STATISTICS IN TRANSITION new series, June 2020
Vol. 21, No. 2, pp. 35–60, DOI 10.21307/stattrans-2020-013
Submitted – 22.03.2019; accepted for publishing – 10.03.2020

On the choice of the number of Monte Carlo iterations and bootstrap replicates in Empirical Best Prediction

Adam Chwila[1], **Tomasz Żądło**[2]

ABSTRACT

Empirical Best Predictors (EBPs) are widely used for small area estimation purposes. In the case of longitudinal surveys, this class of predictors can be used to predict any given population or subpopulation characteristic for any time period, including future periods. Generally, the value of an EBP is computed by means of Monte Carlo algorithms, while its MSE is usually estimated using the parametric bootstrap method. Model-based simulation studies of the properties of the predictors require numerous repetitions of the random generation of population data. This leads to a question about the dependence between the number of iterations in all the procedures and the stability of the results. The aim of the paper is to show this dependence and to propose methods of choosing the appropriate number of iterations in practice, using a set of real economic longitudinal data available at the United States Census Bureau website.

Key words: survey sampling, economic longitudinal data, prediction for future periods.

1. Introduction

Empirical Best Predictors have been used in small area estimation problems for a long time. In papers published by Jiang and Lahiri (2001) and Jiang (2003) prediction problems under generalized linear mixed models were studied. A large number of papers were published after a well-known Molina and Rao (2010) paper, where this class of predictors was used to predict poverty measures. What is more, they presented a special case of the predictor under normality of the transformed variable of interest together with the proposal of a very fast algorithm for a special case of the model called the nested error mixed model. Then, many authors generalized these results relaxing normality assumption (e.g. Elbers and vad der Weide (2014), Diallo (2014) and Diallo and Rao (2018)), considering nonlinear models and usually the prediction of small area

[1] University of Economics in Katowice, Katowice, Poland, E-mail: achwila@gmail.com.
ORCID: https://orcid.org/0000-0003-4671-4298.
[2] University of Economics in Katowice, Katowice, Poland, E-mail: tomasz.zadlo@ue.katowice.pl.
ORCID: https://orcid.org/0000-0003-0638-0748.

fractions (e.g. Berg and Chandra (2014), Boubeta, Lombardía and Morales (2016, 2017), Hobza and Morales (2016), Zimmermann and Münnich (2018)), analyzing the problem of back transformation of the variable of interest (Molina and Martin (2018)) and studying the semi-parametric EBP (Marino et al. (2019)).

In these papers, the authors assume a different number of iterations in the EBP procedure (which will be denoted by L), in the parametric bootstrap method used to estimate MSE (which will be denoted by B) and in Monte Carlo simulation studies (which will be denoted by K). The applications presented in these papers are usually supported by model-based simulation studies. It gives possibility to use additional methods to choose the appropriate number of iterations based on simulation results, such as stability of simulation results or the simulation bias of unbiased Best Predictor, which cannot be computed in practice (for real data).

It is clear that the appropriate choice of the number of iterations is different for different data, different models and different prediction problems and hence we would like to present some examples studied by different authors. Although in practice, as stated by Tzavidis et al. (2018), usually $L = 50$ or $L = 100$ is used, in the small area estimation literature different numbers of iterations L in the EBP procedure are studied:

- in applications: from 50 to 1000 in Molina and Rao (2010), 100 in Guadarrama, Molina and Rao (2018),
- in simulation studies: 50 in Molina and Rao (2010), 100 in Das and Haslett (2019), 500 in Boubeta, Lombardía, Morales (2017) and 2500 in Boubeta, Lombardía, Morales (2016).

Examples of numbers of iterations B taken into account by different authors are:

- in applications: 500 in Molina and Rao (2010), Hobza and Morales (2016), Boubeta, Lombardía and Morales (2017), Guadarrama, Molina and Rao (2018),
- in simulation studies: 500 in Molina and Rao (2010), Boubeta, Lombardía and Morales (2016), Guadarrama, Molina and Rao (2018); and 1000 in González-Manteiga, Lombardía, Molina, Morales and Santamaría (2008).

The numbers of iterations in Monte Carlo simulation studies assumed by different authors equal: 500 in Das and Haslett (2019), 500 and 10 000 and 50 000 for different purposes in Molina and Rao (2010); 500 and 1 000 and 10 000 for different purposes in Guadarrama, Molina and Rao (2018); 1 000 in González-Manteiga, Lombardía, Molina, Morales and Santamaría (2008), Guadarrama, Molina and Rao (2014), Boubeta, Lombardía, Morales (2016, 2017); 5 000 in Diallo and Rao (2018); 10 000 in Hobza and Morales (2016) and Molina and Martín (2018); 50 000 in Jiang and Lahiri (2006).

Based on a real economic longitudinal dataset we analyse three problems which, according to our best knowledge, are not presented in the literature:

- the dependence between the number of iterations L of the EBP procedure and the stability of EBP values,
- the dependence between the number of iterations B in the parametric bootstrap procedure and the stability of values of MSE estimators,
- the dependence between the number of Monte Carlo iterations K and the stability of ratios of MSEs of the predictors: the EBP and the BP.

We also propose:

- two criteria allowing the appropriate choice of L and B, which can be used in practice (based on real sample data),
- a criterion allowing to choose the appropriate number of iterations K in simulation studies.

2. Some remarks on bootstrap procedures

In this section we present the literature review on the convergence of bootstrap procedures taking into account two issues. Firstly, we are interested in analysing how bootstrap estimators under B replications approximate their values when B tends to infinity. Secondly, we show that based on some bootstrap procedures we can obtain asymptotically unbiased estimators of some unknown parameters. Although we are interested in the parametric bootstrap method, we discuss available results for different bootstrap procedures.

Davison and Hinkley (1997) pp. 34–37 study the problem of the decomposition of variances of different bootstrap estimators into the part resulting from data variation and simulation variation. They study nonparametric bootstrap procedure and derive variances and bootstrap variances of the following statistics: bootstrap estimator of the bias of the sample mean, bootstrap estimator of the variance of the sample mean and bootstrap estimator of the variance of the sample quantile. They present bootstrap variances of these statistics as functions of: (i) their unconditional variance and (ii) the simulation variance depending on the number of bootstrap iterations. It gives a direct tool to determine the number of nonparametric bootstrap replicates to obtain the required ratio of the simulation variance and the unconditional variance. Davison and Hinkley (1997) pp. 155–156 study also the problem of the convergence of the parametric bootstrap procedure but in the case of testing hypotheses. They derive powers of tests in two cases: for the given number of bootstrap iterations and when it tends to infinity. Their ratio is a function of bootstrap replicates, which allows one to determine the number of replicates to obtain the required level of the ratio.

Efron and Tibshirani (1986) p. 72 study the number nonparametric bootstrap replications in the case of estimation of the standard error showing that the CV of the bootstrap estimator of the standard error based on B replications is a function of: (i) the CV of the bootstrap estimator of the standard error based on infinite B replications, (ii) the number of bootstrap replications and (iii) the expected value (over the distribution of the variable of interest) of the kurtosis of the bootstrap distribution of the considered estimator. Because the formula is generally not estimable, it is not used to find a specific value of bootstrap replications but to determine a range of acceptable values.

An interesting procedure is proposed and studied by Andrews and Buchinsky (1997, 2000, 2001). They study two cases in bootstrap procedures – firstly, B iterations and, secondly, an infinite number of iterations. They determine the number of bootstrap iterations to obtain the value of the modulus of the percentage deviation between values of bootstrap estimators in these two cases not greater than the specified value with the declared probability close to 1. It can be used for different bootstrap techniques including parametric and nonparametric bootstrap and both for independent and dependent data. Estimation of the MSE is not considered by the authors – they consider estimation of the square root of variance, confidence intervals, test statistics and p-values. In simulation studies they consider properties of their method only for standard nonparametric bootstrap.

Even if a bootstrap estimator accurately approximates its value under infinite number of replications, it does not mean that it is a good estimator of the parameter. Usually bootstrap approximates the population distribution of certain sample statistics but the failure in convergence of the bootstrap distribution to the correct distribution may also occur (e.g. Beran (1997)). Hall and Martin (1988) prove that the nonparametric bootstrap quantile variance estimator converges with the increase of the sample size to the true variance (but slowly). Singh (1981) shows that the nonparametric bootstrap asymptotically (when the sample size tends to infinity) approximates the population distribution of the standardized sample mean and the distribution of the sample quantiles. The parametric bootstrap MSE estimator of the empirical best linear unbiased predictor proposed by Butar and Lahiri (2003) estimates the unknown MSE with the bias of order $o(D^{-1})$, where D is the number of small areas. Chatterjee, Lahiri and Li (2008) use parametric bootstrap to estimate the distribution of the centered and scaled empirical best linear unbiased predictor and show that it accurately approximates the true distribution (and derive the order of the approximation). Hall and Maiti (2006) propose a very accurately parametric bootstrap confidence intervals, that do not depend of the form of small area predictor, with the coverage error $O(D^{-3})$. Hall and Maiti (2006) present also results crucial for our analysis – they prove that the biases of parametric bootstrap MSE estimators (considered in this paper) of both the empirical best linear unbiased predictor and the

empirical best predictor are of order $O(D^{-1})$, where D is the number of small areas. What is more, the double bootstrap MSE estimator of the predictor, not considered in our paper due to very time-consuming computations, is of order $O(D^{-2})$.

3. Empirical Best Predictor

We consider the model-based approach in survey sampling assuming the following longitudinal mixed linear model for population data:

$$Q(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \tag{1}$$

where $Q(\mathbf{Y})$ is the random vector of the variable of interest after transformation $Q(.)$ including random variables for future periods in the case of longitudinal data, $\mathbf{X}$ and $\mathbf{Z}$ are known matrices of full ranks of the auxiliary variables including known or assumed values for future periods, $\boldsymbol{\beta}$ is the unknown vector of fixed effects, $\mathbf{v}$ and $\mathbf{e}$ – called vectors of random effects and random components – are independent, $\mathbf{v} \sim (\mathbf{0}, \mathbf{G}(\boldsymbol{\delta}))$ and $\mathbf{e} \sim (\mathbf{0}, \mathbf{R}(\boldsymbol{\delta}))$, where $\boldsymbol{\delta}$ is a vector of unknown parameters called variance components. Without the loss of the generality, we assume that first elements in the population vector $Q(\mathbf{Y})$, are the random variables which realizations are known from the longitudinal survey, which can be written as $Q(\mathbf{Y}) = \left[Q(\mathbf{Y}_s^T) \quad Q(\mathbf{Y}_r^T) \right]^T$, where subscript "s" is used for the sample and "r" for non-sampled elements. What is more, a similar decomposition can be used for matrices of auxiliary variables: $\mathbf{X} = \left[\mathbf{X}_s^T \quad \mathbf{X}_r^T \right]^T$ and $\mathbf{Z} = \left[\mathbf{Z}_s^T \quad \mathbf{Z}_r^T \right]^T$, for the vector of random components $\mathbf{e} = \left[\mathbf{e}_s^T \quad \mathbf{e}_r^T \right]^T$ and for covariance matrix of random components $\mathbf{R}(\boldsymbol{\delta}) = \begin{bmatrix} \mathbf{R}_{ss}(\boldsymbol{\delta}) & \mathbf{R}_{sr}(\boldsymbol{\delta}) \\ \mathbf{R}_{rs}(\boldsymbol{\delta}) & \mathbf{R}_{rr}(\boldsymbol{\delta}) \end{bmatrix}$. Based on (1), the covariance matrix of $Q(\mathbf{Y})$, denoted by $\mathbf{V}(\boldsymbol{\delta})$, is given by $D^2(Q(\mathbf{Y})) = \mathbf{V}(\boldsymbol{\delta}) = \mathbf{Z}\mathbf{G}(\boldsymbol{\delta})\mathbf{Z}^T + \mathbf{R}(\boldsymbol{\delta})$ and it can be decomposed as follows $\mathbf{V}(\boldsymbol{\delta}) = \begin{bmatrix} \mathbf{V}_{ss}(\boldsymbol{\delta}) & \mathbf{V}_{sr}(\boldsymbol{\delta}) \\ \mathbf{V}_{rs}(\boldsymbol{\delta}) & \mathbf{V}_{rr}(\boldsymbol{\delta}) \end{bmatrix}$, where $\mathbf{V}_{ss}(\boldsymbol{\delta}) = \mathbf{Z}_s\mathbf{G}(\boldsymbol{\delta})\mathbf{Z}_s^T + \mathbf{R}_{ss}(\boldsymbol{\delta})$, $\mathbf{V}_{rr}(\boldsymbol{\delta}) = \mathbf{Z}_r\mathbf{G}(\boldsymbol{\delta})\mathbf{Z}_r^T + \mathbf{R}_{rr}(\boldsymbol{\delta})$, $\mathbf{V}_{sr}(\boldsymbol{\delta}) = \mathbf{Z}_s\mathbf{G}(\boldsymbol{\delta})\mathbf{Z}_r^T + \mathbf{R}_{sr}(\boldsymbol{\delta})$ and $\mathbf{V}_{rs}(\boldsymbol{\delta}) = \mathbf{V}_{sr}^T(\boldsymbol{\delta})$.

To estimate parameters of (1), i.e. vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, different methods can be used including the restricted maximum likelihood method (REML) used in this paper (see e.g. Jiang (2007) pp. 12-15). In the method the value of the estimator of $\boldsymbol{\delta}$, denoted by $\hat{\boldsymbol{\delta}}$, is computed by maximization of the Gaussian likelihood function of $\mathbf{A}^T\mathbf{Y}_s$, where $\mathbf{A}$ is any matrix such that $\mathbf{A}^T\mathbf{X}_s = \mathbf{0}$. The method is robust on non-normality – it gives consistent estimators even if the distribution is not normal (Jiang (1996)). The

estimator of $\boldsymbol{\beta}$ is given by (e.g. Jiang (2007) p. 75): $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_s^T \mathbf{V}_{ss}^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \mathbf{V}_{ss}^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{Y}_s$. The empirical best linear unbiased predictor of $\mathbf{v}$ is as follows (e.g. Jiang (2007) p. 76): $\hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}) = \mathbf{G}(\hat{\boldsymbol{\delta}}) \mathbf{Z}_s^T \mathbf{V}_{ss}^{-1}(\hat{\boldsymbol{\delta}}) (\mathbf{Y}_s - \mathbf{X}_s \hat{\boldsymbol{\beta}})$.

Under (1) the best predictor $\hat{\theta}$ of any function $\theta(Q^{-1}(\mathbf{Y}))$, or shortly θ, minimizing the mean squared error is given by (e.g. Molina and Rao (2010)):

$$\hat{\theta}_{BP} = E(\theta \mid Q(\mathbf{Y}_s)). \tag{2}$$

Special cases of $\theta(Q^{-1}(\mathbf{Y}))$ are population and subpopulation characteristics such as the mean or the median in the current or future period. The value of (2) can be computed if the shape and the parameters of the distribution $Q(\mathbf{Y}_r) \mid Q(\mathbf{Y}_s)$ are known. In practical applications the shape of the multivariate distribution of $Q(\mathbf{Y})$ is assumed, the parameters of the distribution (in the case of (1) - $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$) are estimated based on the known realization of $Q(\mathbf{Y}_s)$ (which gives $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$), and the distribution of $Q(\mathbf{Y}_r) \mid Q(\mathbf{Y}_s)$ is derived (or directly the conditional expectation given by (2)). The two-stage predictor obtained according to this idea is called the Empirical Best Predictor (EBP). Its value can be computed based on the following iterative algorithm, presented originally by Molina and Rao (2010):

- generate L vectors $Q(\mathbf{Y}_r)$ (denoted by $Q(\mathbf{Y}_r^{(l)})$, where $l = 1, 2, ..., L$) based on the empirical distribution of $Q(\mathbf{Y}_r) \mid Q(\mathbf{Y}_s)$ (the distribution of $Q(\mathbf{Y}_r) \mid Q(\mathbf{Y}_s)$ where $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are replaced by $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$),
- construct L population vectors $Q(\mathbf{Y}^{(l)}) = \left[Q(\mathbf{Y}_s^T) \quad Q(\mathbf{Y}_r^{(l)T}) \right]^T$ $(l = 1, 2, ..., L)$, where one realization of $Q(\mathbf{Y}_s)$ available from the sample and different realizations of $Q(\mathbf{Y}_r)$ are used,
- compute the EBP as $\hat{\theta}_{EBP} = L^{-1} \sum_{l=1}^{L} \theta(Q^{-1}(\mathbf{Y}^{(l)}))$ (which means that the back transformation is needed).

If we assume (1) and multivariate normality of the transformed variable of interest, then the distribution $Q(\mathbf{Y}_r) \mid Q(\mathbf{Y}_s)$ is multivariate normal with the following vector of expected values $\mathbf{X}_r \boldsymbol{\beta} + \mathbf{V}_{rs}(\boldsymbol{\delta}) \mathbf{V}_{ss}^{-1}(\boldsymbol{\delta}) \left(Q(\mathbf{Y}_s) - \mathbf{X}_s \boldsymbol{\beta} \right)$ and the following variance-covariance matrix $\mathbf{V}_{rr}(\boldsymbol{\delta}) - \mathbf{V}_{rs}(\boldsymbol{\delta}) \mathbf{V}_{ss}^{-1}(\boldsymbol{\delta}) \mathbf{V}_{sr}(\boldsymbol{\delta})$. Molina and Rao (2010) also propose a very fast algorithm for EBP computation for the special case of (1) called the nested error mixed linear model, where the generation of population vectors from the multivariate normal conditional distribution is replaced by iid generation using the univariate normal distribution.

To estimate the mean squared error of the EBP the parametric bootstrap method can be used. The bootstrap model used to generate the data is given by (Chatterjee, Lahiri and Li (2008), González-Manteiga et al. (2008)):

$$Q(\mathbf{Y}^*) = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\mathbf{v}^* + \mathbf{e}^*, \tag{3}$$

where $\mathbf{v}^* \sim \mathrm{N}(\mathbf{0}, \mathbf{G}(\hat{\boldsymbol{\delta}}))$, $\mathbf{e}_s^* \sim \mathrm{N}(\mathbf{0}, \mathbf{R}(\hat{\boldsymbol{\delta}}))$, $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\beta}}$ are estimators of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$, respectively. We use the restricted maximum likelihood method to estimate the parameters. The MSE estimator is given by (González-Manteiga et al. (2008)):

$$\hat{MSE}(\hat{\theta}_{EBP}) = B^{-1}\sum_{b=1}^{B}\left(\hat{\theta}_{EBP}(Q^{-1}(\mathbf{Y}_s^{*(b)})) - \theta(Q^{-1}(\mathbf{Y}^{*(b)}))\right)^2, \tag{4}$$

where $\hat{\theta}_{EBP}(Q^{-1}(\mathbf{Y}_s^{*(b)}))$ and $\theta(Q^{-1}(\mathbf{Y}^{*(b)}))$ are values of the predictor and the predicted characteristic, respectively, for the bth realization of the bootstrap model.

4. Data and model

Our considerations are based on whole population real economic longitudinal data available at the website of the United States Census Bureau (https://www.census.gov/library/ publications/ 2011/ compendia/ usa-counties-2011.html):

- the number of new private housing units of single-family houses authorized by building permits for years 2007-2009 (the variable of interest),
- the number of births for years 2006-2008 (the first auxiliary variable),
- the private nonfarm annual payroll in USD for years 2006-2008 (the second auxiliary variable)

for 177 counties from the following $D = 4$ states: Washington, Idaho, Oregon and California. We consider a relatively small population because of very time-consuming computations. Auxiliary variables are from the year preceding the construction of housing units. What is more, we assume that values of both auxiliary variables for 2009 are known and they are used to predict population and subpopulation characteristics of the variable of interest in 2010 (treated as the future period).

We mimic a real analysis. Because our further considerations are model-based and conditional (based on the given sample), we draw one sample. It is a stratified sample of counties, where states are strata, with proportional allocation (of size 20% of the population size) in the first period. Then, the same elements in periods 2 and 3 are in the sample, which gives a balanced panel sample. This gives us the division of the whole population dataset into the sample, where both the auxiliary information and the values of the variable of interest are available, and the non-sampled elements for which only auxiliary information is known.

A relatively large sample fraction is considered because: (i) the population size is small due to the complexity of computations and (ii) – at the same time – we must obtain enough sample observations for model parameters estimation purposes. Of course, this specific setting implies a limited generalization of our results for different datasets.

We consider the problem of prediction of the following population and subpopulations characteristics for the future period: means, medians, standard deviations, quartile deviations, moment and quartile skewness coefficients. For all of the variables the log transformation is used (after adding a constant), and hence the back transformation of the variable of interest is used to compute the EBP.

To find the best fitted linear mixed model we use the procedure presented by Verbeke and Molenberghs (2009) pp. 121-132, where firstly the fixed effects models are considered, then different random effects are added, to finally obtain the mixed model (in our case based on the AIC criterion). We have considered about 700 different models for both cases considered below.

The model we have chosen is given by (it will be called **model 1**):

$$Q(Y_{idt}) = \beta_1 + (\beta_2 + v_{1d})x_{1idt} + (\beta_3 x_{2idt} + v_{1i})\ln(t) + v_{2i} + v_{2d} + e_{1idt}, \tag{5}$$

where $Q(Y_{idt})$, x_{1idt}, x_{2idt} are log transformed variables (after adding a constant) $i = 1,2,...,N;$ $d = 1,2,...,D;$ $t = 1,2,...,M;$ $v_{1d} \sim (0, \sigma^2_{v1d})$, $v_{2d} \sim (0, \sigma^2_{v2d})$, $v_{1i} \sim (0, \sigma^2_{v1i})$, $v_{2i} \sim (0, \sigma^2_{v2i})$, $e_{1idt} \sim (0, \sigma^2_{e1})$, random effects and random components are mutually independent, in our case the population size $N = 177$, the number of time periods (including the future one) $M = 4$ and the number of subpopulations $D = 4$.

Additionally, we have chosen the best fitted nested error model with the logarithmic trend (as in (5)) and only one random effect for the purpose of the comparative study (it will be called **model 2**):

$$Q(Y_{idt}) = \beta_4 + \beta_5 x_{1idt} + (\beta_6 x_{1idt} + \beta_7)\ln(t) + v_{3i} + e_{2idt}, \tag{6}$$

where $v_{3i} \sim (0, \sigma^2_{v3i})$, $e_{2idt} \sim (0, \sigma^2_{e2})$, v_{3i} and e_{2idt} are mutually independent and other notations are as in (5). The choice of this class of models is due to the possibility of using the fast algorithm for EBP computation proposed in Molina and Rao (2010).

Based on permutation tests we can claim that parameters of both models are statistically significant. The normality assumption for both models is met for the considered longitudinal sample data (we have used Shapiro-Wilk test and residuals after the Cholesky transformation).

In the next sections we will consider EBPs under model 1 (given by (5)) denoted by EBP1 and under model 2 (given by (6)) denoted by EBP2, and their parametric bootstrap MSE estimators based on (4).

5. Number of iterations in EBP procedure

We consider stability of values of EBP1 and EBP2 computed under different numbers of iterations L (where $L = 100, 200, ..., 1000$) taken into account in the EBP iterative procedure presented in Section 3. Each boxplot in Figure 1 presents $M = 500$ values of EBP1 of one out of six population characteristics computed for different numbers of iterations L. For example, the first boxplot at the top left corner of Figure 1 presents 500 values of EBP1 used to predict the population mean, computed based on $L = 100$ iterations. In Figure 1, we see that results in each out of six considered cases tend to stabilize at around $L = 400$.

Similar plots are prepared for EBP2 and the prediction in the arbitrarily chosen third subpopulation (which gives 3 additional plots not presented in the paper). Then, based on values presented in each boxplot, we compute the value of the CV and present all of the results in Figure 2. The CV is given by:

$$CV_L = CV_{L,M}(\hat{\theta}^L_{EBP}) = \left(M^{-1} \sum_{i=1}^{M} \hat{\theta}^{L,i}_{EBP} \right)^{-1} \left(M^{-1} \sum_{i=1}^{M} \left(\hat{\theta}^{L,i}_{EBP} - M^{-1} \sum_{i=1}^{M} \hat{\theta}^{L,i}_{EBP} \right)^2 \right)^{0.5}, \quad (7)$$

where $CV_{L,M}(\hat{\theta}^L_{EBP})$ is the coefficient of variation computed based on M values of EBP and L is the number of iterations in the case of ith EBP estimation. For example, the star at the top left corner in Figure 2 is the value of the CV computed for values presented in the boxplot at the top left corner in Figure 1. Hence, in Figure 2 we can compare CVs of EBP1 and EBP2 of different population and subpopulation characteristics predicted for the future period. Coefficients of variation decrease from 5.74% ($L = 100$) for the standard deviation to 0.34% ($L = 1000$) for the median.

In six parts of Figure 2 we present the differences for different functions of random variables predicted for the future period. If we compare prediction methods (EBP1 and EBP2 of population characteristics; EBP1 and EBP2 of subpopulation characteristics), the results are similar – the differences are substantial only in the case of prediction of the standard deviation and the mean for small numbers of iterations. The differences between CVs for the third subpopulation and the whole population (EBP1 of population and subpopulation characteristics; EBP2 of population and subpopulation characteristics) are higher, especially for prediction of functions based on quantiles.

The results presented in Figure 2 are based on real sample data and they can be used to choose the number of iterations in practice assuming the maximum acceptable value of the CV (possibly different for different considered cases). For example, if – in the case of EBP computations – we accept values of the CV smaller or equal 3% in all of the considered cases, then $L = 400$ is sufficient. It can be noticed that for all of the considered prediction problems the linear growth in L causes a decrease in the CV

slower than the linear one. However, the exact path of the convergence is dependent on the predicted characteristic as well as on the choice of the population or the subpopulation (which is connected to the sample size). For the quantile measures the required number of iterations (for the certain CV goal) would be greater if the subpopulation was considered instead of the whole population. It can also be noticed that for certain characteristics (the standard deviation, the moment skewness coefficient, the quartile skewness coefficient) the absolute improvement of the results is more tangible, especially for the number of iterations around $L = 100$. Therefore, the incentive of enlarging the number of iterations would be dependent of the above aspects.

Alternatively, we can consider the assumed acceptable value of the change of the CV (see Figure 7 in Appendix), which is given by:

$$RCCV_L^{L+100} = 100 CV_L^{-1} (CV_{L+100} - CV_L), \tag{8}$$

where CV_L is given by (7). For example, if we accept the decrease (comparing L with $L-100$) of the CV smaller or equal 20%, then $L = 400$ is sufficient for all of the considered cases, too. It can be noticed that the relative change of CVs is a measure that, unlike the CV itself, behaves very similarly for all the considered characteristics. For example, the difference between $L = 100$ and $L = 200$ iterations causes the improvement around 30%. It can be also noticed that the relative improvements are independent on the choice of the population or the subpopulation. This means that the chosen measure (the CV or the relative change of CVs) may have an impact on the final conclusion. The difference in the observed convergence between EBP1 (based on the model with 4 random effects) and EBP2 (based on the model with 1 random effect) is negligible for all cases besides the standard deviation and the mean.

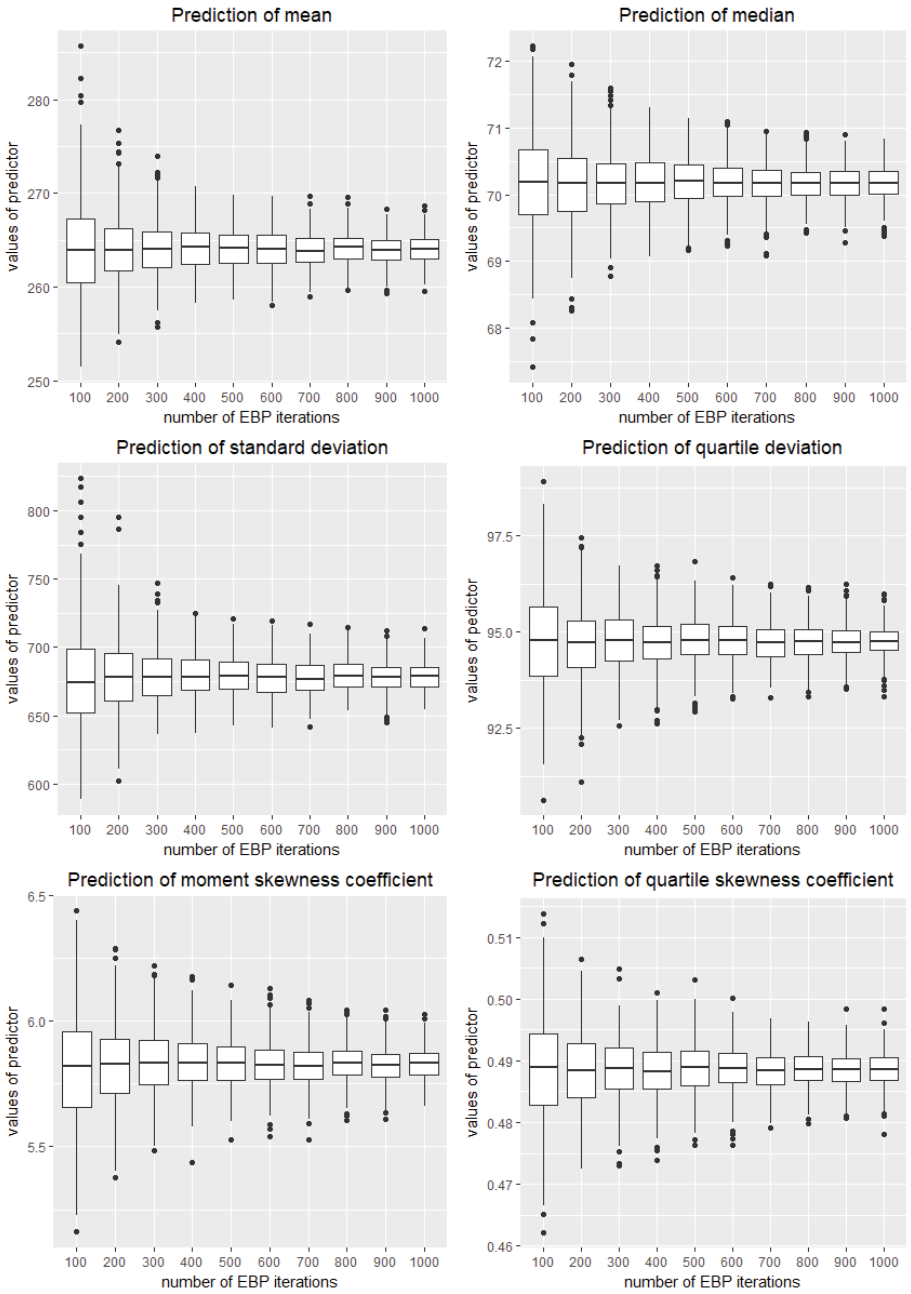

Figure 1. Variability of 500 values of EBP1 of different population characteristics in the future period computed for different numbers of iterations $L = 100, 200, ..., 1000$

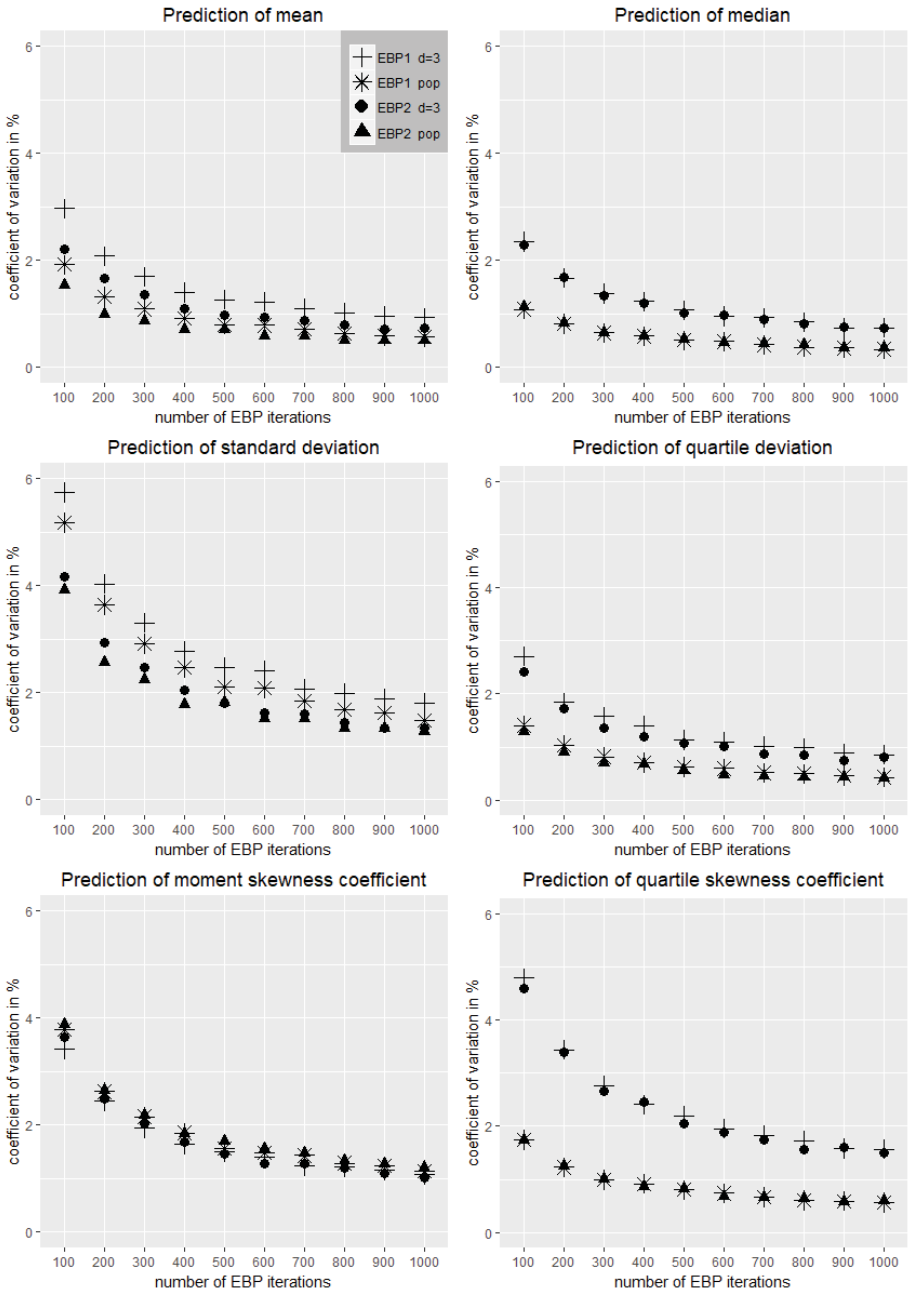

Figure 2. CVs computed based on 500 values of EBP1 and EBP2 of different population and subpopulation characteristics in the future period for different numbers of iterations $L = 100, 200, ..., 1000$

6. Number of iterations for parametric bootstrap MSE estimator

In all of the cases considered in this section EBPs are computed assuming $L = 500$ – higher than suggested in the previous section (i.e. $L = 400$) to obtain more stable results for MSE estimation. We consider MSE estimators of EBP1 under model 1 and EBP2 under model 2 computed for different numbers of iterations B (where $B = 100, 200, ..., 1000$) taken into account in the parametric bootstrap procedure presented in Section 3. Each boxplot in Figure 3 presents 100 values of MSE estimator of EBP1 computed for different numbers of iterations B for one out of six prediction problems. For example, the first boxplot at the top left corner presents 100 values of the MSE estimator of EBP1 of the population mean computed based on $B = 100$ iterations. The results presented in Figure 3 for $B = 100$ and $B = 200$ are generally unstable, at $B = 300$ they start to stabilize, for B from 500 to 1000 are quite similar. Similar figures are created for bootstrap MSE estimators of EBP2 and the third subpopulation (which gives three additional figures not presented in the paper).

Then, based on the values presented in each boxplot we compute the value of the CV and present all of the results in Figure 4. The coefficient of variation, similarly as in the case of (7), is given by:

$$CV_B = CV_{B,M}(\hat{MSE}^B(\hat{\theta}^L_{EBP})) =$$

$$= \left(M^{-1}\sum_{i=1}^{M} \hat{MSE}^{B,i}(\hat{\theta}^L_{EBP}) \right)^{-1} \left(M^{-1}\sum_{i=1}^{M} \left(\hat{MSE}^{B,i}(\hat{\theta}^L_{EBP}) - M^{-1}\sum_{i=1}^{M} \hat{MSE}^{B,i}(\hat{\theta}^L_{EBP}) \right)^2 \right)^{0.5}, \quad (9)$$

where $CV_{B,M}(\hat{MSE}^B(\hat{\theta}^L_{EBP}))$ is the coefficient of variation based on M values of the MSE estimator of EBP, L is the fixed number of EBP iterations, B is the number of bootstrap iterations in the case of ith MSE estimation.

For example, the star symbol at the top left corner in Figure 4 presents the value of the CV computed based on the values presented in the boxplot at the top left corner in Figure 3. Hence, we can compare CVs of the values of MSE estimators of EBP1 and EBP2 for six different prediction problems.

Values in Figure 4 decrease, although the decrease is not as smooth as in the case of the EBP (compare with Figure 2) – possibly due to the additional source of the variability resulting from the computation of EBP values and a smaller number of values per boxplot. Coefficients of variation decrease from 43.5% $(B = 100)$ for the standard deviation to 4.31% $(B = 1000)$ for the quantile skewness coefficient. The results for different models (MSEs estimators under model 1 and under model 2), the population and the third subpopulation and different prediction problems (except the prediction of the standard deviation in the future period) are similar, especially for larger numbers of iterations B.

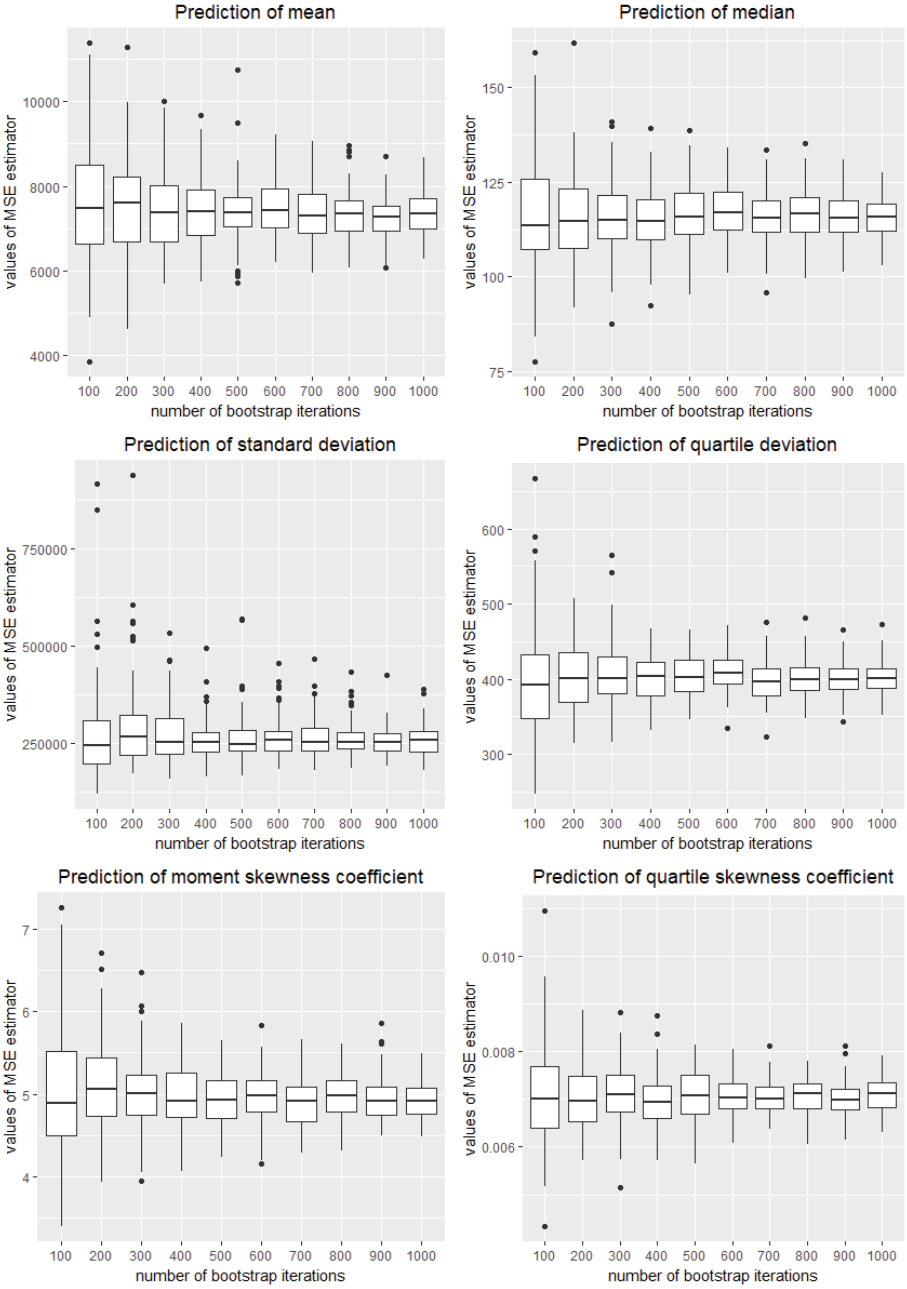

Figure 3. Variability of 100 values of parametric bootstrap MSE estimators of EBP1 of different population characteristics in the future period computed for different numbers of iterations $B = 100, 200, ..., 1000$

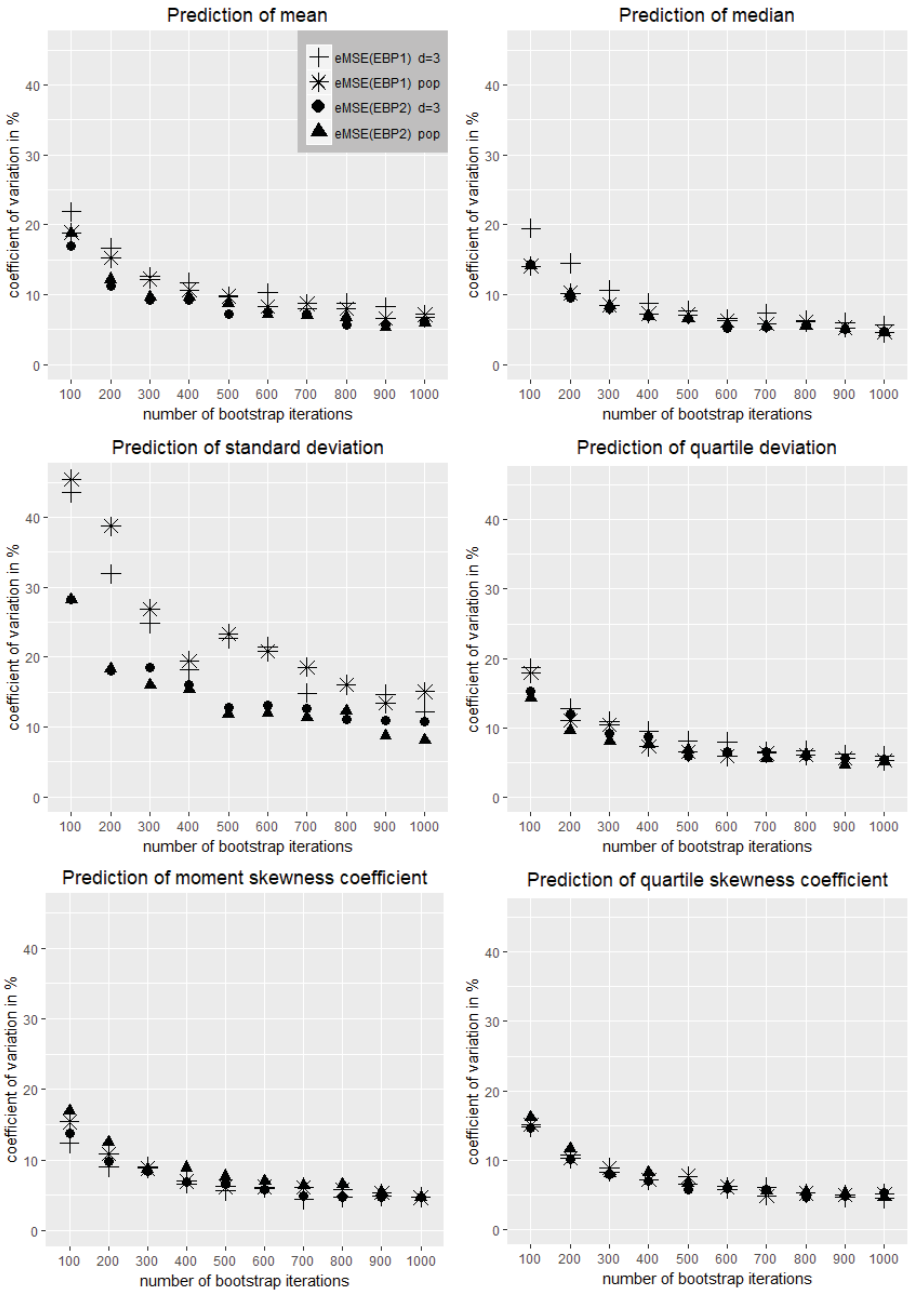

Figure 4. CVs computed based on 100 values of parametric bootstrap MSE estimators of EBP1 and EBP2 of different population and subpopulation characteristics in the future period for different numbers of iterations $B = 100, 200, ..., 1000$

Similarly to the previous section, the choice of the appropriate number of bootstrap iterations can be made based on the maximum acceptable value of the CV of values of MSE estimators. For example, if – in this case – we accept values of the CV smaller or equal 10%, then $B = 500$ is sufficient in most of the considered cases except the problem of prediction of the standard deviation (see Figure 4). Similarly to the results presented in the previous section, the linear growth in the number of bootstrap iterations causes a decrease in the CV slower than the linear one for all the considered prediction problems. The exact paths of convergence vary in dependence on the considered characteristic, the choice between the subpopulation and the population and the considered model (which can be noticed especially for the median).

Alternatively, we can consider the acceptable value of the change of the CV (see Figure 8 in Appendix), which is given by:

$$RCCV_B^{B+100} = 100 CV_B^{-1}(CV_{B+100} - CV_B), \tag{10}$$

where CV_B is given by (9). If we compare Figure 8 with Figure 7, we see that the changes are less stable because of the same reasons, as stated in the case of the comparison of Figure 4 with Figure 2 in the previous paragraph. If we accept the decrease (comparing B with $B-100$) of the CV smaller or equal 20%, then in most of the considered cases $B = 500$ is sufficient. The relative changes of CVs behave similarly for all the considered characteristics, although the results are quite unstable and difficult for more in-depth analysis. The difference in the convergence between EBP1 (based on the model with 4 random effects) and EBP2 (based on the model with 1 random effect) is negligible for all the cases beside standard deviation and mean, similarly as in the previous section.

7. Number of iterations in Monte Carlo simulation studies

Our considerations, proposals and conclusions in two previous sections were based on the real sample data. In this section we study the problem of model-based simulation studies of the properties of EBPs, where values of the variable of interest are generated based on model 1 (see (5)) for EBP1 and model 2 (see (6)) for EBP2. In simulation studies the appropriate number of Monte Carlo iterations is usually chosen based on the accepted value of the absolute simulation biases of unbiased statistics. For example, in design-based experiments Barbiero and Mecatti (2010) accept the relative values of modulus of simulation biases of the unbiased Horvitz-Thompson estimator and the unbiased estimator of its variance equal 1% and 3%, respectively. Similarly, in our case, we can assume the accepted relative value of modulus of simulation biases for the (unbiased) Best Predictors. But in the case of EBPs it is known that the ratio of MSEs of the EBP and the BP is greater than 1, while its simulation value may be lower than 1

even if the simulation bias of the unbiased BP is low. It means that the simulation ratio of MSEs of the EBP and BP may be of greater importance (as the measure of the quality of the Monte Carlo simulation study under the given number of iterations) than the value of the simulation bias of BP. Hence, we propose to check in simulation studies if (i) the simulation ratio of these MSEs is greater than 1 and (ii) to check the stability of values of these ratios. The value of the criterion is computed as:

$$K^{-1}\sum_{k=1}^{K}(\hat{\theta}_{EBP}^{k}-\theta^{k})^{2}\left(K^{-1}\sum_{k=1}^{K}(\hat{\theta}_{BP}^{k}-\theta^{k})^{2}\right)^{-1}, \tag{11}$$

where K is the number of iterations in the simulation study, $\hat{\theta}_{EBP}^{k}$, $\hat{\theta}_{BP}^{k}$ and θ^{k} are the values of the EBP, BP and the predicted characteristic, respectively, in the *k*th iteration of the simulation study.

All results in this section are computed for $L=500$. In Figures 5 and 6 we consider different numbers of iterations because model 1 is more complex than model 2. The results for the simpler model 2 presented in Figure 6 tend to stabilize at $K=5000$. In the case of a more complex model 1 (see Figure 5), results for $K=5000$ are unstable especially in the case of the prediction of the median and the quartile skewness coefficient. To explain these results we should take into account two issues. Firstly, we assume that the number of EBP iterations $L=500$ is acceptable as shown in Section 5. Secondly, the results presented in Figure 5 are obtained based on one simulation study for an assumed number of Monte Carlo iterations K, which can but does not have to show possible instability of one specific result. Hence, the observed peaks for these two cases should be interpreted as a result of too small number of Monte Carlo iterations leading to possible instability of the results. The results for the more complex model (Figure 6) tend to stabilize at $K=15000$ iterations. What is more, in many cases MSEs ratios for the third subpopulation are close to 1, which is an argument for higher K if it is possible. The paths of improvement of the results are different for different prediction problems, however the generalization of the results is quite difficult due to the single execution of the simulation for each K. Predictors of some characteristics (i.e. the standard and the quartile deviations) tend to behave more stable than others, which may indicate a different strategy of the optimal choice of K for the specific simulation conditions like the considered characteristics. The complexity of the model has a significant impact on the simulation stability, which is opposite to the results presented in the two previous sections.

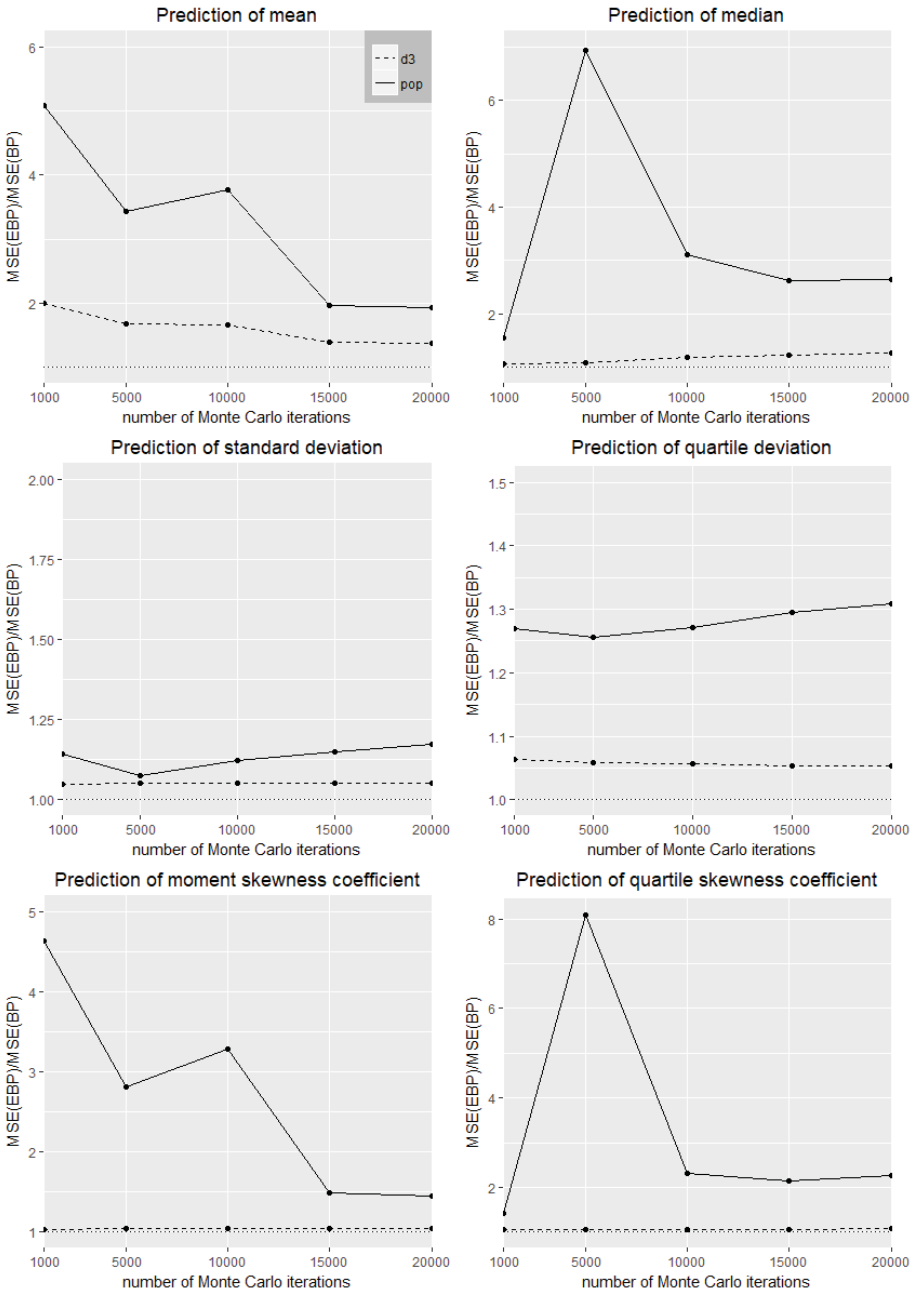

Figure 5. Ratios of MSE(EBP1) and MSE(BP1) of different population and subpopulation characteristics in the future period computed for different numbers of Monte Carlo iterations under model 1

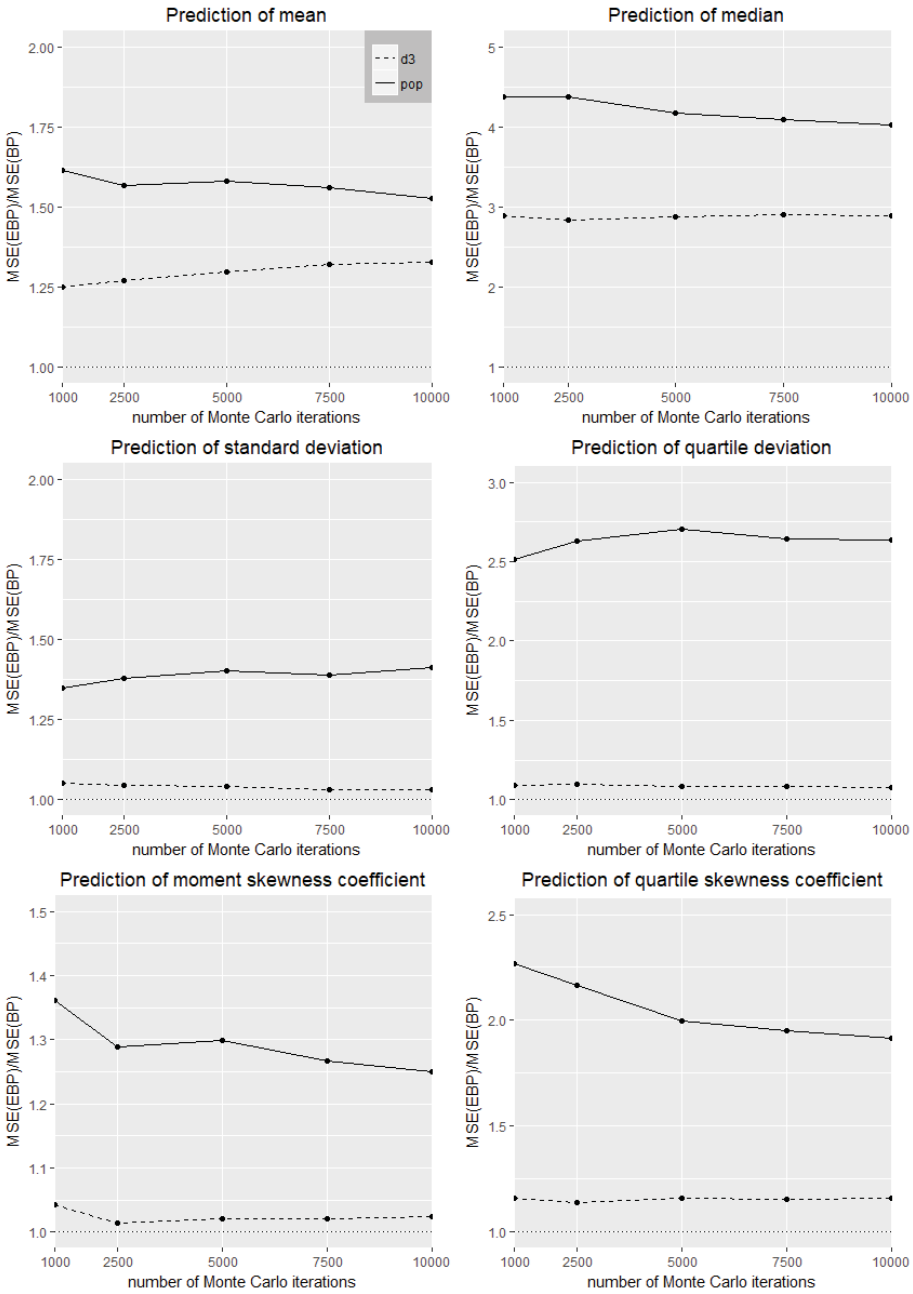

Figure 6. Ratios of MSE(EBP2) and MSE(BP2) of different population and subpopulation characteristics in the future period computed for different numbers of Monte Carlo iterations under model 2

8. Discussion

In this section we present possible generalizations of the proposed criteria, some alternatives and limitations of our procedure.

The CV, as well as the relative change of CVs, can be replaced by a more robust measure, e.g. based on quantiles (like the interquartile range) and the relative change of the chosen measure, respectively. It may be helpful especially in the case where two iterative algorithms are used at the same time as in the case of estimation of the MSE based on B bootstrap replications of the EBP approximated in L iterations (see Section 6), where the simulation variability resulting from the first procedure influences the results of the second procedure. What is more, in the case of consideration of highly volatile characteristics such as the standard deviation, more robust measure may be applied in practice, however in most cases the CV should be sufficient. The adequate measure can be determined by the researcher after the study of some boxplots presented in Figures 1 and 3.

The stopping role assumed in this paper to be the absolute or the relative difference of the appropriate measure of the simulation variability can also be changed. For example, we can assume that the required number of iterations is obtained (that the procedure should be stopped) if two distributions, represented by two adjacent boxplots in Figure 1 or in Figure 3, are the same, which is verified by the appropriate nonparametric test.

The drawback of our procedure results from the necessity of conducting the computations several times per one iteration number to obtain data represented by one boxplot. The alternative, to be developed and studied in further research, could be based on the idea of statistical quality control (e.g. control charts) where only one value is computed per one iteration number. In the statistical quality control it is checked when the monitored process becomes "not in control". In our case, we will have to check, based on the appropriate criteria, when the process becomes "in control" (becomes stable). Although in this approach the number of computations per one iteration will be one, we will have to increase the number of steps and replace, e.g. $L = 100, 200, 300, \ldots$ by $L = 10, 20, 30, \ldots$ but even though the total number of iterations will be smaller.

The methods considered in the paper are in practice highly dependent on the available time, overall complexity of the simulations and the available hardware. Furthermore, the sufficient improvement of the measures is a subjective case that is heavily dependent on the origin of the data (i.e. in the case of some medical simulations even small improvements can be very important). For the considered dataset the convergence of the CV computed for the EBP as well as the MSE estimator may vary, which provides additional difficulties in terms of generalization of the results.

9. Conclusion

We consider the problem of the stability of results in iterative procedures used for the computation of the empirical best predictor and its parametric bootstrap MSE estimator. We show the dependence between the number of iterations and the stability of iteratively obtained values of two predictors based on a simple and more complex model in the case of prediction of different future population and subpopulation characteristics. In the case of the EBP iterative algorithm and the parametric bootstrap procedure used to estimate the MSE we propose two methods of choosing the appropriate number of iterations. The first one is based on the maximum acceptable value of the CV of the results obtained several times for a given number of iterations. The second one is the stability criterion assessed based on the minimum relative decrease in the CV. In the case of Monte Carlo simulation studies we suggest two criteria based on the ratio of MSEs of the EBP and the BP. The number of Monte Carlo iterations should be controlled to obtain simulation ratios of the MSEs: stable and greater than one. All of the considerations are supported by real longitudinal economic data available at the United States Census Bureau website.

REFERENCES

ANDREWS, D. W. K., BUCHINSKY, M., (1997). On the number of bootstrap repetitions for bootstrap standard errors, confidence intervals, and tests, *Cowles Foundation Discussion Paper No. 1141R*, pp. 1–51.

ANDREWS, D. W. K., BUCHINSKY, M., (2000). A three-step method for choosing the number of bootstrap repetitions, *Econometrica*, Vol. 67, pp. 23–51.

ANDREWS, D. W. K., BUCHINSKY, M., (2001). Evaluation of a three-step method for choosing the number of bootstrap repetitions, *Journal of Econometrics,* Vol. 103, pp. 345–386.

BARBIERO, A., MECATTI, F., (2010). *Bootstrap algorithms for variance estimation in πPS* sampling, In: Mantovan, P., Secchi, P. (Eds.), Complex Data Modeling and Computationally Intensive Statistical Methods. Contributions to Statistics, Springer, Milano, pp. 57–69.

BERAN, R., (1997). Diagnosing Bootstrap Success, *Annals of the Institute of Statistical Mathematics*, Vol. 49, pp. 1–24.

BERG, E., CHANDRA, H., (2014). Small area prediction for a unit-level lognormal model, *Computational Statistics and Data Analysis*, Vol. 78, pp.159–175.

BOUBETA, M., LOMBARDÍA, M. J, MORALES, D., (2016). Empirical best prediction under area-level Poisson mixed models, *Test,* Vol. 25, pp. 548–569.

BOUBETA, M., LOMBARDÍA, M. J, MORALES, D., (2017). Poisson mixed models for studying the poverty in small areas, *Computational Statistics and Data Analysis*, Vol. 107, pp. 32–47.

BUTAR, B. F., LAHIRI, P., (2003). On measures of uncertainty of empirical Bayes small-area estimators, *Journal of Statistical Planning and Inference*, Vol. 112, pp. 63–76.

CHATTERJEE, S., LAHIRI, P. LI, H., (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models, *Annals of Statistics*, Vol. 36 (3), pp. 1221–1245.

DAS, S., HASLETT, S., (2019). A comparison of methods for poverty estimation in developing countries, *International Statistical Review*, DOI: 10.1111/insr.12314, available online: https://onlinelibrary.wiley.com/doi/pdf/10.1111/insr.12314.

DIALLO, M. S., (2014). *Small area estimation under skew-normal nested error models,* PhD diss., Carleton University.

DIALLO, M. S., RAO, J. N. K., (2018). Small area estimation of complex parameters under unit-level models with skew-normal errors, *Scandinavian Journal of Statistics*, Vol. 2018, pp.1–25.

DAVISON, A. C., HINKLEY D. V., (1997). *Bootstrap Methods and their Application*, Cambridge University Press.

EFFRON, B., TIBSHIRANI, R., (1986), Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science*, Vol. 1(1), pp. 54–75.

ELBERS, CH., VAD DER WEIDE, R., (2014). Estimation of normal mixtures in a nested error model with an application to small area estimation of poverty and inequality. *World Bank Group. Policy Research Working Paper* 6962, pp. 1–31.

GONZÁLEZ-MANTEIGA W., LOMBARDÍA, M. J., MOLINA, I., MORALES, D., SANTAMARÍA, L., (2008). Bootstrap mean squared error of small-area EBLUP, *Journal of Statistical Computation and Simulation*, Vol. 78(5), pp. 443–462.

GUADARRAMA, M., MOLINA, I., RAO, J. N. K., (2018). Small area estimation of general parameters under complex sampling designs, *Computational Statistics and Data Analysis,* Vol.121, pp. 20–40.

HALL, P., MAITI, T., (2006). On Parametric Bootstrap Methods for Small Area Prediction, *Journal of the Royal Statistical Society. Series B*, Vol. 68(2), pp. 221–238.

HALL, P., MARTIN, M. A., (1988). Exact convergence rate of bootstrap quantile variance estimator, *Probability Theory and Related Fields*, Vol. 80, pp. 261–268.

HOBZA, T., MORALES, D., (2016). Empirical best prediction under unit-level logit mixed models, *Journal of Official Statistics,* Vol. 32(3), pp. 661–692.

JIANG, J., (1996). REML estimation: asymptotic behavior and related topics, *Annals of Statistics*, Vol. 24 (1), pp. 255–286.

JIANG, J., (2003). Empirical best prediction for small-area inference based on generalized linear mixed models, *Journal of Statistical Planning and Inference,* Vol. 111, pp. 117–127.

JIANG, J., (2007). *Linear and Generalized Linear Mixed Models and Their Appliactions,* Springer, New York.

JIANG, J., LAHIRI, P., (2001). Empirical best prediction for small area inference with binary data, *Annals of the Institute of Statistical Mathematics*, Vol. 53(2), pp. 217–243.

JIANG, J., LAHIRI, P., (2006). Estimation of Finite Population Domain Means, *Journal of the American Statistical Association*, Vol. 101(473), pp. 301–311.

MARINO, M. F., RANALLI, M. G., SALVATI, N., ALFÒ, M., (2019). Semi-Parametric Empirical Best Prediction for small area estimation of unemployment indicators, *The Annals of Applied Statistics*, Vol. 13(2), pp. 1166–1197.

MOLINA, I., MARTIN, N., (2018). EBP under a nested error model with log transformation, *Annals of Statistics*, Vol. 46(5), pp. 1961–1993.

MOLINA, I., RAO, J. N. K., (2010). Small area estimation of poverty indicators, *The Canadian Journal of Statistics*, Vol. 38(3), pp. 369–385.

SINGH, K., (1981). On the Asymptotic Accuracy of Efron's Bootstrap, *The Annals of Statistics*, Vol. 9(6), pp. 1187–1195.

TZAVIDIS, N., ZHANG, L.-C., LUNA, A., SCHMID, T., ROJAS-PERILLA, N., (2018). From start to finish: a framework for the production of small area official statistics, *Journal of the Royal Statistical Society A*, Vol. 181(4), pp. 927–979.

VERBEKE, G., MOLENBERGHS, G., (2009). *Linear mixed models for longitudinal data,* Springer-Verlag, New York.

ZIMMERMANN, T., MÜNNICH, R., (2018). Small area estimation with a lognormal mixed model under informative sampling, *Journal of Official Statistics*, Vol. 34(2), pp. 523–542.

APPENDIX

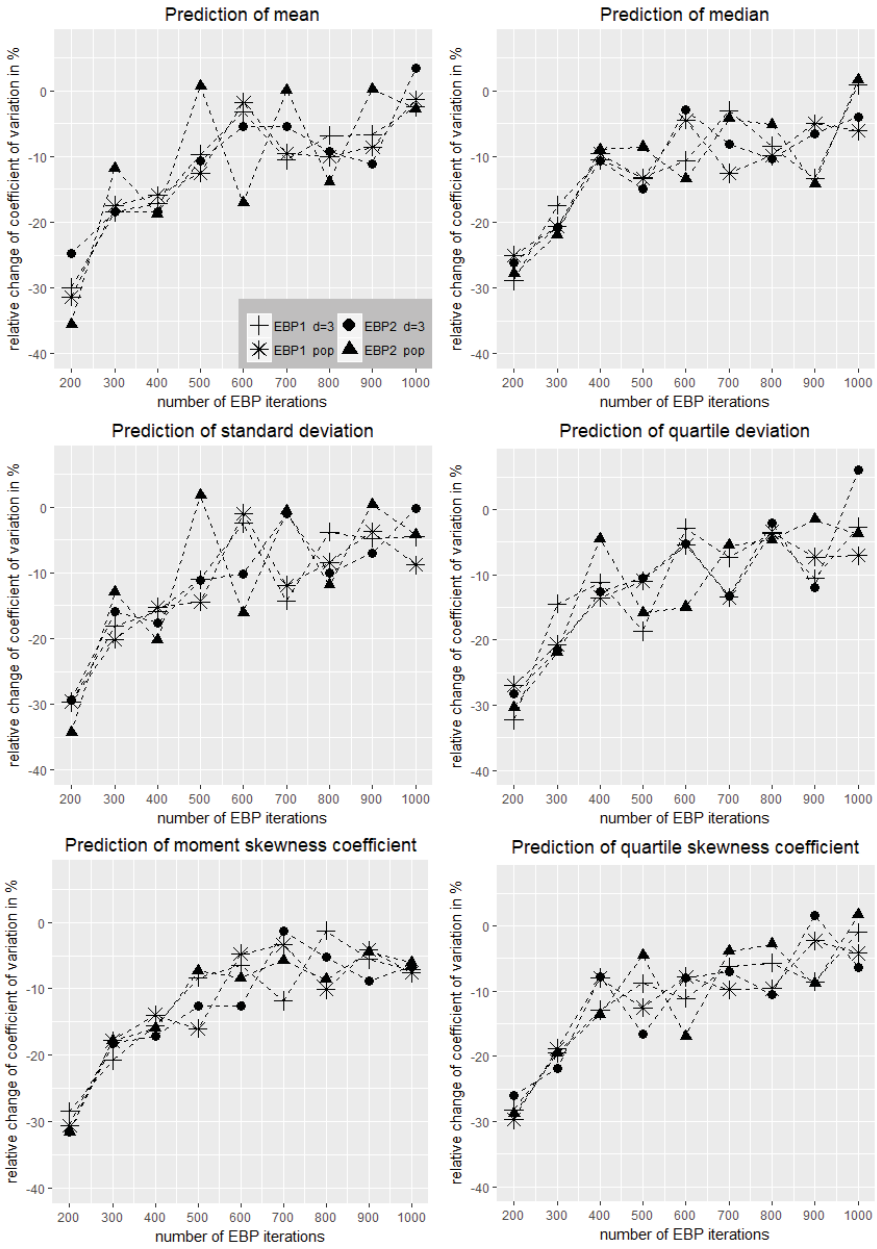

Figure 7. Relative changes of CVs of 500 values of EBP1 and EBP2 of different population and subpopulation characteristics computed as $100(CV_L - CV_{L-100}) / CV_{L-100}$ for $L = 200, 300, ..., 1000$

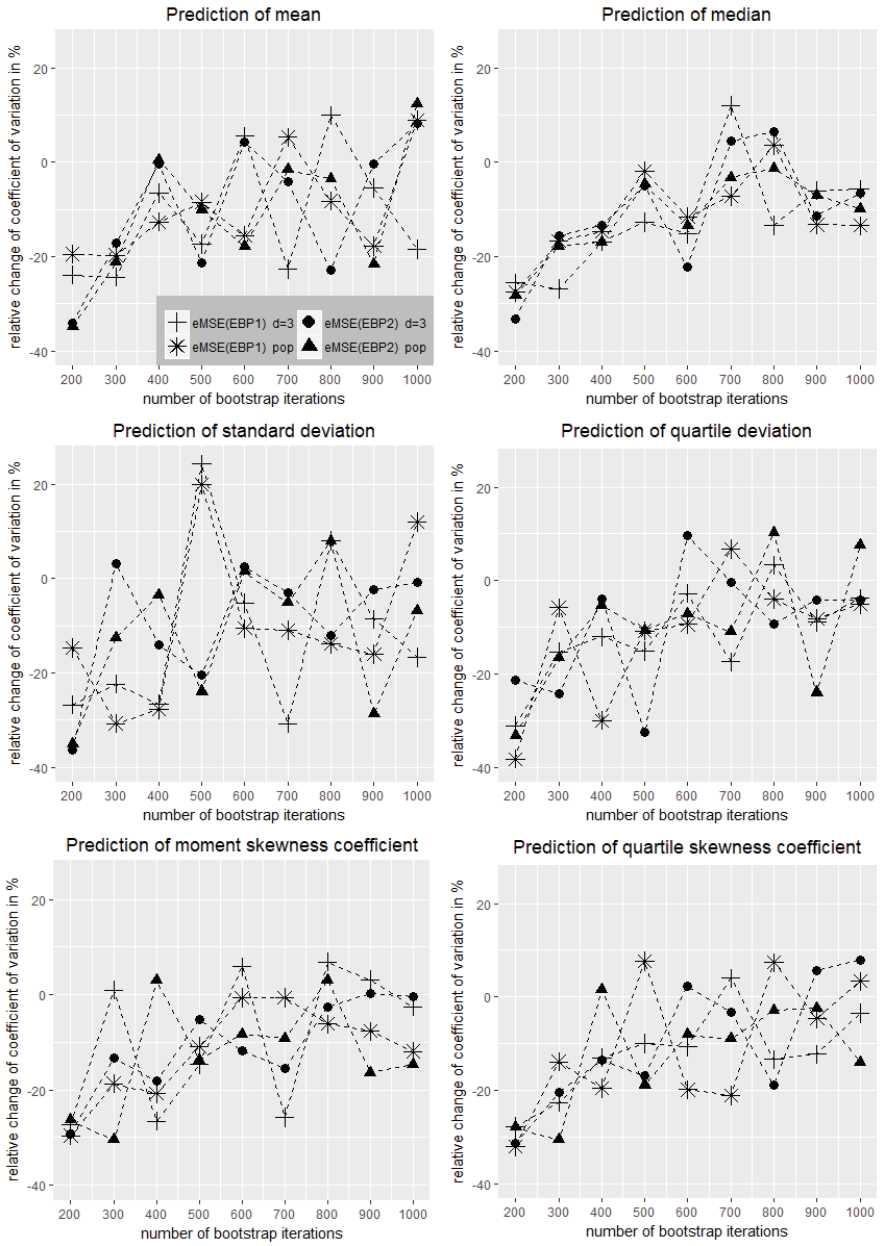

Figure 8. Relative changes of CVs of 100 values of parametric bootstrap MSE estimates of EBP1 and EBP2 of different population and subpopulation characteristics computed as $100(CV_B - CV_{B-100}) / CV_{B-100}$ for $B = 200, 300, ..., 1000$