators and various modes of data contamination.

This paper also applies our bootstrap procedure to the estimation of the total volume and value of cash, debit card and credit card transactions in Canada using data from the Bank of Canada's 2009 and 2013 Method-of-Payments (MOP) surveys. The proposed bootstrap provides useful quality indicators that help to determine the best predictor in practice. In particular, with these data, our bootstrap selected the estimator of Jiongo et al. (2013) (henceforth denoted JHD) as the best predictor from among all the above-mentioned alternatives. The empirical results obtained for the JHD estimator are consistent with those presented by Henry et al. (2015) and Fung et al. (2015) at the national level. However, our paper has the particular advantage of making use of small-area estimation techniques to compute reliable estimates as well as their precisions at the level of the provinces and nested subgroups of household income. In particular, from our estimates and the proposed bootstrap MSE we found a significant annual average decline rate of 3.1 percent in the volume of cash transactions at the national level. Estimates obtained at more disaggregated levels show that the annual decline in cash transactions is actually relatively higher among high-income households living in the most populated provinces such as Ontario and Quebec.

The rest of the paper is organized as follows. Section 2 presents the model and 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 application is provided in Section 5. Concluding remarks are given in Section 6, followed by a technical appendix.

2 Preliminairies

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 and Rao (2009) robust estimator that is used to construct our bootstrap MSE estimation.

2.1 Underlying Model and EBLUP

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

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

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, administrative data, other surveys) are used for small-area estimation. Denote by $x_{ij} = (x_{1ij}, \ldots, x_{pij})^\top$ 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

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

where β is an unknown p-dimensional fixed-effects regression parameter vector, and the regressor x_{ij} is a p-dimensional vector of observed responses. The basic unit-level model assumes that the area-specific random effects v_i follow an independent $N(0, \sigma_v^2)$ and are independent of the unit errors e_{ij}, which are assumed to be independent $N(0, \sigma_e^2)$. Model (1) can be rewritten as a special case of the general linear mixed model with block diagonal covariance structure as follows:

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

where y_i is the 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 a n_i-vector of ones. Denote by $\theta = (\beta^\top, \delta^\top)^\top$ the vector of model parameters, where $\delta = (\sigma_e^2, \sigma_v^2)^\top$ is the vector of variance parameters. The variance-covariance matrix of y_i is given by $V_i = \sigma_e^2 I_{n_i} + \sigma_v^2 1_{n_i} 1_{n_i}^\top$, 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. Under these normality assumptions, the empirical best linear unbiased predictor (EBLUP) of the area mean is given by

$$\bar{Y}_{iEBLUP} = N_i^{-1}\left(\sum_{j\in s_i} y_{ij} + \sum_{j\in\mathcal{U}_i - s_i} \hat{y}_{ij}\right),$$

where s_i is the set of sampled units in area i, and $\hat{y}_{ij} = x_{ij}^\top\hat{\beta} + \hat{v}_i$. Here, $\hat{v}_i$ is the EBLUP of the random effect, given by

$$\hat{v}_i = \hat{\rho}_i(\bar{y}_i - \bar{x}_i^\top\hat{\beta}), \tag{3}$$

where $\hat{\rho}_i = \dfrac{n_i\hat{\sigma}_v^2}{\hat{\sigma}_e^2 + n_i\hat{\sigma}_v^2}$, $\hat{\delta} = (\hat{\sigma}_e^2, \hat{\sigma}_v^2)$, and $\hat{\theta} = (\hat{\beta}^\top, \hat{\delta}^\top)^\top$ are the maximum likelihood estimators of θ_0.

Estimation of MSE of small-area estimators is a challenging problem even in the case of this classical EBLUP estimator, and closed-form expressions do not exist in general. Prasad and Rao (1990) proposed a linearization method; other methods obtained in the literature can be found in Rao and Molina (2015).

2.2 Robust Estimation

EBLUP estimators of small-area totals or means are efficient when the model assumptions hold, but they can be very sensitive to outliers or departures from the assumed normal distributions for the random effects in the model. The above linear mixed model can be robustified by assuming that the random effects and error terms follow a mixture distribution. More specifically, we consider a small amount of contamination by assuming that the v_i and the e_{ij} are generated by independent sequences of mixture distributions given by

$$v_i = \left(1 - \frac{A_i}{\sqrt{k}}\right)\alpha_i + \frac{A_i}{\sqrt{k}}\alpha_i^* \quad \text{and} \quad e_{ij} = \left(1 - \frac{A_{ij}}{\sqrt{k}}\right)\varepsilon_{ij} + \frac{A_{ij}}{\sqrt{k}}\varepsilon_{ij}^*, \tag{4}$$

respectively. The α_i are independently and identically distributed with a Gaussian distribution F_v of mean 0 and variance σ_v^2, while ε_{ij} are assumed to be independent with an identical Gaussian distribution F_e of mean 0 and variance σ_e^2, and are independent to the α_i. The α_i^* and ε_{ij}^* are arbitrary independent random variables with finite first four moments. The A_i and A_{ij} are independent and identically distributed Bernoulli random variables.

Contamination type (4) is considered by Jaeckel (1971) to study robust estimates of location with a model for asymmetric contamination. He provided a useful interpretation that we adapt to model (4) as follows: the amount of contamination is large enough to affect the performance of the estimator (EBLUP) but is too small to be measured accurately at the given number of areas k. Heritier and Ronchetti (1994) and Salibian-Barrera et al. (2016) also studied contamination in expression (4) to examine the robustness of hypothesis-testing procedures for regression coefficients. Here, the choice of the contaminations that converge to 0 at the rate $\sqrt{k}$ is made to obtain asymptotic results, which leads to useful approximation for a finite sample size and a small number of areas as shown in the Monte Carlo simulations. Our contamination scheme is more general than those of Field et al. (2010) and Sinha and Rao (2009), who only consider the case where α_i^* and ε_{ij}^* belong to a Gaussian distribution with mean 0. We consider the more general case of asymmetric contamination including the random intercept and the random slope as well as the contamination with a non-Gaussian distribution.

With this formulation, the variance-covariance matrix of y_i in Equation 2 is given by $V_i = \sigma_e^2 I_{n_i} + \sigma_v^2 1_{n_i} 1_{n_i}^\top + O\left(1/\sqrt{k}\right)$. However, for our proofs, we can neglect the term in $O\left(1/\sqrt{k}\right)$ and still write the variance-covariance matrix as $V_i = \sigma_e^2 I_{n_i} + \sigma_v^2 1_{n_i} 1_{n_i}^\top$, without detracting from generality.

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

$$S(y, X, \theta) \equiv \sum_{i=1}^{k} \Psi(y_i, X_i, \theta) = 0, \tag{5}$$

where $\Psi(y_i, X_i, \theta) = \left(\Psi_1(y_i, X_i, \theta)^\top, \Psi_2(y_i, X_i, \theta)^\top\right)^\top$; Ψ_1 is a p-dimensional vector of estimating functions associated to the regression parameter β; and $\Psi_2 = \left(\Psi_{21}^\top, \Psi_{22}^\top\right)^\top$ is a two-dimensional vector of estimating functions associated to the variance parameters $\delta = (\sigma_e^2, \sigma_v^2)$. Note that, though the distribution is not necessarily normal, the local asymptotic normality is assumed, and that leads to the log likelihood that is approximately similar to that of the normality distribution. Therefore, we can derive the robust version of the EBLUP using the robustifying maximum likelihood estimating equations of the normal distribution given by expression (5).

This class of estimators includes robust maximum likelihood estimators developed by Sinha and Rao (2009), for which the functions Ψ_1 and Ψ_2 are specifically defined by

$$\Psi_1(y_i, X_i, \theta) = X_i^\top V_i^{-1} U_i^{1/2} \Psi_b(r_i) \tag{6}$$

$$\Psi_{2l}(y_i, X_i, \theta) = \Psi_b^\top(r_i) U_i^{1/2} V_i^{-1} \frac{\partial V_i}{\partial \delta_l} V_i^{-1} U_i^{1/2} \Psi_b(r_i) - \mathrm{tr}\left(\mathrm{K_i V_i^{-1}} \frac{\partial \mathrm{V_i}}{\partial \mathrm{\delta_l}}\right), \tag{7}$$

where $l = 1, 2$, $r_i = U_i^{-1/2}(y_i - X_i\beta)$, $\Psi_b(r_i) = (\psi_b(r_{i1}), \ldots, \psi_b(r_{in_i}))^\top$ is an n_i-vector of bounded functions, $U_i = \mathrm{diag}(\mathrm{V_i})$ is a diagonal matrix whose elements are the diagonal elements of the matrix V_i, and $K_i = E\left\{\psi_b^2(r)\right\} I_{n_i}$ where r has the standard normal distribution $r \sim \mathcal{N}(0, 1)$. An example of function ψ_b is the Huber-type function defined by

$$\psi_b(u) = \min\{|b|, \max(-|b|, u)\}, \tag{8}$$

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 (non-robust) maximum likelihood estimation. Other robust small-area estimators are based on M-quantile models, which prevent the problems associated with specification of random effects, and allow for inter-area differences to be characterized by area-specific M-quantile coefficients (see Chambers and Tzavidis 2006 and Tzavidis et al. 2010).

Newton-Raphson algorithms that numerically solve for these robust estimators can be found in Sinha and Rao (2009).[1] From the robustly estimated parameters, $\hat{\theta}_R = (\hat{\beta}_R^\top, \hat{\delta}_R^\top)^\top$, obtained

[1]Note, however, that the Newton-Raphson method can be unstable, especially for the variance parameters estimation. Other methods could be used, for example, the fixed-point iterative method of Anderson (1973). In fact, Tzavidis et al. (2016) suggest combining the two approaches, where the regression parameters are estimated by using a Newton-Raphson algorithm while the variance parameters are estimated by using a fixed-point algorithm.

from (5), the Sinha-Rao REBLUP for the area mean $\bar{Y}_i$, denoted $\hat{\bar{Y}}_{i\text{SR}}$, is of a plug-in type given by

$$\hat{\bar{Y}}_{i\text{SR}} = N_i^{-1}\sum_{j\in s_i} y_{ij} + \left(1-n_iN_i^{-1}\right)\bar{x}_{ic}^\top\hat{\beta}_R + \left(1-n_iN_i^{-1}\right)\hat{v}_{iR}, \tag{9}$$

where $\bar{x}_{ic} = \dfrac{1}{N_i-n_i}\sum_{j\in U_i-s_i} x_{ij}$, 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 $\hat{\theta}_R = (\hat{\beta}_R^\top, \hat{\delta}_R^\top)^\top$:

$$\sigma_e^{-1}\sum_{i=1}^k X_i^\top \Psi\{\sigma_e^{-1}(y_i - X_i\beta - v_i 1_{n_i})\} = 0, \tag{10}$$

$$\sigma_e^{-1}1_{n_i}^\top \Psi\{\sigma_e^{-1}(y_i - X_i\beta - v_i 1_{n_i})\} - \sigma_v^{-1}\psi_b(\sigma_v^{-1}v_i) = 0, \quad (i=1,\ldots,k). \tag{11}$$

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

$$\hat{\bar{Y}}_{i\text{SR}} = N_i^{-1}\sum_{j\in s_i} y_{ij} + \left(1-n_iN_i^{-1}\right)\bar{x}_{ic}^\top\hat{\beta}_R + \left(1-n_iN_i^{-1}\right)\hat{\rho}_{iR}\sum_{j\in s_i}\hat{f}_{ijR}\left(y_{ij} - x_{ij}^\top\hat{\beta}_R\right),$$

where

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

with

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

Denote by $\theta_R = (\beta_R^\top, \delta_R^\top)^\top$ the probability limit of $\hat{\theta}_R = (\hat{\beta}_R^\top, \hat{\delta}_R^\top)^\top$ (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:

$$\begin{aligned}\left(1-n_iN_i^{-1}\right)^{-1}\left(\hat{\bar{Y}}_{iSR} - \bar{Y}_i\right) &= \left(\bar{x}_{ic} - \hat{\rho}_{iR}\bar{x}_{iR}\right)^\top\left(\hat{\beta}_R - \beta_R\right) - (1-\hat{\rho}_{iR})v_i + \hat{\rho}_{iR}\bar{e}_{iR} - \bar{e}_{ic} \\ &+ \left(\bar{x}_{ic} - \hat{\rho}_{iR}\bar{x}_{iR}\right)^\top\left(\beta_R - \beta\right),\end{aligned} \tag{12}$$

where

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

As 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 show that sufficient condition to establish the convergence of our bootstrap using the Bickel and Freedman (1981) approach is to establish 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 maximum likelihood (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 explode.

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

Assumption A1. $\lim_{k\to\infty} \dfrac{k}{n} = c \in [0,1]$

Assumption A2. The covariates X_i are distributed over a bounded support.

Assumption A3. The $p \times p$ matrix J_1 defined by

$$J_1(\theta) = \lim_{k\to\infty} \sum_{i=1}^{k} I_{1k}^{-1/2} X_i^\top V_i^{-1} U_i^{1/2} E_m \left\{ \Psi_b(r_i)\Psi_b(r_i)^\top \right\} U_i^{1/2} V_i^{-1} X_i I_{1k}^{-1/2}$$

exists, is positive definite and continuous in θ, 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 2×2 matrix J_2 defined by

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

exists, is positive definite and continuous in θ, 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 θ, where

$$G_k(\theta) = -\begin{bmatrix} \sum_{i=1}^k I_{1k}^{-1/2} \dfrac{\partial \Psi_1(y_i, X_i, \theta)}{\partial \beta} I_{1k}^{-1/2} & 0 \\ 0 & \sum_{i=1}^k I_{2k}^{-1/2} \dfrac{\partial \Psi_2(y_i, X_i, \theta)}{\partial \delta} I_{2k}^{-1/2} \end{bmatrix};$$

and the above convergence in probability is uniform on compact sets of θ.

Assumption A6.

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

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), Weiss (1973), Weiss (1975) use 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 effect estimate 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 θ 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,*

$$I_k^{1/2}(\hat{\theta}_R - \theta_R) \xrightarrow{d} \mathcal{N}(0, G_R^{-1} \Sigma_R G_R^{-1}), \tag{13}$$

where $\hat{\theta}_R = (\hat{\beta}_R^\top, \hat{\delta}_R^\top)^\top$ is the unique solution to (5), and θ_R is its probability limit.

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

$$I_k^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, G_0^{-1} \Sigma_0 G_0^{-1}), \tag{14}$$

where $\hat{\theta} = (\hat{\beta}^\top, \hat{\delta}^\top)^\top$ is the solution to (5), and $\hat{\theta}$ is the maximum likelihood estimator of the true parameter vector θ_0.

Proof. See Theorem A of Huggins (1993). □

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, θ_R, is possibly different from the true parameter vector θ_0. However, since the classic ML estimator $\hat{\theta}$ is based on the original data that include 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 and 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 non-robust maximum likelihood estimators that are asymptotically unbiased. This allows us 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 the misspecification of the distribution of the random effects or that of the errors (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 from 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 from among $\hat{e}_{lg} - \frac{1}{n}\sum_{l=1}^{k}\sum_{g=1}^{n_l} \hat{e}_{lg}$, $l = 1, \ldots, k$, $g = 1, \ldots, n_l$, respectively; where $\hat{u}_i$ and $\hat{e}_{lg}$ are defined as follows:[2]

$$\hat{u}_i = \frac{1}{\sqrt{\hat{\rho}_i}}\hat{v}_i, \quad i = 1, \ldots, k, \quad \text{and} \quad \hat{e}_{lg} = y_{lg} - x_{lg}^\top \hat{\beta} - \frac{\hat{\tau}_l}{\hat{\rho}_l}\hat{v}_l, \quad l = 1, \ldots, k, g = 1, \ldots, n_l,$$

where

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

and $\hat{v}_i$ is the EBLUP of the random effect given in (27).

[2] Different subscript notations are used here to emphasize the fact that the randomly selected component of e^* for which the coordinate (i, j) is assigned, e_{ij}^*, is independent of the corresponding area and units from the original residual $\hat{e}_{ij}$.

Note that the estimates $\hat{\theta} = (\hat{\beta}^\top, \hat{\delta}^\top)^\top$, where $\hat{\delta} = (\hat{\sigma}_e^2, \hat{\sigma}_v^2)$, used at this step are the (non-robust) maximum likelihood estimators of θ_0.

Although there are similarities between our procedure and those of Chambers and Chandra (2013) and Mokhtarian and Chambers (2013), their method makes correlation assumptions about the random effects v_i whereas ours does not. Thus, their method works better when this type of correlation exists in the data, and ours works better otherwise. The block bootstrap samples carry over correlations to the bootstrap random effects v_i^*, but it is unclear whether their procedure is consistent, especially as theoretical validity has not been established for their approach. On the other hand, our bootstrap is similar to those of Carpenter et al. (2003), Field et al. (2008) and Field et al. (2010), who examine the so-called random effects bootstrap in classical statistics. We go further by demonstrating the validity of the bootstrap for both the ML estimates of the mixed linear model (see theorem 2) and a more general and complex parameter, which is the robust estimator of the small-area mean or total.

Step 2: Compute the mean of the bootstrap population:

$$\bar{Y}_i^* = N_i^{-1} \sum_{j=1}^{N_i} x_{ij}^\top \hat{\beta} + u_i^* + N_i^{-1} \sum_{j=1}^{N_i} e_{ij}^*. \tag{15}$$

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

$$y_{ij}^* = x_{ij}^\top \hat{\beta} + u_i^* + e_{ij}^*, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \tag{16}$$

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 $\mathcal{P}(s)$ that was used to draw the original sample. Equation (16) can be rewritten in the form

$$y_i^* = X_i \hat{\beta} + u_i^* 1_{n_i} + e_i^*, \quad i = 1, \ldots, k. \tag{17}$$

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

$$\hat{\bar{Y}}_{i\text{SR}}^* \quad = \quad N_i^{-1} \sum_{j \in s_i} y_{ij}^* + \left(1 - n_i N_i^{-1}\right) \bar{x}_{ic}^\top \hat{\beta}_R^* + \left(1 - n_i N_i^{-1}\right) \hat{v}_{iR}^*. \tag{18}$$

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 $\hat{\bar{Y}}_{iSR}$ by

$$\widehat{MSE}\left(\hat{\bar{Y}}_{i\text{SR}}\right) = B^{-1} \sum_{b=1}^{B} \left(\hat{\bar{Y}}_{iSR}^{*(b)} - \bar{Y}_i^{*(b)}\right)^2,$$

where $\hat{\bar{Y}}_{iSR}^{*(b)}$ and $\bar{Y}_i^{*(b)}$ correspond to Expressions (18) and (15), respectively, for the b^{th} bootstrap sample.

Although the procedure does not specify the number of bootstrap samples to be generated, we recommend choosing a number sufficiently large such that further increases do not substantially affect the estimated values.[3] The proposed bootstrap method is expected to work reasonably well regardless of the nature of the outliers, i.e., whether they are in the fixed effects, in the random effects or in the error term, and should be robust to non-normality of the random components of the model.

[3]The number of bootstrap samples that insure convergence may depend on the specific application. In ours, 500 bootstraps were satisfactory. So the reader may start with a similar number if their application is similar.

3.2 Validity of the Boostrap Estimator

Denote by $d_t, t = 1, 2, \ldots$, the Mallows (1972) metric for probabilities in $\Re^{p+2}$, relative to the Euclidean norm $||.||$. If μ and ν are two probabilities in $\Re^{p+2}$, then $d_t(\mu, \nu)$ is the infimum of $\left[E\left(||U - V||^t\right)\right]^{1/t}$ over all pairs of random vectors U and V whose distributions are μ and ν, 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 $\hat{u}_i$, $i = 1, \ldots, k$, centred 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 $\hat{e}_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, centred at their mean, and let F_{ek} be the empirical distribution of the e_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$. Define by $\Phi_k(F_{v,e})$ the distribution of $I_k^{1/2}(\hat{\theta}_R - \theta_R)$, and by $\Phi_k(\hat{F}_{u,e})$ the distribution of $I_k^{1/2}(\hat{\theta}_R^* - \hat{\theta}_R)$, where $\hat{\theta}_R^*$ is the robust estimate of $\hat{\theta}$ 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 are given conditionally on the original sample (X_i, y_i), $i = 1, \ldots, k$.

Assumption B3. The $p \times p$ matrix J_1^* defined by

$$J_1^*(\theta) = \lim_{k\to\infty} \sum_{i=1}^{k} I_{1k}^{-1/2} X_i^\top V_i^{-1} U_i^{1/2} E_* \left\{\Psi_b(r_i^*)\Psi_b(r_i^*)^\top\right\} U_i^{1/2} V_i^{-1} X_i I_{1k}^{-1/2}$$

exists, is positive definite and is a continuous function of θ.

Assumption B4. The 2×2 matrix J_2^* defined by

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

exists, is positive definite and is a continuous function of θ.

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 θ, where

$$G_k^*(\theta) = -\begin{bmatrix} \sum_{i=1}^k I_{1k}^{-1/2} \dfrac{\partial \Psi_1(y_i^*, X_i, \theta)}{\partial \hat{\beta}} I_{1k}^{-1/2} & 0 \\ 0 & \sum_{i=1}^k I_{2k}^{-1/2} \dfrac{\partial \Psi_2(y_i^*, X_i, \theta)}{\partial \hat{\delta}} I_{2k}^{-1/2} \end{bmatrix}.$$

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

Assumption B6.

$$I_k^{-1} S(y^*, X, \hat{\theta}_R) \xrightarrow{p} 0.$$

Assumption B7.

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

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

Lemma 2. *Under Assumptions A0–A2, B3–B7, and conditional on the sample,*

$$I_k^{1/2}(\hat{\theta}_R^* - \hat{\theta}_R) \xrightarrow{d} \mathcal{N}(0, {G^*}_R^{-1} \Sigma_R^* {G^*}_R^{-1}); \tag{19}$$

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

$$I_k^{1/2}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{d} \mathcal{N}(0, {G^*}_0^{-1} \Sigma_0^* {G^*}_0^{-1}). \tag{20}$$

To show the validity of our bootstrap, we first 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 and B3–B7 hold. Then, for $k \longrightarrow \infty$, and uniformly over θ,*

$$d_4\left(F_v, \hat{F}_{ku}\right) \xrightarrow{p} 0 \quad \text{and} \quad d_4\left(F_e, \hat{F}_{ke}\right) \xrightarrow{p} 0 \tag{21}$$

$$J_1^*(\theta) \xrightarrow{p} J_1(\theta) \tag{22}$$

$$J_2^*(\theta) \xrightarrow{p} J_2(\theta) \tag{23}$$

$$G^*(\theta) \xrightarrow{p} G(\theta). \tag{24}$$

Proof. Use the Mallows (1972) metric for $t = 4$, and the results from Bickel and Freedman (1981). See the Appendix. □

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

Theorem 1. *Under Assumptions A0–A7 and B3–B7, and conditional on the sample,*

$$d_2^{p+2}\left\{\Phi_k(F_{v,e}), \Phi_k(\hat{F}_{u,e})\right\} \xrightarrow{p} 0 \quad \text{as } k \to \infty.$$

Proof. The proof follows immediately from the above results and Lemma 8.3 of Bickel and Freedman (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 distributions are defined by $\Phi_k(F_{v,e})$ and $\Phi_k(\hat{F}_{u,e})$, respectively. By Lemmas 1 and 2, their asymptotic distributions are given by $\mathcal{N}(0, G_R^{-1}\Sigma_R G_R^{-1})$ and $\mathcal{N}(0, {G^*}_R^{-1}\Sigma_R^* {G^*}_R^{-1})$, respectively. It then follows by Lemma 3 and Levy's Continuity Theorem that, conditional on the sample,

$$\xi_R^* \xrightarrow{d} \xi_R.$$

It also easily follows that conditional on the sample

$$E_* \left[\|\xi^*_R\|^2 \right] \longrightarrow \text{tr} \left({G^*}^{-1}_R \Sigma^*_R {G^*}^{-1}_R \right), \quad \text{and} \quad E_m \left[\|\xi_R\|^2 \right] \longrightarrow \text{tr} \left(G^{-1}_R \Sigma_R G^{-1}_R \right),$$

which, by Lemma 3 and the continuous mapping theorem, implies that

$$E_* \left[\|\xi^*_R\|^2 \right] \xrightarrow{p} E_m \left[\|\xi_R\|^2 \right].$$

It then follows by Lemma 8.3 a) of Bickel and Freedman (1981) that

$$d_2^{p+2} \left\{ \Phi_k(F_{v,e}), \Phi_k(\hat{F}_{u,e}) \right\} \xrightarrow{p} 0 \quad \text{as } k \longrightarrow \infty.$$

□

The following theorem is the main result of this paper. It states that under the conditions given above, the proposed bootstrap MSE estimator of the Sinha and Rao (2009) REBLUP is a consistent estimator of the MSE.

Theorem 2. *Under Assumptions A0 to A7 and B3 to B7, and conditional on the sample,*

$$\left| E_* \left(\hat{\bar{Y}}^*_{iSR} - \bar{Y}^*_i \right)^2 - E_m \left(\hat{\bar{Y}}_{iSR} - \bar{Y}_i \right)^2 \right| \xrightarrow{p} 0 \quad \textit{as} \quad k \to \infty.$$

Proof. By Lemma 8.3 a) of Bickel and Freedman (1981), it is sufficient to show that $d_2 \left(\hat{\bar{Y}}^*_{iSR} - \bar{Y}^*_i, \hat{\bar{Y}}_{iSR} - \bar{Y}_i \right) \xrightarrow{p} 0$. Denote $\hat{\gamma}_R = I_{1k}^{1/2} \left(\hat{\beta}_R - \beta_R \right)$ and $\hat{\gamma}^*_R = I_{1k}^{1/2} \left(\hat{\beta}^*_R - \hat{\beta}_R \right)$. Then, from Equation (12) above, we can write $\left(1 - n_i N_i^{-1}\right)^{-1} \left(\hat{\bar{Y}}_{iSR} - \bar{Y}_i \right)$ as an affine function of $\hat{\gamma}_R$, v_i, $\bar{e}_{iR}$, $\bar{e}_{ic}$. That is, $\left(1 - n_i N_i^{-1}\right)^{-1} \left(\hat{\bar{Y}}_{iSR} - \bar{Y}_i \right) = \Lambda_i \left(\hat{\gamma}_R, v_i, \bar{e}_{iR}, \bar{e}_{ic} \right)$. It then follows from Assumptions A0 and A2 that there exists a positive constant, $M > 0$, such that

$$\|\Lambda_i \left(\hat{\gamma}_R, v_i, \bar{e}_{iR}, \bar{e}_{ic} \right) \|^2 \leq M \left[1 + \| \left(\hat{\gamma}_R, v_i, \bar{e}_{iR}, \bar{e}_{ic} \right)^\top \|^2 \right].$$

Given that, by Theorem 1 and Condition (21) of Lemma 3 above, we must have

$$d_2 \left(\left(\hat{\gamma}^*_R, v^*_i, \bar{e}^*_{iR}, \bar{e}^*_{ic} \right)^\top, \left(\hat{\gamma}_R, v_i, \bar{e}_{iR}, \bar{e}_{ic} \right)^\top \right) \xrightarrow{p} 0.$$

It then follows by Lemma 8.5 of Bickel and Freedman (1981) that

$$d_2 \left(\hat{\bar{Y}}^*_{iSR} - \bar{Y}^*_i, \hat{\bar{Y}}_{iSR} - \bar{Y}_i \right) \xrightarrow{p} 0.$$

□

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 the EBLUP and three small-area robust estimators: the robust estimator of Sinha and Rao (2009), SR; the robust estimator of Chambers et al. (2014), CCST3; and the robust estimator of Jiongo et al. (2013) based on the conditional bias concept of Beaumont et al. (2013), JHD.

For each of these small-area estimators, the performance of the proposed bootstrap MSE procedure, denoted as JNBOOT, is assessed and compared with several alternative MSE estimators. For the small-area estimator JHD, we compare our results with the bootstrap MSE estimators of Sinha and Rao (2009), denoted as SRBOOT, and Jiongo et al. (2013), denoted as 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. (2014), denoted as CCT and CCST, respectively. Finally, for the EBLUP, we compare our results with the above including the estimator of Prasad and Rao (1990), denoted as PR.[4]

4.1 Simulation Design

We consider three values for the number of areas: $k = 40$, $k = 20$ and $k = 10$. For each value of k, we take $N_1 = \ldots = N_k = 50$. The values of the auxiliary variable are generated from a log-normal distribution with mean $E\{\log(x)\} = 1.0$ and standard deviation $\text{var}^{1/2}\{\log(x)\} = 0.5$ in the log-scale. In each area of the population, random samples of size $n_1 = \ldots = n_k = 5$ have been selected by simple random sampling without replacement. The values of the variable of interest are generated as $y_{ij} = 100 + 5x_{ij} + v_i + e_{ij}$ where the random-area and errors are independently generated according to three outlier contamination scenarios:

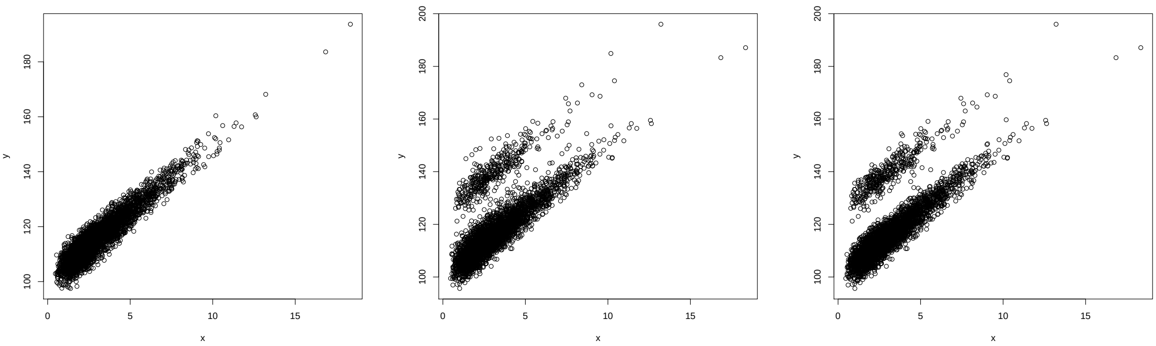
- Scenario 1: no outlier, $v_i \sim \mathcal{N}(0, 4)$ and $e_{ij} \sim \mathcal{N}(0, 6)$.
- Scenario 2: asymmetric contamination of the error terms, $e_{ij} \sim (1 - \frac{A_{ij}}{\sqrt{k}})\mathcal{N}(0, 6) + \frac{A_{ij}}{\sqrt{k}}\mathcal{L}(150 + x_{ij}, 7)$. For the random effects, we specify the distribution of the first $\frac{9k}{10}$ areas as $v_i \sim \mathcal{N}(0, 4)$ and that of the last $\frac{k}{10}$ areas as $v_i \sim \mathcal{L}(9, 5) / \sqrt{k}$. Throughout, $\mathcal{L}(\mu, s)$ denotes the logistic distribution of mean μ and scale s, and the A_{ij} are independent Bernoulli random variables with parameter $p = 0.1$.
- Scenario 3: asymmetric contamination of the errors terms, $e_{ij} \sim (1 - \frac{A_{ij}}{\sqrt{k}})\mathcal{N}(0, 6) + \frac{A_{ij}}{\sqrt{k}}\mathcal{L}(150 + x_{ij}, 7)$. For the random effects, we also have an asymmetric contamination $v_i \sim (1 - \frac{A_i}{\sqrt{k}})\mathcal{N}(0, 4) + \frac{A_i}{\sqrt{k}}\mathcal{L}(57, 5)$, where the A_i are independent Bernoulli random variables with parameter $p = 0.1$.

Figure 1 provides a picture of the simulated data for the three scenarios, where the number of areas is equals to $k = 40$.

For each scenario, we generate $T = 1000$ populations and $B = 1000$ bootstrap replications. The tuning constant for the small-area estimator JHD 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. (2014). 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. Let $\hat{\bar{Y}}_i$ denote an arbitrary estimator of the small-area mean $\bar{Y}_i$. Then the relative bias for the area mean $\bar{Y}_i$ associated to $\hat{\bar{Y}}_i$ is given by

$$\text{RB}(\hat{\bar{Y}}_i) = 100 \times T^{-1} \sum_{t=1}^{T} \frac{\hat{\bar{Y}}_i^{(t)} - \bar{Y}_i^{(t)}}{\bar{Y}_i^{(t)}}, \quad (i = 1, \ldots, k), \tag{25}$$

[4]Note that for the EBLUP, is it not useful to report the bootstrap procedure SRBOOT since it uses robust parameter estimates. Hence, only JHDBOOT and JNBOOT are reported for this case.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

Figure 1: Scatter plots of the populations generated from the three scenarios, where the number of areas is equals to $k = 40$.

and the root relative mean squared error is given by

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

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 $\widehat{\text{MSE}}(\hat{\bar{Y}}_i)$ the estimator of the mean squared error of $\hat{\bar{Y}}_i$. The relative bias associated to $\widehat{\text{MSE}}(\hat{\bar{Y}}_i)$ is given by

$$\text{RB}\left(\widehat{\text{MSE}}(\hat{\bar{Y}}_i)\right) = 100 \times T^{-1} \sum_{t=1}^{T} \frac{\widehat{\text{MSE}}(\hat{\bar{Y}}_i)^{(t)} - \text{MSE}(\hat{\bar{Y}}_i)}{\text{MSE}\hat{\bar{Y}}_i)}, \quad (i = 1, \ldots, k),$$

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

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

Section 4.2 presents simulation results based on all the domains. We use graphs and measures of central tendency such as the median of all the areas. Simulation results are obtained under the three scenarios.

4.2 Simulation Results

The results reported in Table 1 present the percent Monte Carlo relative biases (RB %) 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 for the three values of the number of areas ($k = 40$, $k = 20$ and $k = 10$) show that the estimator JHD proposed by Jiongo et al. (2013) performs 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 effects level or the fixed effects level. On the other hand, as expected, the Sinha and Rao (2009) estimator is biased, and the bias is moderated for the

Chambers et al. (2014) predictors for Scenario 2 and Scenario 3. The former predictor yields a smaller mean squared error than the latter.

For the MSE estimators, results are similar for $k = 40$, $k = 20$ and $k = 10$, except when the EBLUP estimator is considered. Table 2 displays the results for $k = 40$ of 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 small-area means, obtained at the median of the areas. In the absence of outliers (Scenario 1 in Table 2), only the analytical pseudolinearization MSE estimators (CCT) and linearization-based MSE estimators (CCST) of the MSE are biased when the Chambers et al. (2014) robust small-area predictor CCST3 is used. All the other MSE estimators display negligible biases, regardless of the small-area estimator considered. Likewise, for the MSE estimators it can be noted that only the analytical pseudolinearization MSE estimator (CCT) and linearization-based MSE estimator (CCST) are unstable throughout. In contrast, all the bootstrap estimators are stable regardless of the small-area estimator considered.

Scenario 2 is similar to Chambers et al. (2014) except that we use asymmetric contamination of the errors with a random-intercepts-and-slopes model rather than a random-intercepts-only model. For the areas 1 to 36, the simulation results in Table 2 show that our proposed bootstrap JNBOOT works well in terms of bias. Moreover, it outperforms its competitors when the robust predictors SR, CCST3 and JHD are considered. In terms of efficiency, the analytical CCT and CCST MSE estimators have large RRMSE. In contrast, our proposed bootstrap performs well regardless of the robust predictor used. Considering the areas 37 to 40, all the MSE estimators are biased when applied to the robust predictor of Sinha and Rao (2009). For the robust predictor of Chambers et al. (2014), the bias is smaller for the analytical pseudolinearization MSE estimator (CCT) and linearization-based MSE estimator (CCST). Our proposed bootstrap performs well in terms of bias for the robust predictor of Jiongo et al. (2013). In terms of efficiency, our proposed bootstrap also performs well regardless of the robust small-area estimator considered.

The set-up of Scenario 3 is similar to the one in Jiongo et al. (2013) except that, in addition to the contamination with the random intercepts and random slopes model, we also use a contamination of the random effects. Results in Table 2 show that the estimators CCT, CCST, SRBOOT and JHDBOOT are biased, regardless of the robust small-area predictor considered. In contrast, the proposed bootstrap, JNBOOT, is unbiased. In addition the JNBOOT also outperforms its competitors in terms of efficiency. Figure 2 further confirms, for each area, the superiority of the proposed bootstrap method, JNBOOT, over all the existing alternatives considered. Indeed, the curve that depicts the relative biases for the JNBOOT estimator over all areas is more aligned and closer to the horizontal axis than its competitors. Likewise, the root relative mean squared errors curve for the JNBOOT estimator is always below its competitors, regardless of the robust estimator considered.

For the particular case of the EBLUP, Table 2 shows that the Prasad and Rao (1990) MSE estimator (PR) performs well relative to its competitors in all the scenarios. This good performance of the well-known Prasad-Rao second order correct MSE estimator may be due to the contamination models containing $\sqrt{k}$ in the denominator, which make the effect of contamination models small as k increases. Indeed, Table 3 shows that when the number of areas is small ($k = 10$), the bootstrap MSE estimators JHDBOOT and JNBOOT outperform the Prasad-Rao estimator in terms of both bias and efficiency. Moreover, it should be noted that even when $k = 40$, the EBLUP estimator remains affected by the presence of outliers, and the more efficient small-area point-estimator is the robust estimator JHD as shown in Table 1. Therefore, it would be more appropriate in practice to use the proposed bootstrap associated with the robust estimator JHD in the presence of outliers in the sample.

5 Application: Methods of Payment in Canada

We apply our bootstrap procedure to the robust estimation of total volumes and values of Canadian payment choices (cash, debit cards and credit cards) using data from the Bank of Canada's 2009 and 2013 Methods-of-Payment (MOP) surveys. As will become clearer below, these data are particularly interesting not only because of their practical importance for evidence-based monetary policy, but also because they contain a substantial number of outliers and departures from normality, thus making an excellent example to illustrate the usefulness of our proposed method. While these data have been previously used to obtain estimates at the national level, our main focus in this application is to show how reliable estimates and precisions for the same and other quantities of interest can be obtained for smaller geographical areas or subgroups of population. This is particularly useful for the design of well-targeted and heterogeneous economic or monetary policies.

5.1 The Data

The Bank of Canada MOP survey is a detailed and representative investigation of consumer payment behaviour in Canada. It collects data from Canadian residents aged 18 and older about the payment methods they use for day-to-day purchases of goods and services (including at the point of sale, person to person and online), but it excludes mortgage and bill payments, and investment and business transactions. Financial, economic, social and demographic information on individuals and households is collected through a survey questionnaire. In addition, a three-day shopping diary collects detailed and reliable statistics on the use of cash in particular, as well as other main payment methods such as debit and credit cards. The survey design was similar for both the 2009 and the 2013 surveys, which allows for straightforward comparison and trend analysis of the results between these two periods.

The survey results reported by Arango and Welte (2012) for the 2009 survey and by Henry et al. (2015) for the 2013 survey provide regional estimates for each transaction type, where the regions considered are the Atlantic (including Newfoundland and Labrador, Prince Edward Island, Nova Scotia, New Brunswick), Quebec, Ontario, the Prairies (including Manitoba, Saskatchewan, Alberta), and British Columbia. In this paper, we take advantage of the small-area estimation techniques to compute reliable estimates for individual provinces, which generally have smaller numbers of units in the sample (compared to the regions), and whose estimates would otherwise be unreliable if traditional approaches were used. Furthermore, we are able to provide more disaggregated statistics such as estimates of totals and shares of volumes and values of transactions for subgroups of household income at the provincial level. Our bootstrap procedure is then used to compute mean square errors of these estimates that are robust to outliers. Within each province, we consider three household income groups: (i) low-income households, defined as households with annual income less than \$40,000; (ii) medium-income households, defined as households with income between \$40,001 and \$80,000; and (iii) high-income households, defined as households with income above \$80,000.[5] Our focus in this paper is therefore on the computation of indicators of quality for the predictors of total volumes and values of transactions with cash, debit cards and credit cards at the above income subgroups in each province (which we refer to as Canadian household domains).

The MOP survey diary collects for each person in the sample day-to-day purchases with cash, debit cards, credit cards and other methods of payment such as mobile payment applications or personal cheques. From this diary, we are able to derive the total volume and value of transactions, as well as the total consumption spending of each individual over the three-day collection period. Total consumption spending is used as an auxiliary variable. Since one of the assumptions of the small-area estimation method used in this paper is that the mean or total population of the auxiliary variable should exist for each area, we use data from the Survey of Household Spending (SHS), which are annualized and released at the household level (see Statistics Canada 2017 CANSIM Table 203-0021), while the MOP is an individual-level survey. Therefore we convert the three-day individual-level data to the annualized household level by multiplying it with a factor of 365/3 and the number of eligible persons in the household.

In addition to the difference in collection units, there are two other differences between the MOP and the SHS surveys: first, in terms of coverage, the SHS collects information on mortgage and bill payments while the MOP does not; second, the boundaries used to define household income groups in the MOP are slightly different from the Statistics Canada income quintile boundaries in some provinces as explained above. We assume that these discrepancies do not have a major impact on the point estimates. Even if they do, this should not overshadow the scope of this application given that its main objective is to illustrate how the proposed bootstrap can be applied in the context of complex survey data to produce quality indicators of small-area predictors.

5.2 Modeling the Volume and Value of Transactions

We estimate total volumes and total values of transactions from cash, debit cards and credit cards for $k = 30$ Canadian household domains (i.e., we have three domains per each of the ten provinces) for both the 2009 and 2013 MOP surveys. We consider units (i.e., households) with at least one transaction in cash, debit card or credit card, leading to a sample size of 1,542 and 2,428 for the 2009 and 2013 MOP surveys, respectively. The data used include (a) the sample size of each area, n_i; (b) the volume or the value of transactions of each unit, y_{ij}; (c) the total household consumption spending, x_{1ij}; (d) the age of the respondent, x_{2ij}; and (e) the household size, x_{3ij}.

The model is defined by:

$$y_{ij} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3 + v_i + e_{ij}, \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, n_i. \quad (26)$$

where we assume that the random effects v_i are independently and identically distributed as $\mathcal{N}(0, \sigma_v^2)$, the error terms e_{ij} are independently and identically distributed as $\mathcal{N}(0, \sigma_e^2)$, and v_i

[5]Note that these are the cut-offs that we define in this paper for the purpose of matching our variables of interest with the auxiliary information from the Survey of Household Spending (taken from Statistics Canada), as explained in Section 2 above. However, in the 2009 and 2013 MOP surveys, these income groups are defined as less than \$45,000, between \$45,001 and \$85,000, and above \$85,000, for the low-, medium- and high-income households, respectively.

and e_{ij} are independent for all $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$.

Results from the non-robust maximum likelihood (ML) and the robust maximum likelihood (RML) estimations of the model parameters are reported in Tables 4 and 5. The standard errors reported are obtained from the estimated covariance matrices associated with their asymptotic counterparts given in expressions (13) and (14). For all payment methods, the estimated coefficients on household size (β_3) and the error variance (σ_e^2) by the ML are implausibly high and substantially different from the RML estimates, suggesting that these parameters are overestimated, presumably due to the presence of a substantial number of outliers or to non-normality. To assess these conjectures, we examine the standardized predictors of the random effects $\hat{u}_i/s_{\hat{u}_i}$, and residuals $\hat{e}_{ij}/s_{\hat{e}_{ij}}$, where $\hat{u}_{ij}$ and $\hat{e}_{ij}$ are defined in Section 3.1 (Step 1), and $s_{\hat{u}_i}$ and $s_{\hat{e}_{ij}}$ are standard deviations of $\hat{u}_i$ and $\hat{e}_{ij}$, respectively. Specifically, we follow an approach similar to the one proposed by Lange and Ryan (1989), who use the normal probability plot to assess normality in random effect models. Figures 3 and 4 present the normal quantile-quantile (Q-Q) plots for the random effects and for the residuals, respectively, with the MOP 2013 transactions data (results for the 2009 MOP survey are similar and available upon request). The normality assumption for the random effects seems reasonable since the majority of the data points are aligned on the 45 degree line. However, for the residuals, there are a significant number of outliers in the tail, characterized by their departure from the 45 degree line.

Specifically, for each payment method, we identify approximately 1.2 percent as potential outliers with residuals larger than 3.5 standard deviations in absolute value. Outliers mostly occur in the province of Quebec (for each payment method, approximately 44 percent of the total number of outliers). Moreover, we notice that these outliers correspond to individuals who are either cash-intensive users or non-cash users. We define cash-intensive users as those for which the share value of cash transactions over the three-day diary period is at least 80 percent; conversely, the non-cash users are defined as those for which the share value of cash transactions is less than 5 percent. Figures 4(a), 4(c), 4(e) show that the normal Q-Q plots for the residuals with the volumes of cash, debit card and credit card transactions follow non-linear (and seemingly convex) patterns, suggesting that the errors are not normally distributed.[6] In contrast, Figures 4(b), 4(d), 4(f) show that the plots are aligned to the 45 degree line for the points that are at the centre of the distribution, suggesting that for the value of transactions the normality assumption is reasonable, although the presence of outliers is also clearly noticeable (from the nonalignment at the northwest and south east corners of these graphs).

In what follows, we estimate our variables of interest and use the proposed bootstrap as well as other existing procedures to compute their MSE and make comparisons.

5.3 Results of the application

The small-area population total is given by $Y_i = \sum_{j=1}^{N_i} y_{ij}$. The bootstrap MSE estimates for the small-area predictors are based on $B = 500$ bootstrap replications. Taking more bootstrap repli-

[6] Dealing with the non-linearity requires the availability of a known total/average for each auxiliary variable. This is not often the case when the model is non-linear. We try to add the dummies for the age classes where the total Canadians are available from Statistics Canada. However, there is no noticeable change in the results. The Kolmogorov-Smirnov two-samples test of distribution does not rejects the equality of the distribution of the residuals for the models with and without the dummies with p-value $p = 0.8409$; $p = 0.7532$; and $p = 0.7762$ for the cash, debit cards and credit cards, respectively.

cations did not significantly change the results. Traditionally, the MSE are computed for quality indicators such as the confidence intervals or the coefficient of variations. We compare the root of the estimated MSE competitors relative to the proposed method. More precisely, we compute and compare the ratio $\sqrt{\widehat{MSE}^{mse}\left(\hat{Y}_i\right)}\Big/\sqrt{\widehat{MSE}^{JNBOOT}\left(\hat{Y}_i\right)}$, where $\widehat{MSE}^{mse}\left(\hat{Y}_i\right)$ refers to the CCT, CCST, SRBOOT, JHDBOOT or JNBOOT estimator of the MSE of $\hat{Y}_i$. In turn, $\hat{Y}_i$ refers to the CCST3, SR or JHD robust predictors. This ratio gives the relative comparison generated by the MSE estimators using a single number that takes the proposed bootstrap as a reference. Figure 5 depicts the results for the cash payment method with the 2013 MOP survey. The results for the other payment methods and for the 2009 MOP survey are similar and are therefore unreported. The proposed bootstrap is more conservative compared to the Sinha and Rao (2009) bootstrap for all the robust predictors considered. For the robust predictor CCST3, the proposed bootstrap is more conservative when the area-specific sample size is large, while the Jiongo et al. (2013) bootstrap is more conservative when the area-specific sample size is small. For the robust predictor SR, the proposed bootstrap is more conservative when the area-specific sample size is large. However, the difference between the two MSE estimates is smaller when the sample size decreases. For the robust predictor JHD, the Jiongo et al. (2013) bootstrap is generally more conservative than the proposed bootstrap. For the robust predictors CCST3 and SR, the analytical methods are more conservative than the proposed bootstrap when the sample size is small. In addition, the analytical MSE estimators CCT and CCST clearly display large variability in small samples. Clearly, the MSE estimation methods discussed in this application show considerable differences in the estimation results, which could lead to different inferences about the predicted small-area quantities of interest. It is therefore important to compare the accuracy of these methods in a rigorous way. In this regard, our results suggest that the proposed bootstrap should be preferred since its validity is formally established by the main theoretical results of this paper and the simulations outcomes obtained in Section 4 confirm its empirical accuracy.

Our MSE bootstrap procedure can also be used to assess the quality of small-area predictors. This is performed by estimating the coefficient of variation, defined as the ratio of the estimated root MSE over the small-area predictor, $\sqrt{\widehat{MSE}\left(\hat{Y}_i\right)}\Big/\hat{Y}_i$. The smaller this ratio, the better the corresponding small-area predictor in terms of quality. The coefficients of variation are computed for Canada as well as for all $k = 30$ Canadian household domains. Figure 6 presents the estimated coefficients of variation of the volume and value of transactions for the above alternative predictors for all areas (including Canada, indexed as area 1 in the figure). Both the EBLUP and the JHD exhibit the best performances compared to the other predictors (the former both have the smallest coefficients of variations across all the area spectrum). In turn, while the EBLUP and the JHD have similar coefficients of variation in general, the JHD clearly outperforms the EBLUP for the total value of cash transactions (see Figures 6(a) and 6(b)). The JHD can therefore be chosen as the best predictor from among all the alternatives considered in this application, and this superiority is consistent with the simulation results obtained earlier (see Table 1). Hence, for the remaining results, we focus on the JHD predictor only.

Tables 6 and 7 report the estimates of the total volumes of cash, debit card and credit card transactions and their estimated standard errors for the 2009 and 2013 MOP surveys, respectively. Likewise, Tables 8 and 9 report the estimates of the total value of cash, debit card and credit card transactions and their estimated root mean square errors for the 2009 and 2013 MOP surveys, respectively. These results are used to compute the method of payment usage shares in terms of volume and value of transactions. At the national level, the volume of cash share decreased by 7.8 percentage points, while its value share remained nearly stable (0.3 percentage point increase) between 2009 and 2013. Between the two survey periods, the debit card share decreased in terms of both volume (-1.2 percentage points) and value (-5.6 percentage points). In contrast, the credit card share significantly increased in terms of both volume (+8.7 percentage points) and value (+5.3 percentage points). These national-level statistics are similar to those found in Henry et al. (2015) and Fung et al. (2015), who use the design-based approach to compute estimates for large domains. This paper goes further and provides estimates of the total volume and value of transactions at both the national and the household income group levels. At the national level, we note an average annual decline of 3.1 percent in total volume, and the associated RMSE estimate shows that this result is significant. At the more disaggregated level, the annual decline in cash transactions is actually relatively higher among high-income households living in the most populated provinces such as Ontario and Quebec as shown in Figures 7 and 8.

6 Concluding Remarks

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 non-robust 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 non-robust 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. We formally prove the theoretical validity of our proposed bootstrap. Moreover, the semi-parametric nature of the proposed method makes it particularly attractive, as it does not rely on the normality assumption. Our theoretical results are derived using an approach similar to Bickel and Freedman (1981) and Freedman (1981), as well as convergence results established by Huggins (1993). The proofs of the proposed bootstrap MSE estimator are provided for the robust estimator of Sinha and Rao (2009), and Monte Carlo simulation results show that the method also works well for the bias-corrected robust predictors of Jiongo et al. (2013) and Chambers et al. (2014).

We examine the behaviour of the proposed method through Monte Carlo simulations and compare its performance with five other methods: the bootstrap MSE estimator of Sinha and Rao (2009), the analytical pseudolinearization MSE estimator and the linearization-based MSE estimator of Chambers et al. (2014), the bootstrap MSE of Jiongo et al. (2013) and the MSE estimator of Prasad and Rao (1990). The results show that for all the different robust small-area estimators and all the various modes of contamination considered, the proposed bootstrap MSE performs well, in terms of both bias and efficiency. We apply our method to the estimation of the total volumes and values of cash, debit card and credit card transactions in Canada, using data from the Bank of Canada MOP surveys from 2009 and 2013. We found a significant annual decline of 3.1 percent in the average volume of cash transactions at the national level, which is consistent with those that Henry et al. (2015) and Fung et al. (2015) obtained using traditional estimation methods. However, unlike these authors' methods, our method is also able to reliably compute these statistics and their precision at a more desegregated level, such as the provinces and the household income groups within the provinces, using the robust small-area estimation techniques. In addition, the proposed bootstrap can be employed to compute indicators of quality useful for selecting the best predictor, which in our data appears to be the Jiongo et al. (2013) predictor, from among the alternatives listed above.

Finally, we note that the linear mixed model does not fit the total volume of transactions data well. This is not very surprising since the volume of transactions is count data whereas linear mixed models are primarily designed for continuous outcomes. Count data models such as Poisson or negative binomial types should therefore be better alternative approaches to develop robust small-area predictors and bootstrap MSE estimators for count outcomes. It would also be interesting to apply the method to the case of influential covariate values, as indicated in Sinha and Rao (2009), who point out difficulties with non-parametric bootstrap methods where the bootstrap samples may not contain outliers in the same proportion as the original data. Alternatively, it might be worthwhile developing a jackknife or a fast and robust bootstrap following Salibian-Barrera et al. (2008). These are avenues for further research.

References

Anderson, T. (1973). Asymptotically efficient estimation of covariance matrices with linear covariance structure. *Annals of Statistics 1*, 135–141.

Arango, C. and A. Welte (2012). The bank of canada's 2009 methods-of-payment survey: Methodology and key results. Bank of Canada Discussion Paper No. 2012-6.

Beaumont, J.-F., D. Haziza, and A. Ruiz-Gazen (2013). A unified approach to robust estimation in finite population sampling. *Biometrika 100*(3), 555–569.

Bickel, P. J. and D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 1196–1217.

Carpenter, J. R., H. Goldstein, and J. Rasbash (2003). A novel bootstrap procedure for assessing the relationship between class size and achievement. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 52*(4), 431–443.

Chambers, R. and H. Chandra (2013). A random effect block bootstrap for clustered data. *Journal of Computational and Graphical Statistics 22*(2), 452–470.

Chambers, R., H. Chandra, N. Salvati, and N. Tzavidis (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(1), 47–69.

Chambers, R. and N. Tzavidis (2006). M-quantile models for small area estimation. *Biometrika 93*(2), 255–268.

Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics 28*(1), 51–60.

Field, C., Z. Pang, and A. H. Welsh (2010). Bootstrapping robust estimates for clustered data. *Journal of the American Statistical Association 105*(492), 1606–1616.

Field, C. A., Z. Pang, and A. H. Welsh (2008). Bootstrapping data with multiple levels of variation. *Canadian Journal of Statistics 36*(4), 521–539.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics 9*(6), 1218–1228.

Fung, B., K. P. Huynh, and G. Stuber (2015). The use of cash in canada. Bank of Canada Review Spring 2015.

Henry, C. S., K. P. Huynh, and Q. R. Shen (2015). 2013 methods-of-payment survey results. Bank of Canada Discussion Paper No. 2015-4.

Heritier, S. and E. Ronchetti (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association 89*(427), 897–904.

Huggins, R. (1993). On the robust analysis of variance components models for pedigree data. *Australian & New Zealand Journal of Statistics 35*(1), 43–57.

Jaeckel, L. (1971). Robust estimates of location: Symmetry and asymmetric contamination. *The Annals of Mathematical Statistics 42*(3), 1020–1034.

Jiongo, V. D., D. Haziza, and P. Duchesne (2013). Controlling the bias of robust small-area estimators. *Biometrika 100*(4), 843–858.

Lange, N. and L. Ryan (1989). Assessing normality in random effects models. *The Annals of Statistics*, 624–642.

Mallows, C. L. (1972). A note on asymptotic joint normality. *The Annals of Mathematical Statistics*, 508–515.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics*, 746–762.

Mokhtarian, P. and R. Chambers (2013). An outlier robust block bootstrap for small area estimation. Proceedings of the 59th World Statistics Congress of the International Statistical Institute (pp. 798-803). Hague, Netherlands: International Statistical Institute.

Opsomer, J. D., G. Claeskens, M. G. Ranalli, G. Kauermann, and F. Breidt (2008). Nonparametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(1), 265–286.

Prasad, N. G. N. and J. N. K. Rao (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American statistical association 85*(409), 163–171.

Rao, J. and I. Molina (2015). *Small Area Estimation* (Second ed.). John Wiley & Sons, Inc.

Salibian-Barrera, M., S. Van Aels, and G. Willems (2008). Fast and robust bootstrap. *Statistical Methods and Applications 17*, 41–71.

Salibian-Barrera, M., S. Van Aelst, and V. J. Yohai (2016). Robust tests for linear regression models based on tau-estimates. *Computational Statistics & Data Analysis 93*, 436–455.

Sinha, S. K. and J. N. K. Rao (2009). Robust small area estimation. *Canadian Journal of Statistics 37*(3), 381–399.

Stahel, W. A. and A. Welsh (1997). Approaches to robust estimation in the simplest variance components model. *Journal of Statistical Planning and Inference 57*(2), 295–319.

Statistics Canada (2017). Cansim table 203-0021.

Tzavidis, N., S. Marchetti, and R. Chambers (2010). Robust estimation of small-area means and quantiles. *Australian & New Zealand Journal of Statistics 52*(2), 167–186.

Tzavidis, N., N. Salvati, T. Schmid, E. Flouri, and E. Midouhas (2016). Longitudinal analysis of the strengths and difficulties questionnaire scores of the millennium cohort study children in england using mquantile randomeffects regression. *Journal of the Royal Statistical Society, Series A 179*(2), 427–452.

Weiss, L. (1971). Asymptotic properties of maximum likelihood estimators in some nonstandard cases. *Journal of the American Statistical Association 66*(334), 345–350.

Weiss, L. (1973). Asymptotic properties of maximum likelihood estimators in some nonstandard cases, ii. *Journal of the American Statistical Association 68*(342), 428–430.

Weiss, L. (1975). The asymptotic distribution of the likelihood ratio in some nonstandard cases. *Journal of the American Statistical Association 70*(349), 204–208.

Table 1: Monte Carlo relative biases (%) and relative root mean squared error (%) for the predictors of the small-area means (median of the areas)

Scenario	Relative bias				Root relative mean squared error			
	EBLUP	JHD	SR	CCST3	EBLUP	JHD	SR	CCST3
	Number of areas, k = 40							
Scenario 1	0.01	0.01	0.01	0.01	0.83	0.85	0.84	0.91
Scenario 2, areas 1 to 36	0.12	-0.02	-1.37	-0.81	1.65	1.51	1.85	2.38
Scenario 2, areas 37 to 40	-0.77	-0.64	-1.85	-0.99	1.59	1.49	2.18	2.26
Scenario 3	0.04	-0.17	-1.32	-0.74	2.05	1.78	1.88	2.45
	Number of areas, k = 20							
Scenario 1	0.00	0.01	0.01	0.00	0.84	0.86	0.86	0.92
Scenario 2, areas 1 to 18	0.16	-0.05	-2.14	-1.49	2.01	1.88	2.54	3.02
Scenario 2, areas 19 to 20	-1.19	-1.04	-2.78	-1.73	2.31	2.08	3.05	2.94
Scenario 3	0.06	-0.19	-2.01	-1.36	2.77	2.36	2.62	3.14
	Number of areas, k = 10							
Scenario 1	0.01	0.01	0.01	0.01	0.88	0.89	0.90	0.93
Scenario 2, areas 1 to 9	0.23	-0.16	-3.12	-2.41	2.79	2.61	3.57	4.09
Scenario 2, areas 10 to 10	-1.59	-1.54	-3.88	-2.78	3.20	2.88	4.17	3.58
Scenario 3	-0.01	-0.42	-2.95	-2.25	3.71	3.17	3.64	4.14

Note: The best linear unbiased predictor is denoted by EBLUP; the Jiongo et al. (2013) predictor based on the conditional-bias concept is denoted by JHD; the Sinha and Rao (2009) predictor is denoted by SR; and the Chambers et al. (2014) predictor is denoted by CCST3. The three scenarios are described below:

- Scenario 1: no outlier, $v_i \sim \mathcal{N}(0, 4)$ and $e_{ij} \sim \mathcal{N}(0, 6)$.
- Scenario 2: asymmetric contamination of the error terms, $e_{ij} \sim (1 - \frac{A_{ij}}{\sqrt{k}})\mathcal{N}(0, 6) + \frac{A_{ij}}{\sqrt{k}}\mathcal{L}(150 + x_{ij}, 7)$. For the random effects, we specify the distribution of the first $\frac{9k}{10}$ areas as $v_i \sim \mathcal{N}(0, 4)$ and that of the last $\frac{k}{10}$ areas as $v_i \sim \mathcal{L}(9, 5)/\sqrt{k}$. Throughout, $\mathcal{L}(\mu, s)$ denotes the logistic distribution of mean μ and scale s, and the A_{ij} are independent Bernoulli random variables with parameter $p = 0.1$.
- Scenario 3: asymmetric contamination of the errors terms, $e_{ij} \sim (1 - \frac{A_{ij}}{\sqrt{k}})\mathcal{N}(0, 6) + \frac{A_{ij}}{\sqrt{k}}\mathcal{L}(150 + x_{ij}, 7)$. For the random effects, we also have an asymmetric contamination $v_i \sim (1 - \frac{A_i}{\sqrt{k}})\mathcal{N}(0, 4) + \frac{A_i}{\sqrt{k}}\mathcal{L}(57, 5)$, where the A_i are independent Bernoulli random variables with parameter $p = 0.1$.

Table 2: Monte Carlo relative biases (RB %) and root relative mean squared error (RRMSE %) for the MSE estimator of the predictors of small-area means (at the median over the $k = 40$ areas)

SAE	MSE	Scenario 1		Scenario 2				Scenario 3	
		Areas 0–40		Areas 0–36		Areas 37–40		Areas 0–40	
		RB	RRMSE	RB	RRMSE	RB	RRMSE	RB	RRMSE
EBLUP	PR	-0.3	11.0	8.6	37.7	12.1	39.8	-1.9	24.3
	CCT	-0.7	47.2	168.0	363.5	162.7	407.4	46.9	164.6
	CCST	0.8	47.4	170.8	370.0	165.3	417.7	48.4	167.3
	JHDBOOT	-1.8	11.9	-10.6	44.3	-7.7	45.1	-8.4	28.4
	JNBOOT	**-1.9**	**12.0**	**-10.8**	**44.2**	**-8.3**	**44.8**	**-9.1**	**28.5**
SR	CCT	-3.8	55.5	-53.8	72.3	-67.6	76.0	-50.7	72.6
	CCST	-2.5	56.8	-50.9	75.1	-65.9	77.6	-48.4	75.6
	SRBOOT	-0.5	15.3	-72.5	72.8	-80.5	80.6	-61.0	62.3
	JHDBOOT	-1.8	11.9	-29.8	45.4	-50.1	55.7	11.2	35.6
	JNBOOT	**-1.6**	**12.2**	**-5.4**	**28.3**	**-32.8**	**38.3**	**4.7**	**26.4**
CCST3	CCT	47.0	95.4	-17.3	136.0	-10.7	145.1	-13.1	135.9
	CCST	48.1	98.8	-15.5	143.0	-9.1	153.2	-11.7	142.2
	SRBOOT	3.5	15.6	-75.7	75.9	-73.2	73.4	-70.9	71.5
	JHDBOOT	0.6	12.0	34.8	42.1	49.4	56.1	26.9	35.6
	JNBOOT	**0.8**	**12.1**	**5.3**	**26.2**	**16.4**	**32.9**	**10.6**	**28.5**
JHD	SRBOOT	0.0	15.5	-56.8	57.5	-57.2	57.9	-54.8	56.6
	JHDBOOT	-1.2	11.8	15.6	56.1	14.3	55.2	31.5	49.5
	JNBOOT	**-1.0**	**12.1**	**-7.1**	**41.4**	**-8.2**	**41.2**	**5.9**	**32.1**

Note: The Jiongo et al. (2013) predictor based on the conditional-bias concept is denoted by JHD; the Sinha and Rao (2009) predictor is denoted by SR; and the Chambers et al. (2014) predictor is denoted by CCST3.
The proposed bootstrap MSE procedure is denoted JNBOOT; the bootstrap MSE estimator of Sinha and Rao (2009) is denoted by SRBOOT; the bootstrap MSE estimator of Jiongo et al. (2013) is denoted by JHDBOOT; the analytical linearization MSE and linearization-based MSE estimators developed by Chambers et al. (2014) are denoted by CCT and CCST, respectively. Ultimately, the Prasad and Rao (1990) MSE estimator is denoted by PR.
The three scenarios are described below:

- Scenario 1: no outlier, $v_i \sim \mathcal{N}(0,4)$ and $e_{ij} \sim \mathcal{N}(0,6)$.
- Scenario 2: asymmetric contamination of the error terms, $e_{ij} \sim (1 - \frac{A_{ij}}{\sqrt{k}})\mathcal{N}(0,6) + \frac{A_{ij}}{\sqrt{k}}\mathcal{L}(150 + x_{ij}, 7)$. For the random effects, we specify the distribution of the first 36 areas as $v_i \sim \mathcal{N}(0,4)$ and that of the last four areas as $v_i \sim \mathcal{L}(9,5)/\sqrt{k}$. Throughout, $\mathcal{L}(\mu, s)$ denotes the logistic distribution of mean μ and scale s, and the A_{ij} are independent Bernoulli random variables with parameter $p = 0.1$.
- Scenario 3: asymmetric contamination of the errors terms, $e_{ij} \sim (1 - \frac{A_{ij}}{\sqrt{k}})\mathcal{N}(0,6) + \frac{A_{ij}}{\sqrt{k}}\mathcal{L}(150 + x_{ij}, 7)$. For the random effects, we also have an asymmetric contamination $v_i \sim (1 - \frac{A_i}{\sqrt{k}})\mathcal{N}(0,4) + \frac{A_i}{\sqrt{k}}\mathcal{L}(57,5)$, where the A_i are independent Bernoulli random variables with parameter $p = 0.1$.

Table 3: Monte Carlo relative biases (RB %) and root relative mean squared error (RRMSE %) for the MSE estimator of the predictors of small-area means (at the median over the $k = 10$ areas)

SAE	MSE	Scenario 1		Scenario 2				Scenario 3	
		Areas 0–10		Areas 0–9		Areas 10–10		Areas 0–10	
		RB	RRMSE	RB	RRMSE	RB	RRMSE	RB	RRMSE
EBLUP	PR	-3.1	20.7	131.6	156.8	62.7	86.7	31.5	58.8
	CCT	8.7	59.6	278.5	441.5	180.4	333.5	87.0	195.3
	CCST	9.7	60.0	282.0	449.5	182.5	342.3	88.7	197.9
	JHDBOOT	-11.7	26.2	6.3	60.0	-25.5	48.6	-15.6	50.2
	JNBOOT	**-11.5**	**26.2**	**3.8**	**59.1**	**-27.4**	**49.3**	**-18.0**	**50.2**
SR	CCT	12.4	193.7	-59.5	447.1	-67.7	330.7	-63.4	100.0
	CCST	13.4	194.1	-56.5	447.3	-66.1	330.4	-61.3	100.6
	SRBOOT	-12.2	31.7	-91.8	91.9	-94.3	94.3	-81.0	84.7
	JHDBOOT	-12.8	27.0	-34.2	48.9	-54.5	59.6	-6.7	52.5
	JNBOOT	**-12.0**	**27.2**	**3.7**	**66.3**	**-28.4**	**53.8**	**4.3**	**59.5**
CCST3	CCT	40.3	94.1	-52.5	126.6	-49.1	111.5	-42.0	130.9
	CCST	41.1	97.1	-50.6	132.2	-46.9	119.1	-40.3	136.5
	SRBOOT	-2.0	27.8	-91.0	91.3	-88.8	89.3	-82.4	85.8
	JHDBOOT	-3.1	21.7	55.0	80.5	91.7	116.9	49.2	75.9
	JNBOOT	**-2.6**	**21.8**	**10.6**	**68.4**	**37.2**	**92.0**	**22.4**	**71.9**
JHD	SRBOOT	-11.5	30.4	-84.1	84.6	-87.8	88.1	-75.7	82.1
	JHDBOOT	-12.3	25.8	33.4	78.2	1.8	53.4	25.8	73.2
	JNBOOT	**-11.6**	**25.9**	**8.9**	**63.8**	**-16.8**	**50.8**	**3.6**	**61.1**

Note: The Jiongo et al. (2013) predictor based on the conditional-bias concept is denoted by JHD; the Sinha and Rao (2009) predictor is denoted by SR; and the Chambers et al. (2014) predictor is denoted by CCST3.
The proposed bootstrap MSE procedure is denoted JNBOOT; the bootstrap MSE estimator of Sinha and Rao (2009) is denoted by SRBOOT; the bootstrap MSE estimator of Jiongo et al. (2013) is denoted by JHDBOOT; the analytical linearization MSE and linearization-based MSE estimators developed by Chambers et al. (2014) are denoted by CCT and CCST, respectively. Ultimately, the Prasad and Rao (1990) MSE estimator is denoted by PR.
The three scenarios are described below:

- Scenario 1: no outlier, $v_i \sim \mathcal{N}(0,4)$ and $e_{ij} \sim \mathcal{N}(0,6)$.
- Scenario 2: asymmetric contamination of the error terms, $e_{ij} \sim (1 - \frac{A_{ij}}{\sqrt{k}})\mathcal{N}(0,6) + \frac{A_{ij}}{\sqrt{k}}\mathcal{L}(150 + x_{ij}, 7)$. For the random effects, we specify the distribution of the first 9 areas as $v_i \sim \mathcal{N}(0,4)$ and that of the last area as $v_i \sim \mathcal{L}(9,5)/\sqrt{k}$. Throughout, $\mathcal{L}(\mu, s)$ denotes the logistic distribution of mean μ and scale s, and the A_{ij} are independent Bernoulli random variables with parameter $p = 0.1$.
- Scenario 3: asymmetric contamination of the errors terms, $e_{ij} \sim (1 - \frac{A_{ij}}{\sqrt{k}})\mathcal{N}(0,6) + \frac{A_{ij}}{\sqrt{k}}\mathcal{L}(150 + x_{ij}, 7)$. For the random effects, we also have an asymmetric contamination $v_i \sim (1 - \frac{A_i}{\sqrt{k}})\mathcal{N}(0,4) + \frac{A_i}{\sqrt{k}}\mathcal{L}(57,5)$, where the A_i are independent Bernoulli random variables with parameter $p = 0.1$.

Table 4: Model parameter estimates, year 2009

	Cash		Debit		Credit	
Coefficients	Estimates	Std	Estimates	Std	Estimates	Std
Variables of interest are the volume of cash, debit card and credit card transactions						
Robust estimation						
Intercept (β_0)	280.3	43.0	202.2	39.6	57.4	29.0
Spending (β_1)	0.00049	0.00020	0.00062	0.00014	0.00136	0.00010
Age (β_2)	3.14	0.80	-1.57	0.62	2.02	0.54
Household size (β_3)	15.4	9.6	29.8	9.8	-0.7	7.3
σ_e^2	203887	4518	98689	2097	72691	2305
σ_v^2	4098	2642	978	911	2416	799
R square (R^2)	0.029		0.018		0.093	
Non-robust estimation						
Intercept (β_0)	219	51	204	52	63	40
Spending (β_1)	0.00065	0.00025	0.00075	0.00018	0.00137	0.00016
Age (β_2)	4.37	0.93	-1.24	0.74	2.91	0.73
Household size (β_3)	41.3	11.6	47.8	13.6	9.8	9.2
σ_e^2	313056	161416	194902	242796	177502	669514
σ_v^2	3439	3999	2287	2978	5868	22403
R square (R^2)	0.045		0.054		0.135	
Variables of interest are the value of cash, debit card and credit card transactions						
Robust estimation						
Intercept (β_0)	3570.1	867.6	3610.0	1563.0	-3418.7	2393.1
Spending (β_1)	0.01562	0.00323	0.06360	0.00438	0.80425	0.00667
Age (β_2)	29.6	14.1	-5.1	20.3	-52.5	33.6
Household size (β_3)	177.9	165.5	837.1	397.6	-1884.2	537.4
σ_e^2	48.5e6	896046	133.2e6	2.5e6	277.0e6	5.8e6
σ_v^2	589370	327578	78675	1.0e6	930269	3.2e6
R square (R^2)	-0.008		0.062		0.801	
Non-robust estimation						
Intercept (β_0)	2314.4	1727.7	-3275.0	3280.1	792.2	3553.6
Spending (β_1)	0.03797	0.00772	0.14068	0.01313	0.82109	0.01376
Age (β_2)	35.2	26.3	103.1	40.8	-137.0	46.8
Household size (β_3)	1372.6	401.1	2452.4	859.4	-3791.3	885.3
σ_e^2	276e6	145.0e6	1048.0e6	447.6e6	1324.0e6	2905.6e6
σ_v^2	2.9e6	8.8e6	97895	7.7e6	89834	3716.6e6
R square (R^2)	0.064		0.141		0.804	

Note: Std denotes the standard errors; $(\beta_0, \beta_1, \beta_2, \beta_3)^\top$ and $(\sigma_e^2, \sigma_v^2)^\top$ denote the regression coefficient and the variance of the errors and random effects in model (26), respectively:
$y_{ij} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3 + v_i + e_{ij}, \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, n_i.$
Non-robust estimation is the ML method, while robust estimation is the robust ML method developed by Sinha and Rao (2009).
The R squares (R^2) are computed using the formula: $1 - \frac{\sum_{i=1}^{k}\sum_{j\in s_i}(y_{ij}-\hat{y}_{ij})^2}{\sum_{i=1}^{k}\sum_{j\in s_i}(y_{ij}-\bar{y}_{ij})^2}$, where
$\hat{y}_{ij} = \hat{\beta}_0 + x_{1ij}\hat{\beta}_1 + x_{2ij}\hat{\beta}_2 + x_{3ij}\hat{\beta}_3 + \hat{v}_i$ and $\bar{y}_{ij} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j\in s_i} y_{ij}$. For the ML, the results of the R^2 are similar to those computed using the generalized R^2 formula.

Table 5: Model parameter estimates, year 2013

	Cash		Debit		Credit	
Coefficients	Estimates	Std	Estimates	Std	Estimates	Std
Variables of interest are the volume of cash, debit card and credit card transactions						
Robust estimation						
Intercept (β_0)	105.3	31.3	143.0	15.3	126.6	32.1
Spending (β_1)	0.00098	0.00014	0.00051	0.00011	0.00248	0.00009
Age (β_2)	2.03	0.45	-1.19	0.23	0.83	0.36
Household size (β_3)	55.6	10.2	28.1	4.4	5.8	6.1
σ_e^2	125863	2607	54876	1698	90216	4414
σ_v^2	1256	456	353	210	4372	1296
R square (R^2)	0.067		0.03		0.156	
Non-robust estimation						
Intercept (β_0)	111.0	36.0	172.0	20.6	150.7	36.8
Spending (β_1)	0.00134	0.00018	0.00113	0.00019	0.00228	0.00012
Age (β_2)	1.70	0.48	-1.87	0.36	0.98	0.44
Household size (β_3)	80.5	13.4	51.6	8.2	26.3	7.1
σ_e^2	200803	170041	151327	1.0e6	179450	375348
σ_v^2	2631	2079	503	2750	6665	597
R square (R^2)	0.096		0.084		0.177	
Variables of interest are the value of cash, debit card and credit card transactions						
Robust estimation						
Intercept (β_0)	260.4	537.6	3041.2	360.4	-1739.0	860.6
Spending (β_1)	0.02584	0.00252	0.03541	0.00409	0.80488	0.00945
Age (β_2)	43.43	8.62	-12.97	7.76	-28.20	12.05
Household size (β_3)	617.0	125.7	607.4	139.9	-1961.7	301.5
σ_e^2	31.3e6	488130	59.1e6	1.2e6	187.9e6	6463179
σ_v^2	518877	231480	498174	262849	4.1e6	1504233
R square (R^2)	0.032		0.026		0.680	
Non-robust estimation						
Intercept (β_0)	-227.6	928.7	252.6	1096.5	242.1	1354.8
Spending (β_1)	0.12153	0.01270	0.16652	0.01577	0.71147	0.02042
Age (β_2)	2.90	17.25	4.72	17.79	-7.56	14.48
Household size (β_3)	1548.8	317.3	1413.3	390.5	-2992.6	586.1
σ_e^2	369.1e6	197.8e6	604.2e6	393.0e6	917.3e6	646.7e6
σ_v^2	6.9e6	12.2e6	2.7e6	5.0e6	13.4e6	18.6e6
R square (R^2)	0.169		0.176		0.700	

Note: Std denotes the standard errors; $(\beta_0, \beta_1, \beta_2, \beta_3)^\top$ and $(\sigma_e^2, \sigma_v^2)^\top$ denote the regression coefficient and the variance of the errors and random effects in model (26), respectively:
$y_{ij} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3 + v_i + e_{ij}, \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, n_i.$
Non-robust estimation is the ML method, while robust estimation is the robust ML method developed by Sinha and Rao (2009).
The R squares (R^2) are computed using the formula: $1 - \frac{\sum_{i=1}^{k}\sum_{j\in s_i}(y_{ij}-\hat{y}_{ij})^2}{\sum_{i=1}^{k}\sum_{j\in s_i}(y_{ij}-\bar{y}_{ij})^2}$, where
$\hat{y}_{ij} = \hat{\beta}_0 + x_{1ij}\hat{\beta}_1 + x_{2ij}\hat{\beta}_2 + x_{3ij}\hat{\beta}_3 + \hat{v}_i$ and $\bar{y}_{ij} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j\in s_i} y_{ij}$. For the ML, the results of the R^2 are similar to those computed using the generalized R^2 formula.

Table 6: 2009 estimates of the total volume of transactions (predictor) and its standard errors (RMSE) for cash, debit cards and credit cards. Estimates are in millions.

Provinces	Income groups	Population size (in 1,000)	Sample size	Cash Predictor	Cash RMSE	Debit Predictor	Debit RMSE	Credit Predictor	Credit RMSE
Canada	All	13,382	1,542	7,601.7	259.0	3,765.5	210.0	3,752.2	197.0
NL	Low	83	18	43.9	5.1	20.3	4.5	16.4	4.9
NL	Medium	83	35	43.3	4.3	30.9	3.8	23.6	4.3
NL	High	41	36	26.7	2.3	14.3	1.8	17.4	2.2
PEI	Low	23	10	12.1	1.4	5.7	1.3	4.9	1.5
PEI	Medium	23	16	13.8	1.3	8.8	1.1	6.8	1.5
PEI	High	11	17	6.8	0.7	3.9	0.5	3.3	0.7
NS	Low	157	41	95.0	9.2	30.1	7.5	32.7	8.2
NS	Medium	157	91	93.7	6.8	48.3	5.8	39.0	5.9
NS	High	78	92	53.5	3.6	24.9	2.7	37.3	3.0
NB	Low	125	52	67.8	6.6	26.4	5.7	30.9	6.1
NB	Medium	125	62	69.2	5.6	44.8	5.1	37.7	5.5
NB	High	62	49	47.1	3.2	22.5	2.7	24.3	3.0
QC	Low	1,359	59	654.5	75.1	236.8	59.1	216.4	59.9
QC	Medium	1,359	81	686.8	63.6	362.4	53.0	370.9	52.4
QC	High	680	77	441.0	32.6	220.1	25.5	279.0	27.1
ON	Low	1,980	61	1,166.3	104.5	454.6	84.2	342.3	93.0
ON	Medium	1,980	100	1,159.3	87.9	619.7	67.3	612.7	70.6
ON	High	990	92	713.2	44.8	368.4	38.0	408.7	38.1
MB	Low	188	27	88.7	10.4	41.3	9.3	41.3	10.4
MB	Medium	188	44	116.5	10.1	51.9	8.0	57.8	8.8
MB	High	94	37	55.1	4.8	30.8	4.2	38.5	5.0
SK	Low	161	31	77.8	9.3	38.0	8.2	32.1	8.9
SK	Medium	161	46	87.4	7.8	58.7	7.0	41.0	8.0
SK	High	80	61	46.2	4.0	30.1	3.3	38.4	3.5
AB	Low	276	36	133.2	16.2	70.0	12.5	54.2	14.4
AB	Medium	553	53	281.6	26.7	148.3	22.6	144.9	25.5
AB	High	553	54	317.7	28.8	192.7	23.6	229.7	24.9
BC	Low	725	47	377.4	39.7	164.9	34.2	178.1	35.5
BC	Medium	725	67	387.9	34.5	265.0	27.0	240.3	31.4
BC	High	362	50	238.3	19.0	130.8	15.5	151.9	17.0

Note: The numbers for Canada include the 10 provinces but not the 3 territories. The income boundaries are $40,000 for the low- to medium-income households and $80,000 for the medium- to high-income households. Newfoundland and Labrador is denoted by NL; Prince Edward Island is denoted by PEI; Nova Scotia is denoted by NS; New Brunswick is denoted by NB; Quebec is denoted by QC; Ontario is denoted ON; Manitoba is denoted by MB; Saskatchewan is denoted by SK; Alberta is denoted by AB; and British Columbia is denoted by BC.

Table 7: 2013 estimates of the total volume of transactions (predictor) and its standard errors (RMSE) for cash, debit cards and credit cards. Estimates are in millions.

Provinces	Income groups	Population size (in 1,000)	Sample size	Cash Predictor	Cash RMSE	Debit Predictor	Debit RMSE	Credit Predictor	Credit RMSE
Canada	All	13,792	2,428	6,694.1	142.0	3,759.2	119.0	5,300.6	134.0
NL	Low	86	6	34.1	4.4	16.3	1.9	22.9	6.9
NL	Medium	86	8	41.5	4.4	24.6	1.8	31.7	6.2
NL	High	43	10	24.1	2.2	15.3	0.9	25.0	3.1
PEI	Low	23	29	9.3	1.1	4.8	0.5	5.8	1.3
PEI	Medium	23	23	11.4	1.1	6.7	0.5	8.6	1.4
PEI	High	12	13	6.5	0.5	3.8	0.3	5.2	0.8
NS	Low	158	19	56.8	7.7	31.1	3.3	42.1	10.8
NS	Medium	158	6	70.6	8.3	40.9	3.4	71.4	11.5
NS	High	79	3	43.8	4.2	26.8	1.8	43.1	6.1
NB	Low	125	18	45.6	6.1	23.0	2.8	36.9	8.3
NB	Medium	125	21	65.1	5.9	34.5	2.8	42.8	7.9
NB	High	63	9	37.7	3.2	21.1	1.5	32.7	4.4
QC	Low	1,386	295	564.8	32.3	249.8	21.6	265.4	32.3
QC	Medium	1,397	231	653.1	37.8	410.4	24.7	563.3	35.9
QC	High	696	96	342.7	26.6	231.3	14.5	368.7	25.4
ON	Low	2,052	335	998.4	48.1	472.6	32.3	480.2	43.3
ON	Medium	2,058	348	1,108.5	46.9	656.5	33.9	799.2	44.6
ON	High	1,028	236	594.0	28.9	340.6	20.2	719.5	28.5
MB	Low	187	47	65.6	8.0	39.1	3.7	47.0	9.1
MB	Medium	187	55	105.0	8.0	53.2	3.9	73.2	8.7
MB	High	94	29	52.2	4.6	33.3	2.3	54.1	5.2
SK	Low	167	20	63.8	8.0	32.6	3.6	43.0	10.7
SK	Medium	167	26	84.6	7.9	48.1	3.5	59.8	9.7
SK	High	84	15	47.8	4.2	29.2	1.8	37.3	5.6
AB	Low	296	49	97.7	12.7	57.6	5.8	115.0	14.7
AB	Medium	597	77	295.7	23.1	164.7	12.1	230.0	24.8
AB	High	596	71	309.8	23.7	220.6	13.2	336.2	25.9
BC	Low	727	132	308.9	23.7	148.3	13.5	178.1	22.8
BC	Medium	728	116	329.7	25.2	196.4	14.4	333.6	24.9
BC	High	364	85	225.5	14.6	125.9	8.1	228.7	14.3

Note: The numbers for Canada include the 10 provinces but not the 3 territories. The income boundaries are $45,000 for the low- to medium-income households and $85,000 for the medium- to high-income households. Newfoundland and Labrador is denoted by NL; Prince Edward Island is denoted by PEI; Nova Scotia is denoted by NS; New Brunswick is denoted by NB; Quebec is denoted by QC; Ontario is denoted ON; Manitoba is denoted by MB; Saskatchewan is denoted by SK; Alberta is denoted by AB; and British Columbia is denoted by BC.

Table 8: 2009 estimates of the total value of transactions (predictor) and its standard errors (RMSE) for cash, debit cards and credit cards. Estimates are in millions.

	Income	Population	Sample	Cash		Debit		Credit	
Provinces	groups	size (in 1,000)	size	Predictor	RMSE	Predictor	RMSE	Predictor	RMSE
Canada	All	13,382	1,542	126,112.0	7,470.0	201,388.5	13,438.0	350,363.8	13,930.0
NL	Low	83	18	628.0	149.7	870.8	128.5	396.5	133.7
NL	Medium	83	35	671.9	114.4	1,222.3	100.0	1,807.2	101.1
NL	High	41	36	431.2	56.7	822.4	52.1	1,790.9	48.5
PEI	Low	23	10	169.3	36.7	237.8	33.4	178.8	33.7
PEI	Medium	23	16	211.7	33.6	342.4	25.9	484.1	23.8
PEI	High	11	17	129.9	17.6	210.8	12.7	420.9	13.5
NS	Low	157	41	1,558.8	228.7	1,632.0	253.5	1,187.3	248.0
NS	Medium	157	91	1,535.4	181.8	2,284.7	189.7	3,587.7	220.0
NS	High	78	92	799.6	85.2	1,574.3	90.4	3,692.1	105.0
NB	Low	125	52	964.1	179.3	1,332.6	191.7	993.7	195.4
NB	Medium	125	62	1,065.1	157.9	1,814.6	138.0	2,964.0	161.7
NB	High	62	49	892.1	86.3	1,265.0	72.0	2,930.4	78.3
QC	Low	1,359	59	11,598.8	1,890.2	13,578.6	2,149.6	10,286.2	2,269.5
QC	Medium	1,359	81	13,494.8	1,672.1	19,003.9	1,581.6	30,036.7	1,628.1
QC	High	680	77	7,892.7	825.5	13,728.0	876.3	30,383.2	879.9
ON	Low	1,980	61	16,978.0	2,529.1	23,120.9	2,780.8	20,974.3	2,986.0
ON	Medium	1,980	100	18,498.7	2,162.8	32,385.5	2,408.1	58,693.5	2,446.7
ON	High	990	92	12,874.3	1,188.1	22,866.9	1,357.1	58,597.3	1,407.1
MB	Low	188	27	1,245.8	284.0	1,966.4	287.4	1,566.8	281.6
MB	Medium	188	44	1,827.2	267.2	2,891.0	212.3	4,962.7	210.6
MB	High	94	37	1,096.3	123.4	1,919.5	116.2	4,405.4	121.4
SK	Low	161	31	1,056.9	245.0	1,754.9	241.8	1,680.7	248.4
SK	Medium	161	46	1,373.4	213.3	2,459.8	191.1	4,394.9	187.1
SK	High	80	61	840.5	102.8	1,698.3	96.0	4,078.4	107.1
AB	Low	276	36	1,904.4	406.4	2,801.2	428.6	2,331.2	420.7
AB	Medium	553	53	4,699.6	777.8	7,917.1	631.3	13,553.3	667.3
AB	High	553	54	5,563.5	731.0	11,528.8	626.1	29,894.1	660.4
BC	Low	725	47	5,284.4	1,015.6	8,057.3	1,100.2	8,633.7	1,106.8
BC	Medium	725	67	6,150.9	903.0	11,473.0	834.1	22,590.2	857.4
BC	High	362	50	4,674.5	474.4	8,627.7	461.4	22,867.5	496.5

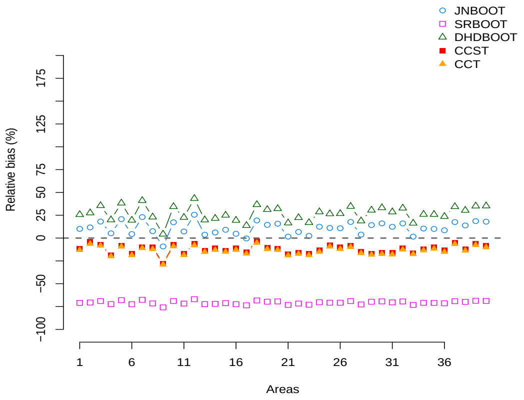
Note: The numbers for Canada include the 10 provinces but not the 3 territories. The income boundaries are $40,000 for the low- to medium-income households and $80,000 for the medium- to high-income households. Newfoundland and Labrador is denoted by NL; Prince Edward Island is denoted by PEI; Nova Scotia is denoted by NS; New Brunswick is denoted by NB; Quebec is denoted by QC; Ontario is denoted ON; Manitoba is denoted by MB; Saskatchewan is denoted by SK; Alberta is denoted by AB; and British Columbia is denoted by BC.

Table 9: 2013 estimates of the total value of transactions (predictor) and its standard errors (RMSE) for cash, debit cards and credit cards. Estimates are in millions.

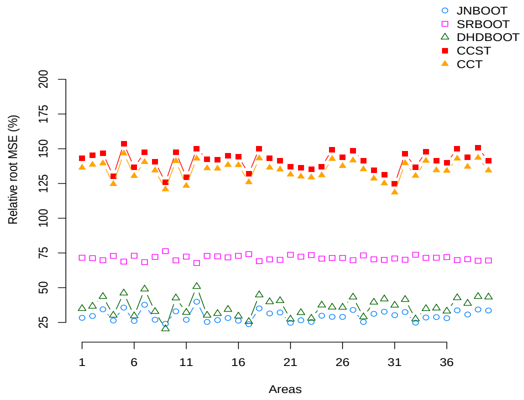
	Income	Population	Sample	Cash		Debit		Credit	
Provinces	groups	size (in 1,000)	size	Predictor	RMSE	Predictor	RMSE	Predictor	RMSE
Canada	All	13,792	2,428	149,010.8	6,260.0	190,014.0	7,852.0	462,244.2	9,366.0
NL	Low	86	6	689.2	276.8	676.9	197.6	1,307.9	290.4
NL	Medium	86	8	862.8	227.2	1,208.1	130.1	2,901.0	278.3
NL	High	43	10	708.0	105.8	877.1	71.1	2,627.9	147.7
PEI	Low	23	29	149.5	44.4	222.4	34.2	329.7	62.7
PEI	Medium	23	23	230.2	49.7	303.7	35.3	607.4	69.4
PEI	High	12	13	161.0	27.5	191.2	18.5	492.8	36.3
NS	Low	158	19	1,078.8	317.4	1,315.5	231.8	2,715.3	467.6
NS	Medium	158	6	1,565.9	500.7	1,972.1	380.3	4,990.7	540.2
NS	High	79	3	1,166.1	309.1	1,480.4	136.7	4,232.1	295.8
NB	Low	125	18	738.1	276.1	978.9	200.4	2,023.4	367.3
NB	Medium	125	21	1,497.6	266.7	1,612.8	194.6	3,285.9	380.4
NB	High	63	9	887.7	146.7	1,126.6	96.5	3,269.8	212.6
QC	Low	1,386	295	13,826.4	1,231.9	11,764.1	1,391.4	16,667.7	2,017.0
QC	Medium	1,397	231	15,551.8	1,490.0	18,539.5	1,549.7	42,173.1	2,402.8
QC	High	696	96	8,191.9	951.8	12,898.1	976.7	36,923.6	1,412.1
ON	Low	2,052	335	18,835.5	1,785.4	22,002.2	1,906.7	32,808.3	2,911.5
ON	Medium	2,058	348	23,730.7	1,759.4	33,602.4	2,084.8	72,705.8	2,924.1
ON	High	1,028	236	14,718.8	1,107.1	19,912.9	1,258.5	69,634.4	1,748.5
MB	Low	187	47	1,056.1	312.9	1,657.9	259.1	3,489.8	462.7
MB	Medium	187	55	1,993.9	301.7	2,428.1	260.2	5,899.0	470.5
MB	High	94	29	1,388.6	175.7	1,815.9	138.4	5,185.8	255.6
SK	Low	167	20	1,035.5	352.0	1,364.3	247.0	2,981.7	472.6
SK	Medium	167	26	2,316.7	328.7	2,560.5	256.9	6,161.8	493.6
SK	High	84	15	1,341.9	191.9	1,659.6	140.7	4,841.4	262.6
AB	Low	296	49	1,469.2	500.5	2,618.6	409.3	8,148.0	735.7
AB	Medium	597	77	5,897.3	912.5	8,271.3	771.4	20,407.1	1,349.6
AB	High	596	71	7,663.6	942.4	13,163.1	861.3	37,914.4	1,383.0
BC	Low	727	132	7,204.4	907.1	6,658.8	866.4	15,552.9	1,398.9
BC	Medium	728	116	7,407.1	965.0	10,171.1	950.3	29,096.1	1,401.3
BC	High	364	85	5,646.7	510.4	6,959.6	524.4	22,869.7	842.3

Note: The numbers for Canada include the 10 provinces but not the 3 territories. The income boundaries are $45,000 for the low- to medium-income households and $85,000 for the medium- to high-income households. Newfoundland and Labrador is denoted by NL; Prince Edward Island is denoted by PEI; Nova Scotia is denoted by NS; New Brunswick is denoted by NB; Quebec is denoted by QC; Ontario is denoted ON; Manitoba is denoted by MB; Saskatchewan is denoted by SK; Alberta is denoted by AB; and British Columbia is denoted by BC.

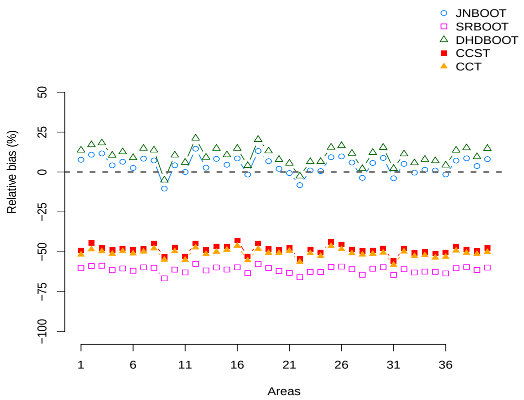
Figure 2: Scenario 3, plots of the relative biases of the MSE estimators



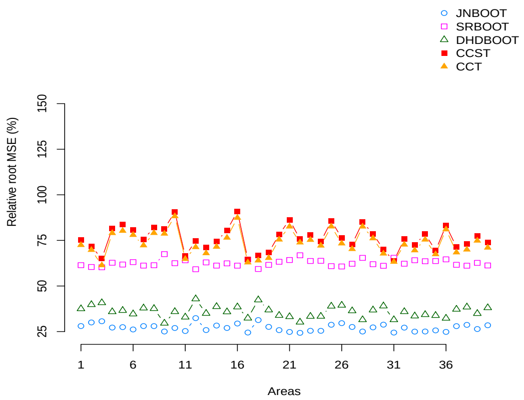
(a) CCST3



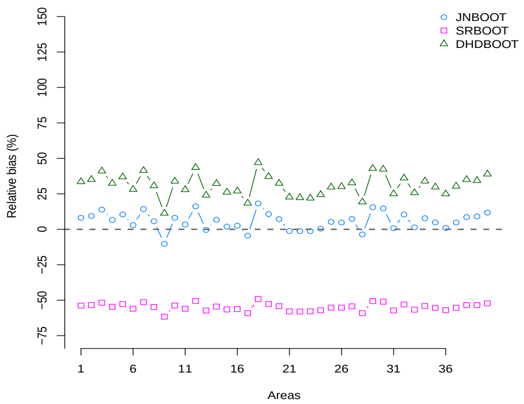
(b) CCST3



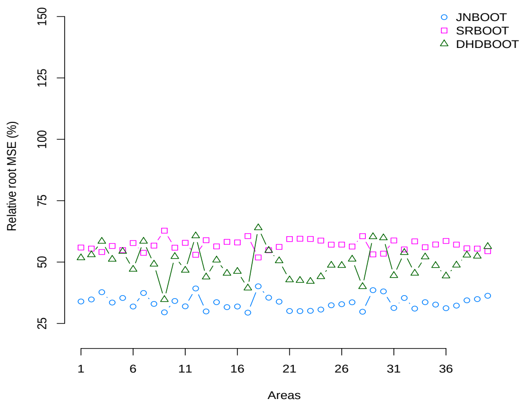
(c) SR



(d) SR



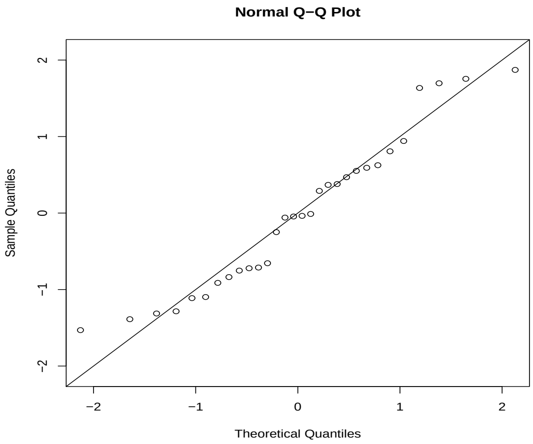
(e) JHD



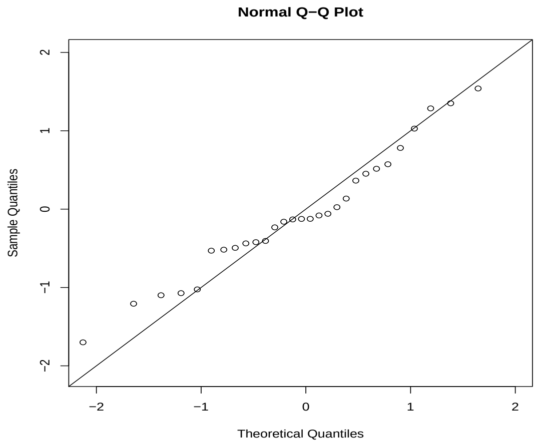
(f) JHD

Note: The proposed bootstrap MSE procedure is in blue and denoted by JNBOOT. The bootstrap MSE estimator of Sinha and Rao (2009) is in pink and denoted by SRBOOT. The bootstrap MSE estimator of Jiongo et al. (2013) is in green and denoted by JHDBOOT. The analytical linearization MSE and linearization-based MSE estimators developed by Chambers et al. (2014) are in yellow and red and denoted by CCT and CCST, respectively.

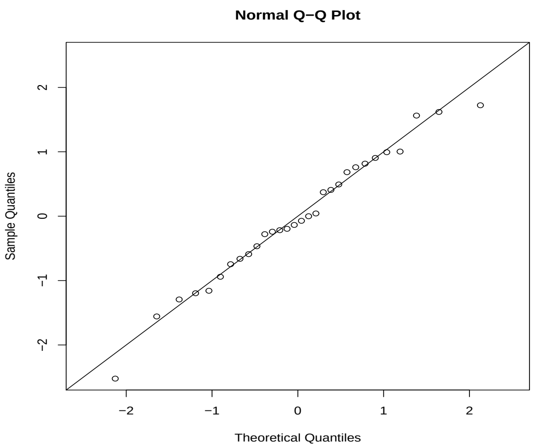
Figure 3: Normal plots for the random effects with the 2013 volumes and values of cash, debit card and credit card transaction models



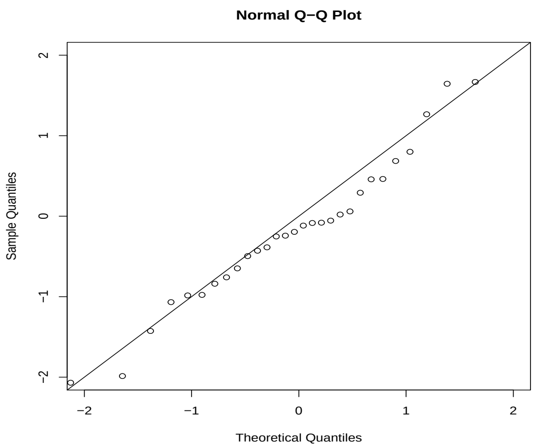
(a) cash volume



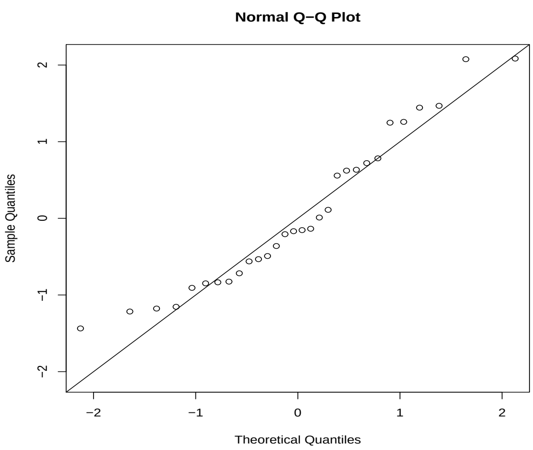
(b) cash value



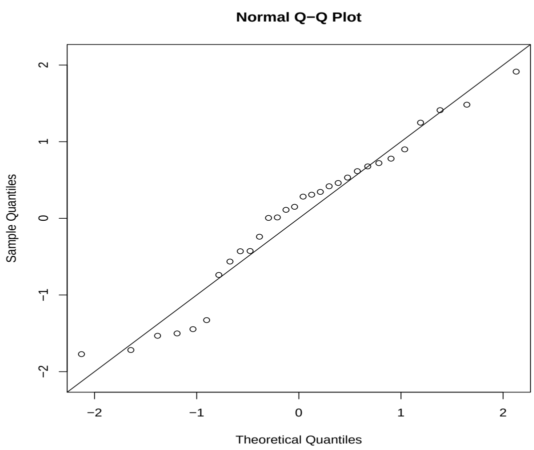
(c) debit volume



(d) debit value



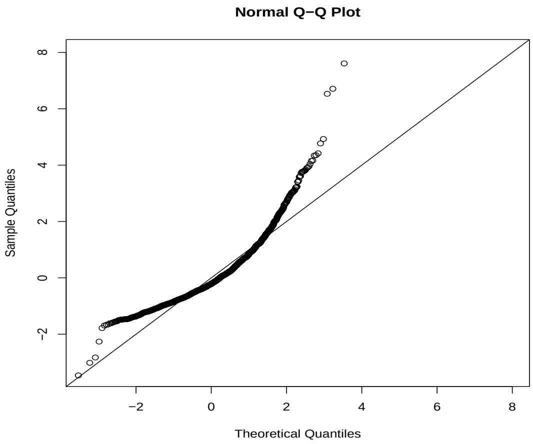
(e) credit volume



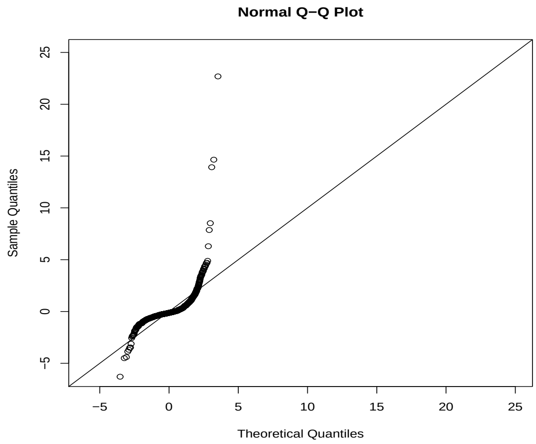
(f) credit value

Note: The plots are computed for the model (26) given by
$y_{ij} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3 + v_i + e_{ij}, \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, n_i,$
where the errors and random effects are assumed to be an independent normal distribution with variance parameters σ_e^2 and σ_v^2, respectively.

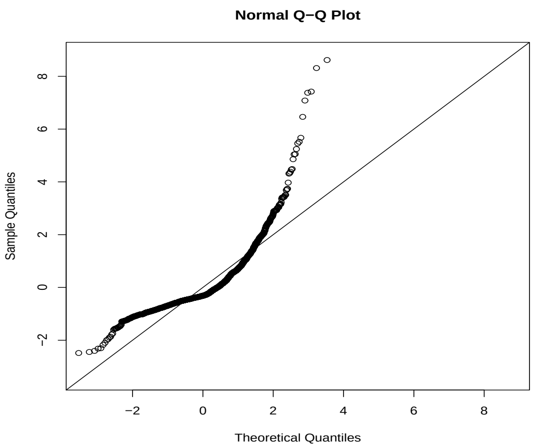
Figure 4: Normal plots for the residuals with the 2013 volumes and values of cash, debit card and credit card transaction models



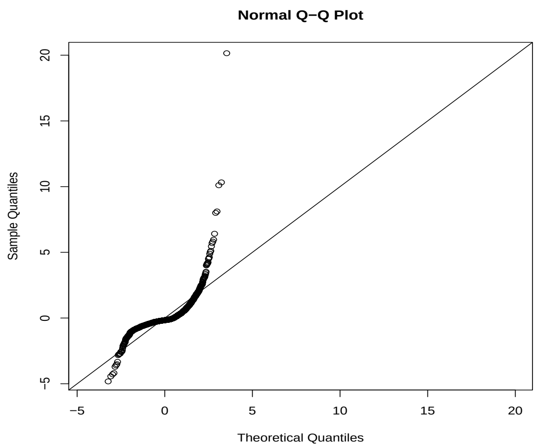
(a) cash volume



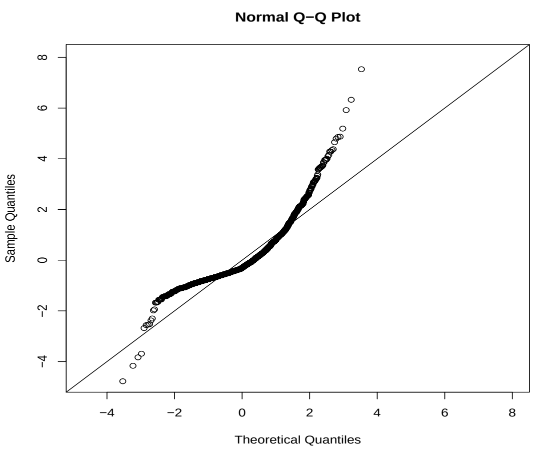
(b) cash value



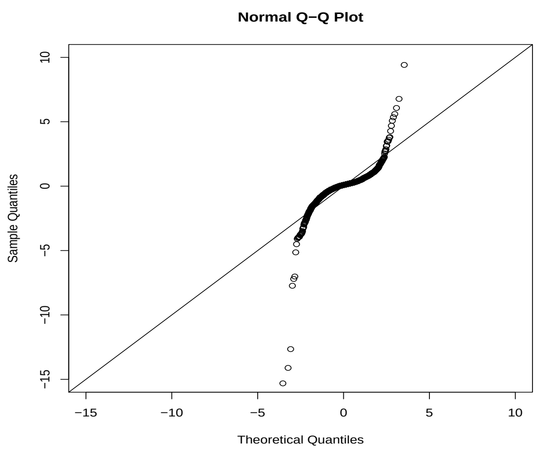
(c) debit volume



(d) debit value



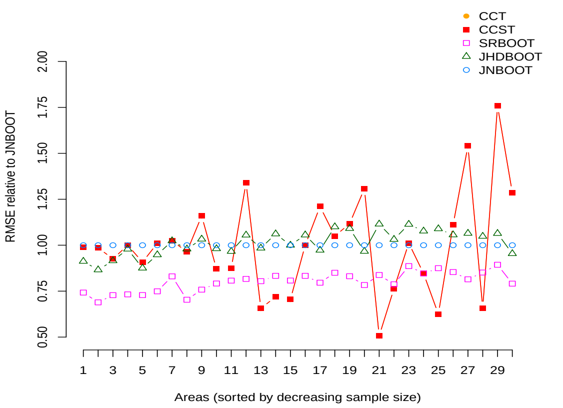
(e) credit volume



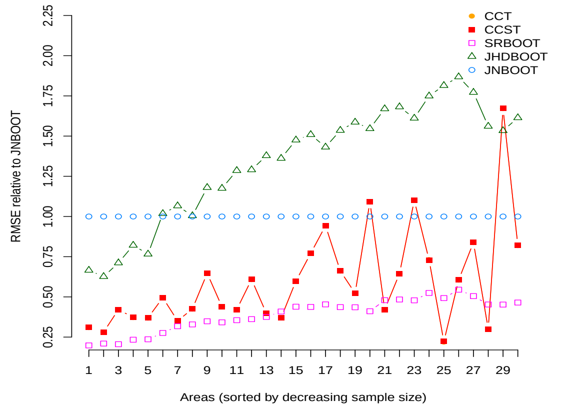
(f) credit value

Note: The plots are computed for the model (26) given by
$y_{ij} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3 + v_i + e_{ij}, \quad i = 1, \ldots, k$ and $j = 1, \ldots, n_i,$
where the errors and random effects are assumed to be an independent normal distribution with variance parameters σ_e^2 and σ_v^2, respectively.

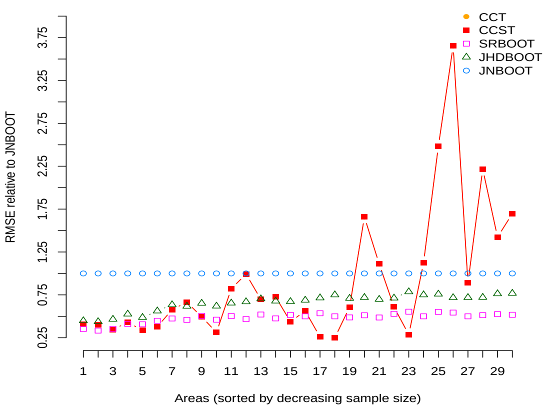
Figure 5: Result for the cash payment method in the 2013 MOP: ratio of the RMSE relative to the JNBOOT RMSE



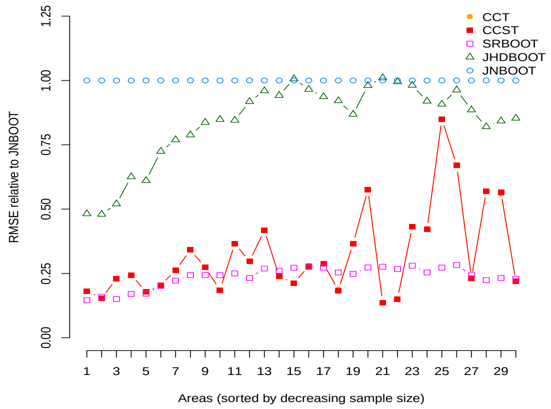
(a) CCST3 volume



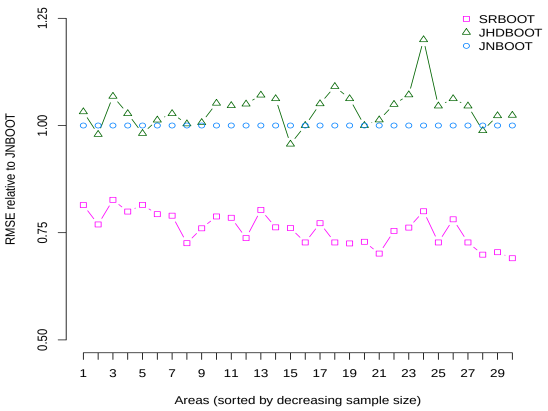
(b) CCST3 value



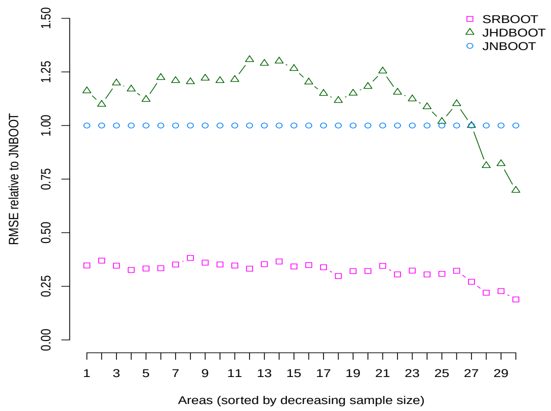
(c) SR volume



(d) SR value



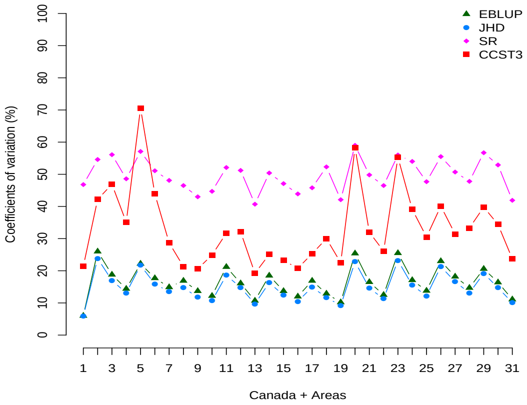
(e) JHD volume



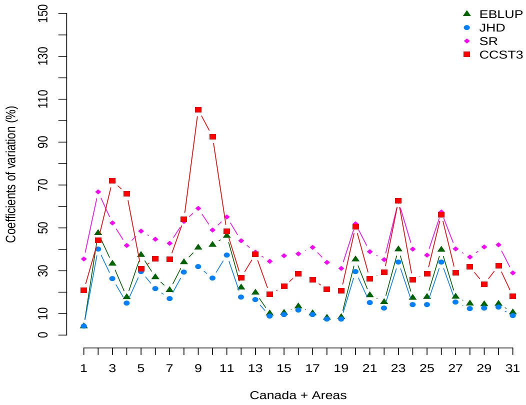
(f) JHD value

Note: The areas are sorted by decreasing sample size. The ratio is computed as $\frac{\sqrt{\widehat{MSE}(\hat{Y}_i)}}{\sqrt{\widehat{MSE}^{JNBOOT}(\hat{Y}_i)}}$, where $\widehat{MSE}\left(\hat{Y}_i\right)$ is a generic notation for CCT, CCST, JHDBOOT, SRBOOT and JNBOOT of point estimates $\hat{Y}_i$; and $\widehat{MSE}^{JNBOOT}$ denotes the proposed estimated MSE.

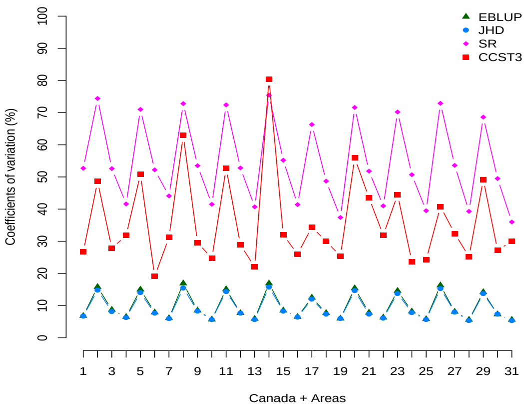
Figure 6: Coefficients of variation of the predictors of the small-area total value of transactions of cash, debit cards and credit cards.



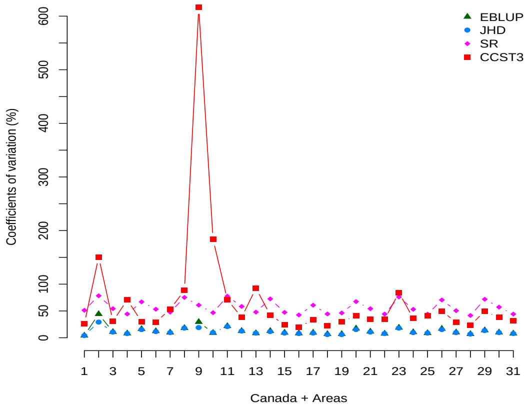
(a) cash 2009



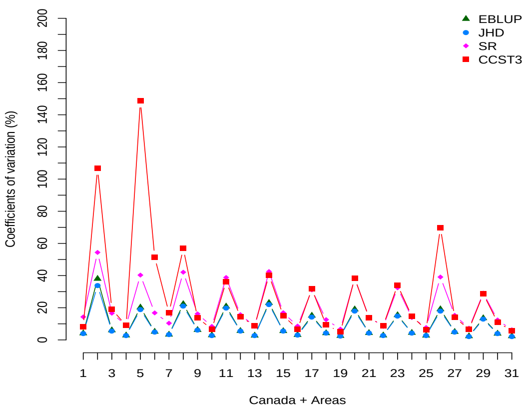
(b) cash 2013



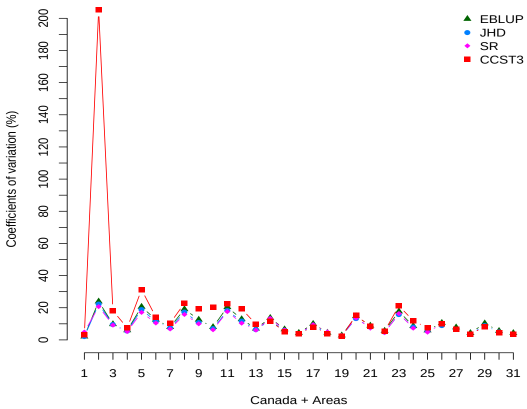
(c) debit 2009



(d) debit 2013



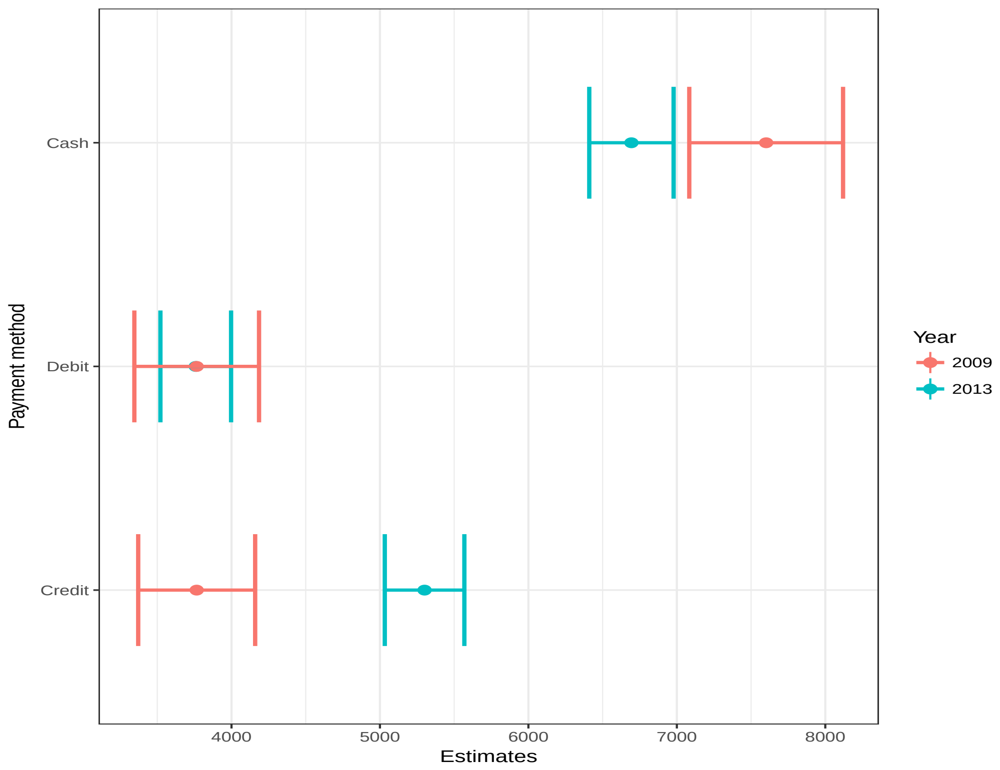
(e) credit 2009



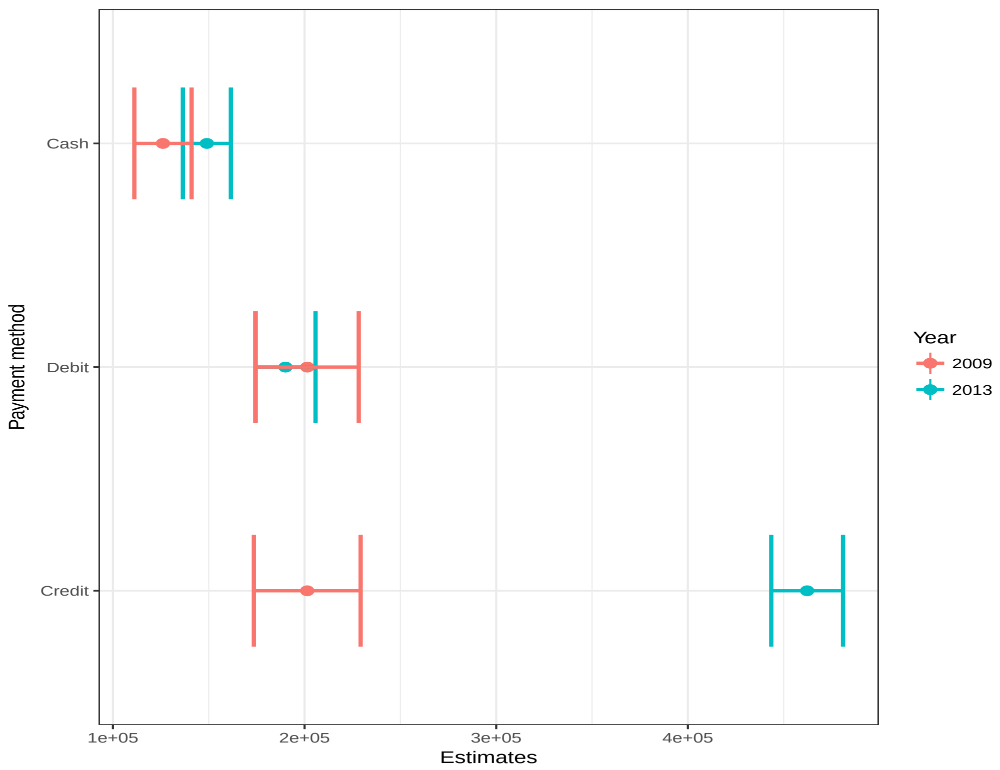
(f) credit 2013

Note: The coefficient of variation is computed by $CV_i = \frac{\widehat{MSE}(\hat{Y}_i)}{\hat{Y}_i}$, where $\widehat{MSE}\left(\hat{Y}_i\right)$ and $\hat{Y}_i$ denote the estimated MSE and total value of transactions with a given payment method (cash, debit or credit) for the household income group i within each province, respectively.

Figure 7: Volume and value of transactions at the national level with the associated confidence intervals



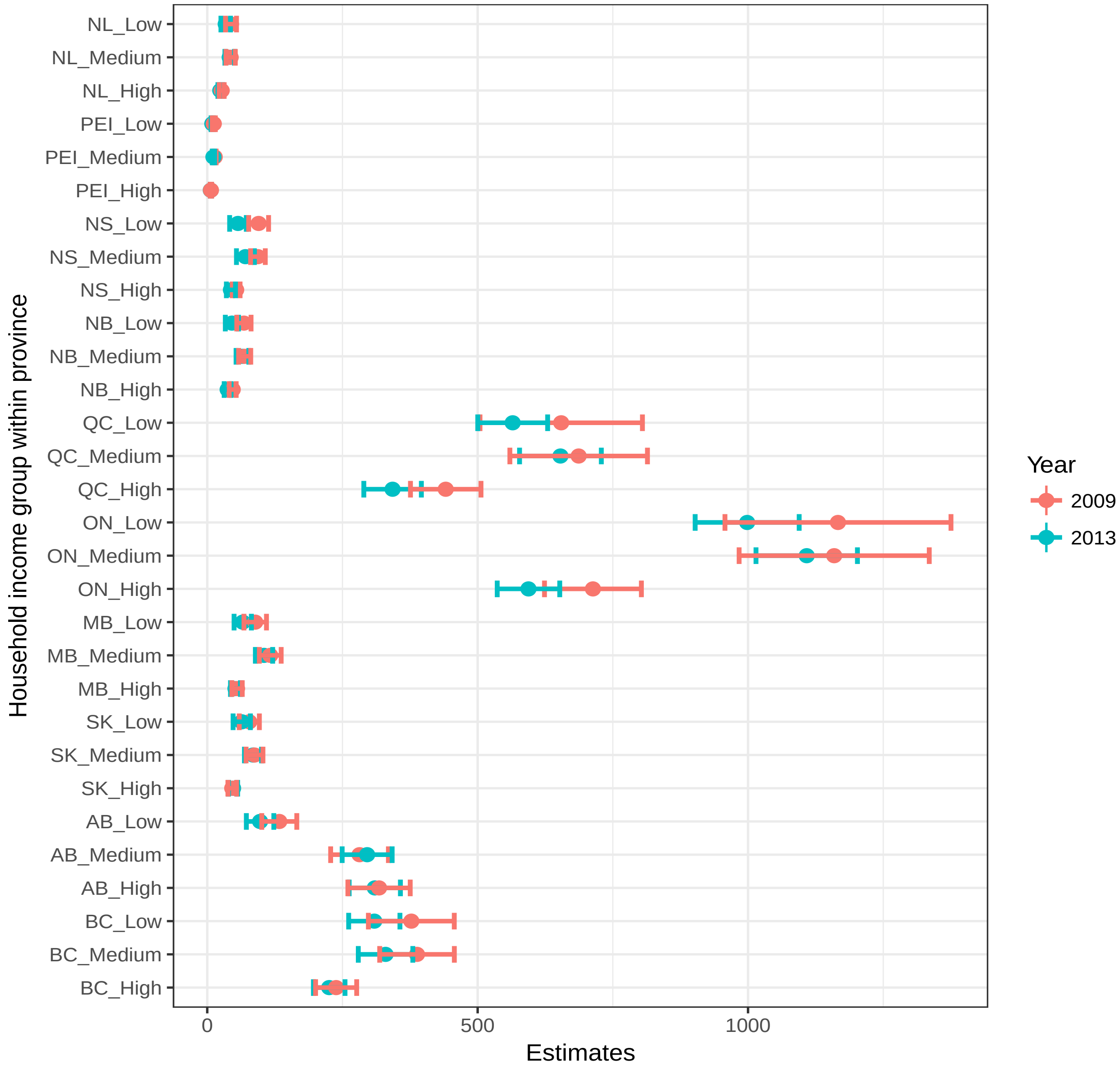
(a) Volume



(b) Value

Note: The confidence intervals are computed as $\hat{Y}_i \pm 1.96\sqrt{\widehat{MSE}\left(\hat{Y}_i\right)}$, where $\widehat{MSE}\left(\hat{Y}_i\right)$ and $\hat{Y}_i$ denote the estimated MSE and total value of transactions with a given payment method (cash, debit or credit), respectively.

Figure 8: Volume of transactions at the nested province household income group level with the associated confidence intervals



Note: The confidence intervals are computed as $\hat{Y}_i \pm 1.96\sqrt{\widehat{MSE}\left(\hat{Y}_i\right)}$, where $\widehat{MSE}\left(\hat{Y}_i\right)$ and $\hat{Y}_i$ denote the estimated MSE and total value of transactions with a given payment method (cash, debit or credit), respectively. Newfoundland and Labrador is denoted by NL; Prince Edward Island is denoted by PEI; Nova Scotia is denoted by NS; New Brunswick is denoted by NB; Quebec is denoted by QC; Ontario is denoted ON; Manitoba is denoted by MB; Saskatchewan is denoted by SK; Alberta is denoted by AB; and British Columbia is denoted by BC.

Appendix: Proofs

This section provides the proofs of Conditions (21) to (24) stated in Lemma 3.

Proof of Lemma 3

Proof of (21): $d_4\left(F_v, \hat{F}_{uk}\right) \xrightarrow{p} 0$ as $k \longrightarrow \infty$ and $d_4\left(F_e, \hat{F}_{ek}\right) \xrightarrow{p} 0$ as $k \longrightarrow \infty$.
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

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

and this implies that

$$d_4\left(F_v, \hat{F}_{uk}\right)^4 \leq 8 d_4\left(F_v, F_u\right)^4 + 64 d_4\left(F_u, F_{uk}\right)^4 + 64 d_4\left(F_{uk}, \hat{F}_{uk}\right)^4.$$

Notice that v_i and e_{ij} are in the local neighborhood of normal distribution given by (4); we have: $\hat{u}_i = u_i + O_p(k^{-1/2})$, where $u_i = \sqrt{\rho_i}\left(\alpha_i + \bar{\varepsilon}_i\right)$. By the stability of the normal distribution of α_i and ε_{ij}, 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 and Freedman (1981) we also have $d_4\left(F_u, F_{uk}\right)^4 \xrightarrow{p} 0$ as $k \longrightarrow \infty$.

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

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

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

Likewise, we have

$$d_4\left(F_e, \hat{F}_{ek}\right)^4 \leq 8 d_4\left(F_e, F_\omega\right)^4 + 64 d_4\left(F_\omega, F_{\omega k}\right)^4 + 64 d_4\left(F_{\omega k}, \hat{F}_{ek}\right)^4,$$

where $\omega_{ij} = (1-\tau_i)\alpha_i + \varepsilon_{ij} - \tau_i \bar{\varepsilon}_i, i = 1, \ldots, k \quad j = 1, \ldots, n_i$.
Note that the sampling residuals are $\tilde{e}_{ij} = \hat{e}_{ij} - \frac{1}{n}\sum_{g=1}^{k}\sum_{l \in s_g} \hat{e}_{gl}$ and that since the v_i and e_{ij} are in the local neighborhood of normal distribution given by (4), the ω_{ij} are independent and identically distributed with the same distribution F_e. Then $d_4\left(F_e, F_\omega\right) = 0$, and by Lemma 8.4 of Bickel and Freedman (1981), $d_4\left(F_\omega, F_{\omega k}\right)^4 \xrightarrow{p} 0$.

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

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

$$\begin{aligned} d_4\left(F_{\omega k}, \hat{F}_{ek}\right)^4 &\leq \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\tilde{e}_{ij} - \omega_{ij}\right)^4 \\ &= O_p(k^{-2}) \\ & \quad \xrightarrow{p} 0 \quad \text{as} \quad k \longrightarrow \infty. \end{aligned}$$

Hence, $d_4\left(F_e, \hat{F}_{ek}\right)^4 \xrightarrow{p} 0$ as $k \longrightarrow \infty$. $\square$

Proof of (22): $J_1^* \xrightarrow{p} J_1$, where

$$J_1^* = \lim_{k\to\infty} \sum_{i=1}^{k} I_{1k}^{-1/2} X_i^\top V_i^{-1} U_i^{1/2} E_* \left\{\Psi_b(r_i^*)\Psi_b(r_i^*)^\top\right\} U_i^{1/2} V_i^{-1} X_i I_{1k}^{-1/2}.$$

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

$$J_{1k} = \sum_{i=1}^{k} I_{1k}^{-1/2} X_i^\top V_i^{-1} U_i^{1/2} E_m \left\{\Psi_b(r_i)\Psi_b(r_i)^\top\right\} U_i^{1/2} V_i^{-1} X_i I_{1k}^{-1/2},$$

$$a_2 = E_m\left\{\psi_b^2(r_{ij})\right\}, \quad a_{11} = E_m\left\{\psi_b(r_{ij_1})\psi_b(r_{ij_2})\right\}, \quad \text{for} \quad j_1 \neq j_2.$$

Then, a straightforward calculation shows that

$$J_{1k} = \begin{pmatrix} J_{111k} & J_{112k} \\ J_{112k}^\top & J_{122k} \end{pmatrix},$$

where

$$\begin{aligned}
J_{111k} &= \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11} + \frac{\sigma_e^2}{\sigma_v^2}(\rho_i - \rho_i^2)(a_2 - a_{11})\right\}, \\
J_{112k} &= \sqrt{\frac{k}{n}}\left[\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11} + \frac{\sigma_e^2}{\sigma_v^2}(\rho_i - \rho_i^2)(a_2 - a_{11})\right\}\bar{\tilde{X}}_i\right], \\
J_{122k} &= \frac{k}{n}\left[\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11} + \frac{\sigma_e^2}{\sigma_v^2}(\rho_i - \rho_i^2)(a_2 - a_{11})\right\}\bar{\tilde{X}}_i^\top \bar{\tilde{X}}_i\right] + \sum_{i=1}^{k}\frac{n_i}{n}\left\{\frac{1}{n_i}\tilde{X}_i^\top \tilde{X}_i - \bar{\tilde{X}}_i^\top \bar{\tilde{X}}_i\right\},
\end{aligned}$$

and $\rho_i = \dfrac{n_i\sigma_v^2}{\sigma_e^2 + n_i\sigma_v^2}$.

Taking the expression to the limit as $k \longrightarrow \infty$ yields

$$J_1 = \begin{pmatrix} J_{111} & J_{112} \\ J_{112}^\top & J_{122} \end{pmatrix},$$

where

$$\begin{aligned}
J_{111} &= \frac{\sigma_e^4}{\sigma_v^4}\nu_1 a_{11} + \frac{\sigma_e^2}{\sigma_v^2}(\nu_1 - \nu_2)(a_2 - a_{11}), \\
J_{112} &= \sqrt{c}\lim_{k\to\infty}\left[\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11} + \frac{\sigma_e^2}{\sigma_v^2}(\rho_i - \rho_i^2)(a_2 - a_{11})\right\}\bar{\tilde{X}}_i\right], \\
J_{122} &= \lim_{k\to\infty}\left[c\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11} + \frac{\sigma_e^2}{\sigma_v^2}(\rho_i - \rho_i^2)(a_2 - a_{11})\right\}\bar{\tilde{X}}_i^\top \bar{\tilde{X}}_i + \sum_{i=1}^{k}\frac{n_i}{n}\left\{\frac{1}{n_i}\tilde{X}_i^\top \tilde{X}_i - \bar{\tilde{X}}_i^\top \bar{\tilde{X}}_i\right\}\right],
\end{aligned}$$

with

$$\nu_1 = \lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\rho_i, \quad \text{and} \quad \nu_2 = \lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\rho_i^2.$$

Using the same reasoning, we can obtain the bootstrap version J_1^* of J_1 defined by

$$J_1^* = \begin{pmatrix} J_{111}^* & J_{112}^* \\ J_{112}^{*\top} & J_{122}^* \end{pmatrix},$$

where

$$\begin{aligned}
J_{111}^* &= \frac{\hat{\sigma}_e^4}{\hat{\sigma}_v^4}\hat{\nu}_1 a_{11}^* + \frac{\hat{\sigma}_e^2}{\hat{\sigma}_v^2}(\hat{\nu}_1 - \hat{\nu}_2)(a_2^* - a_{11}^*), \\
J_{112}^* &= \sqrt{c}\lim_{k\to\infty}\left[\frac{1}{k}\sum_{i=1}^k \left\{\frac{\hat{\sigma}_e^4}{\hat{\sigma}_v^4}\hat{\rho}_i^2 a_{11}^* + \frac{\hat{\sigma}_e^2}{\hat{\sigma}_v^2}(\hat{\rho}_i - \hat{\rho}_i^2)(a_2^* - a_{11}^*)\right\}\bar{\tilde{X}}_i\right], \\
J_{122}^* &= \lim_{k\to\infty}\left[c\frac{1}{k}\sum_{i=1}^k \left\{\frac{\hat{\sigma}_e^4}{\hat{\sigma}_v^4}\hat{\rho}_i^2 a_{11}^* + \frac{\hat{\sigma}_e^2}{\hat{\sigma}_v^2}(\hat{\rho}_i - \hat{\rho}_i^2)(a_2^* - a_{11}^*)\right\}\bar{\tilde{X}}_i^\top \bar{\tilde{X}}_i + \sum_{i=1}^k \frac{n_i}{n}\left\{\frac{1}{n_i}\tilde{X}_i^\top \tilde{X}_i - \bar{\tilde{X}}_i^\top \bar{\tilde{X}}_i\right\}\right],
\end{aligned}$$

and

$$\hat{\nu}_1 = \frac{1}{k}\sum_{i=1}^k \hat{\rho}_i, \quad \hat{\nu}_2 = \frac{1}{k}\sum_{i=1}^k \hat{\rho}_i^2, \quad a_2^* = E_*\left\{\psi_b^2(r_{ij}^*)\right\}, \quad a_{11}^* = E_*\left\{\psi_b(r_{ij_1}^*)\psi_b(r_{ij_2}^*)\right\}, \quad j_1 \neq j_2.$$

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 $(\hat{\sigma}_v^2, \hat{\sigma}_e^2, \hat{\nu}_1, \hat{\nu}_2)$ converge in probability to $(\sigma_v^2, \sigma_e^2, \nu_1, \nu_2)$. Hence, by continuity, $J_1^* \xrightarrow{p} J_1$. □

Proof of (23): $J_2^* \xrightarrow{p} J_2$, where

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

The proof of (23) proceeds exactly as for (22). The derivation is, however, more cumbersome because it requires calculation of the fourth-order moments. Thus, after a lengthy algebraic expansion, one could write

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

with

$$J_{211k} = \frac{1}{\sigma_e^4}\left(\frac{1-\eta}{\eta}\right)^2\left(\frac{a_{11} - \eta a_2}{1-\eta}\right)^2\left(\bar{\rho} - \frac{1}{\eta}\bar{\rho^2}\right)^2 k + A_{11k},$$

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

$$\begin{aligned}
A_{11k} &= \frac{1}{\sigma_e^4\eta^2}\left(a_4 - 4a_{31} - 4a_{22} - 12a_{211} - 6a_{1111}\right)\left(\bar{\rho} - 3\bar{\rho^2} + 3\bar{\rho^3} - \bar{\rho^4}\right) \\
&+ \frac{1}{\sigma_e^4\eta^2}\left\{4a_{31} + 4a_{22} - 18a_{211} + 11a_{1111} - (a_2 - a_{11})^2\right\}\left(\bar{\rho^2} - 2\bar{\rho^3} + \bar{\rho^4}\right) \\
&+ \frac{1-\eta}{\sigma_e^4\eta^3}\left\{6a_{211} - 6a_{1111} - 2a_{11}(a_2 - a_{11})\right\}\left(\bar{\rho^3} - \bar{\rho^4}\right) \\
&+ \frac{(1-\eta)^2}{\sigma_e^4\eta^4}\left(a_{1111} - a_{11}^2\right)\bar{\rho^4} \\
&= A_{11}\left(\sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \bar{\rho}, \bar{\rho^2}, \bar{\rho^3}, \bar{\rho^4}\right).
\end{aligned}$$

The numbers $a_4, a_{31}, a_{22}, a_{211}, a_{1111}$ are fourth-order moments defined by

$$a_4 = E_m\left\{\psi_b^4(r_{ij})\right\}, \quad a_{31} = E_m\left\{\psi_b^3(r_{ij_1})\psi_b(r_{ij_2})\right\}, \quad a_{22} = E_m\left\{\psi_b^2(r_{ij_1})\psi_b^2(r_{ij_2})\right\},$$
$$a_{211} = E_m\left\{\psi_b^2(r_{ij_1})\psi_b(r_{ij_2})\psi_b(r_{ij_3})\right\}, \quad a_{1111} = E_m\left\{\psi_b(r_{ij_1})\psi_b(r_{ij_2})\psi_b(r_{ij_3})\psi_b(r_{ij_4})\right\},$$
$$\text{where} \quad j_1 \neq j_2, \quad j_1 \neq j_3, \quad j_1 \neq j_4,, \quad j_2 \neq j_3, \quad j_2 \neq j_4, \quad \text{and} \quad j_3 \neq j_4,$$

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

Note that $A_{11}(\cdot)$, as defined above, is a continuous function of its arguments.

Likewise,

$$\begin{aligned} J_{212k} &= \frac{1}{\sigma_e^4}\left(\frac{1-\eta}{\eta}\right)\left(\frac{a_{11}-\eta a_2}{1-\eta}\right)^2\left(\bar{\rho} - \frac{1}{\eta}\bar{\rho^2}\right)\left(1 - \bar{\rho}\sqrt{\frac{k}{n}}\right)\sqrt{nk} \\ &+ \frac{1}{\sigma_e^4\eta^2}\left\{a_{211} - a_{1111} + a_{11}(a_2 - a_{11})\right\}\sqrt{\frac{n}{k}} + A_{12k}\sqrt{\frac{k}{n}}, \end{aligned}$$

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

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

$$\begin{aligned} J_{222k} &= \frac{1}{\sigma_e^4}\left(\frac{a_{11}-\eta a_2}{1-\eta}\right)^2\left\{1 - \left(\bar{\rho} + \frac{1}{\eta}\bar{\rho} - \frac{1}{\eta}\bar{\rho^2}\right)\frac{k}{n}\right\}^2 n \\ &+ \frac{1}{\sigma_e^4(1-\eta)^2}\left\{a_{22} - 2a_{211} + a_{1111} - (a_2 - a_{11})^2\right\}\left(\frac{1}{n}\sum_{i=1}^k n_i^2\right) \\ &+ \frac{1}{\sigma_e^4(1-\eta)^2}\left\{a_4 - 3a_{22} - 4a_{31} + 12a_{211} - a_{1111} + 2(a_2 - a_{11})^2\right\} + A_{22k} \times \frac{k}{n}, \end{aligned}$$

where, as above, A_{22k} is a bounded sequence of real numbers and can be written as $A_{22k} = A_{22}\left(\sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \bar{\rho}, \bar{\rho^2}, \bar{\rho^3}, \bar{\rho^4}\right)$, where $A_{12}(\cdot)$ 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

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

It follows that

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

where

$$
\begin{aligned}
J_{211} &= \frac{1}{\sigma_e^4 \eta^2}\left(a_4 - 4a_{31} - 4a_{22} - 12a_{211} - 6a_{1111}\right)\left(\nu_1 - 3\nu_2 + 3\nu_3 - \nu_4\right) \\
&+ \frac{1}{\sigma_e^4 \eta^2}\left\{4a_{31} + 4a_{22} - 18a_{211} + 11a_{1111} - (a_2 - a_{11})^2\right\}\left(\nu_2 - 2\nu_3 + \nu_4\right) \\
&+ \frac{1-\eta}{\sigma_e^4 \eta^3}\left\{6a_{211} - 6a_{1111} - 2a_{11}(a_2 - a_{11})\right\}\left(\nu_3 - \nu_4\right) \\
&+ \frac{(1-\eta)^2}{\sigma_e^4 \eta^4}\left(a_{1111} - a_{11}^2\right)\nu_4, \\
&= A_{11}\left(\sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \nu_1, \nu_2, \nu_3, \nu_4\right),
\end{aligned}
$$

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

and

$$
\begin{aligned}
J_{222} &= \frac{1}{\sigma_e^4(1-\eta)^2}\left\{a_4 - 3a_{22} - 4a_{31} + 12a_{211} - a_{1111} + 2(a_2 - a_{11})^2\right\} \\
&+ c A_{22}\left(\sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \nu_1, \nu_2, \nu_3, \nu_4\right).
\end{aligned}
$$

Likewise, the bootstrap version J_2^* of J_2 is given by

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

where

$$
\begin{aligned}
J_{211}^* &= A_{11}\left(\hat{\sigma}_e^2, \hat{\eta}, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, \hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\nu}_4\right), \\
J_{212}^* &= \sqrt{c} A_{12}\left(\hat{\sigma}_e^2, \hat{\eta}, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, \hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\nu}_4\right), \\
J_{222}^* &= \frac{1}{\hat{\sigma}_e^4(1-\hat{\eta})^2}\left\{a_4^* - 3a_{22}^* - 4a_{31}^* + 12a_{211}^* - a_{1111}^* + 2(a_2^* - a_{11}^*)^2\right\} \\
&+ c A_{22}\left(\hat{\sigma}_e^2, \hat{\eta}, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, \hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\nu}_4\right).
\end{aligned}
$$

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

Proof of (24): $G^*(\theta) \xrightarrow{p} G(\theta)$ where

$$
G^*(\hat{\theta}) = \lim_{k\to\infty} \sum_{i=1}^{k} \begin{bmatrix} -I_{1k}^{-1/2} E_* \left\{\frac{\partial \Psi_1(y_i^*, X_i, \hat{\theta})}{\partial \hat{\beta}}\right\} I_{1k}^{-1/2} & 0 \\ 0 & -I_{2k}^{-1/2} E_* \left\{\frac{\partial \Psi_2(y_i^*, X_i, \hat{\theta})}{\partial \hat{\delta}}\right\} I_{2k}^{-1/2} \end{bmatrix},
$$

and

$$
G(\theta) = \lim_{k\to\infty} \sum_{i=1}^{k} \begin{bmatrix} -I_{1k}^{-1/2} E_m \left\{\frac{\partial \Psi_1(y_i, X_i, \theta)}{\partial \beta}\right\} I_{1k}^{-1/2} & 0 \\ 0 & -I_{2k}^{-1/2} E_m \left\{\frac{\partial \Psi_2(y_i, X_i, \theta)}{\partial \delta}\right\} I_{2k}^{-1/2} \end{bmatrix}.
$$

We use the same reasoning as for the proofs of (22) and (23). A straightforward expansion of the components of $G_k(\theta)$ allows us to express it as a sum of two components, one of which is a bounded sequence and another of which depends on a_2, a_{11}, d_2, d_{11}, where $d_2 = E_m\left\{r_{ij}\psi_b'(r_{ij})\psi_b(r_{ij})\right\}$ and $d_{11} = E_m\left\{r_{ij_1}\psi_b'(r_{ij_1})\psi_b(r_{ij_2})\right\}, j_1 \neq j_2$.
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

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

where

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

where $d_1 = E_m\left\{\psi_b'(r_{ij})\right\}$.

Likewise, we have

$$G^*(\hat{\theta}) = \begin{bmatrix} G_{111}^* & G_{112}^* & 0 & 0 \\ G_{112}^{*\top} & G_{122}^* & 0 & 0 \\ 0 & 0 & G_{211}^* & G_{212}^* \\ 0 & 0 & G_{212}^* & G_{222}^* \end{bmatrix},$$

where

$$
\begin{aligned}
G_{111}^{*} &= \frac{d_1^*}{\hat{\sigma}_e^2}\frac{1-\hat{\eta}}{\hat{\eta}}\hat{\nu}_1, \\
G_{112}^{*} &= \frac{d_1^*}{\hat{\sigma}_e^2}\frac{1-\hat{\eta}}{\hat{\eta}}\sqrt{c}\lim_{k\to\infty}\left[\frac{1}{k}\sum_{i=1}^{k}\hat{\rho}_i\bar{\tilde{X}}_i\right], \\
G_{122}^{*} &= \frac{d_1^*}{\hat{\sigma}_e^2}\lim_{k\to\infty}\left[\frac{1-\hat{\eta}}{\hat{\eta}}c\frac{1}{k}\sum_{i=1}^{k}\hat{\rho}_i\bar{\tilde{X}}_i^\top\bar{\tilde{X}}_i+\sum_{i=1}^{k}\frac{n_i}{n}\left\{\frac{1}{n_i}\tilde{X}_i^\top\tilde{X}_i-\bar{\tilde{X}}_i^\top\bar{\tilde{X}}_i\right\}\right], \\
G_{211}^{*} &= \frac{1}{\hat{\sigma}_e^4}\left(\frac{1-\hat{\eta}}{\hat{\eta}}\right)^2\{(1-\hat{\eta})a_2^*+d_{11}^*\}\,\hat{\nu}_2, \\
G_{212}^{*} &= \frac{1}{\hat{\sigma}_e^4}\left(\frac{1-\hat{\eta}}{\hat{\eta}}\right)\sqrt{c}\,\{(1-\hat{\eta})a_2^*+d_{11}^*\}\,(\hat{\nu}_1-\hat{\nu}_2), \\
G_{222}^{*} &= \frac{a_2^*}{\hat{\sigma}_e^2}+\frac{c}{\hat{\sigma}_e^2}\left[a_2^*\{2\hat{\nu}_2-3\hat{\nu}_1+(\hat{\nu}_1-\hat{\nu}_2)\hat{\eta}\}+\frac{1-\hat{\eta}}{\hat{\eta}}(\hat{\nu}_1-\hat{\nu}_2)d_{11}^*\right],
\end{aligned}
$$

with

$$
d_1^* = E_*\left\{\psi_b^{'}(r_{ij}^*).\right\},\;\; d_2^* = E_*\left\{r_{ij}^*\psi_b^{'}(r_{ij}^*)\psi_b(r_{ij}^*)\right\},\;\; d_{11}^* = E_*\left\{r_{ij_1}^*\psi_b^{'}(r_{ij_1}^*)\psi_b(r_{ij_2}^*)\right\}, j_1 \neq j_2.
$$

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