South African Statistical Journal
2018, Vol. 52, No. 1, 73–92

A NONPARAMETRIC POINT ESTIMATION TECHNIQUE USING THE m-OUT-OF-n BOOTSTRAP

Leonard Santana [1]
Department of Statistics, North-West University, Potchefstroom, South Africa
e-mail: *Leonard.Santana@nwu.ac.za*

Jan W. H. Swanepoel
Department of Statistics, North-West University, Potchefstroom, South Africa

We investigate a method which can be used to improve an existing point estimator by a modification of the estimator and by using the m-out-of-n bootstrap. The estimation method used, known as bootstrap robust aggregating (or BRAGGing) in the literature, will be applied in general to the estimators that satisfy the smooth function model (for example, a mean, a variance, a ratio of means or variances, or a correlation coefficient), and then specifically to an estimator for the population mean. BRAGGing estimators based on both a naive and corrected version of the m-out-of-n bootstrap will be considered. We conclude with proposed data-based choices of the resample size, m, as well as Monte-Carlo studies illustrating the performance of the estimators when estimating the population mean for various distributions.

Key words: BRAGGing, Cornish-Fisher expansion, m-out-of-n bootstrap, Monte-Carlo study, Point estimation, Resample size, Smooth function model.

1. Introduction

A modification to the traditional bootstrap proposed by, among others, Bickel and Freedman (1981), Bretagnolle (1983), and Swanepoel (1986), has been shown to remedy many of the inconsistency problems associated with the bootstrap's non-regular cases as discussed in, for example, Shao and Tu (1995). This method, called the m-out-of-n bootstrap, has also been shown to be useful not only in cases where the traditional bootstrap fails, but also in cases where it is valid (see, e.g., Lee, 1999; Chung and Lee, 2001; Janssen, Swanepoel and Veraverbeke, 2001; Arcones, 2003; Cheung and Lee, 2005). Papers related to the selection of the resample size include Sakov and Bickel (1999), Chung and Lee (2001), Allison, Santana and Swanepoel (2011), and Alin, Martin, Beyaztas and Pathak (2017).

We will investigate a method which can be used to improve the performance of an existing point estimator by a modification of the estimator and by using the m-out-of-n bootstrap. The estimation method which we will use has come to be known as bootstrap robust aggregating (or BRAGGing) in the literature, which is a robust version of bootstrap aggregating (or BAGGing). The name BAGGing was coined in the field of machine learning by Breiman (1996), but the concept applied to statistical point estimators is slightly older, having first appeared in Swanepoel (1988) and Swanepoel (1990)

---
[1] Corresponding author.
MSC2010 subject classifications. 62G05, 62G09.

(these papers referred to the technique as *an approximating functional approach*). In recent years this topic has been the recipient of renewed interest with articles concerning BAGGing being published by Bühlmann (2002), Buja and Stuetzle (2006) and Croux, Joossens and Lemmens (2007), and articles concerning BRAGGing being published by Bühlmann (2003) and Berrendero (2007).

This paper is organized as follows. In Section 2 we will briefly discuss the two core techniques employed in this article namely the m-out-of-n bootstrap and BRAGGing, focusing primarily on the work done which appeared in Swanepoel (1988) and Berrendero (2007). The view held by Swanepoel (1988) allows a more general approach to the work. In Section 3 we will look at a simple variant of these BRAGGing estimators developed by making use of *corrected* m-out-of-n bootstrap concepts which will also be briefly discussed. Since these estimation techniques are based on m-out-of-n bootstrap ideas, we will be interested in data-based choices of the resample size m. Section 4 begins the theoretical development of the optimal sample size for general statistics satisfying the smooth function model through the use of Cornish-Fisher expansions. For illustrative purposes, we develop in Section 5 estimators for these theoretical choices of the resample size when estimating the population mean. In Section 6 we provide the results of a Monte-Carlo simulation to empirically compare the various techniques considered in the article.

2. The m-out-of-n bootstrap and BRAGGing

The modifications of the bootstrap discussed here typically involve sampling with replacement fewer than n observations from $X_1, X_2, \ldots, X_n$. The notation that will be employed to denote the resulting bootstrap sample is $X_1^*, X_2^*, \ldots, X_m^*$, where $m \leq n$, and we refer to this bootstrap procedure as the *m-out-of-n bootstrap*. We will also distinguish between a *naive* application of the m-out-of-n bootstrap and a *corrected* version of the m-out-of-n bootstrap.

The purpose of the m-out-of-n bootstrap is twofold (Bickel and Sakov, 2002):

- Obtaining consistency when the traditional bootstrap is inconsistent.
- When the traditional bootstrap is consistent, then the m-out-of-n bootstrap is used to attain equivalent behaviour, but with second (or higher) order accuracy, with reduced computational time (Bickel and Yahav, 1988; Bickel, Götze and van Zwet, 1997; Beran, 1997; Sakov, 1998).

In this article we apply the m-out-of-n bootstrap in the point estimation of a parameter using a technique called bootstrap robust aggregating or *BRAGGing*. The definition of a BRAGGing point estimator of a parameter as given in Swanepoel (1988) is now briefly described.

The BRAGGing point estimator

Let θ be a parameter of interest which can be expressed as some functional t of an unknown distribution F, i.e., $\theta = t(F)$. Suppose also that $t(F)$ can be approximated by a sequence of functionals $t_m(F)$, i.e., $t_m(F) \approx t(F)$, with the approximation becoming increasingly accurate as $m \to \infty$. The proposed estimator for θ makes use of this functional sequence and the empirical distribution function F_n to create the plug-in estimator

$$\tilde{\theta} = t_m(F_n).$$

As discussed in Swanepoel (1990) and Berrendero (2007), one possibility for this estimator is

$$\tilde{\theta}_{brag} = t_m(F_n) = \mathrm{Med}^*(\widehat{\theta}_m^*),$$

where $\widehat{\theta}_n \equiv \widehat{\theta}_n(X_1, X_2, \ldots, X_n)$ is some preliminary estimator for θ and $\widehat{\theta}^*_m \equiv \widehat{\theta}_m(X^*_1, X^*_2, \ldots, X^*_m)$. Med* refers to the median over the conditional probability law of $X^*_1, X^*_2, \ldots, X^*_m$ given $X_1, X_2, \ldots, X_n$. This estimator is known as the BRAGGing estimator in the literature, and it can easily be approximated by a simple Monte-Carlo simulation.

3. A variant of the BRAGGing estimator

The BRAGGing estimator discussed in the previous section was based on the naive application of the m-out-of-n bootstrap to the median of $\widehat{\theta}_n$. Using a corrected version of the m-out-of-n bootstrap makes it possible to derive a new version of this BRAGGing estimator.

- *First version of the estimator:* To distinguish between the new BRAGGing estimator and the original one we will adopt a new notation for these estimators. Let the original BRAGGing estimator, $\tilde{\theta}_{brag}$, be renamed $\tilde{\theta}_{brag,1}$.

- *Second version of the estimator:* The new version of the BRAGGing estimator, denoted by $\tilde{\theta}_{brag,2}$, is derived by first explaining the procedure for conducting a corrected version of the m-out-of-n bootstrap:

  Assume that $n^{\alpha}(\widehat{\theta}_n - \theta)$ has a non-degenerate limiting distribution, where the normalizing constant n^{α} is known with $\alpha > 0$. Then rewrite the statistic $\widehat{\theta}_n$ as

$$\widehat{\theta}_n = \frac{1}{n^{\alpha}}\left[n^{\alpha}(\widehat{\theta}_n - \theta)\right] + \theta. \tag{1}$$

  When this technique is coupled with the m-out-of-n bootstrap we will refer to the result as the *corrected* m-out-of-n bootstrap.

  For ease of exposition we consider only the case where $\alpha = 0.5$ (a value which is appropriate for estimators related to the smooth function model), and so applying the m-out-of-n bootstrap to the median of the expression given in (1) we get the following estimator

$$\tilde{\theta}_{brag,2} := \frac{1}{\sqrt{n}}\text{Med}^*\left(\sqrt{m}(\widehat{\theta}^*_m - \widehat{\theta}_n)\right) + \widehat{\theta}_n = \sqrt{\frac{m}{n}}\tilde{\theta}_{brag,1} + \left(1 - \sqrt{\frac{m}{n}}\right)\widehat{\theta}_n. \tag{2}$$

  Note that this corrected m-out-of-n estimator is a convex combination between the original BRAGGing estimator, $\tilde{\theta}_{brag,1}$, and the estimator $\widehat{\theta}_n$.

4. The choice of m for a general statistic associated with the smooth function model

We will now consider the various ways of selecting an optimal value of m when using the BRAGGing technique applied to estimators associated with the smooth function model. The derivation of these choices of m will make use of Cornish-Fisher expansions discussed in detail in Hall (1992) and Chung and Lee (2001). The resulting theoretical choices will then facilitate the development of data-dependent choices of m.

4.1 Cornish-Fisher expansion of a general statistic

Considering the smooth function model, let θ be defined as some parameter which is a function of a d-dimensional mean $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and where $\boldsymbol{X}$ is a d-dimensional column vector whose i^{th} component is denoted by $X^{(i)}$. In other words, we have that $\theta = g(\boldsymbol{\mu})$, where g is defined such that $g : \mathbb{R}^d \to \mathbb{R}$. An estimator for θ, based on a random sample $\boldsymbol{X}_j = (X_j^{(1)}, X_j^{(2)}, \ldots, X_j^{(d)})$, $j = 1, 2, \ldots, n$, from $\boldsymbol{X}$, is then the simple plug-in estimator given by $\widehat{\theta}_n = g(\bar{\boldsymbol{X}}_n)$, where $\bar{\boldsymbol{X}}_n$ is a d-dimensional vector defined as

$$\bar{\boldsymbol{X}}_n = \left(\frac{1}{n} \sum_{j=1}^{n} X_j^{(1)}, \frac{1}{n} \sum_{j=1}^{n} X_j^{(2)}, \ldots, \frac{1}{n} \sum_{j=1}^{n} X_j^{(d)} \right)^T . \tag{3}$$

Define the standardized version of $\widehat{\theta}_n$ as the statistic T_n in the following way:

$$T_n := \frac{\sqrt{n} \left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu}) \right)}{h(\boldsymbol{\mu})},$$

where $\{h(\boldsymbol{\mu})\}^2$ is the asymptotic variance of $\sqrt{n} g(\bar{\boldsymbol{X}}_n)$. We assume therefore that T_n satisfies the assumptions of the smooth function model. The m-out-of-n bootstrap version of this statistic is then

$$T_m^* = \frac{\sqrt{m} \left(g(\bar{\boldsymbol{X}}_m^*) - g(\bar{\boldsymbol{X}}_n)) \right)}{h(\bar{\boldsymbol{X}}_n)}.$$

The Cornish-Fisher expansion of the median of T_m^* is then (the details of this expansion as well as the definitions of $\widehat{k}_{3,1}$ and $\widehat{k}_{1,2}$ can be found in Hall (1992) and in Appendix A of this article)

$$\sqrt{m} \left(\tilde{\theta}_{brag,1} - g(\bar{\boldsymbol{X}}_n)) \right) / h(\bar{\boldsymbol{X}}_n) = -m^{-1/2} \left[\frac{1}{6} \widehat{k}_{3,1} - \widehat{k}_{1,2} \right] + O_p(m^{-3/2}).$$

Solving for $\tilde{\theta}_{brag,1}$ we get

$$\tilde{\theta}_{brag,1} = g(\bar{\boldsymbol{X}}_n) - m^{-1} h(\bar{\boldsymbol{X}}_n) \left[\frac{1}{6} \widehat{k}_{3,1} - \widehat{k}_{1,2} \right] + O_p(m^{-2}). \tag{4}$$

The expression in (4) can now be used to determine an asymptotically optimal choice of the sample size m. From the leading terms in (4) we can therefore approximate $\tilde{\theta}_{brag,1}$ by $\tilde{\theta}_{brag,1}^A$, where

$$\tilde{\theta}_{brag,1}^A := g(\bar{\boldsymbol{X}}_n) - m^{-1} h(\bar{\boldsymbol{X}}_n) \left[\frac{1}{6} \widehat{k}_{3,1} - \widehat{k}_{1,2} \right]. \tag{5}$$

Therefore, the rule for selecting m based on $\tilde{\theta}_{brag,1}^A$ will involve finding the m value that minimizes the mean squared error (MSE) of this estimator.

4.2 The optimal choice of m when estimating θ using $\tilde{\theta}_{brag,1}^A$

We will now obtain the MSE of $\tilde{\theta}_{brag,1}^A$ using the definition given in (5):

$$\begin{aligned}
q(m) &:= \mathrm{MSE}(\tilde{\theta}_{brag,1}^A) \\
&= \mathrm{E} \left\{ \left(g(\bar{\boldsymbol{X}}_n) - m^{-1} h(\bar{\boldsymbol{X}}_n) \left[\frac{1}{6} \widehat{k}_{3,1} - \widehat{k}_{1,2} \right] - g(\boldsymbol{\mu}) \right)^2 \right\} \\
&= \mathrm{E} \left\{ \left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu}) \right)^2 \right\} - 2m^{-1} \mathrm{E} \left\{ \left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu}) \right) h(\bar{\boldsymbol{X}}_n) \left[\frac{1}{6} \widehat{k}_{3,1} - \widehat{k}_{1,2} \right] \right\}
\end{aligned}$$

$$+ m^{-2}\mathrm{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]^2\right\}.$$

The first derivative of $q(m)$ is

$$\frac{dq(m)}{dm} = 2m^{-2}\mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]\right\} - 2m^{-3}\mathrm{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]^2\right\}.$$

Setting $dq(m)/dm$ to zero and solving for m we get:

$$m_0 = \frac{\mathrm{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]^2\right\}}{\mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]\right\}}. \tag{6}$$

Note that the global optimal solution for m given in expression (6) can be negative, which can only occur if the denominator of (6) is negative, i.e., if $\Delta := \mathrm{E}\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}]\} < 0$. Naturally, a negative sample size is infeasible, so we restrict our optimisation procedure to only consider those solutions where $m \geq 1$. However, since the sign of the optimal solution is entirely dictated by the sign of Δ, one can argue the following: if $\Delta > 0$, then the optimal solution of m is given by the expression in (6), but if $\Delta < 0$, then the optimal solution is $m_0 = n$. The latter solution is obtained by noting from $dq(m)/dm$ that $q(m)$ is monotone decreasing for $m > 0$ when $\Delta < 0$, and so the optimal solution for m will occur at the upper bound of m. Therefore, with these restrictions in place, one can express the feasible optimal solution as follows

$$m_1 := \begin{cases} \min\left\{\max(n_0, m_0), n\right\}, & \text{if } \Delta > 0 \\ n, & \text{if } \Delta \leq 0, \end{cases} \tag{7}$$

where $n_0 \geq 1$ is a prescribed lower bound for the values of m_1. Note also that the solution for the optimal sample size is truncated to ensure that it lies between n_0 and n. We will estimate m_1 by linear approximations and bootstrap methods.

4.3 The optimal choice of m when estimating θ using $\tilde{\theta}^A_{brag,2}$

We now derive the optimal resample size, m, when performing the estimation using $\tilde{\theta}^A_{brag,2}$, which is defined by replacing $\tilde{\theta}_{brag,1}$ with $\tilde{\theta}^A_{brag,1}$ in equation (2):

$$\tilde{\theta}^A_{brag,2} = \sqrt{\frac{m}{n}}\tilde{\theta}^A_{brag,1} + \left(1 - \sqrt{\frac{m}{n}}\right) g(\bar{\boldsymbol{X}}_n). \tag{8}$$

We begin by deriving the MSE of $\tilde{\theta}^A_{brag,2}$:

$$\begin{aligned} r(m) &:= \mathrm{MSE}(\tilde{\theta}^A_{brag,2}) \\ &= \mathrm{E}\left\{\left(\sqrt{\frac{m}{n}}\tilde{\theta}^A_{brag,1} + \left(1 - \sqrt{\frac{m}{n}}\right) g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)^2\right\} \\ &= \mathrm{E}\Bigg\{\Bigg(\sqrt{\frac{m}{n}}\left\{g(\bar{\boldsymbol{X}}_n) - m^{-1}h(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]\right\} + \left(1 - \sqrt{\frac{m}{n}}\right) g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\Bigg)^2\Bigg\} \end{aligned}$$

$$= \mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)^2\right\} - 2m^{-\frac{1}{2}}n^{-\frac{1}{2}}\mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]\right\}$$
$$+ m^{-1}n^{-1}\mathrm{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]^2\right\}.$$

The first derivative of $r(m)$ is

$$\frac{dr(m)}{dm} = m^{-\frac{3}{2}}n^{-\frac{1}{2}}\mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]\right\}$$
$$- m^{-2}n^{-1}\mathrm{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]^2\right\}.$$

Note that the above expression is non-positive when $\Delta = \mathrm{E}\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}]\}$ is less than or equal to zero. Therefore, since this implies that $r(m)$ is a monotone non-increasing function for positive m, the minimum value for $r(m)$ is obtained at $m = \infty$. However, since we do not want to choose $m > n$, the optimal ‘practical’ solution is simply $m = n$ in the case where $\Delta \leq 0$. In the case where $\Delta > 0$, we can obtain the optimal positive solution for m by setting $dr(m)/dm$ to zero and solve for m, yielding:

$$m_2 := \frac{1}{n}\left[\frac{\mathrm{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]^2\right\}}{\mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2}\right]\right\}}\right]^2 = \frac{1}{n}\left[m_0\right]^2. \tag{9}$$

A solution for m which incorporates the solution derived in (9) as well as the practical considerations discussed above when $\Delta \leq 0$, is

$$m_3 := \begin{cases} \min\left\{\max(n_0, m_2), n\right\}, & \text{if } \Delta > 0 \\ n, & \text{if } \Delta \leq 0, \end{cases} \tag{10}$$

where $n_0 \geq 1$ is once again some prescribed lower bound for the values of m_2, and Δ is defined as above.

We now illustrate the above techniques by considering the example of estimating the population mean.

5. Estimation of the population mean

As a means of illustrating the techniques already derived we will now focus on the estimation of the population mean, μ. However, it should be emphasized that it is possible to apply these techniques to any estimator derived from the smooth function model, since Hall (1992) obtained Cornish-Fisher expansions for these statistics.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations and let $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. Set $\mathbf{x} = (x^{(1)}, x^{(2)})^T$, where $x^{(i)}$ is the i^{th} component of $\mathbf{x}$. Then define $g(\mathbf{x}) = x^{(1)}$ and $h^2(\mathbf{x}) = x^{(2)} - (x^{(1)})^2$. Now, recall that $\boldsymbol{X}_j = (X_j^{(1)}, X_j^{(2)})$, for $d = 2$, and choosing $X_j^{(1)} = X_j$ and $X_j^{(2)} = X_j^2$ we find $g(\bar{\boldsymbol{X}}_n) = \bar{X}_n$, and $h^2(\bar{\boldsymbol{X}}_n) = n^{-1}\sum_{i=1}^{n} X_i^2 - (n^{-1}\sum_{i=1}^{n} X_i)^2 =: \widehat{\mu}_2$, where $\bar{\boldsymbol{X}}_n$ is defined as in equation (3). Finally, the quantities stated in equation (5) are then given by (see Hall, 1992, pp. 52–55)

$$g(\bar{\boldsymbol{X}}_n) = \bar{X}_n, \qquad g(\boldsymbol{\mu}) = \mu, \qquad h^2(\bar{\boldsymbol{X}}_n) = \widehat{\mu}_2, \qquad \widehat{k}_{3,1} = \widehat{\kappa}_3, \qquad \widehat{k}_{1,2} = 0.$$

Therefore,

$$\frac{1}{6}\widehat{k}_{3,1} - \widehat{k}_{1,2} = \frac{1}{6}\widehat{\kappa}_3 = \frac{1}{6}\frac{\widehat{\mu}_3}{\widehat{\mu}_2^{3/2}},$$

where $\widehat{\kappa}_3 = \widehat{\mu}_3/\widehat{\mu}_2^{3/2}$ and $\widehat{\mu}_\nu = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^\nu$, so the approximate Cornish-Fisher expansion of the median of the bootstrap sample mean is now given by (see (5)):

$$\tilde{\theta}^A_{brag,1} = \bar{X}_n - \frac{1}{6}m^{-1}\widehat{\mu}_2^{1/2}\widehat{\kappa}_3. \tag{11}$$

Note that the corresponding expression for $\tilde{\theta}^A_{brag,2}$ can then be defined by substituting (11) into (8).

The approximate optimal choice of m in this case is then obtained by substituting the appropriate values into (6). This gives us the expression

$$m_0 = \frac{1}{6}\cdot\frac{\mathrm{E}\left[\widehat{\kappa}_3^2\widehat{\mu}_2\right]}{\mathrm{E}\left[\left(\bar{X}_n - \mu\right)\widehat{\kappa}_3\widehat{\mu}_2^{1/2}\right]}.$$

Now, if we define $Y_i = X_i - \mathrm{E}(X_i)$ we have the following form of m_0:

$$m_0 = \frac{1}{6}\cdot\frac{\mathrm{E}\left[\widehat{\mu}_3^2/\widehat{\mu}_2^2\right]}{\mathrm{E}\left[\bar{Y}_n\widehat{\mu}_3/\widehat{\mu}_2\right]}. \tag{12}$$

5.1 Data-based choices of m_0

Using (12) as a theoretical starting point we will now attempt to estimate m_0 using various strategies. The strategies which will be followed are:

1. A Taylor series expansion of the numerator and denominator of (12) (individually) followed by the estimation of the products of population moments using a bias correction approach (to order $1/n^2$).

2. An unbiased estimator of the numerator in (12) and then applying a Taylor series expansion to the denominator followed by a bias correction approach for the estimation of population moment product terms (to the order $1/n$ and $1/n^2$).

3. A bootstrap approximation of the expression in (12).

Techniques 1 and 2 rely on Taylor series expansions of the numerator and denominator in (12), derived in Appendix B, given by (20) and (22) respectively. Combining these expansions of the numerator and the denominator terms in the expression (12) and simplifying such that the terms $1/\mu_2^2$ and $1/\mu_2^4$ no longer appear, we get the following approximation of m_0:

$$m_0^A = \frac{n}{6}\cdot\frac{\tilde{A}(\tilde{\boldsymbol{\mu}}) + \frac{1}{n}\tilde{B}(\tilde{\boldsymbol{\mu}})}{\tilde{C}(\tilde{\boldsymbol{\mu}}) + \frac{1}{n}\tilde{D}(\tilde{\boldsymbol{\mu}})}, \tag{13}$$

where

$$\tilde{A}(\tilde{\boldsymbol{\mu}}) = \mu_2^2\mu_3^2,$$
$$\tilde{B}(\tilde{\boldsymbol{\mu}}) = \mu_2^2\mu_6 + 9\mu_2^5 + 8\mu_2^2\mu_3^2 - 6\mu_2^3\mu_4 + 3\mu_3^2\mu_4 - 4\mu_2\mu_3\mu_5,$$
$$\tilde{C}(\tilde{\boldsymbol{\mu}}) = \mu_2^3\mu_4 - 3\mu_2^5 - \mu_2^2\mu_3^2,$$
$$\tilde{D}(\tilde{\boldsymbol{\mu}}) = 3\mu_2^3\mu_4 - \mu_2^2\mu_6 - 3\mu_2^5 + 3\mu_2\mu_3\mu_5 + \mu_2\mu_4^2 - 3\mu_3^2\mu_4 - 5\mu_2^2\mu_3^2,$$

and $\tilde{\boldsymbol{\mu}}$ is the vector consisting of the products of population moments $\{\mu_2^2\mu_3^2, \mu_2^2\mu_6, \mu_2^3\mu_4, \mu_2^5, \mu_3^2\mu_4, \mu_2\mu_3\mu_5, \mu_2\mu_4^2\}$.

Remark. Notice that m_0^A can take on a wide range of values, depending on the complexity of the underlying distribution (i.e., it is dependent on the behaviour of the central moments).

Expression (13) allows us to formally define the proposed *estimators* for m_0 defined in (12). We will now present a list of various estimators for the resample size when estimating the population mean with the statistic $\tilde{\theta}^A_{brag,1}$. The letters in parentheses indicate the abbreviations which will be used to represent these estimators.

5.1.1 The bias corrected (BC) estimator

The first technique makes use of the expression for m_0^A given in (13). Instead of naively substituting sample moments for population moments we will use estimators for the product of population moments that are corrected for bias up to order $1/n^2$ and $1/n^3$. The resulting estimator for m_0^A is denoted by $\widehat{m}_{0,BC}$.

We will now present the different bias corrected products of sample moments. The estimators corrected by removing the $1/n$ order bias terms are denoted by the subscript BC1, and those corrected by removing the $1/n$ and $1/n^2$ bias terms are denoted by the subscript BC2. The products of population moments that we are required to correct for bias in this estimator are $\mu_2^2\mu_3^2$, $\mu_2^2\mu_6$, $\mu_2^3\mu_4$, μ_2^5, $\mu_3^2\mu_4$, $\mu_2\mu_3\mu_5$, and $\mu_2\mu_4^2$.

Remark. The method of bias correction used to obtain these estimators consists of three parts:

1. Obtain the naive plug-in estimator for the product of population moments. For example, if we wish to estimate $\mu_2^2\mu_3^2$ then we obtain $\widehat{\mu}_2^2\widehat{\mu}_3^2$.
2. Determine the expected value of this naive estimator up to order n^{-2}. For example, the expected value of $\widehat{\mu}_2^2\widehat{\mu}_3^2$ is

$$\begin{aligned}\mathrm{E}(\widehat{\mu}_2^2\widehat{\mu}_3^2) = \mu_2^2\mu_3^2 &+ \frac{1}{n}\{\mu_2^2\mu_6 + 4\mu_2\mu_3\mu_5 + \mu_3^2\mu_4 - 26\mu_2^2\mu_3^2 - 6\mu_2^3\mu_4 + 9\mu_2^5\} \\ &+ \frac{1}{n^2}\{2\mu_2\mu_8 + \mu_4\mu_6 - 23\mu_2^2\mu_6 + 2\mu_3\mu_7 + 2\mu_5^2 - 74\mu_2\mu_3\mu_5 - 31\mu_3^2\mu_4 \\ &+ 354\mu_2^2\mu_3^2 + 174\mu_2^3\mu_4 - 22\mu_2\mu_4^2 - 180\mu_2^5\} + O(n^{-3}).\end{aligned}$$

   Obtaining expressions for these expected values is a *very* long and *incredibly* tedious process. All of the relevant expected values have been derived by the authors and are available on request.

3. To remove the first order bias terms from the estimator we simply subtract the plug-in estimator of the $1/n$ order terms from the original estimator. The resulting estimator's bias is of the order $1/n^2$ and is referred to as the BC1 type. To remove the first and second order bias terms we subtract the plug-in estimators of both the $1/n$ and $1/n^2$ order terms from the estimator. The resulting estimator's bias is then of the order $1/n^3$ and is referred to as the BC2 type.

   For example, to correct the naive estimator $\widehat{\mu}_2^2\widehat{\mu}_3^2$ by removing the $1/n$ order terms, we subtract the plug-in estimator of the $1/n$ terms in $\mathrm{E}(\widehat{\mu}_2^2\widehat{\mu}_3^2)$ from $\widehat{\mu}_2^2\widehat{\mu}_3^2$. Since we remove the order $1/n$ bias terms, this estimator is of the BC1 type, and will be denoted by

$$\widehat{\left(\mu_2^2\mu_3^2\right)}_{BC1} := \widehat{\mu}_2^2\widehat{\mu}_3^2 - \frac{1}{n}\{\widehat{\mu}_2^2\widehat{\mu}_6 + 4\widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 + \widehat{\mu}_3^2\widehat{\mu}_4 - 26\widehat{\mu}_2^2\widehat{\mu}_3^2 - 6\widehat{\mu}_2^3\widehat{\mu}_4 + 9\widehat{\mu}_2^5\}.$$

   A similar expression can be obtained for the BC2 type estimator.

The *BC1* estimators for products of population moments: The relevant corrected estimators with the order $1/n$ bias terms removed (as explained above) are then:

$$\begin{aligned}
\widehat{\left(\mu_2^2\mu_3^2\right)}_{BC1} &:= \widehat{\mu}_2^2\widehat{\mu}_3^2 - \frac{1}{n}\{\widehat{\mu}_2^2\widehat{\mu}_6 + 4\widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 + \widehat{\mu}_3^2\widehat{\mu}_4 - 26\widehat{\mu}_2^2\widehat{\mu}_3^2 - 6\widehat{\mu}_2^3\widehat{\mu}_4 + 9\widehat{\mu}_2^5\},\\
\widehat{\left(\mu_2^2\mu_6\right)}_{BC1} &:= \widehat{\mu}_2^2\widehat{\mu}_6 - \frac{1}{n}\{\widehat{\mu}_4\widehat{\mu}_6 + 2\widehat{\mu}_2\widehat{\mu}_8 - 11\widehat{\mu}_2^2\widehat{\mu}_6 - 12\widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 + 15\widehat{\mu}_2^3\widehat{\mu}_4\},\\
\widehat{\left(\mu_2^3\mu_4\right)}_{BC1} &:= \widehat{\mu}_2^3\widehat{\mu}_4 - \frac{1}{n}\{-13\widehat{\mu}_2^3\widehat{\mu}_4 + 3\widehat{\mu}_2^2\widehat{\mu}_6 + 3\widehat{\mu}_2\widehat{\mu}_4^2 - 12\widehat{\mu}_2^2\widehat{\mu}_3^2 + 6\widehat{\mu}_2^5\},\\
\widehat{\left(\mu_2^5\right)}_{BC1} &:= \widehat{\mu}_2^5 - \frac{1}{n}\{-15\widehat{\mu}_2^5 + 10\widehat{\mu}_2^3\widehat{\mu}_4\},\\
\widehat{\left(\mu_3^2\mu_4\right)}_{BC1} &:= \widehat{\mu}_3^2\widehat{\mu}_4 - \frac{1}{n}\{-21\widehat{\mu}_3^2\widehat{\mu}_4 + 2\widehat{\mu}_3\widehat{\mu}_7 + \widehat{\mu}_4\widehat{\mu}_6 + 30\widehat{\mu}_2^2\widehat{\mu}_3^2 - 18\widehat{\mu}_2\widehat{\mu}_4^2 - 18\widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 + 9\widehat{\mu}_2^3\widehat{\mu}_4\},\\
\widehat{\left(\mu_2\mu_3\mu_5\right)}_{BC1} &:= \widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 - \frac{1}{n}\{\widehat{\mu}_3\widehat{\mu}_7 + \widehat{\mu}_2\widehat{\mu}_8 - 15\widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 + 15\widehat{\mu}_2^3\widehat{\mu}_4 + \widehat{\mu}_5^2 - 5\widehat{\mu}_3^2\widehat{\mu}_4 - 5\widehat{\mu}_2\widehat{\mu}_4^2\\
&\quad + 10\widehat{\mu}_2^2\widehat{\mu}_3^2 - 3\widehat{\mu}_2^2\widehat{\mu}_6\},\\
\widehat{\left(\mu_2\mu_4^2\right)}_{BC1} &:= \widehat{\mu}_2\widehat{\mu}_4^2 - \frac{1}{n}\{-12\widehat{\mu}_2\widehat{\mu}_4^2 + 2\widehat{\mu}_4\widehat{\mu}_6 + \widehat{\mu}_2\widehat{\mu}_8 - 8\widehat{\mu}_3^2\widehat{\mu}_4 - 8\widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 + 12\widehat{\mu}_2^3\widehat{\mu}_4 + 16\widehat{\mu}_2^2\widehat{\mu}_3^2\}.
\end{aligned}$$

The *BC2* estimators for products of population moments: The relevant corrected estimators with the order $1/n$ and $1/n^2$ bias terms removed (as explained above) are then:

$$\begin{aligned}
\widehat{\left(\mu_2^2\mu_3^2\right)}_{BC2} &:= \widehat{\mu}_2^2\widehat{\mu}_3^2 - \frac{1}{n}\{\widehat{\mu}_2^2\widehat{\mu}_6 + 4\widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 + \widehat{\mu}_3^2\widehat{\mu}_4 - 26\widehat{\mu}_2^2\widehat{\mu}_3^2 - 6\widehat{\mu}_2^3\widehat{\mu}_4 + 9\widehat{\mu}_2^5\}\\
&\quad - \frac{1}{n^2}\{2\widehat{\mu}_2\widehat{\mu}_8 + \widehat{\mu}_4\widehat{\mu}_6 - 23\widehat{\mu}_2^2\widehat{\mu}_6 + 2\widehat{\mu}_3\widehat{\mu}_7 + 2\widehat{\mu}_5^2 - 74\widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 - 31\widehat{\mu}_3^2\widehat{\mu}_4\\
&\quad + 354\widehat{\mu}_2^2\widehat{\mu}_3^2 + 174\widehat{\mu}_2^3\widehat{\mu}_4 - 22\widehat{\mu}_2\widehat{\mu}_4^2 - 180\widehat{\mu}_2^5\},\\
\widehat{\left(\mu_2^3\mu_4\right)}_{BC2} &:= \widehat{\mu}_2^3\widehat{\mu}_4 - \frac{1}{n}\{-13\widehat{\mu}_2^3\widehat{\mu}_4 + 3\widehat{\mu}_2^2\widehat{\mu}_6 + 3\widehat{\mu}_2\widehat{\mu}_4^2 - 12\widehat{\mu}_2^2\widehat{\mu}_3^2 + 6\widehat{\mu}_2^5\} - \frac{1}{n^2}\{228\widehat{\mu}_2^2\widehat{\mu}_3^2 + 158\widehat{\mu}_2^3\widehat{\mu}_4\\
&\quad - 18\widehat{\mu}_3^2\widehat{\mu}_4 - 30\widehat{\mu}_2\widehat{\mu}_4^2 - 48\widehat{\mu}_2\widehat{\mu}_3\widehat{\mu}_5 - 30\widehat{\mu}_2^2\widehat{\mu}_6 + 3\widehat{\mu}_2\widehat{\mu}_8 + 4\widehat{\mu}_4\widehat{\mu}_6 - 123\widehat{\mu}_2^5\},\\
\widehat{\left(\mu_2^5\right)}_{BC2} &:= \widehat{\mu}_2^5 - \frac{1}{n}\{-15\widehat{\mu}_2^5 + 10\widehat{\mu}_2^3\widehat{\mu}_4\} - \frac{1}{n^2}\{115\widehat{\mu}_2^5 - 110\widehat{\mu}_2^3\widehat{\mu}_4 + 15\widehat{\mu}_2\widehat{\mu}_4^2 + 10\widehat{\mu}_2^2\widehat{\mu}_6 - 60\widehat{\mu}_2^2\widehat{\mu}_3^2\}.
\end{aligned}$$

Finally, the estimator is then given by:

$$\widehat{m}_{0,BC} := \frac{n}{6} \cdot \frac{\tilde{A}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC2}) + \frac{1}{n}\tilde{B}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC1})}{\tilde{C}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC2}) + \frac{1}{n}\tilde{D}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC1})}, \tag{14}$$

where

$$\begin{aligned}
\tilde{A}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC2}) &= \widehat{\left(\mu_2^2\mu_3^2\right)}_{BC2}, \\
\tilde{B}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC1}) &= \widehat{(\mu_2^2\mu_6)}_{BC1} + 9\widehat{(\mu_2^5)}_{BC1} + 8\widehat{\left(\mu_2^2\mu_3^2\right)}_{BC1} - 6\widehat{(\mu_2^3\mu_4)}_{BC1} + 3\widehat{\left(\mu_3^2\mu_4\right)}_{BC1} - 4\widehat{(\mu_2\mu_3\mu_5)}_{BC1}, \\
\tilde{C}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC2}) &= \widehat{(\mu_2^3\mu_4)}_{BC2} - 3\widehat{(\mu_2^5)}_{BC2} - \widehat{\left(\mu_2^2\mu_3^2\right)}_{BC2}, & (15) \\
\tilde{D}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC1}) &= 3\widehat{(\mu_2^3\mu_4)}_{BC1} - \widehat{(\mu_2^2\mu_6)}_{BC1} - 3\widehat{(\mu_2^5)}_{BC1} + 3\widehat{(\mu_2\mu_3\mu_5)}_{BC1} \\
&\quad + \widehat{(\mu_2\mu_4^2)}_{BC1} - 3\widehat{\left(\mu_3^2\mu_4\right)}_{BC1} - 5\widehat{\left(\mu_2^2\mu_3^2\right)}_{BC1}, & (16)
\end{aligned}$$

and $\widehat{\tilde{\boldsymbol{\mu}}}_{BC1}$ is the vector consisting of the bias corrected products of sample moments with the first order bias terms removed, that is,

$$\widehat{\tilde{\boldsymbol{\mu}}}_{BC1} = \left(\widehat{\left(\mu_2^2\mu_3^2\right)}_{BC1}, \widehat{(\mu_2^2\mu_6)}_{BC1}, \widehat{(\mu_2^3\mu_4)}_{BC1}, \widehat{(\mu_2^5)}_{BC1}, \widehat{\left(\mu_3^2\mu_4\right)}_{BC1}, \widehat{(\mu_2\mu_3\mu_5)}_{BC1}, \widehat{(\mu_2\mu_4^2)}_{BC1}\right).$$

Likewise, $\widehat{\tilde{\boldsymbol{\mu}}}_{BC2}$ is the vector consisting of the bias corrected products of sample moments with the first and second order bias terms removed,

$$\widehat{\tilde{\boldsymbol{\mu}}}_{BC2} = \left(\widehat{\left(\mu_2^2\mu_3^2\right)}_{BC2}, \widehat{(\mu_2^2\mu_6)}_{BC2}, \widehat{(\mu_2^3\mu_4)}_{BC2}, \widehat{(\mu_2^5)}_{BC2}, \widehat{\left(\mu_3^2\mu_4\right)}_{BC2}, \widehat{(\mu_2\mu_3\mu_5)}_{BC2}, \widehat{(\mu_2\mu_4^2)}_{BC2}\right).$$

5.1.2 The unbiased numerator and bias corrected denominator (UNBC) estimator

The second estimator makes use of the fact that the ratio $\widehat{\mu}_3^2/\widehat{\mu}_2^2$ (see (12)) is an unbiased estimator of the parameter $\mathrm{E}\left(\widehat{\mu}_3^2/\widehat{\mu}_2^2\right)$. The denominator in (12) is then estimated using the bias correction techniques discussed in the previous method. The estimator is then given by:

$$\widehat{m}_{0,UNBC} := \frac{n}{6} \cdot \frac{\widehat{\mu}_3^2/\widehat{\mu}_2^2}{\tilde{C}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC2}) + \frac{1}{n}\tilde{D}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC1})}, \tag{17}$$

where $\tilde{C}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC2})$, $\tilde{D}(\widehat{\tilde{\boldsymbol{\mu}}}_{BC1})$, $\widehat{\tilde{\boldsymbol{\mu}}}_{BC1}$, and $\widehat{\tilde{\boldsymbol{\mu}}}_{BC2}$ are defined as above.

5.1.3 The bootstrap (BS) estimator

The bootstrap estimator of m_0 will be based on the quantity given in (12). This estimator is derived by estimating the two expected values which appear in (12) by using the bootstrap. The estimator is given by:

$$\widehat{m}_{0,BS} := \frac{1}{6} \cdot \frac{\mathrm{E}^*\left[\widehat{\mu}_3^{*2}/\widehat{\mu}_2^{*2}\right]}{\mathrm{E}^*\left[\bar{Y}_n^*\widehat{\mu}_3^*/\widehat{\mu}_2^*\right]}, \tag{18}$$

where the terms $\widehat{\mu}_j^*$ are the sample moments based on the resampled bootstrap data, $X_1^*, X_2^*, \ldots, X_n^*$, and $Y_i^* = X_i^* - \bar{X}_n$, $i = 1, 2, \ldots, n$. When we apply the bootstrap to calculate (18) we can once again sample fewer than n observations to determine the value. That is, we can use a k-out-of-n bootstrap but, for the sake of simplicity, we will choose $k = n$.

5.2 Estimators for m_1 and m_3

In the preceding sections, various estimators for m_0, defined in (6), were discussed. However, the term m_0 was only required to obtain the expressions for m_1 and m_3 (i.e., the theoretical expressions for the resample size given in (7) and (10)).

Therefore, in the light of the above discussion, the proposed estimators for m_1 and m_3 will now be provided when making use of $\tilde{\theta}_{brag,1}$ and $\tilde{\theta}_{brag,2}$ to estimate the population mean μ.

5.2.1 *Estimator for* m_1

The estimator for m_1, based on the expression given in (7), is given by

$$\widehat{m}_{1,\star} := \begin{cases} \min\left\{\max(n_0, \widehat{m}_{0,\star}), n\right\}, & \text{if } \widehat{\Delta}_\star > 0 \\ n, & \text{if } \widehat{\Delta}_\star \leq 0, \end{cases}$$

where n_0 is the lower bound for the resample size, the term $\widehat{m}_{0,\star}$ is the estimated resample size obtained from either (14), (17), or (18), i.e., $\star$ can be BC, UNBC, or BS, and $\widehat{\Delta}_\star$ represents the denominator for the expressions for $\widehat{m}_{1,\star}$, i.e.,

$$\widehat{\Delta}_{BC} = \widehat{\Delta}_{UNBC} = \tilde{C}(\widehat{\boldsymbol{\mu}}_{BC2}) + \frac{1}{n}\tilde{D}(\widehat{\boldsymbol{\mu}}_{BC1})$$

and

$$\widehat{\Delta}_{BS} = \mathrm{E}^*\left[\bar{Y}_n^*\widehat{\mu}_3^*/\widehat{\mu}_2^*\right],$$

where $\tilde{C}(\widehat{\boldsymbol{\mu}}_{BC2})$ and $\tilde{D}(\widehat{\boldsymbol{\mu}}_{BC1})$ are defined in (15) and (16), respectively.

A possible choice for n_0 is $n_0 = \max(np, 1)$ with $0 < p < 1$. According to, among others, del Barrio, Cuesta-Albertos and Matrán (2002), the choice of the estimated resample size should not be too small because it can lead to instability of the estimators and so, for the practical purposes of the Monte-Carlo studies, we will set the value p to be equal to the constant value 0.1 (i.e., the lower bound for the truncation is 10% of the original sample size, n).

5.2.2 *Estimator for* m_3

The estimator for the resample size to be used when calculating $\tilde{\theta}_{brag,2}$ is slightly different from the one used when calculating $\tilde{\theta}_{brag,1}$. The estimator for m_3, based on the expression given in (10), is given by

$$\widehat{m}_{3,\star} := \begin{cases} \min\left\{\max(n_0, \widehat{m}_{2,\star}), n\right\}, & \text{if } \widehat{\Delta}_\star > 0 \\ n, & \text{if } \widehat{\Delta}_\star \leq 0, \end{cases}$$

where (see (9))

$$\widehat{m}_{2,\star} := \frac{1}{n}\left[\widehat{m}_{0,\star}\right]^2.$$

6. Monte-Carlo simulation results

The performance of the data-dependent choices of m discussed in the previous section is evaluated in this section by using Monte-Carlo simulations. The performance of $\tilde{\theta}_{brag,j}$, $j = 1, 2$ (using the various estimators of m discussed above) is measured by calculating the ratio

$$\zeta(\tilde{\theta}_{brag,j}) := \frac{\mathrm{MSE}(\bar{X}_n)}{\mathrm{MSE}(\tilde{\theta}_{brag,j})}, \quad j = 1, 2,$$

where $\tilde{\theta}_{brag,j}$ is calculated using one of the three proposed estimators for m. If $\zeta(\tilde{\theta}_{brag,j})$ is greater than 1, then it indicates that $\tilde{\theta}_{brag,j}$ performs better (in a mean squared error sense) than the sample mean when estimating the population mean.

The configurations for this set of Monte-Carlo simulations are:

- The number of Monte-Carlo simulations performed for each entry in the tables is $MC = 5\,000$. The BS-estimator, defined in (18) is based on $B = 1\,000$ bootstrap replications for each Monte-Carlo trial.
- The sample sizes used are $n = 20, 50, 100, 500$ and $1\,000$.
- Data were drawn from the double exponential distribution, F-distribution, normal distribution and the contaminated normal distribution, with densities defined respectively by:

  (i) $f(x) = (1/2\sigma)\exp(-|x-\mu|/\sigma)$, $-\infty < x < \infty$,
  (ii) $f(x) = \left((n/m)^{n/2} \big/ B(n/2, m/2)\right) x^{(n-2)/2}(1 + nx/m)^{-(n+m)/2}$, $x \geq 0$,
  (iii) $f(x) = (1/\sigma)\phi((x-\mu)/\sigma)$, $-\infty < x < \infty$,
  (iv) $f(x) = ((1-p)/\sigma_1)\phi((x-\mu_1)/\sigma_1) + (p/\sigma_2)\phi((x-\mu_2)/\sigma_2)$, $-\infty < x < \infty$,

  where $B(\cdot,\cdot)$ is the beta-function and $\phi(\cdot)$ the standard normal density function. The specific parameter choices for each of these distributions are given in each table's caption.

Displayed in each of the tables is the value $\zeta(\tilde{\theta}_{brag,j})$, $j = 1, 2$, for the different estimators of m as well as the mean over the Monte-Carlo trials of these estimates of m (denoted by $\bar{m}$) and their standard errors (denoted by SE($\bar{m}$)). Note that in these tables it is possible to obtain standard errors for $\bar{m}$ equal to zero. This occurs whenever the procedure calculates a bootstrap resample size which is smaller than the prescribed lower bound of $n_0 = np$ in each Monte-Carlo iteration. Indeed, one can see in Tables 1 and 4 that this occurs for some of the larger sample sizes.

To aid readability, the largest $\zeta(\tilde{\theta}_{brag,1})$ and $\zeta(\tilde{\theta}_{brag,2})$ values are highlighted in the tables (in bold) for each sample size across the three different methods of obtaining the resample size data-dependently.

6.1 Conclusions drawn from the tables

The conclusions drawn from the tables will be broken down into two sections; the first section deals with the performance of the m-out-of-n estimator, $\tilde{\theta}_{brag,1}$, using the associated proposed estimators for m, and the second section deals with the performance of the corrected m-out-of-n estimator, $\tilde{\theta}_{brag,2}$, using the associated proposed estimators for m. We make conclusions concerning the classes of distributions used and comment on the overall performance of the estimators. An indication of which ones perform the best will also be given.

6.1.1 The performance of $\tilde{\theta}_{brag,1}$ using the proposed estimators for m

When we consider the ratio $\zeta(\tilde{\theta}_{brag,1})$ for the m-out-of-n estimators in Tables 1–4 we find that the performance of the estimators is very good (i.e., almost always producing values larger than 1) for small to moderate sample sizes, and that it converges to 1 as the sample size becomes larger. However, the tendency differs for the various distributions considered. When we categorize the distributions according to general properties we find the following:

- *Symmetric, heavy-tailed distributions:* The distributions that are symmetric and heavy-tailed include the double exponential distribution and the unimodal contaminated normal distribution (displayed in Tables 1 and 4). These distributions show significant improvements over the sample mean when estimating the population mean. In Table 4, we find that the maximum recorded improvement is 89.4% for $n = 20$.

- *Skewed, heavy-tailed distributions:* The distribution we considered that is both skewed and heavy-tailed is the $F(8, 5)$-distribution (found in Table 2). The improvement over the sample mean for these estimators applied to this distribution ranges between 31.7% and 106.9% for all sample sizes considered, with the largest gains being made with the smaller sample sizes. Here we see that the BS procedure for determining the resample size is preferable for all sample sizes considered.

- *Standard normal distribution:* The results for the standard normal distribution are displayed in Table 3. Since the sample mean is admissible as an estimator for the mean of a normal distribution we cannot expect to see much improvement over the sample mean in this case. However, the estimators using the various choices of m for this distribution routinely achieve at least 99% of the performance of the sample mean in the worst cases (even managing to match the sample mean in other cases). All of the estimators seem to perform equally well for this distribution.

In general, we see a marked improvement over the sample mean when the underlying distribution has heavier tails. This improvement is most acute when working with small to moderate sample sizes, but is less impressive when one has larger samples (except in the case of the F-distribution where we still see improvements of up to 31.7% for samples as large as 1000).

6.1.2 *The performance of $\tilde{\theta}_{brag,2}$ using the proposed estimators for m*

When we consider the ratio $\zeta(\tilde{\theta}_{brag,2})$ for the corrected m-out-of-n estimators in Tables 1–4 we can make similar conclusions to those made concerning the m-out-of-n estimators. Once again the estimators are found to perform very well for small to moderate sample sizes, but the results are not as impressive as those observed in the m-out-of-n estimators' results. The tendency of the estimators' performance is that, as the sample size increases, the performance starts to match the performance of the sample mean. Once again, when we categorize the distributions according to general properties we find that we can make the following conclusions:

- *Symmetric, heavy-tailed distributions:* From Tables 1 and 4 it is clear that, in the majority of the cases presented, $\tilde{\theta}_{brag,2}$ significantly outperforms $\bar{X}_n$, especially for the contaminated normal distribution. Here the maximum recorded improvement is 45.1% for $n = 20$.

- *Skewed, heavy-tailed distributions:* From Table 2 it follows that the improvement in mean squared error over $\bar{X}_n$ ranges from 11.3% up to 65.7%, once again achieved by the BS smoothing method.

- *Standard normal distribution:* The results in Table 3 show that $\tilde{\theta}_{brag,2}$ and $\bar{X}_n$ have almost identical performance.

Table 1. MSE ratios and mean values of the estimated resample sizes using the double exponential distribution, $\mu = 0,\ \sigma = \sqrt{2}$.

n		$brag, 1$			$brag, 2$		
		BS	**BC**	**UNBC**	**BS**	**BC**	**UNBC**
20	ζ	**1.082**	1.013	1.048	**1.056**	1.021	1.031
	$\bar{m}$	10.5	15.0	11.0	9.5	13.1	10.3
	$SE(\bar{m})$	0.115	0.084	0.118	0.122	0.111	0.123
50	ζ	1.056	**1.089**	1.040	1.021	**1.022**	1.017
	$\bar{m}$	11.6	8.9	10.4	10.3	8.9	8.8
	$SE(\bar{m})$	0.202	0.162	0.175	0.2	0.173	0.169
100	ζ	1.081	**1.095**	1.076	**1.052**	1.051	1.049
	$\bar{m}$	12.9	11.4	12.2	12.2	11.4	11.4
	$SE(\bar{m})$	0.197	0.148	0.161	0.19	0.157	0.152
500	ζ	0.978	0.977	**0.985**	**0.983**	0.982	0.97
	$\bar{m}$	50	50	50	50	50	50
	$SE(\bar{m})$	0	0	0	0	0	0
1000	ζ	**0.979**	**0.979**	0.975	0.979	**0.979**	0.975
	$\bar{m}$	100	100	100	100	100	100
	$SE(\bar{m})$	0	0	0	0	0	0

Table 2. MSE ratios and mean values of the estimated resample sizes using the $F(8, 5)$-distribution.

n		$brag, 1$			$brag, 2$		
		BS	**BC**	**UNBC**	**BS**	**BC**	**UNBC**
20	ζ	**2.069**	1.202	1.203	**1.657**	1.204	1.205
	$\bar{m}$	14.5	20.0	19.9	12.5	20.0	19.9
	$SE(\bar{m})$	0.089	0.007	0.016	0.117	0.01	0.02
50	ζ	**1.913**	1.363	1.445	**1.715**	1.441	1.450
	$\bar{m}$	30.9	29.9	47.7	23.9	33.6	46.2
	$SE(\bar{m})$	0.214	0.291	0.08	0.277	0.285	0.128
100	ζ	**1.669**	1.041	1.118	**1.404**	1.109	1.115
	$\bar{m}$	52.7	47.6	91.5	36.0	59.4	85.7
	$SE(\bar{m})$	0.388	0.575	0.194	0.488	0.594	0.317
500	ζ	**1.405**	1.106	1.129	**1.296**	1.128	1.132
	$\bar{m}$	185.5	187.9	372.8	99.0	210.5	302.8
	$SE(\bar{m})$	1.421	2.558	1.576	1.568	2.908	2.358
1000	ζ	**1.317**	0.906	0.914	**1.113**	0.915	0.916
	$\bar{m}$	317.6	344.8	653.4	161.8	381.9	495.1
	$SE(\bar{m})$	2.488	4.92	3.693	2.57	5.729	5.212

Table 3. MSE ratios and mean values of the estimated resample sizes using the standard normal distribution.

		$brag, 1$			$brag, 2$		
n		**BS**	**BC**	**UNBC**	**BS**	**BC**	**UNBC**
20	ζ	0.990	**0.992**	0.984	0.992	0.991	**0.993**
	$\bar{m}$	16.9	17.8	15.3	16.5	16.9	14.9
	$SE(\bar{m})$	0.089	0.064	0.107	0.099	0.087	0.113
50	ζ	**1.020**	1.019	1.017	**1.023**	1.023	1.022
	$\bar{m}$	38.5	32.4	31.0	37.7	31.2	30.3
	$SE(\bar{m})$	0.265	0.294	0.305	0.280	0.309	0.313
100	ζ	**1.014**	1.013	1.011	1.015	**1.019**	1.018
	$\bar{m}$	72.6	62.4	61.6	71.4	61.2	60.7
	$SE(\bar{m})$	0.569	0.612	0.618	0.588	0.626	0.628
500	ζ	**1.019**	1.011	1.009	1.017	1.018	**1.020**
	$\bar{m}$	316.7	297.6	297.5	313.8	295.9	295.8
	$SE(\bar{m})$	3.090	3.145	3.145	3.125	3.163	3.162
1000	ζ	1.007	**1.010**	1.006	**1.014**	**1.014**	1.013
	$\bar{m}$	317.6	344.8	653.4	161.8	381.9	495.1
	$SE(\bar{m})$	2.488	4.92	3.693	2.57	5.729	5.212

Table 4. MSE ratios and mean values of the estimated resample sizes using the contaminated normal distribution, $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 0,\ \sigma_2 = 8,\ p = 0.1$.

		$brag, 1$			$brag, 2$		
n		**BS**	**BC**	**UNBC**	**BS**	**BC**	**UNBC**
20	ζ	**1.894**	1.166	1.200	**1.451**	1.170	1.180
	$\bar{m}$	10.1	28.6	36.6	7.1	32.1	34.5
	$SE(\bar{m})$	0.131	0.311	0.261	0.117	0.304	0.287
50	ζ	1.657	**1.696**	1.188	**1.348**	1.228	1.162
	$\bar{m}$	10.1	28.6	36.6	7.1	32.1	34.5
	$SE(\bar{m})$	0.131	0.311	0.261	0.117	0.304	0.287
100	ζ	1.548	**1.767**	1.218	**1.274**	1.203	1.170
	$\bar{m}$	14.1	38.0	53.5	11.0	47.3	48.1
	$SE(\bar{m})$	0.147	0.588	0.566	0.105	0.609	0.599
500	ζ	**1.132**	1.129	1.130	1.048	1.049	**1.051**
	$\bar{m}$	50.3	50.1	51.2	50.0	50.3	50.2
	$SE(\bar{m})$	0.059	0.092	0.166	0.005	0.137	0.129
1000	ζ	1.082	1.083	**1.084**	1.046	**1.047**	1.046
	$\bar{m}$	100	100	100	100	100	100
	$SE(\bar{m})$	0	0	0.002	0	0	0

In general, we see that these estimators do not perform nearly as well as the $\tilde{\theta}_{brag,1}$ estimators. This can be ascribed to the fact that $\tilde{\theta}_{brag,2}$ is a convex combination of both $\tilde{\theta}_{brag,1}$ and $\bar{X}_n$, and it is well known that the estimator $\bar{X}_n$ is not robust against outliers.

7. Concluding remarks

Similar results were obtained for different choices of the parameters appearing in the densities defined in (i) – (iv) on page 84. Simulations were also done by generating data from other distributions, such as the uniform, centered exponential, Weibull, and Pareto, which yielded the same satisfactory results as reported in Tables 1–4.

Based on all of our simulation results we would make the suggestion of favouring the $\tilde{\theta}_{brag,1}$ estimator when estimating the population mean, especially for heavy-tailed distributions. It is also quite evident that the BS-procedure for choosing the resample size data-dependently generally has the best performance. This finding is encouraging because, while it may be slightly more expensive in terms of computational time, it is much easier to implement and more generally applicable than the other methods presented in this paper to obtain the data-dependent choice of m.

In this paper we considered estimating the mean of a population so that we could present our findings in a comprehensible manner. However, similar results hold for estimators associated with the smooth function model because valid Cornish-Fisher expansions hold in this more general setup (Hall, 1992).

Acknowledgements. The second author would like to thank the National Research Foundation of South Africa for financial support.

Appendix

A. Details of the Cornish-Fisher expansion of the general statistic

Let the median of the bootstrap distribution of the statistic $g(\bar{\boldsymbol{X}}_m^*)$ be denoted by $\text{Med}^*(g(\bar{\boldsymbol{X}}_m^*)) = \tilde{\theta}_{brag,1}$. Using a Cornish-Fisher expansion we can obtain an expansion for $\tilde{\theta}_{brag,1}$. We proceed by first noting that

$$\mathrm{P}^*(g(\bar{\boldsymbol{X}}_m^*) \le \tilde{\theta}_{brag,1}) \approx \frac{1}{2},$$

$$\text{i.e.,} \qquad \mathrm{P}^*\left(T_m^* \le \frac{\sqrt{m}\left(\tilde{\theta}_{brag,1} - g(\bar{\boldsymbol{X}}_n)\right)}{h(\bar{\boldsymbol{X}}_n)}\right) \approx \frac{1}{2}.$$

Now, the Cornish-Fisher expansion of the α^{th} quantile of the bootstrap distribution of T_m^*, denoted by $v(\alpha)$, is given by

$$v(\alpha) = z(\alpha) + m^{-1/2}\widehat{p}_1^{cf}(z(\alpha)) + m^{-1}\widehat{p}_2^{cf}(z(\alpha)) + O_p(m^{-3/2}).$$

However, since we are interested in calculating the median we choose $\alpha = 0.5$ and obtain the following expression:

$$\sqrt{m}\left(\tilde{\theta}_{brag,1} - g(\bar{\boldsymbol{X}}_n)\right)/h(\bar{\boldsymbol{X}}_n) = m^{-1/2}\widehat{p}_1^{cf}(0) + m^{-1}\widehat{p}_2^{cf}(0) + O_p(m^{-3/2})$$

$$= -m^{-1/2}\widehat{p}_1(0) + m^{-1}\left\{\widehat{p}_1(0)\,\frac{d}{dx}\widehat{p}_1(x)\Big|_{x=0} - \widehat{p}_2(0)\right\} + O_p(m^{-3/2})$$
$$= -m^{-1/2}\widehat{p}_1(0) + m^{-1}\widehat{p}_1(0)\,\frac{d}{dx}\widehat{p}_1(x)\Big|_{x=0} - m^{-1}\widehat{p}_2(0) + O_p(m^{-3/2}).$$

Note the following (see Hall, 1992, pp. 88–89):

- $\widehat{p}_1^{cf}(x) = -\widehat{p}_1(x)$, so that $\widehat{p}_1^{cf}(0) = -\widehat{p}_1(0)$,
- $\widehat{p}_2^{cf}(x) = \widehat{p}_1(x)\left[\frac{d}{dx}\widehat{p}_1(x)\right] - \frac{1}{2}x\{\widehat{p}_1(x)\}^2 - \widehat{p}_2(x)$, so that $\widehat{p}_2^{cf}(0) = \widehat{p}_1(0)\left[\frac{d}{dx}\widehat{p}_1(x)\big|_{x=0}\right] - \widehat{p}_2(0)$,
- $\widehat{p}_1(x) = -(\widehat{k}_{1,2} + \frac{1}{6}\widehat{k}_{3,1}(x^2-1))$ so that $\widehat{p}_1(0) = -\widehat{k}_{1,2} + \frac{1}{6}\widehat{k}_{3,1}$,
- $\frac{d}{dx}\widehat{p}_1(x) = -\frac{1}{3}\widehat{k}_{3,1}x$ so that $\frac{d}{dx}\widehat{p}_1(x)\big|_{x=0} = 0$,
- $\widehat{p}_2(x) = -x\left[\frac{1}{2}(\widehat{k}_{2,2}+\widehat{k}_{1,2}^2) + \frac{1}{24}(\widehat{k}_{4,1} + 4\widehat{k}_{1,2}\widehat{k}_{3,1})(x^2-3) + \frac{1}{72}\widehat{k}_{3,1}^2(x^4-10x+15)\right]$ so that $\widehat{p}_2(0) = 0$.

It should also be noted that the $\widehat{k}_{i,j}$ terms are defined as the *estimated* versions of the polynomials appearing in the expansion of the j^{th} cumulant of T_n, κ_{j,T_n}, i.e.,

$$\kappa_{j,T_n} = n^{-(j-2)/2}\left(k_{j,1} + n^{-1}k_{j,2} + n^{-2}k_{j,3} + \ldots\right).$$

B. Taylor series expansions of the numerator and denominator of m_1

B.1 Expression for the numerator

Using a multivariable Taylor series expansion, we find that the numerator of (12) can be written as the expected value of the function $f(x,y) = (x/y)^2$ evaluated in the points $x = \widehat{\mu}_3$ and $y = \widehat{\mu}_2$, expanded about the values μ_3 and μ_2. The expansion is provided below:

$$\begin{aligned}
\mathrm{E}\left\{\left(\frac{\widehat{\mu}_3}{\widehat{\mu}_2}\right)^2\right\} &\approx \mathrm{E}\Bigg\{\left(\frac{\mu_3}{\mu_2}\right)^2 + 2\frac{\mu_3}{\mu_2^2}(\widehat{\mu}_3-\mu_3) - 2\frac{\mu_3^2}{\mu_2^3}(\widehat{\mu}_2-\mu_2) \\
&\quad + \frac{1}{\mu_2^2}(\widehat{\mu}_3-\mu_3)^2 + 3\frac{\mu_3^2}{\mu_2^4}(\widehat{\mu}_2-\mu_2)^2 - 4\frac{\mu_3}{\mu_2^3}(\widehat{\mu}_3-\mu_3)(\widehat{\mu}_2-\mu_2)\Bigg\} \\
&= \left(\frac{\mu_3}{\mu_2}\right)^2 + 2\frac{\mu_3}{\mu_2^2}\mathrm{E}(\widehat{\mu}_3) - 2\frac{\mu_3^2}{\mu_2^3}\mathrm{E}(\widehat{\mu}_2) + \frac{1}{\mu_2^2}\mathrm{E}\left(\widehat{\mu}_3^2\right) - 2\frac{\mu_3}{\mu_2^2}\mathrm{E}(\widehat{\mu}_3) + 3\frac{\mu_3^2}{\mu_2^4}\mathrm{E}\left(\widehat{\mu}_2^2\right) \\
&\quad - 6\frac{\mu_3^2}{\mu_2^3}\mathrm{E}(\widehat{\mu}_2) - 4\frac{\mu_3}{\mu_2^3}\mathrm{E}(\widehat{\mu}_3\widehat{\mu}_2) + 4\frac{\mu_3}{\mu_2^2}\mathrm{E}(\widehat{\mu}_3) + 4\frac{\mu_3^2}{\mu_2^3}\mathrm{E}(\widehat{\mu}_2). \qquad (19)
\end{aligned}$$

The expected values in this expression can be simplified through tedious calculation (the full form of these derivations can be quite long and so have been omitted. They are available on request from the authors). The expressions become:

$$\mathrm{E}(\widehat{\mu}_2) = \mu_2 + \frac{1}{n}\{-\mu_2\},$$
$$\mathrm{E}(\widehat{\mu}_3) = \mu_3 + \frac{1}{n}\{-3\mu_3\} + \frac{1}{n^2}\{2\mu_3\},$$

$$
\begin{aligned}
\mathrm{E}(\widehat{\mu}_2^2) &= \mu_2^2 + \frac{1}{n}\left\{\mu_4 - 3\mu_2^2\right\} + \frac{1}{n^2}\left\{-2\mu_4 + 5\mu_2^2\right\} + \frac{1}{n^3}\left\{\mu_4 - 3\mu_2^2\right\},\\
\mathrm{E}(\widehat{\mu}_3^2) &= \mu_3^2 + \frac{1}{n}\left\{\mu_6 + 9\mu_2^3 - 7\mu_3^2 - 6\mu_2\mu_4\right\} + O\left(\frac{1}{n^2}\right),\\
\mathrm{E}(\widehat{\mu}_2\widehat{\mu}_3) &= \mu_2\mu_3 + \frac{1}{n}\left\{\mu_5 - 8\mu_2\mu_3\right\} + O\left(\frac{1}{n^2}\right).
\end{aligned}
$$

Once all of the above expressions have been substituted into equation (19) we then obtain the final expression for the expansion of the numerator of (12):

$$
\mathrm{E}\left\{\left(\frac{\widehat{\mu}_3}{\widehat{\mu}_2}\right)^2\right\} = \left(\frac{\mu_3}{\mu_2}\right)^2 + \frac{1}{n}\left\{\frac{1}{\mu_2^4}\left[\mu_2^2\mu_6 + 9\mu_2^5 + 8\mu_2^2\mu_3^2 - 6\mu_2^3\mu_4 + 3\mu_3^2\mu_4 - 4\mu_2\mu_3\mu_5\right]\right\} + O(n^{-2}). \quad (20)
$$

B.2 Expression for the denominator

The expression for the denominator is obtained in a similar fashion to the numerator except that the Taylor expansion is carried out on a function of three variables as opposed to the two variable Taylor expansion used for the numerator.

The denominator of (12) (without the multiplier 6) can now be written as a multivariable Taylor series expansion of the expected value of the function $f(x, y, z) = xz/y$ evaluated in the points $x = \bar{Y}_n$, $y = \widehat{\mu}_2$ and $z = \widehat{\mu}_3$, expanded about the values $\mu_1 = 0$, μ_2 and μ_3. The expansion is given below:

$$
\begin{aligned}
\mathrm{E}\left(\bar{Y}_n\frac{\widehat{\mu}_3}{\widehat{\mu}_2}\right) \approx \mathrm{E}\Bigg\{ & \frac{\mu_1\mu_3}{\mu_2} + \frac{\mu_1}{\mu_2}(\widehat{\mu}_3 - \mu_3) - \frac{\mu_1\mu_3}{\mu_2^2}(\widehat{\mu}_2 - \mu_2) + \frac{\mu_3}{\mu_2}(\bar{Y}_n - \mu_1) + \frac{\mu_1\mu_3}{\mu_2^3}(\widehat{\mu}_2 - \mu_2)^2\\
& - \frac{\mu_3}{\mu_2^2}(\bar{Y}_n - \mu_1)(\widehat{\mu}_2 - \mu_2) + \frac{1}{\mu_2}(\bar{Y}_n - \mu_1)(\widehat{\mu}_3 - \mu_3) - \frac{\mu_1}{\mu_2^2}(\widehat{\mu}_2 - \mu_2)(\widehat{\mu}_3 - \mu_3)\\
& - \frac{\mu_1\mu_3}{\mu_2^4}(\widehat{\mu}_2 - \mu_2)^3 + \frac{\mu_3}{\mu_2^3}(\bar{Y}_n - \mu_1)(\widehat{\mu}_2 - \mu_2)^2 + \frac{\mu_1}{\mu_2^3}(\widehat{\mu}_2 - \mu_2)^2(\widehat{\mu}_3 - \mu_3)\\
& - \frac{1}{\mu_2^2}(\bar{Y}_n - \mu_1)(\widehat{\mu}_2 - \mu_2)(\widehat{\mu}_3 - \mu_3) + \frac{\mu_1\mu_3}{\mu_2^5}(\widehat{\mu}_2 - \mu_2)^4 - \frac{\mu_3}{\mu_2^4}(\bar{Y}_n - \mu_1)(\widehat{\mu}_2 - \mu_2)^3\\
& - \frac{\mu_1}{\mu_2^4}(\widehat{\mu}_2 - \mu_2)^3(\widehat{\mu}_3 - \mu_3) + \frac{1}{\mu_2^3}(\bar{Y}_n - \mu_1)(\widehat{\mu}_2 - \mu_2)^2(\widehat{\mu}_3 - \mu_3)\Bigg\}\\
= & - \frac{\mu_3}{\mu_2^2}\mathrm{E}\left\{\bar{Y}_n(\widehat{\mu}_2 - \mu_2)\right\} + \frac{1}{\mu_2}\mathrm{E}\left\{\bar{Y}_n(\widehat{\mu}_3 - \mu_3)\right\} + \frac{\mu_3}{\mu_2^3}\mathrm{E}\left\{\bar{Y}_n(\widehat{\mu}_2 - \mu_2)^2\right\}\\
& - \frac{1}{\mu_2^2}\mathrm{E}\left\{\bar{Y}_n(\widehat{\mu}_2 - \mu_2)(\widehat{\mu}_3 - \mu_3)\right\} - \frac{\mu_3}{\mu_2^4}\mathrm{E}\left\{\bar{Y}_n(\widehat{\mu}_2 - \mu_2)^3\right\}\\
& + \frac{1}{\mu_2^3}\mathrm{E}\left\{\bar{Y}_n(\widehat{\mu}_2 - \mu_2)^2(\widehat{\mu}_3 - \mu_3)\right\}. \qquad (21)
\end{aligned}
$$

After laborious calculations the expected values in the above equation simplify as follows:

$$
\mathrm{E}\left(\bar{Y}_n(\widehat{\mu}_2 - \mu_2)\right) = \frac{1}{n}\left\{\mu_3\right\} + \frac{1}{n^2}\left\{-\mu_3\right\},
$$

$$\begin{aligned}
\mathrm{E}\left(\bar{Y}_n(\widehat{\mu}_3-\mu_3)\right) &= \frac{1}{n}\left\{\mu_4-3\mu_2^2\right\}+\frac{1}{n^2}\left\{-3\mu_4+9\mu_2^2\right\}+O(n^{-3}),\\
\mathrm{E}\left(\bar{Y}_n(\widehat{\mu}_2-\mu_2)^2\right) &= \frac{1}{n^2}\left\{2\mu_3\mu_5+\mu_4^2-4\mu_2^2\mu_4-8\mu_3^2\mu_2+3\mu_2^4\right\}+O(n^{-3}),\\
\mathrm{E}\left(\bar{Y}_n(\widehat{\mu}_3-\mu_3)(\widehat{\mu}_2-\mu_2)\right) &= \frac{1}{n^2}\left\{\mu_6-10\mu_2\mu_4-7\mu_3^2+15\mu_2^3\right\}+O(n^{-3}),\\
\mathrm{E}\left(\bar{Y}_n(\widehat{\mu}_2-\mu_2)^3\right) &= \frac{1}{n^2}\left\{3\mu_3\mu_4-3\mu_2^2\mu_3\right\}+O(n^{-3}),\\
\mathrm{E}\left(\bar{Y}_n(\widehat{\mu}_2-\mu_2)^2(\widehat{\mu}_3-\mu_3)\right) &= \frac{1}{n^2}\left\{2\mu_3\mu_5+\mu_4^2-4\mu_2^2\mu_4-8\mu_3^2\mu_2+3\mu_2^4\right\}+O(n^{-3}).
\end{aligned}$$

Once all of these expressions have been substituted into (21) and the terms have been collected, we find that the denominator becomes

$$\begin{aligned}
\mathrm{E}\left(\bar{Y}_n\frac{\widehat{\mu}_3}{\widehat{\mu}_2}\right) &= \frac{1}{n\mu_2^2}\left[\mu_2\mu_4-3\mu_2^3-\mu_3^2\right]\\
&\quad+\frac{1}{n^2\mu_2^4}\left[3\mu_2^3\mu_4-\mu_2^2\mu_6-3\mu_2^5+3\mu_2\mu_3\mu_5+\mu_2\mu_4^2-3\mu_3^2\mu_4-5\mu_2^2\mu_3^2\right]+O(n^{-3}). \quad (22)
\end{aligned}$$

Once again, the details of the above derivations are too lengthy to be reported here, but they are available from the authors on request.

References

ALIN, A., MARTIN, M. A., BEYAZTAS, U., AND PATHAK, P. K. (2017). Sufficient m-out-of-n (m/n) bootstrap. *J. Stat. Comput. Sim.*, **87**, 1742–1753.

ALLISON, J. S., SANTANA, L., AND SWANEPOEL, J. W. H. (2011). Two new data-dependent choices of m when applying the m-out-of-n bootstrap to hypothesis testing. *J. Stat. Comput. Sim.*, **81**, 2107–2120.

ARCONES, M. A. (2003). On the asymptotic accuracy of the bootstrap under arbitrary resample size. *Ann. I. Stat. Math.*, **55**, 563–583.

BERAN, R. (1997). Diagnosing bootstrap success. *Ann. I. Stat. Math.*, **49**, 1–24.

BERRENDERO, J. R. (2007). The bagged median and the bragged mean. *Am. Stat.*, **61**, 325–330.

BICKEL, P., GÖTZE, F., AND VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Stat. Sinica*, **7**, 1–31.

BICKEL, P. J. AND FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Stat.*, **9**, 1196–1217.

BICKEL, P. J. AND SAKOV, A. (2002). Extrapolation and the bootstrap. *Sankhyā. Ser. A*, **64**, 640–652.

BICKEL, P. J. AND YAHAV, J. A. (1988). Richardson extrapolation and the bootstrap. *J. Am. Stat. Assoc.*, **86**, 387–393.

BREIMAN, L. (1996). Bagging predictors. Technical Report 421, University of California, Berkley, California.

BRETAGNOLLE, J. (1983). Lois limites du Bootstrap de certaines fonctionnelles. *Ann. I. H. Poincaré B.*, **19**, 281–296.

Bühlmann, P. (2002). Bootstraps for time series. *Stat. Sci.*, **17**, 52–72.

Bühlmann, P. (2003). Bagging, subagging and bragging for improving some prediction algorithms. *In* Akritas, M. G. and Politis, D. N. (Editors) *Recent Advances and Trends in Nonparametric Statistics*. Elsevier, Amsterdam.

Buja, A. and Stuetzle, W. (2006). Observations on bagging. *Stat. Sinica*, **16**, 323–351.

Cheung, K. Y. and Lee, S. M. S. (2005). Variance estimation for sample quantiles using the m out of n bootstrap. *Ann. I. Stat. Math.*, **57**, 279 – 290.

Chung, K. H. and Lee, S. M. S. (2001). Optimal bootstrap sample size in construction of percentile confidence bounds. *Scand. J. Stat.*, **28**, 225–239.

Croux, C., Joossens, K., and Lemmens, A. (2007). Trimmed bagging. *Comput. Stat. Data An.*, **52**, 362–368.

del Barrio, E., Cuesta-Albertos, J. A., and Matrán, C. (2002). Asymptotic stability of the bootstrap sample mean. *Stoch. Proc. Appl.*, **97**, 289–306.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. Springer Verlag, New York.

Janssen, P., Swanepoel, J. W. H., and Veraverbeke, N. (2001). Modified bootstrap consistency rates for U-quantiles. *Stat. Probabil. Lett.*, **54**, 261–268.

Lee, S. M. S. (1999). On a class of m out of n bootstrap confidence intervals. *J. Roy. Stat. Soc. B.*

Sakov, A. (1998). *Using the m out of n Bootstrap in Hypothesis Testing*. Ph.D. thesis, University of California, Berkley.

Sakov, A. and Bickel, P. J. (1999). Choosing m in the m-out-of-n bootstrap. *Proc. Am. Stat. Assoc.*, 125–128.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer, New York.

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works. *Commun. Stat. A – Theor.*, **15**, 3193–3203.

Swanepoel, J. W. H. (1988). A new nonparametric estimation technique under asymmetry using the bootstrap. Technical report, Potchefstroom University for Christian Higher Education, Potchefstroom, South Africa.

Swanepoel, J. W. H. (1990). A review of bootstrap methods. *S. Afr. Stat. J.*, **24**, 1–34.

Manuscript received 2017-06-07, revised 2017-09-19, accepted 2017-10-16.