Estimation of the inequality indices based on the well-known rank-based sampling schemes

Najmeh Nakhaei Rad [a,d,e], Mahdi Salehi [b,e], and Yaser Mehrali[c]

[a]Department of Mathematics and Statistic, Mashhad Branch, Islamic Azad University, Mashhad, Iran;
[b]Department of Mathematics and Statistics, University of Neyshabur, Neyshabur, Iran;
[c]Department of Statistics, Faculty of Science, University of Isfahan, Isfahan, Iran;
[d]DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa;
[e]Department of Statistics, University of Pretoria, Pretoria, South Africa

* CONTACT: Najmeh Nakhaei Rad. Department of Mathematics and Statistic, MashhadBranch, Islamic Azad University, Mashhad, Iran. Email: najme.rad@gmail.com

Abstract

Using the simple random sampling (SRS) for collecting the income data may results poor estimators specially when the sample size is not enough large. Since under this circumstance, it may be difficult to obtain a representative subset from the income population based on SRS. Ranked set sampling (RSS) and its simplified versions overcome to this shortcoming. These sampling schemes work based on judgment ranking of the sample units. Moreover, the judgment post-stratification sampling (JPS) is also another rank-based sampling plan that can be considered as a competitor of RSS. This paper is organized in order to find the most appropriate sampling scheme among the SRS, RSS, JPS and some more, for estimating of some well-known inequality indices. Comparison of the estimators is carried out through a simulation study based on both perfect and imperfect ranking mechanisms. Results show that the suggested scheme is different for each inequality index. Finally, a real data set is analyzed.

Keywords: Atkinson index; Gini index; Judgment post stratification sampling; Median ranked set sampling; MLD index; Modified rank set sampling; Monte Carlo simulation; Ranked set sampling; Theil index

1. Introduction

Ranked set sampling (RSS) was first introduced by McIntyre (1952) as a competitor of simple random sampling (SRS), the most common tool in the statistical methods. Since, when the sample size is not enough large, it may be difficult to obtain a representative subset from the population based on SRS. He actually proposed a new efficient and cost saving sampling method to estimate the unknown population mean which was based on ranked sets. After him, Halls and Dell (1966) used RSS for estimating the forage yields. They named this method as ranked set sampling (RSS). To choose a ranked set sample, Takahashi (1970) applied a different approach from McIntyre's one. The median ranked set sampling (MedRSS) and the modified ranked set sampling (MRSS) as two simplified versions of RSS were respectively defined by Muttlak (1997) and Stokes (1980b). There are also many studies about comparing RSS with SRS. Among of them, we refer to Takahasi and Wakimoto (1968), Dell and Clutter (1972), Stokes (1977, 1980a) and Stokes and Sager

(1988). Dey, Salehi, and Ahmadi (2017) compared the estimation of the parameter of Rayleigh distribution based on SRS, RSS, MedRSS and MRSS. MacEachern, Stasny, and Wolfe (2004) introduced another tank-based sampling scheme called the judgment post-stratification sampling (JPS), as an alternative to the RSS. Dastbaravarde et al. (2016) obtained the non-parametric estimation of $E(g(X))$ using JPS and compared it with SRS and RSS.

In economics, SRS is commonly used to estimate the inequality indices. However, recently, some authors have investigated the RSS plan in economics area specially for estimating the inequality indexes. Al-Talib and Al-Nasser (2008), compared the estimate of Gini index from continuous distribution based on RSS and SRS. Bansal, Arora, and Mahajan (2013) compared the estimates of Gini index, Bonferroni index and Absolute Lorenz index based on RSS with SRS and systematic sampling in Parametric case. Nakhaei Rad, Mohtashami Borzadaran, and Yari (2016) compared the estimates of Gini index, Theil index, MLD index and Atkinson index based on RSS and SRS in the non-parametric case.

This paper tries to find an answer to the question-that is 'among the well-known rank-based sampling methods RSS, MedRSS, MRSS and JPS, which one outperforms the others in estimating the inequality indexes?'. Thus, the rest of the paper is set as follows. In Sec. 2, the non-parametric estimates of inequality indices including Gini index, Theil index, MLD index and Atkinson index based on the mentioned sampling schemes are derived. The Gini index, MLD index and Theil index are respectively as:

$$G = \frac{1}{2\mu}\int_0^\infty \int_0^\infty \left|x - y\right| f(x) f(y)\, dx dy$$

(1.1)

$$T = \int_0^\infty \frac{x}{\mu} \log \frac{x}{\mu} dF(x)$$

(1.2)

$$MLD = -\int_0^\infty \log \frac{x}{\mu} dF(x),$$

(1.3)

where *X* and *Y* are two independent random variables come from the income distribution *f*(*x*) and $\mu = E(X)$. The Atkinson family, is defined as below

$$A(\varepsilon) = 1 - \left(\int_0^\infty \left(\frac{x}{\mu}\right)^{1-\varepsilon} dF(x)\right)^{\frac{1}{(1-\varepsilon)}}, \varepsilon > 0, \varepsilon \neq 1,$$

(1.4)

where **ε** controls the inequality aversion. For the case $\varepsilon = 1$, the Atkinson index is defined as

$$A(1) = 1 - \frac{1}{\mu} \exp\left(\int_0^\infty \log (x) dF(x)\right).$$

(1.5)

Sec. 3 is dedicated to obtain the value of these indices for generalized beta distribution of second kind (GB_2). The GB_2 is the most famous income distribution in economics which includes several other income distributions as special or limiting cases. Sec. 4 provides a simulation study in order to make a comparison between the estimators obtained based on the various schemes (gathered from both perfect and imperfect ranking procedures) via the mean squared error (MSE) and the bias criteria. In Sec. 5, a real data set is analyzed. Finally, and Sec. 6 concludes.

2. Estimations of inequality indices

In this section, the non-parametric estimation of the mentioned inequality indices are obtained based on SRS, RSS, MedRSS, MRSS and JPS, respectively.

2.1. Estimation based on the SRS

Suppose that $\mathbf{X}_{SRS} = (X_1, \ldots, X_n)$, is a simple random sample of size *n* drawn from a continuous population with the cumulative distribution function *F*(*x*) and the probability density function *f*(*x*). The estimations of indices (1.1)–(1.5) based on SRS are obtained as follow

$$\widehat{G}_{SRS} = \frac{1}{2n^2\overline{X}_{SRS}} \sum_{j=1}^{n}\sum_{i=1}^{n}\left|X_i - X_j\right|$$

(2.1)

$$\widehat{T}_{SRS} = \frac{1}{n\overline{X}_{SRS}} \sum_{i=1}^{n} X_i \ \log\ X_i - \ \log\ \overline{X}_{SRS},$$

(2.2)

$$\widehat{MLD}_{SRS} = -\frac{1}{n}\sum_{i=1}^{n} \log\ X_i + \ \log\ \overline{X}_{SRS},$$

(2.3)

$$\widehat{A(\varepsilon)}_{SRS} = 1 - \frac{1}{\overline{X}_{SRS}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}},$$

(2.4)

$$\widehat{A(1)}_{SRS} = 1 - \frac{1}{\overline{X}_{SRS}}\exp\left(\frac{1}{n}\sum_{i=1}^{n} \log\ X_i\right),$$

(2.5)

respectively, where $\overline{X}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} X_i$. It is to be noted that (2.1) can be also written as (Rizzo 2002)

$$\widehat{G}_{SRS} = \frac{2\sum_{i=1}^{n} iY_i}{n^2\overline{X}_{SRS}} - \frac{n+1}{n},$$

(2.6)

where Y_i denotes the *i*th ordered observation.

2.2. Estimation based on the RSS

To obtain a ranked set sample of size $n = r * m$, the number of *m* simple random samples of size *m* should be chosen and ordered by visual inspection. Then, the smallest observation from the first sample, the second smallest observation from the second sample are selected and this procedure is continued until the largest observation of the *m*th sample is selected as follows:

$$\begin{array}{ccccccc} 1: & \underline{X_{1:m}^{(1)}} & X_{2:m}^{(1)} & \ldots & X_{m:m}^{(1)} & \rightarrow & X_{(1,1)} = X_{1:m}^{(1)} \\ 2: & X_{1:m}^{(2)} & \underline{X_{2:m}^{(2)}} & \ldots & X_{m:m}^{(2)} & \rightarrow & X_{(2,2)} = X_{2:m}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ m: & X_{1:m}^{(m)} & X_{2:m}^{(m)} & \ldots & \underline{X_{m:m}^{(m)}} & \rightarrow & X_{(m,m)} = X_{m:m}^{(m)} \end{array}$$

(2.7)

By repeating this procedure *r* times, an *r*-cycle ranked set sample with the size *n* = *rm* will be derived. Let us denote the resulting units by $\mathbf{X}_{RSS} = \left\{X_{(i,i)k},\ i = 1,\ldots,m,\ k = 1,\ldots,r\right\}$. Here, the subscript *k* stands for the cycle number. The estimations of indices (1.1)–(1.5) based on RSS are obtained as follow

$$\widehat{G}_{RSS} = \frac{1}{2n^2\overline{X}_{RSS}}\sum_{k,l=1}^{r}\sum_{i,j=1}^{m}\left|X_{(i,i)k} - X_{(j,j)l}\right|$$

(2.8)

$$\widehat{T}_{RSS} = \frac{1}{n\overline{X}_{RSS}}\sum_{k=1}^{r}\sum_{i=1}^{m} X_{(i,i)k}\ \log\ X_{(i,i)k} - \log\ \overline{X}_{RSS},$$

(2.9)

$$\widehat{MLD}_{RSS} = -\frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m}\log\ X_{(i,i)k} + \log\ \overline{X}_{RSS},$$

(2.10)

$$\widehat{A(\varepsilon)}_{RSS} = 1 - \frac{1}{\overline{X}_{RSS}}\left(\frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m} X_{(i,i)k}^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}},$$

(2.11)

$$\widehat{A(1)}_{RSS} = 1 - \frac{1}{\overline{X}_{RSS}} \exp\left(\frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m} \log\, X_{(i,i)k}\right),$$

(2.12)

respectively, where $\overline{X}_{RSS} = \frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m} X_{(i,i)k}$.

2.3. Estimation based on the MRSS

To obtain a modified ranked set sample of size *n* = *rm*, the number of *m* simple random samples of size *m* should be selected and ranked by judgment or a visual inspection. Then, the smallest observation from the all samples is taken. By repeating this procedure, *r* times, $\mathbf{X}_{MRSS} = \{X_{(1,i)k},\ i = 1, \ldots, m,\ k = 1, \ldots, r\}$, a modified ranked set sample of size $n = r * m$, is obtained. The estimations of indices (1.1)–(1.5) based on MRSS are derived by using $(X_{(1,i)k}, \overline{X}_{MRSS})$ rather than $(X_{(i,i)k}, \overline{X}_{RSS})$ in (2.8)–(2.12), where $\overline{X}_{MRSS} = \frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m} X_{(1,i)k}$. Note that in MRSS, the largest observation from the all simple random samples also can be chosen. Since the income distributions are mostly skewed to the right, here, only the smallest observations are considered.

2.4. Estimation based on the MedRSS

The following algorithm may be applied in order to collect an *r*-cycle MedRSS with set size *m*: Firs *m* SRS's of size *m* are drawn. When *m* is an odd number, then the sample is selected by measuring the judgment median of each SRS's. For the even set sizes, suppose that we obtain $\frac{m}{2} + \frac{m}{2}$ SRS's of size *m*. Then, the largest judgment ranked units are measured from the first half of the SRS's, while the smallest jadgment ranked units are measured from the second half of the SRS's. By replicating the above method for *r* times, the desired sample size *n* = *mr* will be accessible. More specifically, if $\mathbf{X}_{MedRSS}$ stands for the mentioned sample, then, it will be of the form

$$\left\{X_{\left(\frac{m+1}{2},i\right)j},\ i = 1, \ldots, m,\ j = 1, \ldots, r\right\},$$

when *m* is odd, and

$$\left\{X_{\left(\frac{m}{2},i\right)j},\ i = 1, \ldots, \frac{m}{2},\ j = 1, \ldots, r\right\} \cup \left\{X_{\left(\frac{m}{2}+1,i\right)j},\ i = \frac{m}{2} + 1, \ldots, m,\ j = 1, \ldots, r\right\},$$

when *m* is even. Therefore, the estimators of the indices (1.1)–(1.5) based on MedRSS are obtained by substituting $X_{(i,i)k}$ by $X_{(c,i)k}$ and $\overline{X}_{RSS}$ by $\overline{X}_{MedRSS}$ in (2.8)–(2.12), where $\overline{X}_{MedRSS} = \frac{1}{n}\sum_{k=1}^{r}\sum_{i=1}^{m} X_{(c,i)k}$.

2.5. Estimation based on the JPS

To select a sample of size n under JPS, a simple random sample size n, $X_1, \ldots, X_n$ should be chosen and measured. For each X_i, $i = 1, \ldots, n$, an auxiliary SRS of size $H - 1$, namely $X_{i2}, \ldots, X_{iH}$, is taken from the same population. Note that the units in the auxiliary samples are not measured. Then each sample $X_i, X_{i2}, \ldots, X_{iH}, i = 1, \ldots, n$, is ordered by judgment. Suppose that R_i is the rank of X_i in the ith sample which is determined by judgment. Then, the JPS with the sample size n is obtained as $\mathbf{X}_{JPS} = \{(X_i, R_i), i = 1, \ldots, n\}$. This data set is called full rank if for each $t = 1, \ldots, H$, there exists at least one R_i which is equal to t. Note that R_i's have discrete uniform distributions on the set of ranks $\{1, 2, \ldots, H\}$. Under these kind of observations, the estimators of indices (1.1)–(1.5) are derived as follows

$$\widehat{G}_{JPS} = \frac{1}{2\overline{X}_{JPS} h'_n} \sum_{l=1}^{H} \sum_{t=1}^{H} \frac{1}{N_{tl}} \sum_{j=1}^{n} \sum_{i=1}^{n} \left| X_i - X_j \right| I_{it} I_{jl},$$

(2.13)

$$\widehat{T}_{JPS} = \frac{1}{\overline{X}_{JPS} h_n} \sum_{t=1}^{H} \frac{1}{N_t} \sum_{i=1}^{n} (X_i \log X_i) I_{it} - \log \overline{X}_{JPS},$$

(2.14)

$$\widehat{MLD}_{JPS} = -\frac{1}{h_n} \sum_{t=1}^{H} \frac{1}{N_t} \sum_{i=1}^{n} (\log X_i) I_{it} + \log \overline{X}_{JPS},$$

(2.15)

$$\widehat{A(\varepsilon)}_{JPS} = 1 - \frac{1}{\overline{X}_{JPS}} \left(\frac{1}{h_n} \sum_{t=1}^{H} \frac{1}{N_t} \sum_{i=1}^{n} X_i^{1-\varepsilon} I_{it} \right)^{\frac{1}{1-\varepsilon}},$$

(2.16)

$$\widehat{A(1)}_{JPS} = 1 - \frac{1}{\overline{X}_{JPS}} \exp \left(\frac{1}{h_n} \sum_{t=1}^{H} \frac{1}{N_t} \sum_{i=1}^{n} (\log X_i) I_{it} \right),$$

(2.17)

respectively, where $\overline{X}_{JPS} = \frac{1}{h_n} \sum_{t=1}^{H} \frac{1}{N_t} \sum_{i=1}^{n} X_i I_{it}$.

$I_{it} = 1$ if $R_i = t$ for $t = 1, \ldots, H$ otherwise, $I_{it} = 0$.

$N_t = \sum_{i=1}^{n} I_{it}$.

$N_{tl} = \sum_{j=1}^{n} \sum_{i=1}^{n} I_{it} I_{jl} = \sum_{i=1}^{n} I_{it} \sum_{i=1}^{n} I_{il} = N_t N_l$.

$h_n = \sum_{t=1}^{H} I_t$, where $I_t = 1$ if $N_t > 0$ (i.e. there is at least one measured observation in the rth post-stratum) otherwise, $I_t = 0$, for $t = 1, \ldots, H$.

$h'_n = \sum_{l=1}^{H} \sum_{t=1}^{H} I_{tl} = \sum_{l=1}^{H} \sum_{t=1}^{H} I_t I_l = \sum_{t=1}^{H} I_t \sum_{l=1}^{H} I_l = h_n^2$, where $I_{tl} = 1$ if $N_{tl} > 0\,(N_t N_l > 0)$ otherwise, $I_{tl} = 0\,(I_{tl} = I_t I_l)$.

3. Inequality indices for the GB_2 distribution

To compute the MSE's and the biases of the estimators of inequality indices under different sampling schemes proposed in Sec. 2, it is necessary to have the true values of these indices when the income distribution is GB_2. The random variable X is said to have $GB_2(a,b,p,q)$ distribution, if its cumulative distribution function is

$$F_{GB_2}(x;a,b,p,q)=1-\frac{B_{\left(1+\left(\frac{x}{b}\right)^{a}\right)^{-1}}(p,q)}{\beta(p,q)}$$

(3.1)

where $a\in\mathbb{R},b,p,q>0$ and $B_x(p,q)$ and $\beta(p,q)$ stand for the incomplete and the complete beta functions, respectively.

The Gini, Theil, MLD and Atkinson indices in (1.1)–(1.5) for GB_2 distribution are derived as (see, McDonald and Ransom 2008, and Nakhaei Rad, Mohtashami Borzadaran, and Yari 2016)

$$G_{GB_2}=\frac{\beta\left(2q-\frac{1}{a},2p+\frac{1}{a}\right)}{\beta(p,q)\beta\left(p+\frac{1}{a},q-\frac{1}{a}\right)}\left\{\frac{1}{p}\times_3F_2\left(1,p+q,2p+\frac{1}{a};1;p+1,2(p+q)\right)-\frac{1}{p+\frac{1}{a}}\right.$$
$$\left.\times_3F_2\left(1,p+q,2p+\frac{1}{a};1;p+\frac{1}{a}+1,2(p+q)\right)\right\}.$$

(3.2)

$$T_{GB_2}=\frac{1}{a}\left(\varphi\left(p+\frac{1}{a}\right)-\varphi\left(q-\frac{1}{a}\right)\right)-\log\frac{\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(q-\frac{1}{a}\right)}{\Gamma(p)\Gamma(q)},$$

(3.3)

$$MLD_{GB_2}=-\frac{1}{a}(\varphi(p)-\varphi(q))+\log\frac{\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(q-\frac{1}{a}\right)}{\Gamma(p)\Gamma(q)},$$

(3.4)

$$A(\varepsilon)_{GB_2}=1-\frac{\Gamma^{\frac{1}{1-\varepsilon}}\left(p+\frac{1-\varepsilon}{a}\right)\Gamma^{\frac{1}{1-\varepsilon}}\left(q-\frac{1-\varepsilon}{a}\right)\Gamma^{1-\frac{1}{1-\varepsilon}}(p)\Gamma^{1-\frac{1}{1-\varepsilon}}(q)}{\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(q-\frac{1}{a}\right)},\varepsilon\neq1,$$

(3.5)

$$A(1)_{GB_2}=1-\frac{\Gamma(p)\Gamma(q)}{b\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(q-\frac{1}{a}\right)}\exp\left(\frac{\varphi(p)}{a}-\frac{\varphi(q)}{a}+\log b\right),$$

(3.6)

respectively, where $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt$ and

$${}_mF_n(a_1,\ldots,a_m;z;b_1,\ldots,b_n)=\sum_{i=0}^{\infty}\frac{(a_1)_i\ldots(a_m)_i}{(b_1)_i\ldots(b_n)_i}\frac{z^i}{i!},$$

with $(a)_i = a(a+1)\dots(a+i-1), \varphi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ and $\Gamma'(z) = \frac{d}{dz}\Gamma(z)$.

4. Simulation study

In this section, the efficiency of the estimators of the inequality indices including Gini, Theil, MLD and Atkinson based on mentioned sampling schemes are compared. The inverse cdf simulation technique is employed in order to generate data from GB_2 distribution. Sample sizes n =10, 30, 45, 60, 80, 100 are considered. In more details, the pairs (m, r) = (2, 5), (3–5, 10, 15, 20) are taken, where m is the set size and r is the number of cycles. The m is chosen to be up to 5 for reducing the ranking error and so approaching to the perfect ranking. In JPS scheme, H is considered to be 5. For analyzing the sensitivity of the results, the simulation is performed for GB_2 with two different parameter sets: $GB_2(2,4,5,8)$ as well as $GB_2(0.5,2,10,20)$ which demonstrate a (an almost) symmetric density and a right-skewed one, respectively. Moreover, in order to analyze the performance of the estimators obtained based on the rank-based sampling designs under the presence of ranking error, imperfect ranking model is applied as well. There are some approaches in order to cosider the effect of imperfect ranking (see e.g. Vock and Balakrishnan 2011). Here, we have used *fraction of random rankings* in which the distribution of $X_{(i,i)}$ given by (2.7) is a mixture of the distribution of $X_{i:m}$, the true ith order statistic, and the underlying distribution, more precisely

$$F_{(i,i)} = (1-\lambda)F_{i:m} + \lambda F, \qquad i = 1,\dots,m,\ 0 < \lambda < 1$$

(4.1)

where $F_{(i,i)}$ and $F_{i:m}$ are the cdfs of $X_{(i,i)}$ and $X_{i:m}$, respectively, and λ is the mixing parameter. Obviously, if $\lambda = 0$, then we have the perfect ranking model. Here we take the values of λ in (4.1) to be 0.1 and 0.4 that correspond to small and moderate ranking errors.

The MSE's and the biases of the estimators obtained based on both perfect and imperfect procedures are computed using a Monte Carlo simulation with 10000 replications. Then, the relative efficiency (RE) is calculated as follows

$$RE(\widehat{\hat{\theta}}, \hat{\theta}_{SRS}) = \frac{MSE(\widehat{\hat{\theta}})}{MSE(\hat{\theta}_{SRS})},$$

(4.2)

where $\widehat{\hat{\theta}}$ is an estimator under each mentioned rank-based sampling schemes. The results for RE are shown in Figures 1–4 while the biases of estimators are plotted in Figures 5–8 for all mentioned inequality indices. In each Figure, the plots in the top panel are drawn for $GB_2(2,4,5,8)$ and the bottom one are drawn for $GB_2(0.5,2,10,20)$. From Figures 1–4, the following general conclusions can be observed.

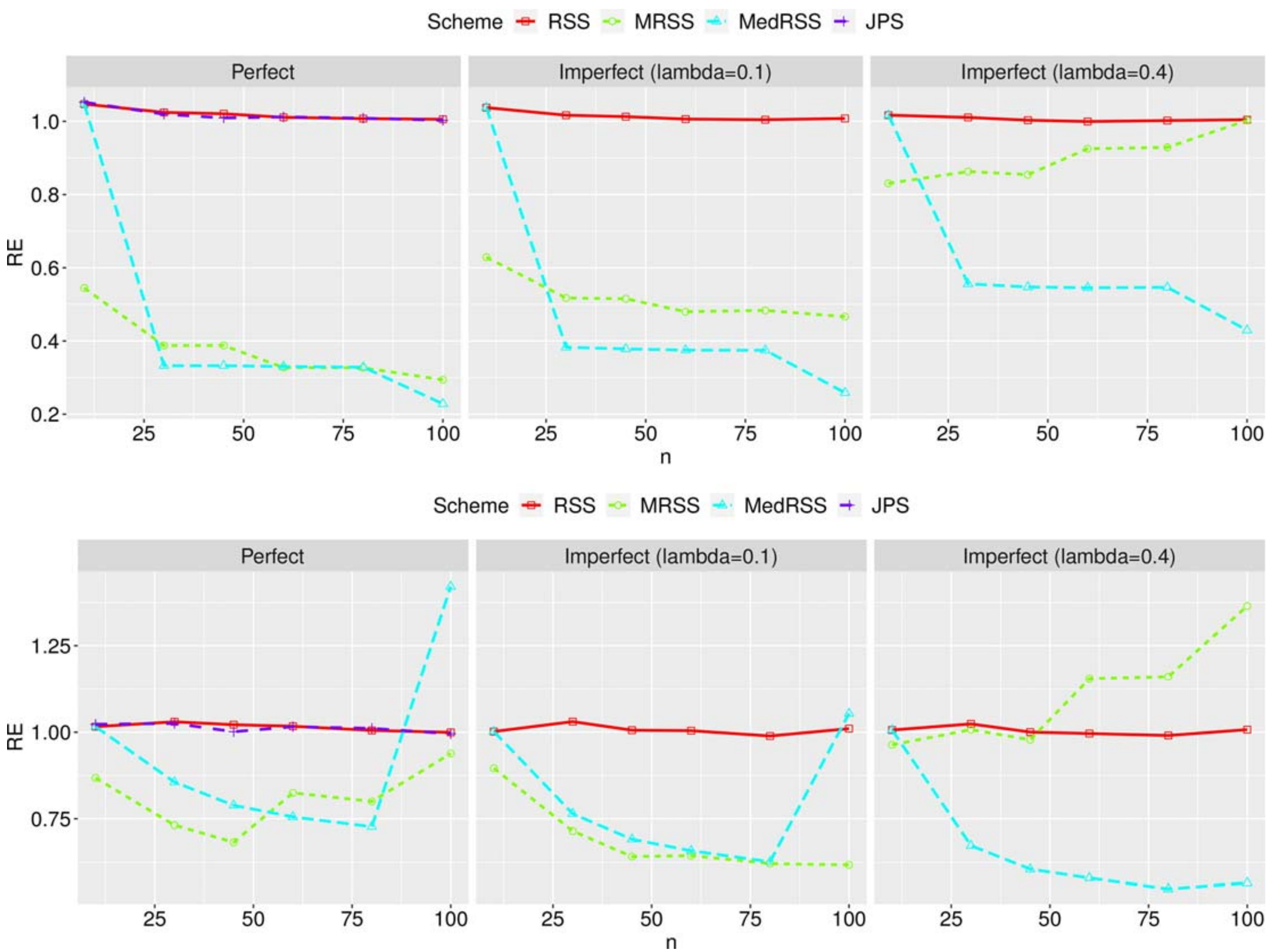


Figure 1. The plot of RE of the estimations of G versus the sample size *n*.

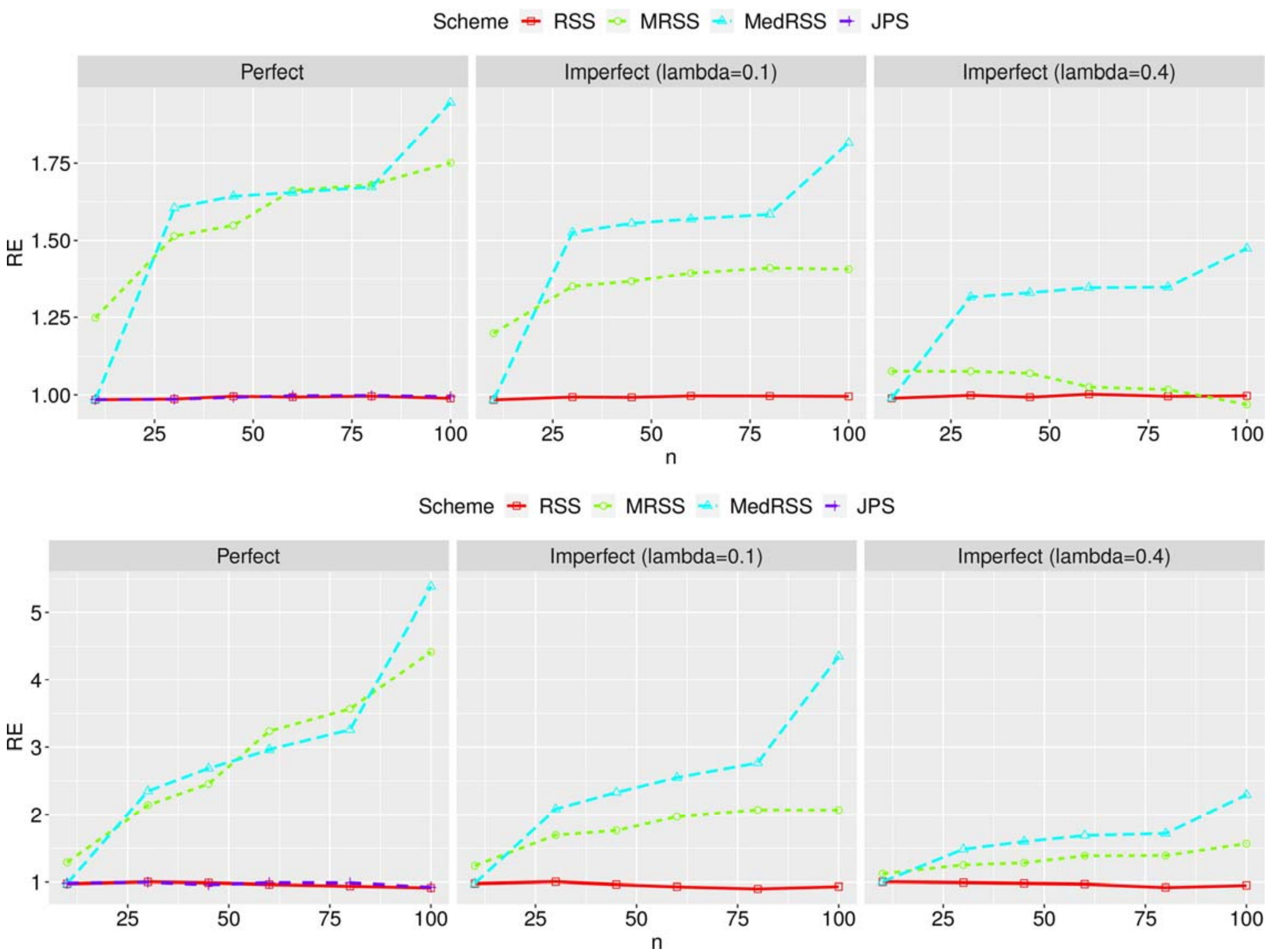


Figure 2. The plot of RE of the estimations of T versus the sample size *n*.

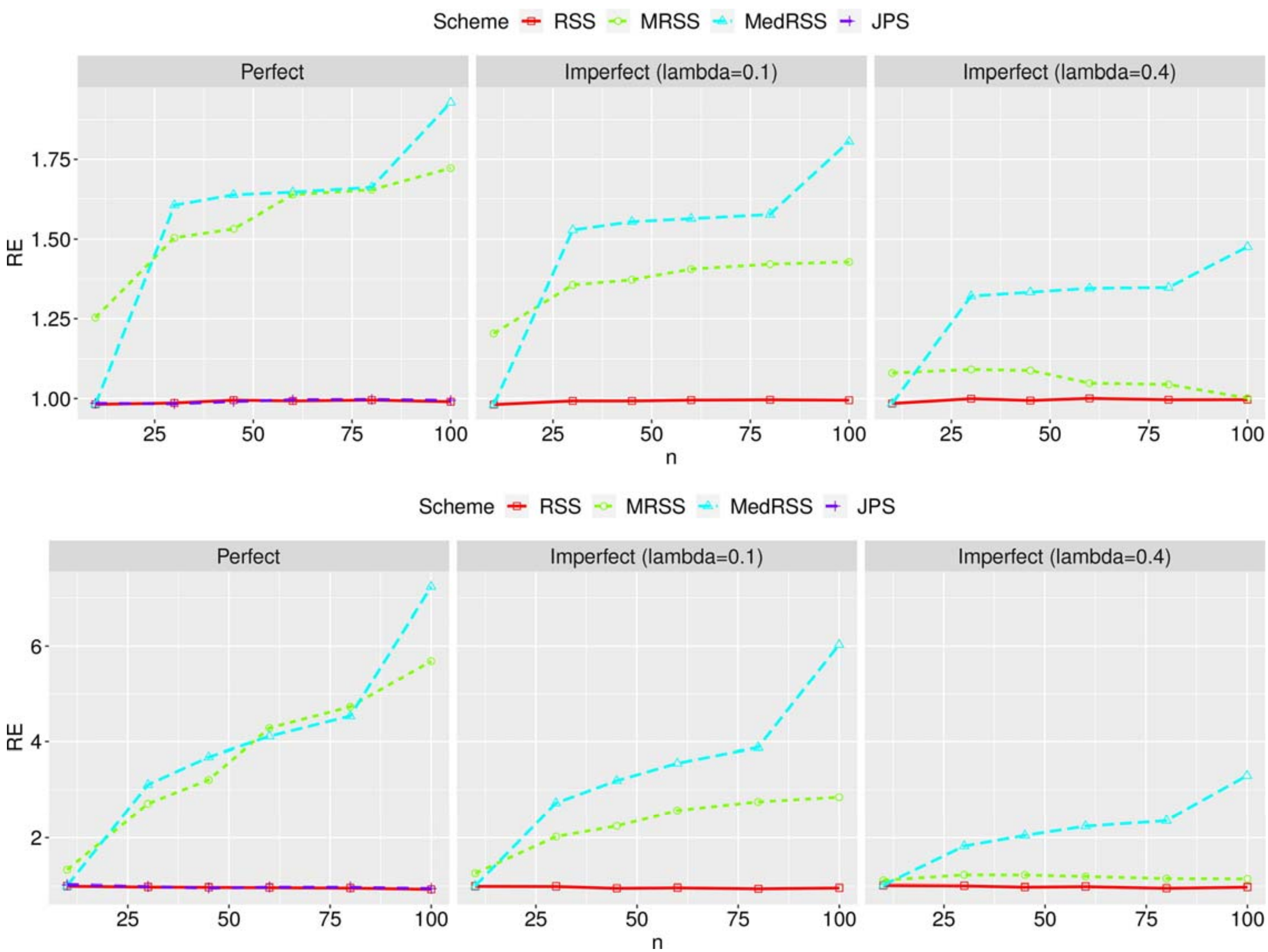


Figure 3. The plot of RE of the estimations of MLD versus the sample size *n*.

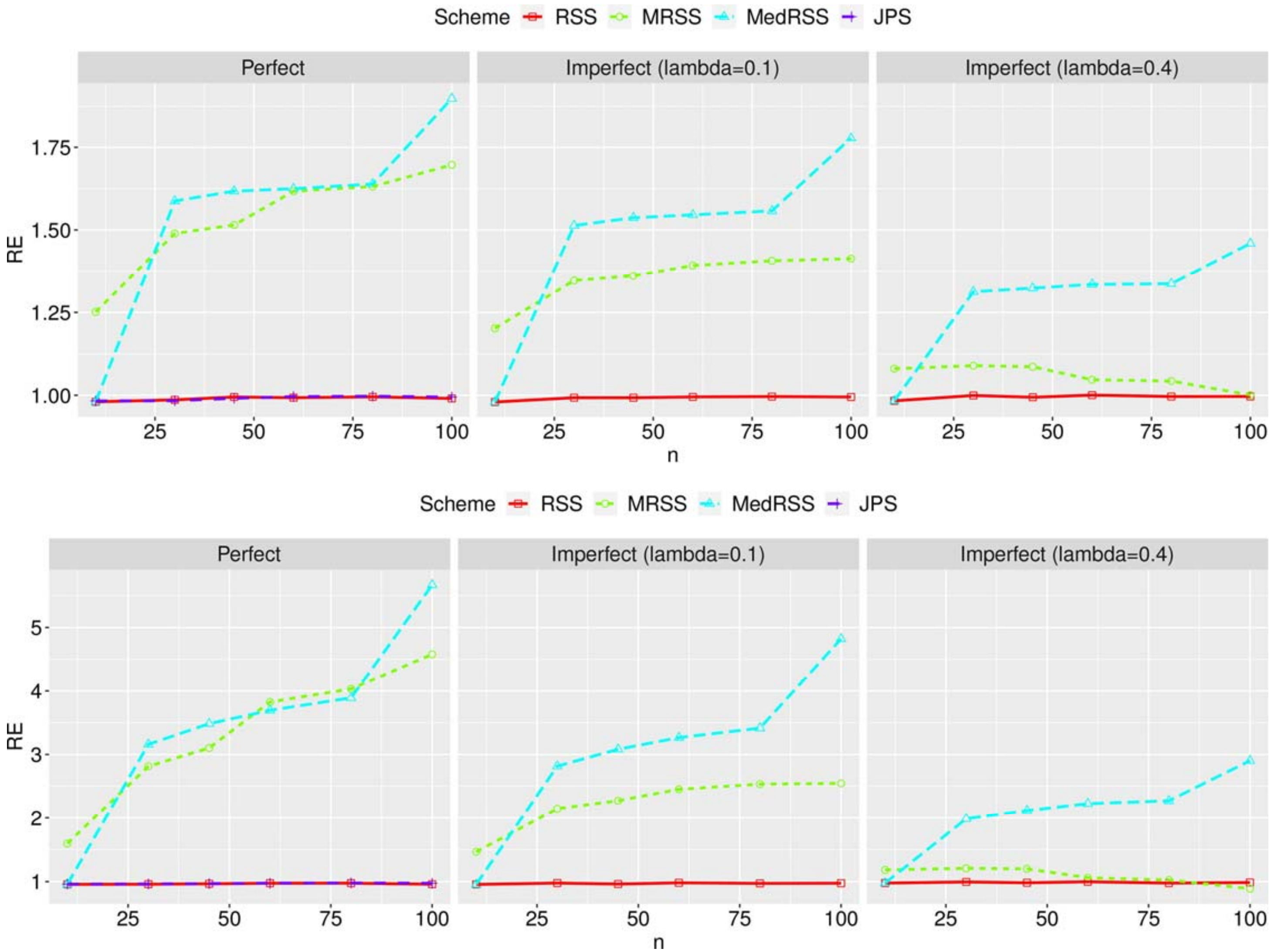


Figure 4. The plot of RE of the estimations of $A(\varepsilon)$ versus the sample size n.

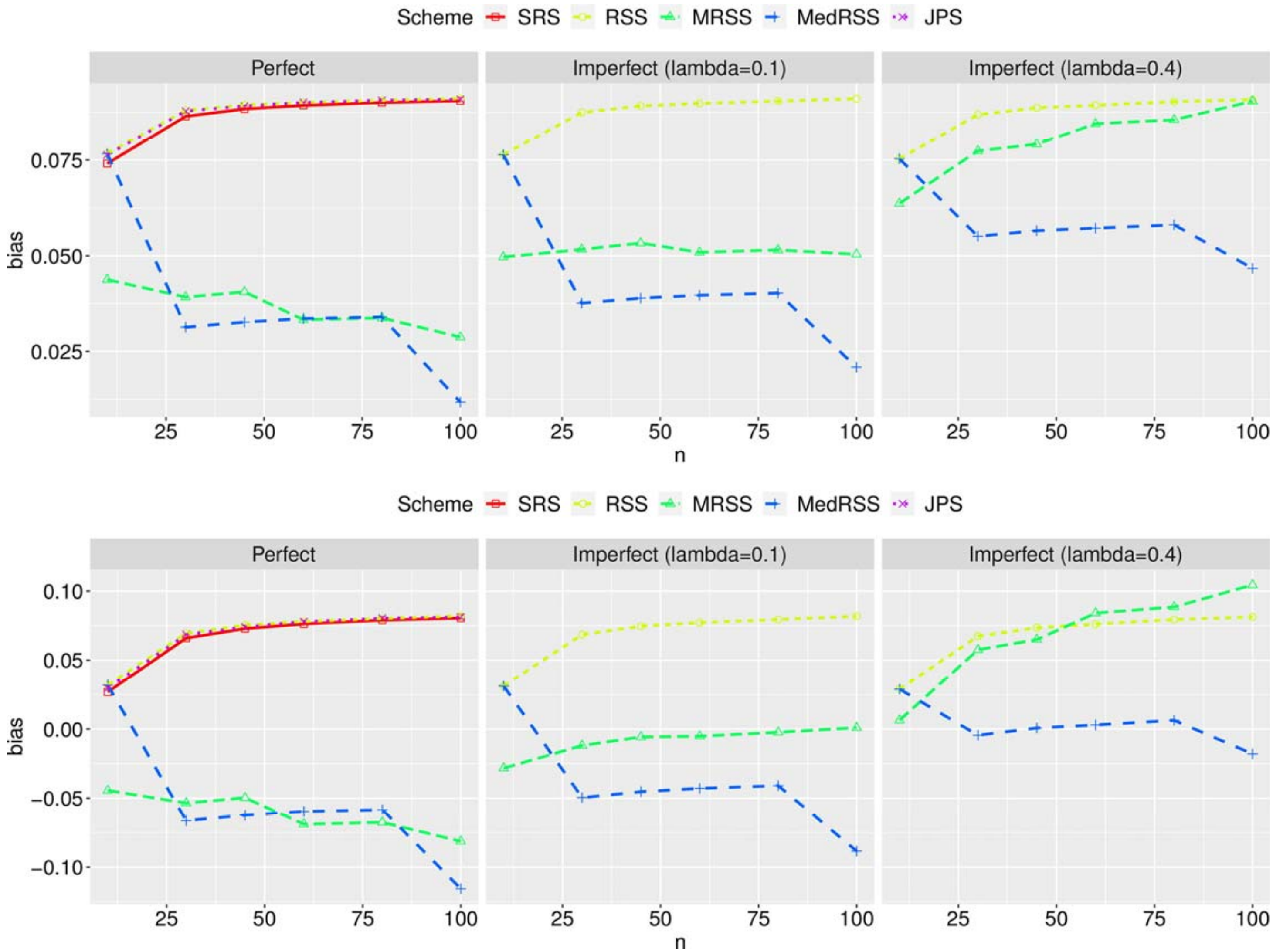
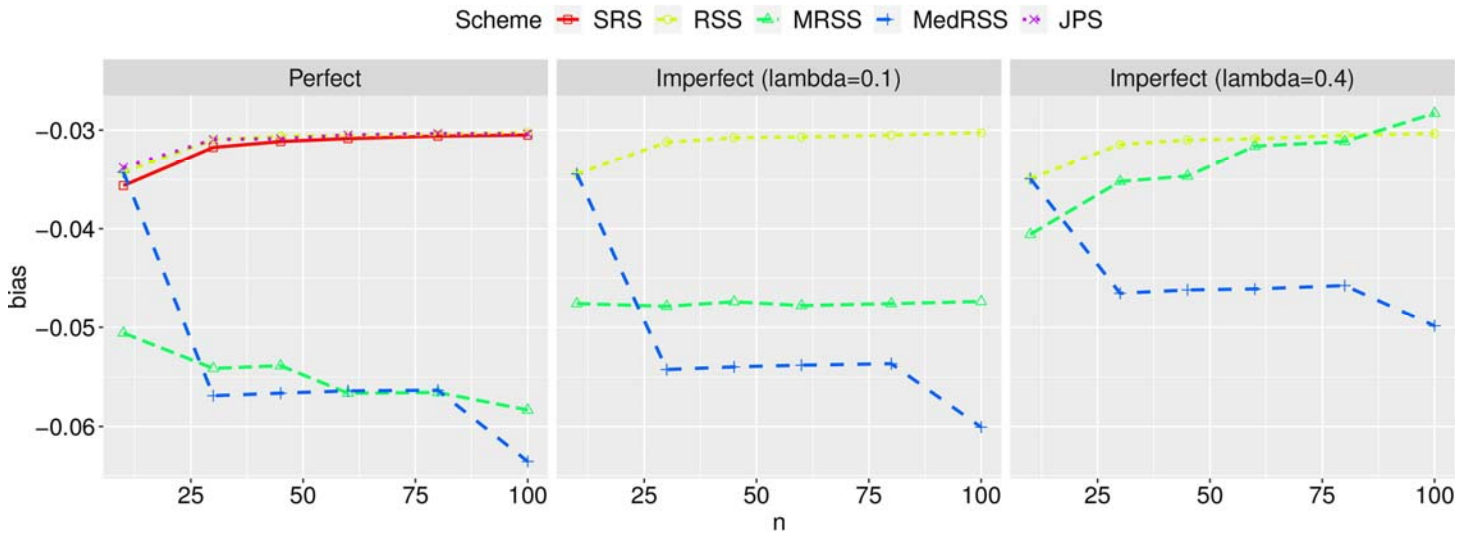


Figure 5. The plot of bias of the estimations of G versus the sample size *n*.



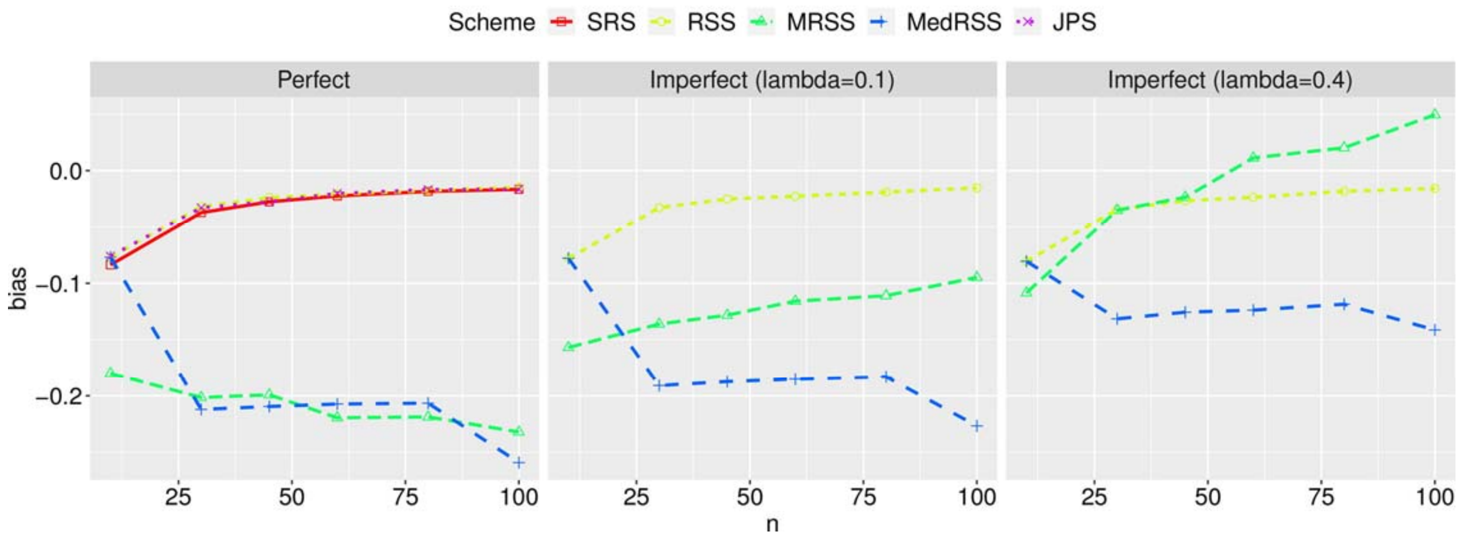


Figure 6. The plot of bias of the estimations of T versus the sample size *n*.

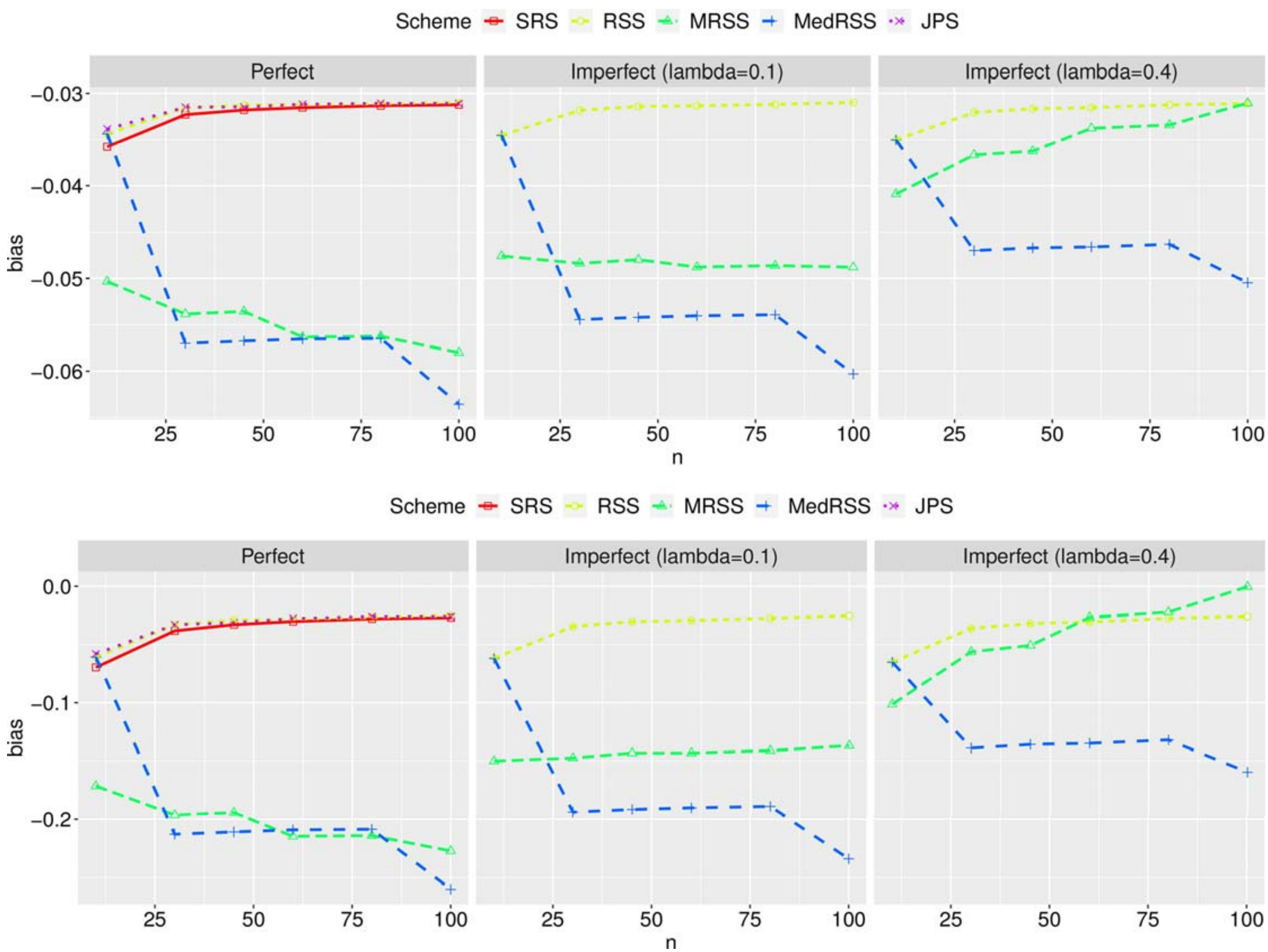


Figure 7. The plot of bias of the estimations of MLD versus the sample size *n*.

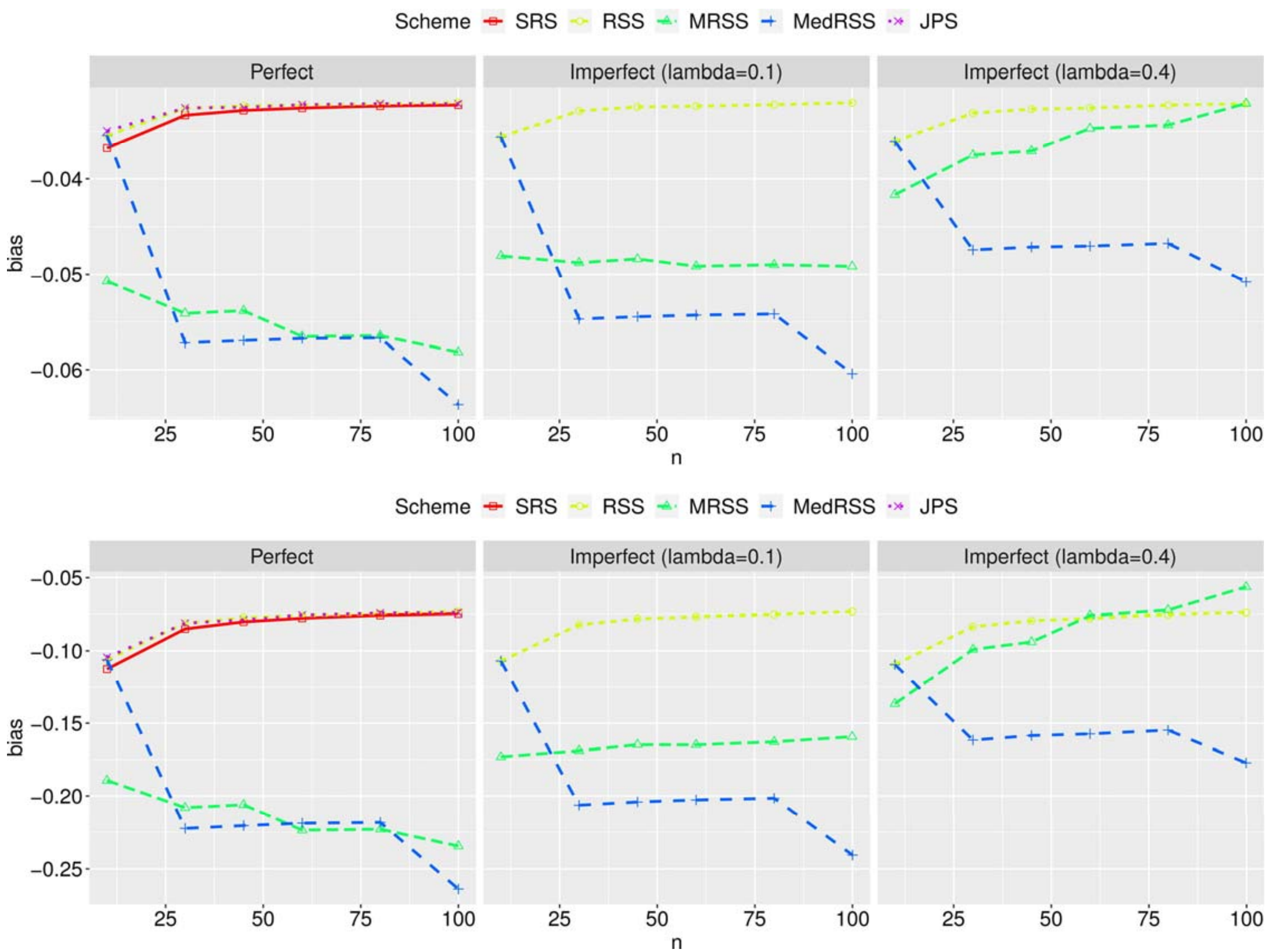


Figure 8. The plot of bias of the estimations of $A(\varepsilon)$ versus the sample size *n*.

- A point which is clear in all circumstances is that both of prefect and imperfect mechanisms have almost similar effects on the performances of the estimators. However, MRSS shows more sensitivity than its counterparts with respect to the magnitude of the error occurred in the ranking process.
- As the ranking error decreases, or equivalently λ tends to 0, the schemes having better performances become even more appropriate and in contrast those having weaker performances get worse. Hence, for a given inequality index, choosing the best scheme and increasing the precision in the ranking process are very crucial here.
- As it is observed, RSS and JPS have a close behavior in almost all of the scenarios.

Furthermore, the following specific points can be perceived.

- For estimating the Gini index under a symmetric GB2 distribution, RSS and JPS schemes perform better than SRS, while MRSS and MedRSS have weaker performance than it. When income distribution is right-skewed, MRSS outperforms the other schemes.
- RSS and JPS work better than SRS in estimating the Theil, MLD and $A(\varepsilon)$ indices. MRSS and MedRSS exhibit worse performance than SRS as *n* increases.
- Furthermore, as it can be seen in Figures 5–8, the (absolute value of the) biases of the estimators under SRS, RSS and JPS schemes are less than the other sampling designs. However, SRS has the least bias.

5. Real data

To illustrate the results obtained in the previous section, a real data set consisting of 7200 GDP (million dollars) per capita of 172 countries in 1970-2012 has been considered (the data were extracted from http://www.unctadstat.unctad.org). GB_2 distribution is fitted to the mentioned data set with parameters $a = 0.0509, b = 15159400, p = 140.9051, q = 94.3788$ (log-likelihood=-79243.91, AIC =158495.8, BIC = 158523). Figure 9 shows the goodness-of-fit plots of the data.

The values of the inequality indices for $GB_2(0.0509, 15159400, 140.9051, 94.3788)$ are calculated from (3.2)–(3.6) as $G = 0.9418, T = 0.5632, MLD = 0.5481, A(1) = 0.9708, A(2) = 0.9989.$ In order to check the sensitivity of the chosen sample, the parametric as well as the non-parametric bootstrap techniques are employed. The MSE's and the biases are reported in Tables 1–2. As it can be seen, RSS and JPS have less MSE's and biases than other sampling methods, similar to the simulation results. Also box-plots of the $\widehat{T}$ and $\widehat{G}$ drawn in Figures 10 and 11 confirm the adequacy of the RSS and JPS plans.

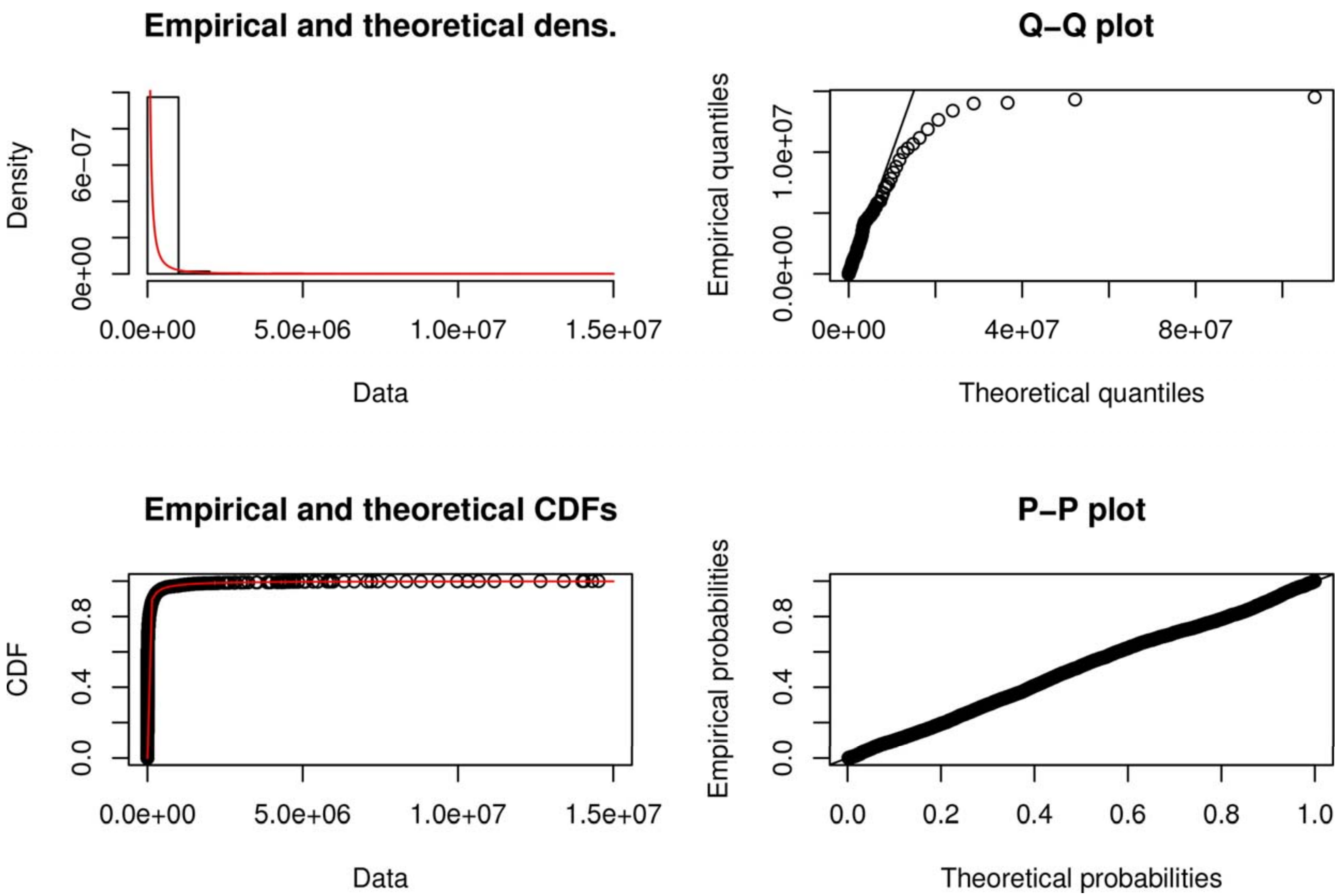


Figure 9. The empirical and theoretical density function and distribution function, P-P and Q-Q plots.

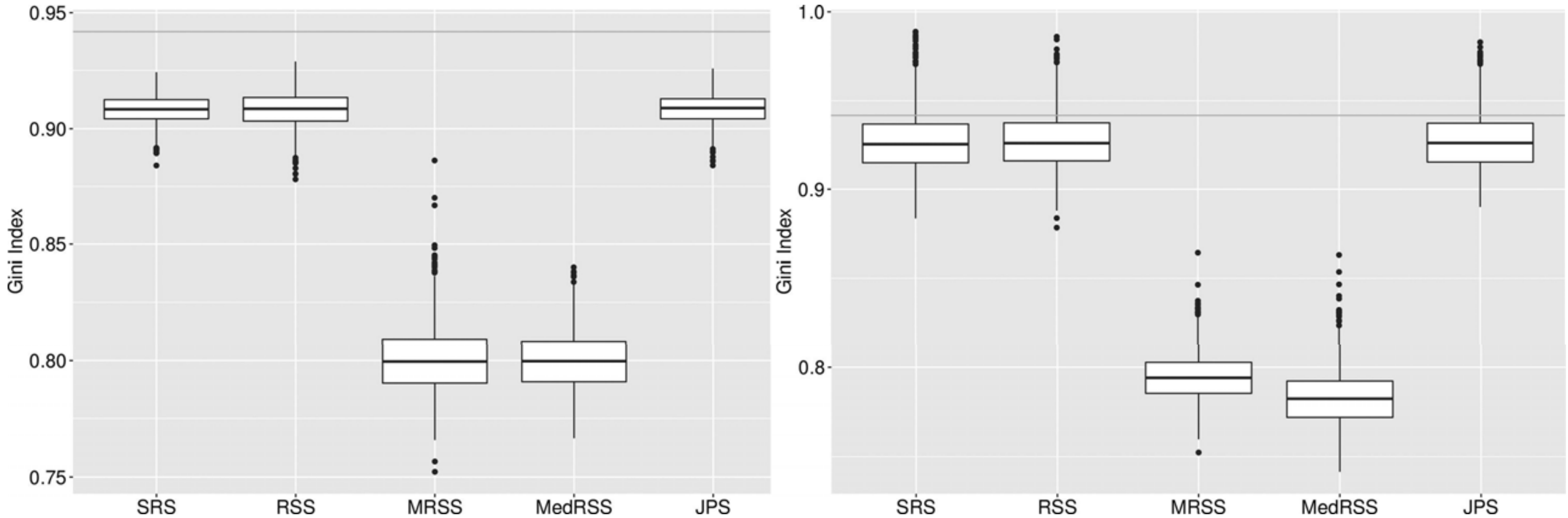


Figure 10. The boxplots of $\widehat{G}$ in parametric (right panel) and non-parametric (left panel) cases.

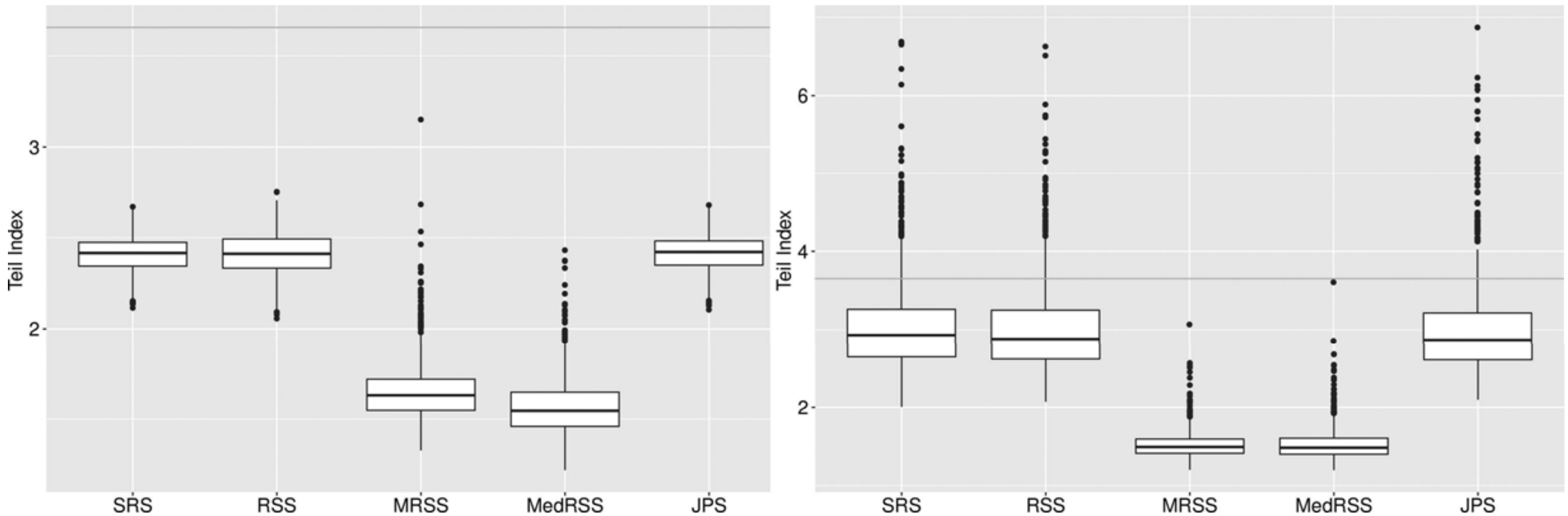


Figure 11. The boxplots of the $\widehat{T}$ in parametric (right panel) and non-parametric (left panel) cases.

Table 1. The MSE and bias for Gini estimator in parametric and non-parametric cases.

Gini	SRS	RSS	MRSS	MedRSS	JPS
MSE (Non-parametric)	0.00117	0.00118	0.02023	0.02028	**0.00115**
MSE (Parametric)	0.00050	0.00047	0.02184	0.02537	**0.00046**
Bias (Non-parametric)	−0.03374	−0.03350	−0.14144	−0.14182	**−0.03333**
Bias (Parametric)	−0.01470	−0.01443	−0.14714	−0.15846	**−0.01438**

Table 2. The MSE and bias for Theil estimator in parametric and non-parametric cases.

Theil	SRS	RSS	MRSS	MedRSS	JPS
MSE (Non-parametric)	1.56421	1.56589	4.05293	4.38125	**1.55553**
MSE (Parametric)	0.74756	**0.72781**	4.58689	4.56083	0.78968
Bias (Non-parametric)	−1.24714	−1.24631	−2.00592	−2.08707	**−1.24370**
Bias (Parametric)	−0.64426	**−0.61361**	−2.13528	−2.12455	−0.64466

6. Conclusion

In this article, in order to find the most appropriate sampling schemes for estimating different inequality indices such as Gini, Theil, MLD, Atkinson, under the GB_2 parent distribution, the SRS was compared with four rank-based sampling plans including RSS, median RSS, modified RSS and JPS. The performance of the resulting estimators were compared by carrying out a simulation study under both perfect and imperfect ranking mechanisms based on two criteria 'bias' and 'relative efficiency'. The simulation results and real data analysis showed that both of prefect and imperfect mechanisms have almost similar impacts on the performances of the estimators. However, as the ranking error decreases, the schemes having better performances become even more appropriate and in contrast those having weaker performances get worse. Hence, choosing the appropriate scheme for estimating a specific inequality index and also increasing the precision of the ranking stage play important roles here.

Acknowledgement

This work was based upon research supported in part by the National Research Foundation (NRF) of South Africa, SARChI Research Chair UID: 71199 and DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa. The Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the CoE-MaSS and the NRF.

References

Al-Talib, M. M., and A. D. Al-Nasser. 2008. Estimation of Gini index from continuous distribution based on ranked set sampling. *Electronic Journal of Applied Statistical Analysis* 1:33–41.

Bansal, P., Arora, S. Mahajan. K., 2013. Estimation of inequality indices based on simple random, ranked set and systematic sampling. *ISRN Probability and Statistics* 2013:1–14. doi:10.1155/2013/659580.

Dastbaravarde, A., N. R. Arghami, and M. Sarmad. 2016. Some theoretical results concerning non parametric estimation by using a judgment poststratification sample. *Communications in Statistics - Theory and Methods* 45 (8):2181–203. doi:10.1080/03610926.2013.878355.

Dell, T. R., and J. L. Clutter. 1972. Ranked set sampling theory with order statistics background. *Biometrics* 28 (2):545–53. doi:10.2307/2556166.

Dey, S., M. Salehi, and J. Ahmadi. 2017. Rayleigh distribution revisited via ranked set sampling. *Metron* 75 (1):69–85. doi:10.1007/s40300-016-0099-2.

Halls, L. K., and T. R. Dell. 1966. Trial of ranked set sampling for forage yeilds. *Forest Science* 12 (1):22–6.

MacEachern, S. N., E. A. Stasny, and D. A. Wolfe. 2004. Judgement post-stratification with imprecise rankings. *Biometrics* 60 (1):207–15. doi:10.1111/j.0006-341X.2004.00144.x.

McDonald, J. B., and M. Ransom. 2008. The generalized beta distribution as a model for the distribution of income: Estimation of related measure of inequality. *Economic Studies in Equality, Social Exclusion and Well-Being* 5:147–66.

McIntyre, G. A. 1952. A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research* 3 (4):385–90. doi:10.1071/AR9520385.

Muttlak, H. A. 1997. Median ranked set sampling. *Journal of Applied Statistical Science* 6:577–86.

Nakhaei Rad, N.,. G. R. Mohtashami Borzadaran, and G. H. Yari. 2016. Estimation of inequality indices based on ranked set sampling. *Hacettepe Journal of Mathematics and Statistics* 46 (134):1–720. doi:10.15672/HJMS.201613419772.

Rizzo, M. L. 2002. A test of homogeneity for two multivariate populations. Proceedings of the American Statistical Association, Physical and Engineering Sciences Section.

Stokes, S. L. 1980a. Estimation of variance using judgement ordered ranked set samples. *Biometrics* 36 (1):35–42. doi:10.2307/2530493.

Stokes, S. L. 1980b. Inferences on the correlation coefficient in bivariate normal populations from ranked set samples. *Journal of the American Statistical Association* 75 (372):989–95. doi:10.1080/01621459.1980.10477584.

Stokes, S. L., and T. W. Sager. 1988. Characterization of a ranked-set sample with application to estimating distribution functions. *Journal of the American Statistical Association* 83:35–42.

Takahashi, K. 1970. Practical note on estimation of population means based on samples stratified by mean of ordering. *Annals of the Institute of Statistical Mathematics* 22:421–8.

Takahasi, K., and K. Wakimoto. 1968. On unbiased estimates of the population mean based on the sample stratified by mean of ordering. *Annals of the Institute of Statistical Mathematics* 20 (1):1–31. doi:10.1007/BF02911622.

Vock, M., and N. Balakrishnan. 2011. A Jonckheere–Terpstra-type test for perfect ranking in balanced ranked set sampling. *Journal of Statistical Planning and Inference* 141 (2):624–30. doi:10.1016/j.jspi.2010.07.005.