Flexible variance estimation in complex sample surveys: rescaled bootstrap in multistage, pps surveys - DRAFT

Malgorzata Osiewicz¹, Sébastien Pérez-Duarte²
¹European Central Bank, e-mail: malgorzata.osiewicz@ecb.europa.eu
²European Central Bank, e-mail: sebastien.perez-duarte@ecb.europa.eu

Abstract

Estimating sampling variance in cross-country surveys is, in general, a complicated task. Sample designs are usually different across countries and are often complex. In order to provide users with tools to estimate sampling variance, the Eurosystem Household Finance and Consumption Survey (HFCS) has implemented the rescaled bootstrap. This method is able to accommodate most user requirements, and is applicable to many of the statistics that will be produced by users, as well as to most of the sampling designs used in the HFCS. Nevertheless, the standard rescaled bootstrap is not appropriate for multistage sampling, in which the one of the stages is drawn with probabilities proportional to size. Moreover, often only the first sampling stage is explicitly covered by the standard bootstrap. We explore several extensions of the rescaled bootstrap to multi-stage samples, including stages with probabilities proportional to size. These extensions strive to ease the task of the final user of the data. The replicate weights allow the data producer to trade off the detailed knowledge of the sample design that the user would need to apply linearization techniques with a large set of replicate weights. Simulations are used to investigate the effectiveness of the different methods.

Keywords: variance estimation, probability proportional to size, replicate weights

Acknowledgements: This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

1. Motivation

Appropriate sampling variance estimation should be an imperative for researchers working with survey data. Variance estimation has been a very active area of research in the past half-century, though the practice may lag somewhat behind the theory, particularly among non-expert users. Additionally, variance estimation is not always made easy for ultimate users of the data, who then treat the data as if it had been collected through a simple random sample from an infinite population.

There are two main difficulties with the existing tools and methods. The first one is that the calculation of standard errors suffers in general from the complexity of the sample
design, and that the more complex the survey, the more difficult it is to appropriately reflect this in the variance estimation. Most social surveys have multistage designs, possibly with some stages selected with probability proportional to size (pps), and rotational patterns in the sample.

The second general difficulty is the availability of sample design information to final users. In many cases the stratification and clustering reflect geographical characteristics (for example the regions, municipalities, urban areas, blocks) which are mostly suppressed from research utility files because of confidentiality issues. Information on strata and primary sampling units (PSU) is thus suppressed in most cases, and the linearization approach cannot be used.

Our objective in this paper is to focus on the applicability of the replicate weights/bootstrap approach to pps sampling, first in the context of with-replacement sampling, and second in the Rao-Hartley-Cochran (1962, hereafter RHC) pps sampling procedure, which can be used with the Rao and Wu (1988) rescaled bootstrap. With the exception of Beaumont and Patak’s (2012) generalized bootstrap in the case of Poisson sampling, to the best of our knowledge the implication of the RHC pps sampling on the rescaled bootstrap weights have so far not been explicitly pointed out and put in practice. Section 2 describes first the standard Rao and Wu rescaled bootstrap before presenting pps alternatives, while section 3 describes a simulation exercise for one-stage pps sampling, to be extended in a revised version of this document to multistage pps.

2. Rescaled bootstrap in stratified and pps surveys

In this section, we first present the “standard” rescaled bootstrap as described by Rao and Wu (1988) in the case of stratified simple random samples. The method is then extended to pps sampling, first in the case of with-replacement sampling, then to Rao-Hartley-Cochran pps sampling, and finally to multistage pps sampling.

2.1 The Rao and Wu (1988) bootstrap

The Eurosystem Household Finance and Consumption Network decided to use the Rao-Wu bootstrap, known as the “rescaled bootstrap” (RSB), for the first wave of the HFCS. Although other bootstrap methods could have performed equally well, the RSB was selected because it translates well to replicate weights, is easy to setup in stratified simple random samples, and is used by some statistical institutes (see Girard 2009 for example). In addition, some recent comparisons through simulations carried out in the case of social surveys (e.g. Münnich et al. 2012) show that the rescaled bootstrap (with its Without-Replacement variant) is the top performer, and the With-replacement version comes very close to it.

The RSB was introduced by Rao and Wu (1988) but its applicability through weight adjustment was only noticed by Rao, Wu and Yue (1992). We describe quickly the setup in a stratified, one-stage survey. We consider the case of strata indexed by $h = 1, \ldots, H$, 

...
with \( N_h \) units in each of them, out of which \( n_h \) are sampled without replacement. The sampling fraction is thus \( f_h = n_h / N_h \). To each unit \((h,i)\) there is a variable of interest \( y_{hi} \) and a weight \( w_{hi} = N_h / n_h \). The total of this variable is \( Y = \sum_{h=1}^{H} \sum_{i=1}^{N_h} y_{hi} \) which is estimated without bias by \( \hat{Y} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi} y_{hi} \). For the RSB, the following is done \( B \) times:

1. A sample of size \( m_h \) is taken with replacement from each stratum.
2. Writing \( r_{hi}^* \) the number of times unit \((h,i)\) is resampled, the weights are adjusted as follows: \( w_{hi}^* = (1 - \lambda_h + \lambda_h \frac{n_h}{m_h} r_{hi}^*) w_{hi} \) with \( \lambda_h = \sqrt{m_h (1 - f_h) / (n_h - 1)} \).
3. The bootstrap total is computed \( \hat{Y}_b = \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi}^* y_{hi} \).

The bootstrap variance is then calculated as \( V_b (\hat{Y}) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{Y}_b - \bar{\hat{Y}})^2 \), where \( \bar{\hat{Y}} \) is the mean of the bootstrap total over all \( B \) iterations.

### 2.2 Extension of the RSB to probability-proportional-to-size sampling

While the RSB has shown its usefulness in several practical cases, its applicability in the context of the general surveys needed to be studied further. The case of multistage samples, with (stratified) simple random sampling in each stage, has been developed by Preston (2009) in the case of without-replacement-bootstrap, and extended to with-replacement-bootstrap in Osiewicz and Perez-Duarte (2012).

A remaining issue is the case of probability proportional to size sampling (hereafter pps), and more particularly in the case of the HFCS, multistage samples where the first stage is drawn with unequal probabilities, as is often the case in household surveys, where the first stratification level is at the district/municipality level, and simple random sampling (or systematic sampling) is carried out in each primary sampling unit.

The literature on pps sampling is too wide to be summarized here. While the framework offered by the Horvitz-Thompson estimator and the corresponding variance estimator (either the Horvitz-Thompson one or the Sen-Yates-Grundy version) is conceptually pleasing, in practice pps sampling is difficult to implement with fixed-size samples while respecting a few desirable properties, except when drawing one or two units by stratum.

We first consider the simple case of with-replacement pps sampling, as this type of sampling has a very simple analog in terms of replicate weights.

#### 2.2.1 RSB with with-replacement pps sampling

In this simplest scenario, units are sampled with pps, with replacement (hereafter SIR PPS). Although more sophisticated forms of sampling will tend to be used in practice, we believe that when the sampling fraction is low (under unequal probability sampling, this

---

1 We will not make an explicit difference in the writing between pps [implying without replacement] and pps [with replacement].
could be taken to be the highest probability; in that case the probability of selecting one unit more than once is limited) the approach should give sensible results.

We let $p_1, \ldots, p_N$ the probabilities of selecting each of the elements in the population (for example, $p_k = x_k / \sum_{i=1}^N x_i$, where $x$ is the “size”). We draw $n$ units with replacement, with the given probabilities. Some units may be sampled more than once; in that case they are kept separate and a different index $k_i$ is used, $i = 1, \ldots, n$.

An estimator of the total of the variable $y$ is $\hat{Y} = \frac{1}{n} \sum_{i=1}^n \frac{y_{k_i}}{p_{k_i}}$. This is called the pwr estimator in Särndal, Swensson and Wretman (1992). Its variance is given by:

$$V(\hat{Y}) = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_{k_i}}{p_{k_i}} - \bar{Y} \right)^2 = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_{k_i}^2}{p_{k_i}} - \bar{Y}^2 \right),$$

which is estimated without bias by:

$$\hat{V}(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^n \left( \frac{y_{k_i}}{p_{k_i}} - \bar{Y} \right)^2.$$

Since the $y_{k_i}$ are i.i.d., the bootstrap is applicable, and just needs to be adapted to take into account the sampling. We therefore resample with replacement and equal probabilities $m$ units $y_1^*, \ldots, y_m^*$, along with their initial selection probabilities $p_1^*, \ldots, p_m^*$. We code this in a framework with weights.

By setting $w_k = \frac{1}{n p_k}$ we can write the estimator of the total $\hat{Y} = \sum_{i=1}^m w_{k_i} y_{k_i}$. Then the bootstrap estimate is $\hat{Y}^* = \sum_{i=1}^m w_{k_i}^* y_{k_i}^*$. With the same rescaling as in the Rao-Wu bootstrap above, we get the correct variance when the bootstrap weights are:

$$w_i^* = \left(1 - \lambda + \lambda r_i^* \frac{2}{n} \right) w_i, \text{ with } \lambda = \left( \frac{1}{n} \right)^{1/2},$$

and $r_i^*$ is the number of times that unit $i$ was resampled.

The question of the optimal $m$ is open; similarly to the stratified simple random sampling, the choice of $m = n - 1$ is natural. The rescaled weights are then always positive and the formula simplifies to $w_i^* = \frac{1}{n} r_i^* w_i$.

### 2.2.2 Extension of with-replacement pps to multistage samples

In the case of multistage samples, when the first sampling stage is with replacement pps, there is a very useful extension of the previous results, as described in Särndal, Swensson and Wretman (1992). Under some conditions on how the with-replacement sampling is done, the only thing that matters is that it is possible to estimate a total $\hat{Y}_i$ from each

---

2 Namely: (i) the sampling in stages 2 and onwards is invariant, i.e. the subsampling of a particular unit is always the same one and independent; (ii) if a PSU is sampled more than once, it is independently subsampled as many times as it is drawn.
PSU, and that the variance coming from the subsampling can be expressed as $V_i$. Then the variance of $\hat{t}$ is:

$$V(\hat{t}) = \frac{1}{n} \sum_{k=1}^{N} p_k \left( \frac{t_k}{p_k} - \hat{t} \right)^2 + \frac{k}{n} \sum_{k=1}^{N} \frac{V_k}{p_k}. \quad (4)$$

The standard unbiased estimator of this variance has a particularly simple expression, with the variance of the second and later stages only entering through the first stage totals:

$$\hat{V}(\hat{t}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left( \frac{\hat{t}}{p_i} - \hat{t} \right)^2. \quad (5)$$

In with-replacement pps sampling, the second and following stages enter the variance formula only through the PSU totals. In terms of replicate weights, it is possible to resample the first stage only and construct the weights as before, or to also resample the second stage and construct replicate weights consistent with both sampling stages.

2.3 The Rao-Hartley-Cochran pps sampling

Rao, Hartley and Cochran (1962, hereafter RHC) proposed a method which ensured without-replacement pps sampling; however, this loses the strict proportionality between the size of the units and their inclusion probability.

To do this, we partition the population at random in $n$ groups of sizes $N_1,...,N_n$. Then we draw one unit from each group, with probability of selecting a unit proportional to $p$ inside each group.

Set $P_k = \sum_{i \in \text{group}_k} p_i$, then the estimator of the total is $\hat{Y} = \sum_{k=1}^{N} \frac{y_k}{P_k} = \sum_{k=1}^{N} P_k \frac{y_k}{p_k}$. It is unbiased, and its variance is

$$V(\hat{Y}) = \frac{n}{N(N-1)} \left( \sum_{i=1}^{n} \frac{y_i^2}{n p_i} - \frac{Y^2}{n} \right). \quad (6)$$

The first factor in the variance above is the reduction compared to the with-replacement pps of equation (1). Variance is reduced overall if the groups can be made of identical sizes. The second factor is the variance of a with-replacement pps sample, as shown in the previous section. This variance can be estimated by

$$\hat{V}(\hat{Y}) = \frac{\left( \sum_{i=1}^{n} \frac{N_k^2 - N}{N^2 - \sum_{i=1}^{n} N_k^2} \right)}{\left( \sum_{k=1}^{N} P_k \left( \frac{y_k}{p_k} - \hat{Y} \right)^2 \right)}. \quad (7)$$

2.3.1 The Rao-Wu resampling for the Rao-Hartley-Cochran approach

The idea is to resample $m$ elements (with replacement) with probabilities $P_k$ (the one attached to the group) in step 1 of the RSB procedure described above.
**Theorem 1:** The bootstrap estimator of the Rao-Wu resampling for Rao-Hartley-Cochran pps samples is:

\[ \hat{Y}^* = \sum_{k=1}^{m} \left( 1 - \mu m^{1/2} P_k + \mu m^{-1/2} r_k^* \right) \frac{Y_k}{P_k} \], with \( \mu^2 = \frac{\left( \sum_{i=1}^{n} N_i^2 - N \right)}{\left( \sum_{i=1}^{n} N_i^2 \right)} \). (8)

This can be rewritten in terms of weights, by setting \( w_k = P_k / p_k \), so that the replicate weights are \( w_k^* = \left( 1 - \mu m^{1/2} + \mu m^{-1/2} r_k^* / P_k \right) w_k \).

The choice of \( m = n - 1 \) ensures positive replicate weights, as long as the units are partitioned in groups of approximately equal size. However, the question is open whether this is the best choice, or if a factor depending on \( \mu \) or the distribution of \( P_k \) would not match better the higher moments of the estimator. This is left for future research.

### 2.3.2 Rao-Wu rescaling for multistage pps samples

The formula above can be extended to cover two-stage samples, where the first stage is (potentially stratified) pps with the Rao-Hartley-Cochran procedure, and the second is (potentially stratified) simple random sampling without replacement. Then variance is estimated through the formula:

\[ V(\hat{Y}) = \mu^2 \sum_{k=1}^{m} \left( \frac{\hat{Y}_k}{P_k} - \hat{Y} \right)^2 + \sum_{k=1}^{m} \frac{P_k}{P} \left( \frac{1}{N_k} - \frac{1}{n_k} \right) s_k^2 \] (9)

where \( \mu \) is defined as in (8), \( N_k \) is the number of SSUs in PSU \( k \), \( n_k \) is the number of SSUs sampled in that PSU, \( \hat{Y}_k \) is the total of \( y \) in PSU \( k \) and \( s_k^2 \) is the sample variance of \( y \) in PSU \( k \).

**Theorem 2:** The Rao-Wu bootstrapping procedure for Rao-Wu rescaled bootstrap for multistage samples, with a RHC pps first stage and SRS second stage, is as follows:

1. Resample \( m \) PSUs (with replacement) with probabilities \( P_k \).
2. In each PSU \( k \) resampled (and independently each time a PSU was resampled more than once), sample with replacement \( m_k \) SSUs.
3. Writing \( r_k^* \) the number of times the PSU \( k \) is resampled, and \( r_{ki}^* \) the number of times the SSU \( i \) in PSU \( k \) is resampled, the weights are adjusted as follows:

\[ w_{ki}^* = \left( 1 - \lambda + \lambda_{r_k^*} - \frac{1}{\sqrt{mP_k}} \lambda_{r_k^*} \right) w_{ki} \] (10)

with \( \lambda^2 = m \mu, \lambda = \sqrt{m} \left( (1 - f_{h_k}) / p_k \right), \) and where \( f_{h_k} = n_k / N_k \) is the sampling fraction in PSU \( k \).
The parallel of replicate weight (10) with both the replicate weight formula for RHC sampling (8) and the standard Rao-Wu rescaling bootstrap in section 2.1 is worth mentioning.3

3. Simulation

A numerical simulation exercise is carried out on the NUTS3 geographical level available through European statistics (Nomenclature of Units for Territorial Statistics). There are 1462 such regions available in Eurostat, covering the European Union as well as candidate, acceding, and EFTA countries.4 The size of these regions ranges from 13km² (Melilla) to over 100,000km² (Norrbotten County in Sweden; Iceland excluding the capital), and these regions cover each from 10,500 (one of the Canary Islands) to 13 million inhabitants (province of Istanbul).

The variables of interest are: total population, total female population, number of births and deaths, total area, and death rate. Total population is highly correlated with the number of women, and is also correlated with the number of births and deaths, while the correlation is lower with the total area and even lower for the death rate (see Figure 1).

The size variable is total population, and n units are drawn (with n equal to 50, 100 or 200), once with with-replacement pps, and once according to the RHC pps sampling scheme. 500 replicate weights are computed, and variance is computed for the total, the mean, and the 10th, 50th, and 90th percentiles. This is repeated 5,000 times. Empirical variances for these estimators are computed by repeating the sampling 500,000 times.

The accuracy of the bootstrap variance estimators in the two sampling scenarios was assessed based on the relative bias (RB) and the relative root mean square error (RRMSE) according to the following formulas:

\[
RB = \frac{1}{\text{Var}(\hat{Y})} \left[ \frac{1}{S} \sum_{s=1}^{S} (\text{Var}(\hat{Y}_s) - \hat{\text{Var}}(\hat{Y})) \right], \quad \text{RRMSE} = \frac{1}{\text{Var}(\hat{Y})} \sqrt{\frac{1}{S} \sum_{s=1}^{S} (\text{Var}(\hat{Y}_s) - \hat{\text{Var}}(\hat{Y}))^2}
\]

One difficulty is obtaining a good estimate of \( \hat{\text{Var}}(\hat{Y}) \). Initial investigations show that RHC pps sampling has a larger variance of the variance, i.e. the variance of the total through repeated samples is poorly estimated, and would require a larger number of iterations. Tables 1 and 2 below show the relative bias and the relative RMSE for the totals, as in that estimator the true variance of the total is available through formula (6).

---

3 If an equal probability sample was implemented through the RHC sampling scheme, with \( N / n = R \) an integer, then \( p_i = 1/N \), \( P_i = 1/n \), and formula (10) simplifies to the two-stage rescaling bootstrap as in Osiewicz and Perez-Duarte (2012).

4 These statistics were downloaded from Eurostat’s website, and are the demo_r_d3area, demo_r_d3natmo, and demo_r_d3avg databases. http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database.
Both sampling schemes show little relative bias of the total (Table 1 and Table 3). This is not surprising, as the bootstrap estimators are unbiased. The bias shown is the results of the randomness of the bootstrap variance estimators, which has two sources: the initial sampling and the bootstrap itself. The number of replicate weights (500 in the simulations below) could be increased, at the expense of additional computational time. Nevertheless, some variance estimators, such as for total area, are highly skewed, as the presence of a few outliers complicates the convergence of the central limit theorem.

As expected, relative RMSEs are low when the variable of interest is correlated with the size variable (total population), e.g. female population or number of births (Table 2 and Table 4). With uncorrelated variables (e.g. birth and death rates), the RMSEs are higher, and are highest for the total area, which suffers from some outlying values. Comparing the with-replacement pps with RHC pps, the RRMSEs are of similar magnitude, and are slightly lower for RHC. With higher sampling fraction, the advantage of the without-replacement RHC shows off compared to the with-replacement pps sampling. In line with the theory and equation (6), the variance is lower with RHC sampling; the larger the sample size, the bigger the gain. Interestingly, the magnitude of the gain translates in a similar way to non-linear estimators, e.g. the median or the first decile.

4. Conclusion

In this paper we extended the rescaled bootstrap to multistage samples where the first stage follows the Rao-Hartley-Cochran pps sampling. Due to its implementation in terms of replicate weights, this approach is transparent to users, since only the widely-implemented use of replicate weights is needed. Extensions to other pps sampling methods would increase the applicability of the method.
Table 1: Relative bias for the estimate of the variance of the total

<table>
<thead>
<tr>
<th></th>
<th>pps with replacement</th>
<th>Rao Hartley Cochran pps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50 100 200</td>
<td>50 100 200</td>
</tr>
<tr>
<td>Area</td>
<td>4.5% 0.1% -1.2%</td>
<td>1.2% -1.7% -2.7%</td>
</tr>
<tr>
<td>Birth rate</td>
<td>-0.8% 1.8% -0.1%</td>
<td>-1.4% 0.8% -1.6%</td>
</tr>
<tr>
<td>Death rate</td>
<td>-0.6% 1.2% 0.4%</td>
<td>-1.3% 0.1% 1.1%</td>
</tr>
<tr>
<td>Female population</td>
<td>0.0% 0.5% 0.0%</td>
<td>-0.1% -0.1% -0.2%</td>
</tr>
<tr>
<td>Male population</td>
<td>0.0% 0.5% -0.1%</td>
<td>-0.1% -0.1% 0.1%</td>
</tr>
<tr>
<td>Births</td>
<td>-0.3% 0.2% 0.0%</td>
<td>0.0% -0.3% 0.2%</td>
</tr>
<tr>
<td>Deaths</td>
<td>-0.3% 0.4% 0.3%</td>
<td>0.1% 0.3% 1.0%</td>
</tr>
</tbody>
</table>

Table 2: Relative root mean square error for the estimate of the variance of the total

<table>
<thead>
<tr>
<th></th>
<th>pps with replacement</th>
<th>Rao Hartley Cochran pps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50 100 200</td>
<td>50 100 200</td>
</tr>
<tr>
<td>Area</td>
<td>301% 207% 146%</td>
<td>290% 201% 131%</td>
</tr>
<tr>
<td>Birth rate</td>
<td>99% 80% 53%</td>
<td>96% 74% 49%</td>
</tr>
<tr>
<td>Death rate</td>
<td>83% 60% 43%</td>
<td>78% 57% 41%</td>
</tr>
<tr>
<td>Female population</td>
<td>30% 22% 16%</td>
<td>30% 21% 15%</td>
</tr>
<tr>
<td>Male population</td>
<td>30% 22% 16%</td>
<td>30% 21% 15%</td>
</tr>
<tr>
<td>Births</td>
<td>46% 33% 24%</td>
<td>46% 31% 22%</td>
</tr>
<tr>
<td>Deaths</td>
<td>23% 17% 13%</td>
<td>24% 17% 12%</td>
</tr>
</tbody>
</table>

Table 3: Relative bias for the estimate of the variance of the median

<table>
<thead>
<tr>
<th></th>
<th>pps with replacement</th>
<th>Rao Hartley Cochran pps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50 100 200</td>
<td>50 100 200</td>
</tr>
<tr>
<td>Area</td>
<td>17.2% 19.8% 14.9%</td>
<td>20.3% 19.4% 15.0%</td>
</tr>
<tr>
<td>Birth rate</td>
<td>3.3% 4.9% 2.6%</td>
<td>4.9% 4.6% 4.9%</td>
</tr>
<tr>
<td>Death rate</td>
<td>18.9% 20.7% 12.5%</td>
<td>19.1% 17.0% 15.0%</td>
</tr>
<tr>
<td>Female population</td>
<td>14.9% 14.0% 9.6%</td>
<td>17.6% 13.4% 10.2%</td>
</tr>
<tr>
<td>Male population</td>
<td>16.1% 15.1% 9.4%</td>
<td>19.0% 14.3% 10.3%</td>
</tr>
<tr>
<td>Births</td>
<td>17.3% 12.0% 7.0%</td>
<td>19.5% 12.0% 10.1%</td>
</tr>
<tr>
<td>Deaths</td>
<td>11.5% 12.4% 9.2%</td>
<td>13.5% 12.1% 9.0%</td>
</tr>
</tbody>
</table>
Table 4: Relative root mean square error for the estimate of the variance of the median

<table>
<thead>
<tr>
<th></th>
<th>pps with replacement</th>
<th>Rao Hartley Cochran pps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Area</td>
<td>89%</td>
<td>84%</td>
</tr>
<tr>
<td>Birth rate</td>
<td>73%</td>
<td>83%</td>
</tr>
<tr>
<td>Death rate</td>
<td>107%</td>
<td>131%</td>
</tr>
<tr>
<td>Female population</td>
<td>75%</td>
<td>65%</td>
</tr>
<tr>
<td>Male population</td>
<td>79%</td>
<td>64%</td>
</tr>
<tr>
<td>Births</td>
<td>79%</td>
<td>64%</td>
</tr>
<tr>
<td>Deaths</td>
<td>76%</td>
<td>66%</td>
</tr>
</tbody>
</table>

Figure 2: Box plot of the relative variance. (a) Totals; (b) Medians

Figure 3: Box plot of the relative variance of the 10th percentile, over different variables and sample sizes
References


