Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v30i19.3765

DEVELOPMENT OF A NOVEL SELECTION CRITERION FOR OPTIMUM CHOICE OF "m" IN THE "m out of n" BOOTSTRAP

Inayat Ullah[1*], Alamgir[2], Shahid Iqbal[3]

[1*,2]Department of Statistics, University of Peshawar
[3]Center for Disaster Preparedness and Management, University of Peshawar

***Corresponding author:** Inayat Ullah
***** Department of Statistics, University of Peshawar. Email: inayat81271@gmail.com

Abstract
Efron (1979) introduced the *n-out-of-n* bootstrap, which is indeed an important tool for statistical inference and has wide spread applications. However, there are situations, where the *n-out-of-n* bootstrap is not consistent. Thus, the *m-out-of-n* bootstrap was introduced to overcome the problem. It reduces the computational burden associated with bootstrapping. But, the problem with *m-out-of-n* bootstrap is the choice of *m*, which is one of the important aspects in bootstrapping. In this paper, we study criteria for choosing best value of *m* in *m-out-of-n* bootstrapping in linear regression. This is a pure computational study that gives general criteria for optimizing *m* in *m-out of-n* bootstrap, under which the chosen *m* ($\hat{m}$) behaves properly.

Key Words: Bootstrap; Optimization; Simulation; resampling methods; consistency

Introduction

Researchers occasionally treat the bootstrap as a magic bullet for statistical inference. It was invented by Efron (1979) and is based on drawing n observations from the empirical distribution of the data. This method is known as the naïve bootstrap. In fact, it has many applications; for examples, see (Hall, 1992) and (Efron and Tibshirani, 1993). The bootstrap is inconsistent in some instances, though. We have numerous examples of this kind of bootstrap antagonism. Although the counterexamples are very simple, the generalization holds for a large range of estimate issues that are crucial in their applications. The counterexample should act as a helpful reminder that there are limitations to the bootstrap's applicability to issues with statistical inference.

Alternative strategies are required when the conventional resampling techniques for estimating sampling distributions fail. For instance, a different bootstrap based on smaller-sized resamples was adopted if the n-out-of-n bootstrap (naive bootstrap) fails and the classical central limit theorem is violated(Silvia and Timothy, 2011).The m-out-of-n bootstrap was introduced as an alternative to the naive bootstrap(Bickel, P., Götze, F. & van Z, W. (1997). This bootstrap strategy, which was just recently introduced as a way to lessen the computing burden associated with bootstrapping, makes use of the different observations in a bootstrap sample. As long as naive bootstrap performs, it is also effective. However, this is the adequate bootstrap in the event that the naive bootstrap fails (Abadie

and Imbens, 2008). The next issue in the m-out-of-n bootstrap is the selection of m(Peter, J. B. and Anat, S (2008)) and (Račkauskas, F. Götze & A. (2001)).
While solving least square problems, Cholesky decomposition method has been used in the algorithm for the factorization of positive symmetric matrix. This method was introduces by a mathematician Higham (1990) in his article "Analysis of the Cholesky decomposition of a semi-definite matrix.

The Proposed Method

We suggested the best-choices of m in m-out-of-n bootstrap criteria. Different values of $m_i,\ (i=1,2,...,11)$ are taken in the analysis given as follow:

$$m_1 = (0.63/pi) * n \approx 20\%, \quad m_2 = (1.1/pi) * n \approx 35\%, \quad m_3 = (3.15/7) * n \approx 45\%$$
$$m_4 = (1/2) * n \approx 50\%, \quad m_5 = (1.73/pi) * n \approx 55\%, \quad m_6 = (1.89/pi) * n \approx 60\%$$
$$m_7 = (2.045/pi) * n \approx 65\%, \quad m_8 = (4.9/7) * n \approx 70\%, \quad m_9 = (3/4) * n \approx 75\%$$
$$m_{10} = (4/5) * n \approx 80\%, \quad m_{11} = (4.25/5) * n \approx 85\%, \quad and\ m = n = 100\%$$

Our objective is to determine m's ideal value. This is a pure computational study that offers broad guidelines for optimizing m in an m-out-of-n bootstrap, under which the selected m ($\hat{m}$) acts as intended, i.e., $\hat{m}/n \to 0$ and $\hat{m} \to \infty$ when the bootstrap is inconsistent.

1.1 Algorithm for picking the best value for m in an m-out-of-n bootstrap:

Suppose we have $\{y_i, x_{i1}, x_{i2}, ..., x_{ip}\}$ for $i = 1,2,3,...,n$ statistical units. Linear relation between the dependent variable $Y = (y_1, y_2, y_3..., y_n)^t$ and p-vector of regressors $x_i = (x_{i1}, x_{i2}, ..., x_{ip})\}$ for $i = 1,2,3,...,n$ is looks as $Y = X\beta + \varepsilon$

Where $\beta = (\beta_0, \beta_1, ..., \beta_p)^t$ are the parameters of the model and $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ are independent and identically distributed random variables with $\mathrm{E}(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \delta^2$ .Suppose the estimate of β is $\hat{\beta}$, where $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p)^t$. $\hat{\beta}$ is estimated by the Least Square Estimator (LSE) $\hat{\beta} = (X^t X)^{-1} X^t Y$
Our group of m-out-of-n bootstrap schemes includes $m = \{m_0 \equiv n, m_1, m_2, ...m_k\}$. It is to be noted that $m_0 = n$ is also a scheme used as a standard for all the other schemes. In order to determine the optimal m, we compare the $T^{*(m)}$ matrices of alternative m-out-of-n bootstrap techniques with that of the standard scheme $m_0 = n$.

The following algorithm is used to determine the best scheme out of all possible k schemes:

1. Generate data for linear regression on using a distribution, and then estimate $\hat{\beta}$ by the least squares estimator, where the $(i, j)th$ entry is $\hat{\beta}_{ij}$, i.e.

$$\hat{\beta}_{ij} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{.j})}{\sum_{i=1}^{n}(x_{ij} - \bar{x}_{.j})^2}$$

Take into consideration that there are m balls and n boxes with probabilities $p_1, p_2, ...p_n$, where

$$p_i \geq 0 \ \text{ and } \sum_{i=1}^{n} pi = 1$$

Throw balls into these boxes at random. Allow the likelihood that a ball will enter in the ith box to be p_i. Consider w_i be the total number of balls in the ith box, $w_i \geq 0$ and $\sum_{i=1}^{n} w_i = m$, then an $(n \times n)$ matrix W^* is given by:

$W^* = (\underline{w}_1^*, \underline{w}_2^*, ..., \underline{w}_n^*) \sim Multinomial(m, (\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}))$

2. Generate $W^* = (\underline{w}_1^*, \underline{w}_2^*, ..., \underline{w}_n^*)$ from multinomial distribution.

3. For each scheme generate a $B \times P$ matrix $T^{*(m_i)}$, whose $(b, j)th$ entry is $T^{*}{}_{bj}^{(m_i)}$, i.e.

$$T_{bj}^{*(m)} = \frac{\left(\hat{\beta}_{bj}^* - \hat{\beta}_j\right)^2}{1/\sqrt{m}} \quad \begin{matrix} b = 1, 2, ..., B \\ j = 1, 2, ..., p \end{matrix}$$

where $\hat{\beta}^{*(m_t)} = (z^t W^{*(d)} z)^{-1} z^t W^{*(d)} y$ is the bootstrap analog of $\hat{\beta}$ at m_t for $t = 1, 2, ..., k$. $W^{*(d)}$ for $i = 1, 2, ..., n$ is an $n \times n$ diagonal matrix, whose diagonal entries are the column vectors of W^* matrix.

4. For each $m = \{m_0 \equiv n, m_1, m_2, ..., m_k\}$, get column means $\overline{x}_{m_t}^*$ and variance $S_{m_t}^{*2}$ from $T^{*(m_i)}$ matrix, where

$$\overline{x}_{(m_t)}^* = \frac{1}{B} \sum_{b=1}^{B} T^{*}{}_{b.}$$

and $$S_{m_t}^{*2} = \frac{1}{B} \sum_{b=1}^{B} (T^{*}{}_{b.} - \overline{x}^*)^2 \quad \text{for} \quad t = 1, 2, ..., k$$

5. Do the cholesky decomposition of each $S^{*2}{}_{(m_i)}$ as follow:

$$p^{*}{}_{11} = \sqrt{a^{*}{}_{11}}$$

$$p^{*}{}_{j1} = \frac{a^{*}{}_{j1}}{p^{*}{}_{11}}, \qquad j \in (1, 2, ..., n)$$

$$p^{*}{}_{ii} = \sqrt{a^{*}{}_{ii} - \sum_{l=1}^{i-1} p^{*}{}_{il}{}^2} \quad i \in [2, n]$$

$$p^{*}{}_{ji} = \left(a^{*}{}_{ji} - \sum_{l=1}^{i-1} p^{*}{}_{il} p^{*}{}_{jl} \right) \Big/ p^{*}{}_{ii} \quad i \in [2, n-1],\ j \in [i+1, n]$$

where a_{ij} are the elements of $S_{m_t}^{*2}$, p_{ij} are the elements of lower triangular matrix P^* and $S_{m_t}^{*2} = P^*.P^{*T}$

6. Compute the result of $G^*(m_t)$ for $t = 1, 2, ..., k$,

where

$$G^*(m_t) = \left(\overline{x}_{(m_t)}^* - \overline{x}_{(m_0)}^*\right) S^{*2}{}_{(m_0)} \left(\overline{x}_{(m_t)}^* - \overline{x}_{(m_0)}^*\right) + P^*.S^{*2}{}_{(m_0)}.P^{*T}$$

To obtain the result of $G^*(m_t)$ for m_t $(t = 1, 2, ..., k)$ from each simulation, repeat these steps a lot of times (such as 10,000 times). take average of $G^*(m_t)$ after each simulation, where m_t $(t = 1, 2, ..., k)$.

Select that value of m_t, whose $G^*(m_t)$ value is minimum. The best choice will be that value of m_t $(t = 1, 2, ..., k)$ whose $G^*(m_t)$ is minimum.

Simulation Studies

In order to evaluate the effectiveness of the criteria used in the m-out-of-n bootstrap, considering the complete model with *p* predictors $(x_{i1}, x_{i2}, ..., x_{ip})$ and the dependent variable y given as follow:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \qquad i = 1, 2, 3, \ldots n \tag{1}$$

Where ε_i's are the independently and identically distributed (iid) standard normal errors. The intercept term is 1 expressed by the first component of x, while the other components are drawn, respectively, from the *Normal, t, exponential, and Laplace* distributions. The simulation studies were conducted by using n = 500, 1000, 1500, and 2000 as the sample sizes.The bootstrap samples are produced using the model (1), which has three, five, seven, nine, and eleven predictors $(i.e.\ p = 3,\ 5,\ 7,\ 9\ and\ 11)$.

For each sample, an m-out-of-n bootstrap was developed using eleven different values of with $m = \{m_0 \equiv n, m_1, m_2, ...m_k\}$.Here $m_0 = n$ is serves as the benchmark and all the other $m_i\,(i = 1, 2, \ldots, 11)$ are compared to it. For various *m*, *S = 1000* Monte Carlo (MC) simulations with *B = 1000* bootstrap replications are used to get the estimated results for the value of G^* at different sample sizes.The results for the value of G^* are *given* in the *Table (1) to Table (4)*. The results are summarized as follow:

- *Table 1* shows the simulation results using data produced by the normal distribution. There are four alternative sample sizes used: n = 500, 1000, 1500, and 2000. For each m_i $(i = 1, 2...11)$, the values are computed. Now for *n =500* and *p = 3*, the value of $G^*(n = 500, p = 3, m_5) < G^*(n = 500, p = 3, m_i)$ *for*all m_i. Similarly, for *n =1000* and *p = 7*, $G^*(n = 1000,\ p = 7,\ m_5) < G^*(n = 1000,\ p = 7,\ m_i)$ where i = 1, 2, . . .11.The value of G^* increases in each sample size as the number of predictors increases. For instance, when *n=1000*and *p=3*, the values of G^* at m_1 is *4.291952*, while the value of G^* for m_1, at *p=5* is *6.758223*, which shows increase in the value of G^*. Similarly, the value of G^* decreases with the increase of the sample size.For instance, the value of G^* is 4.450994 at *n=500* and *p=3*,whereas the respective values of G^* *at p=3* are 4.291952, 4.211217 and 3.85135 forthe sample sizes of*n=1000*, *1500 and 2000*.In short, we can say that for all $p = 3,\ 7,\ 9,$ and 11 $G^*(n = 500,\ m_i) < G^*(n = 1000,\ m_i) < G^*(n = 1500,\ m_i) < G^*(n = 2000,\ m_i)$ $i = 1,\ 2, \ldots 11$

It is evident from the data that for all four sample sizes, $m_5 = (1.73\ /\ pi) * n$ produces a smaller value of G^* than the other values of m_i ,s $(i = 1, 2..., 4, 6...11)$. This suggests that, in the given situation, m_5 is the best option for m in the m-out-of-n bootstrap.

- The simulation results using the data generated from the t-distribution are shown in Table 2. For each m_i $(i = 1, 2...11)$, four different samples of sizes *n=500, 1000, 1500 and 2000 are used to compute the value* of G^*. Now at*n=500* and *p=7*, we have $G^*(m_5) \le G^*(m_i)$ for $i = 1, 2...11$

.Similarly, at*n=1500* and *p=9,* $G^*(n=1500,\ p=9,\ m_5) \le G^*(n=1500,\ p=9, m_i)$ for each $i = 1, 2, \ldots, 11$. Just like in *Table 1* asthenumber of predictors increases, the value of G^* also increases, for instance, at*n=1500* and *p=3*, the values of G^* at m_1 is 4.211217, on the other hand at *p=5, 7, 9,* and *11 and n=1500,* the value of G^* for m_1, are *6.575076, 8.360561, 10.34293, 12.09127* respectively. Likewise, the G^* value decreases with the increase of the sample size, i.e. for instance, at *p=7 and n=500* and, the G^*value at m_7 is 8.541859, whereas at*p=7 and* n=*1000, 1500, 2000,* the corresponding G^*values are 8.390287, 8.078263 and 7.590608 respectively. The resultsshows that in all the four sample sizes, m_5 gives the smaller result for G^* compared to all m_i $(i = 1, 2, ..., 4, 6, ...11)$. This suggests that in t-distribution, m_5 is the best option for m in the m-out-of-n bootstrap as well.

- The simulation results using the data generated from the Laplace distribution are shown in Table 3. For each m_i $(i = 1, 2...11)$, four different samples of sizes *n=500, 1000, 1500 and 2000 are used to compute the value* of G^*. Now at*n=500* and *p=7*, we have $G^*(m_5) \le G^*(m_i)$ for $i = 1, 2...11$. Similarly, at *n=1500* and *p=9,* $G^*(n=1500,\ p=9,\ m_5) \le G^*(n=1500,\ p=9, m_i)$ for each $i = 1, 2, \ldots, 11$. Just like in*Table 1 and 2,* asthenumber of predictors increases, the value of G^* also increases, for instance, at *n=1500* and *p=5,* the values of G^* at $m_6$is 4.813858, on the other hand at *p= 3, 7, 9,* and *11 and n=1500,* the value of G^* for m_6, are 4.012566, 5.353734, 6.647686, 8.538686 respectively. Likewise, the G^* value decreases with the increase of the sample size, i.e. for instance, at *p=5 and n=500* and, the G^* value at m_7 is,5.794934, whereas at *p=5 and* **n**=*1000, 1500, 2000,* the corresponding G^*values are4.913191, 4.906442, 4.620683respectively. The results shows that in all the four sample sizes, m_5 gives the smaller result for G^* compared to all m_i $(i = 1, 2, ..., 4, 6, ...11)$. This suggests that in Laplace distribution, m_5 is also a best option for m in the m-out-of-n bootstrap.
- The results of *Table 4* are computed from the data generated from the exponential distribution. For each m_i $(i = 1, 2, ...11)$, the G^*value is calculatedby using all the sample sizes. Like the other distributions, in exponential distribution also, we can see that $G^*(m_5) \le G^*(m_i)$ for all $i = 1, 2...11$. Similarly, in this case, the value of G^* increases, when the number of predictors increases, for instance, i.e. at m_6 , when *n=2000* and *p=3,* the values of G^* is 3.464049, whereas at *n=2000* and *p=3,* the value of G^* for m_6 is 6.375435 and $G^*(m_5, p=3, 2000) < G^*(m_5, p=7, 2000)$. Similarly, the value of G^* decreases with the increase of sample size, i.e. at n=*1000* and *p=5,* the value ofG^* for m_7 is 5.821993, whereas the respective values ofG^* for m_7 , *n=500, 1500, 2000* and *p=5,* are 6.427996, 5.71038 and 5.594662. The main difference between the exponential distribution is that in other distributions the value of G^* decreases form m_1 to m_5 and then increases from m_6 to m_{11} in all sample cases. On the other handin exponential distribution the value of G^* decreases at m_5 and m_9, where $m_5 < m_9$. Therefore, the best choice for m in *m-out-of-m* bootstrap is m_5 in the exponential as well as other distributions.

Table 1: *Simulation results of G^* based on data generating from the normal distribution, with B=1000 and S=1000*

	n=500					n=1000				
	P = 3	**P = 5**	**P = 7**	**P = 9**	**P = 11**	**P = 3**	**P = 5**	**P = 7**	**P = 9**	**P = 11**
m1	[1,] 4.450994	[1,] 6.859249	[1,] 8.932928	[1,] 10.82672	[1,] 12.57565	[1,] 4.291952	[1,] 6.758223	[1,] 8.791855	[1,] 10.56567	[1,] 12.41823
m2	[1,] 4.385355	[1,] 6.557438	[1,] 8.710148	[1,] 10.56944	[1,] 12.12532	[1,] 4.303821	[1,] 6.404375	[1,] 8.534334	[1,] 10.32361	[1,] 12.02952
m3	[1,] 4.158518	[1,] 6.315677	[1,] 8.448624	[1,] 10.26417	[1,] 11.86821	[1,] 4.031641	[1,] 6.175937	[1,] 8.135773	[1,] 10.03542	[1,] 11.51794
m4	[1,] 4.020746	[1,] 6.144174	[1,] 8.279195	[1,] 10.06524	[1,] 11.60098	[1,] 3.893818	[1,] 6.048667	[1,] 7.991838	[1,] 9.81787	[1,] 11.27811
m5	[1,] 3.795128	[1,] 5.902272	[1,] 8.015066	[1,] 9.725019	[1,] 11.20594	[1,] 3.628518	[1,] 5.7620635	[1,] 7.701745	[1,] 9.641315	[1,] 11.09962
m6	[1,] 4.039592	[1,] 6.11943	[1,] 8.282167	[1,] 10.04397	[1,] 11.44869	[1,] 3.924479	[1,] 6.028114	[1,] 8.164681	[1,] 9.846633	[1,] 11.31035
m 7	[1,] 4.166447	[1,] 6.394613	[1,] 8.541859	[1,] 10.40398	[1,] 11.64597	[1,] 3.947256	[1,] 6.284258	[1,] 8.390287	[1,] 10.11558	[1,] 11.46757
m8	[1,] 4.355988	[1,] 6.593449	[1,] 8.722153	[1,] 10.67628	[1,] 11.81589	[1,] 4.059728	[1,] 6.417235	[1,] 8.519176	[1,] 10.26575	[1,] 11.61176
m9	[1,] 4.387595	[1,] 6.684006	[1,] 8.795709	[1,] 10.82364	[1,] 12.03797	[1,] 4.177301	[1,] 6.533693	[1,] 8.572863	[1,] 10.46143	[1,] 11.81738
m10	[1,] 4.447112	[1,] 6.690366	[1,] 8.843942	[1,] 10.92284	[1,] 12.33514	[1,] 4.240471	[1,] 6.618733	[1,] 8.591013	[1,] 10.642325	[1,] 12.25851
m11	[1,] 4.467508	[1,] 6.984497	[1,] 8.963113	[1,] 10.98395	[1,] 12.60591	[1,] 4.312759	[1,] 6.769328	[1,] 8.763187	[1,] 10.67583	[1,] 12.59221
	n=1500					**n=2000**				
	P = 3	**P = 5**	**P = 7**	**P = 9**	**P = 11**	**P = 3**	**P = 5**	**P = 7**	**P = 9**	**P = 11**
m 1	[1,] 4.211217	[1,] 6.575076	[1,] 8.360561	[1,] 10.34293	[1,] 12.09127	[1,] 3.85135	[1,] 6.453429	[1,] 7.949509	[1,] 10.18522	[1,] 11.85752
m2	[1,] 4.08079	[1,] 6.273096	[1,] 8.104185	[1,] 10.17108	[1,] 11.90766	[1,] 3.63201	[1,] 5.946137	[1,] 7.649421	[1,] 10.06957	[1,] 11.76563
m3	[1,] 3.924626	[1,] 6.096365	[1,] 7.980491	[1,] 9.906361	[1,] 11.45096	[1,] 3.823145	[1,] 5.815247	[1,] 7.417088	[1,] 9.761621	[1,] 11.28974
m4	[1,] 3.789202	[1,] 5.917037	[1,] 7.730235	[1,] 9.727789	[1,] 11.08779	[1,] 3.564816	[1,] 5.67673	[1,] 7.298204	[1,] 9.521187	[1,] 10.95164
m5	[1,] 3.489865	[1,] 5.626336	[1,] 7.512083	[1,] 9.484641	[1,] 10.96053	[1,] 3.372642	[1,] 5.499133	[1,] 7.328783	[1,] 9.353323	[1,] 10.87822
m6	[1,] 3.600914	[1,] 5.888214	[1,] 7.658195	[1,] 9.606461	[1,] 11.20949	[1,] 3.587923	[1,] 5.678765	[1,] 7.492455	[1,] 9.509447	[1,] 11.06936
m 7	[1,] 3.756236	[1,] 6.175139	[1,] 8.078263	[1,] 9.817807	[1,] 11.33958	[1,] 3.725949	[1,] 5.88386	[1,] 7.590608	[1,] 9.653202	[1,] 11.21679
m8	[1,] 3.896229	[1,] 6.270742	[1,] 8.227132	[1,] 10.01748	[1,] 11.57499	[1,] 3.879123	[1,] 5.989333	[1,] 7.77281	[1,] 9.87156	[1,] 11.44806
m9	[1,] 4.051006	[1,] 6.428718	[1,] 8.293832	[1,] 10.15079	[1,] 11.98182	[1,] 4.00979	[1,] 6.111602	[1,] 7.851647	[1,] 9.948592	[1,] 11.52521
m10	[1,] 4.151546	[1,] 6.572415	[1,] 8.386757	[1,] 10.27162	[1,] 12.11137	[1,] 4.170981	[1,] 6.342212	[1,] 7.997841	[1,] 10.201533	[1,] 11.74466
m11	[1,] 4.273664	[1,] 6.78382	[1,] 8.488474	[1,] 10.52794	[1,] 12.32497	[1,] 4.267987	[1,] 6.709107	[1,] 8.292226	[1,] 10.50673	[1,] 11.98696

Table2: *Simulation results of G^* based on data generating from the t - distribution, with B=1000and S=1000*

	n=500					n=1000				
	P = 3	**P = 5**	**P = 7**	**P = 9**	**P = 11**	**P = 3**	**P = 5**	**P = 7**	**P = 9**	**P = 11**
m 1	[1,] 12.29996	[1,] 20.49585	[1,] 28.63909	[1,] 34.02711	[1,] 43.04391	[1,] 12.01013	[1,] 19.59089	[1,] 26.22013	[1,] 33.11502	[1,] 39.81181
m2	[1,] 12.12982	[1,] 20.13359	[1,] 28.39983	[1,] 33.86294	[1,] 42.59902	[1,] 11.81753	[1,] 19.23834	[1,] 26.12188	[1,] 32.78114	[1,] 39.01307
m3	[1,] 11.76876	[1,] 19.13332	[1,] 28.00455	[1,] 33.53833	[1,] 41.98735	[1,] 11.5199	[1,] 18.90936	[1,] 25.86214	[1,] 32.61491	[1,] 38.54238
m4	[1,] 11.42619	[1,] 18.82501	[1,] 27.66542	[1,] 33.06729	[1,] 41.35479	[1,] 11.28878	[1,] 18.62719	[1,] 25.57805	[1,] 32.2982	[1,] 37.04229
m5	[1,] 11.11694	[1,] 18.44336	[1,] 27.11985	[1,] 32.63337	[1,] 40.58867	[1,] 11.02295	[1,] 18.23609	[1,] 25.18977	[1,] 31.07513	[1,] 36.54629
m6	[1,] 11.47841	[1,] 18.57905	[1,] 27.36186	[1,] 32.82941	[1,] 41.26673	[1,] 11.30727	[1,] 18.35769	[1,] 25.41594	[1,] 31.26838	[1,] 37.09329
m 7	[1,] 11.62728	[1,] 18.75889	[1,] 27.64374	[1,] 32.97378	[1,] 41.57285	[1,] 11.53266	[1,] 18.56225	[1,] 25.62621	[1,] 31.42327	[1,] 37.51142
m8	[1,] 11.85255	[1,] 19.21694	[1,] 27.85827	[1,] 33.18606	[1,] 41.87646	[1,] 11.67192	[1,] 18.87674	[1,] 25.84921	[1,] 31.60889	[1,] 37.98801
m9	[1,] 11.89033	[1,] 19.39411	[1,] 28.14385	[1,] 33.51872	[1,] 42.17695	[1,] 11.75808	[1,] 18.99001	[1,] 26.08242	[1,] 31.83831	[1,] 38.27127
m10	[1,] 11.98498	[1,] 19.67846	[1,] 28.30541	[1,] 33.84753	[1,] 42.43717	[1,] 11.80232	[1,] 19.20105	[1,] 26.11993	[1,] 32.05104	[1,] 38.54155
m11	[1,] 12.06065	[1,] 19.87633	[1,] 28.48575	[1,] 33.98255	[1,] 42.71775	[1,] 11.98942	[1,] 19.41368	[1,] 26.19449	[1,] 32.55432	[1,] 38.975703
	n=1500					**n=2000**				
	P = 3	**P = 5**	**P = 7**	**P = 9**	**P = 11**	**P = 3**	**P = 5**	**P = 7**	**P = 9**	**P = 11**
m 1	[1,] 11.82191	[1,] 18.2191	[1,] 25.39645	[1,] 29.19677	[1,] 35.38118	[1,] 11.57704	[1,] 17.35953	[1,] 22.81335	[1,] 27.27796	[1,] 33.17221
m2	[1,] 11.40695	[1,] 18.00695	[1,] 25.16909	[1,] 28.84319	[1,] 34.51317	[1,] 11.39113	[1,] 17.19853	[1,] 22.69995	[1,] 26.83599	[1,] 32.75913
m3	[1,] 11.32319	[1,] 17.86319	[1,] 24.78766	[1,] 28.52062	[1,] 34.12332	[1,] 11.09147	[1,] 16.97308	[1,] 22.19084	[1,] 26.42764	[1,] 32.27916
m4	[1,] 11.17071	[1,] 17.57071	[1,] 24.41044	[1,] 28.26803	[1,] 33.64247	[1,] 11.01516	[1,] 16.67075	[1,] 21.93583	[1,] 26.09346	[1,] 31.87327
m5	[1,] 10.83632	[1,] 17.30363	[1,] 23.84876	[1,] 27.75213	[1,] 32.99885	[1,] 10.59438	[1,] 16.18385	[1,] 21.45446	[1,] 25.69823	[1,] 31.28907
m6	[1,] 11.09162	[1,] 17.47162	[1,] 24.22937	[1,] 28.02498	[1,] 33.17972	[1,] 10.82791	[1,] 16.44866	[1,] 21.73245	[1,] 25.97966	[1,] 31.54352
m 7	[1,] 11.26305	[1,] 17.57305	[1,] 24.69212	[1,] 28.46553	[1,] 33.44042	[1,] 11.14947	[1,] 16.67603	[1,] 22.06259	[1,] 26.29839	[1,] 31.93697
m8	[1,] 11.39967	[1,] 17.83967	[1,] 24.98024	[1,] 28.68439	[1,] 33.86205	[1,] 11.25834	[1,] 16.87138	[1,] 22.52962	[1,] 26.53711	[1,] 32.45633
m9	[1,] 11.57046	[1,] 17.97046	[1,] 25.04767	[1,] 28.87317	[1,] 34.23658	[1,] 11.34438	[1,] 17.04778	[1,] 22.57845	[1,] 26.81328	[1,] 32.71663
m10	[1,] 11.58302	[1,] 18.08302	[1,] 25.17548	[1,] 29.01084	[1,] 34.77987	[1,] 11.47992	[1,] 17.254859	[1,] 22.73669	[1,] 26.98452	[1,] 32.97087
m11	[1,] 11.91491	[1,] 18.27491	[1,] 24.46822	[1,] 29.23554	[1,] 35.44901	[1,] 11.62342	[1,] 17.472782	[1,] 22.98094	[1,] 27.42144	[1,] 33.43854

Table3: *Simulation results of G^* based on data generating from the t - distribution, with B=1000and S=1000*

	n=500					n=1000				
	P = 3	**P = 5**	**P = 7**	**P = 9**	**P = 11**	**P = 3**	**P = 5**	**P = 7**	**P = 9**	**P = 11**
m 1	[1,] 4.730702	[1,] 6.440587	[1,] 8.698666	[1,] 9.66992	[1,] 10.4798	[1,] 4.609944	[1,] 5.561923	[1,] 6.69085	[1,] 7.671317	[1,] 10.05457
m2	[1,] 4.595279	[1,] 6.151284	[1,] 8.539571	[1,] 9.35593	[1,] 10.29245	[1,] 4.520492	[1,] 5.36879	[1,] 6.316035	[1,] 7.498607	[1,] 9.716908
m3	[1,] 4.461353	[1,] 6.110688	[1,] 8.345909	[1,] 9.16764	[1,] 10.00694	[1,] 4.432886	[1,] 5.293875	[1,] 6.080492	[1,] 7.291392	[1,] 9.577121
m4	[1,] 4.328387	[1,] 6.034178	[1,] 8.139764	[1,] 8.90753	[1,] 9.892805	[1,] 4.27058	[1,] 5.053182	[1,] 5.703817	[1,] 7.175401	[1,] 9.366097
m5	[1,] 3.987519	[1,] 5.600553	[1,] 7.810292	[1,] 8.64025	[1,] 9.589901	[1,] 3.868556	[1,] 4.867688	[1,] 5.232602	[1,] 7.037998	[1,] 8.828811
m6	[1,] 4.252604	[1,] 5.744107	[1,] 7.988735	[1,] 8.74749	[1,] 9.705455	[1,] 4.215528	[1,] 4.676916	[1,] 5.494466	[1,] 7.165356	[1,] 9.129322
m 7	[1,] 4.418364	[1,] 5.794934	[1,] 8.225702	[1,] 8.82848	[1,] 9.760235	[1,] 4.387074	[1,] 4.913191	[1,] 5.820618	[1,] 7.52911	[1,] 9.551239
m8	[1,] 4.60133	[1,] 5.907977	[1,] 8.480665	[1,] 8.899577	[1,] 9.830085	[1,] 4.519237	[1,] 4.947751	[1,] 5.845812	[1,] 7.659692	[1,] 9.685334
m9	[1,] 4.692101	[1,] 6.121915	[1,] 8.499383	[1,] 9. 18692	[1,] 9.996931	[1,] 4.615608	[1,] 5.042868	[1,] 6.096629	[1,] 7.808017	[1,] 9.773828
m10	[1,] 4.862112	[1,] 6.341362	[1,] 8.526994	[1,] 9.308153	[1,] 10.23854	[1,] 4.778817	[1,] 5.185046	[1,] 6.230367	[1,] 7.871182	[1,] 9.818891
m11	[1,] 4.926256	[1,] 6.381733	[1,] 6.671348	[1,] 9.46992	[1,] 10.402342	[1,] 4.838515	[1,] 5.313173	[1,] 6.420219	[1,] 7.989251	[1,] 9.997517
	n=1500					**n=2000**				
	P = 3	**P = 5**	**P = 7**	**P = 9**	**P = 11**	**P = 3**	**P = 5**	**P = 7**	**P = 9**	**P = 11**
m1	[1,] 4.556269	[1,] 5.420502	[1,] 6.282068	[1,] 7.248254	[1,] 9.201874	[1,] 4.436577	[1,] 5.112589	[1,] 6.006328	[1,] 6.98534	[1,] 8.741026
m2	[1,] 4.487011	[1,] 5.316781	[1,] 6.098826	[1,] 7.014837	[1,] 9.019786	[1,] 4.161104	[1,] 4.989444	[1,] 5.842222	[1,] 6.752992	[1,] 8.506969
m3	[1,] 4.33909	[1,] 5.28281	[1,] 5.889712	[1,] 6.861502	[1,] 8.734327	[1,] 4.057935	[1,] 4.701508	[1,] 5.591513	[1,] 6.531637	[1,] 8.39859
m4	[1,] 4.035132	[1,] 4.965479	[1,] 5.536939	[1,] 6.624582	[1,] 8.542564	[1,] 3.839037	[1,] 4.552857	[1,] 5.258528	[1,] 6.399047	[1,] 8.02845
m5	[1,] 3.829321	[1,] 4.734714	[1,] 4.992794	[1,] 6.419783	[1,] 8.363817	[1,] 3.515356	[1,] 4.357941	[1,] 4.854547	[1,] 6.072221	[1,] 7.805854
m6	[1,] 4.012566	[1,] 4.813858	[1,] 5.353734	[1,] 6.647686	[1,] 8.538686	[1,] 3.749715	[1,] 4.504114	[1,] 4.990862	[1,] 6.219625	[1,] 8.023471
m7	[1,] 4.203146	[1,] 4.906442	[1,] 5.642925	[1,] 6.837038	[1,] 8.691689	[1,] 3.852751	[1,] 4.620683	[1,] 5.171835	[1,] 6.348828	[1,] 8.400572
m8	[1,] 4.282724	[1,] 5.138298	[1,] 5.791747	[1,] 6.884102	[1,] 8.762432	[1,] 3.993814	[1,] 4.760302	[1,] 5.282013	[1,] 6.542426	[1,] 8.557887
m9	[1,] 4.323839	[1,] 5.364348	[1,] 5.968705	[1,] 7.149629	[1,] 8.901804	[1,] 4.16095	[1,] 4.835873	[1,] 6.384314	[1,] 6.615984	[1,] 8.73519
m10	[1,] 4.414005	[1,] 5.584578	[1,] 6.098948	[1,] 7.396317	[1,] 9.119673	[1,] 4.276459	[1,] 4.991057	[1,] 6.466869	[1,] 6.775572	[1,] 8.968502
m11	[1,] 4.530739	[1,] 5.67977	[1,] 6.229051	[1,] 7.566835	[1,] 9.193722	[1,] 4.320113	[1,] 5.193247	[1,] 6.585414	[1,] 6.871095	[1,] 9.064575

Table 4: *Simulation results of G^* based on data generating from the exponential distribution, with B=1000 and S=1000*

	n=500					n=1000				
	P = 3	**P = 5**	**P = 7**	**P = 9**	**P = 11**	**P = 3**	**P = 5**	**P = 7**	**P = 9**	**P = 11**
m 1	[1,] 4.236631	[1,] 6.956833	[1,] 9.569459	[1,] 11.23963	[1,] 14.24727	[1,] 4.215379	[1,] 6.331224	[1,] 8.487398	[1,] 10.32016	[1,] 12.39258
m2	[1,] 4.165496	[1,] 6.793815	[1,] 9.358805	[1,] 11.05677	[1,] 13.93626	[1,] 4.088507	[1,] 6.089767	[1,] 8.171939	[1,] 10.11203	[1,] 12.28597
m3	[1,] 4.033459	[1,] 6.580075	[1,] 9.171016	[1,] 10.87054	[1,] 13.64903	[1,] 3.910457	[1,] 5.898515	[1,] 7.879436	[1,] 9.913041	[1,] 12.07288
m4	[1,] 3.836493	[1,] 6.40576	[1,] 8.960666	[1,] 10.60927	[1,] 13.43274	[1,] 3.750463	[1,] 5.671107	[1,] 7.650116	[1,] 9.754258	[1,] 11.86045
m5	[1,] 3.550946	[1,] 6.264377	[1,] 8.635836	[1,] 10.46271	[1,] 12.00247	[1,] 3.466626	[1,] 5.343708	[1,] 7.456127	[1,] 9.528158	[1,] 11.59592
m6	[1,] 3.678461	[1,] 6.375435	[1,] 8.751865	[1,] 10.67306	[1,] 12.58112	[1,] 3.527584	[1,] 5.651321	[1,] 7.572214	[1,] 9.790123	[1,] 11.70483
m 7	[1,] 3.784227	[1,] 6.427996	[1,] 8.830345	[1,] 10.75091	[1,] 12.82983	[1,] 3.650268	[1,] 5.821993	[1,] 7.636368	[1,] 9.876039	[1,] 11.83081
m8	[1,] 3.881453	[1,] 6.54723	[1,] 8.95723	[1,] 10.79156	[1,] 13.04419	[1,] 3.750722	[1,] 5.971328	[1,] 7.786575	[1,] 9.922792	[1,] 11.94807
m9	[1,] 3.625201	[1,] 6.415476	[1,] 8.714753	[1,] 10.50851	[1,] 12.34503	[1,] 3.503387	[1,] 5.649658	[1,] 7.556911	[1,] 9.730282	[1,] 11.68453
m10	[1,] 3.795092	[1,] 6.530354	[1,] 8.852891	[1,] 10.71955	[1,] 12.64686	[1,] 3.747957	[1,] 5.897735	[1,] 7.730122	[1,] 9.838093	[1,] 11.91049
m11	[1,] 3.967638	[1,] 6.660809	[1,] 8.920967	[1,] 10.82984	[1,] 12.99325	[1,] 3.907451	[1,] 6.159524	[1,] 7.951289	[1,] 10.04907	[1,] 12.07427

	n=1500					n=2000				
	P = 3	**P = 5**	**P = 7**	**P = 9**	**P = 11**	**P = 3**	**P = 5**	**P = 7**	**P = 9**	**P = 11**
m 1	[1,] 4.14023	[1,] 6.228563	[1,] 8.174299	[1,] 10.09423	[1,] 12.3191	[1,] 3.94672	[1,] 6.053964	[1,] 8.093469	[1,] 9.571414	[1,] 11.85193
m2	[1,] 4.004789	[1,] 5.980665	[1,] 8.049061	[1,] 9.745907	[1,] 12.14962	[1,] 3.704133	[1,] 5.903224	[1,] 7.947977	[1,] 9.404284	[1,] 11.61094
m3	[1,] 3.860593	[1,] 5.750262	[1,] 7.727888	[1,] 9.544103	[1,] 11.86175	[1,] 3.684994	[1,] 5.671498	[1,] 7.693783	[1,] 9.211966	[1,] 11.49781
m4	[1,] 3.669744	[1,] 5.571535	[1,] 7.617948	[1,] 9.385922	[1,] 11.65812	[1,] 3.567931	[1,] 5.493114	[1,] 7.557738	[1,] 9.006736	[1,] 11.18713
m5	[1,] 3.413879	[1,] 5.298799	[1,] 7.392169	[1,] 9.233775	[1,] 11.44286	[1,] 3.331057	[1,] 5.236276	[1,] 7.310364	[1,] 8.78768	[1,] 10.86967
m6	[1,] 3.491447	[1,] 5.515552	[1,] 7.548119	[1,] 9.562412	[1,] 11.64721	[1,] 3.464049	[1,] 5.476489	[1,] 7.392948	[1,] 8.895508	[1,] 11.17473
m 7	[1,] 3.582478	[1,] 5.710385	[1,] 7.592593	[1,] 9.617596	[1,] 11.78102	[1,] 3.526669	[1,] 5.594662	[1,] 7.529141	[1,] 8.981402	[1,] 11.32276
m8	[1,] 3.657484	[1,] 5.768587	[1,] 7.672802	[1,] 9.782667	[1,] 11.84738	[1,] 3.597154	[1,] 5.723161	[1,] 7.578119	[1,] 9.07928	[1,] 11.42323
m9	[1,] 3.479269	[1,] 5.472085	[1,] 7.513611	[1,] 9.547046	[1,] 11.58061	[1,] 3.448512	[1,] 5.412652	[1,] 7.470007	[1,] 8.811331	[1,] 11.10998
m10	[1,] 3.609051	[1,] 5.75502	[1,] 7.682839	[1,] 9.763699	[1,] 11.79559	[1,] 3.559228	[1,] 5.528202	[1,] 7.646152	[1,] 8.960704	[1,] 11.39236

Applications on Real Data

To check the method of selecting best choice of m in *m-out-of-n*bootstrap, Ten (10) real data sets are taken from the R-DATASETS. The data sets are 'fgl', 'Boston', 'Fishing', 'Crime', 'Student', 'College', 'Forest fires', 'RiceFarms', 'Weather' and PatentsHGH. Details of the Variables of the data sets are given in the appendix. Short information related to the data sets are as below.

Boston: -The data is related to the Housing values in Greenbelts of Boston (Harrison, and Rubinfeld. 1978). The data contains 506 observations. It has 13 independent variables with the "rate of crime per capita" as dependent variable.

Crime: -It is a linear regression data shows Crime in North Carolina (Baltagi, 2006). The data is collected in the United States of America from 1981 to 1987. It contains 630 regional observations. The data has 23 independent variables with one dependent variable "crimes committed per person".

fgl: - It is a linear regression data shows the Measurements of Forensic Glass Fragments (Venables, and Ripley, 2002). The data has 214 observations and 10 variables. The data is collected in the United States of America from 1981 to 1987. It contains 630 regional observations. The data has 09 independent variables with one dependent variable "refractive index". The data was collected by B. German on fragments of glass collected in forensic work.

Fishing: -The data is related to the choice of fishing mode(Herriges, and Kling, 1999). It is linear regression data collected in United States of America. The data contains 1182 observations. The data has 11 independent variables and one dependent variable "monthly income".

Forest fires: -One of the major environmental concern to occur is the Forest fires(Amatulli, Peréz-Cabello, de la Riva, 2007). It is also called wildfires. Due to wildfire the forest preservation is affected. It also ecological, economic damage, and cause for human suffering. The data has 513 observations with 13 variables. The only dependent variable is "the burned area".

Student: Source of the data is the Education Longitudinal Study of 2002(Ingels, 2002). The data is collected from the 10^{th} grade students. Data contains 752 schools for checking its Hypothetical student-level. There are 9,679observations with 17 variables. There is no missing observation in the data. Dependent variable of the data is "math score".

RiceFarms: - The data is related to the production of Rice in the Indonesia country (Mariyono, 2014). 1026 observations were collected from*langan, malausma, wargabinangun, sukaambit, gunungwangi*

*and ciwangi*of Indonesia. 17 variables are included in the data. "Price of rough rice per kg" is the dependent variable of the data.

PatentsHGH: -This data indicate the Dynamic Relation between R & D and Patents. The data contains 1730 observations with 15 variables. The data were collectedin United States from 1975 to 1979.

College: - The College data is collected from many US Colleges. The data is related to the US News and World Report of 1995 issue. A data matrix of 777 observations with 18 variables are included in the data. "Graduation rate" is taken as the dependent variable. This dataset was used for ASA Statistical Graphics Section's 1995 Data Analysis Exposition.

Weather: -The weather dataset is collected in 2016-17 in various cities of United States of America. The data has 3655 observations and 15 variables. In the data "High Temperature" is the dependent variable. The data was downloaded from Weather Underground in January 2018.

Simulation results for each of the data set are given in *Table.5*. All the results for $m_i \ \ (i=1,2...11)$ are compared to the slandered result of $m = n$.For each dataset B = 1000 bootstrap replications are used and for each $m_i \ \ (i=1,2...11)$, the results of G^* are computed. In *Table.5*,the results of G^* for all the nine data sets are given. The results are summarized as follow,

- The *results* of G^* for 11 different choices of m for fgl dataset are given in the first column of *Table. 5*. The data has 10 variables with 214 observations with no missing observation.From the results we can see that for each $m_i \ \ (i=1,2,3,4,5)$, the value of G^* decreases. The value of G^*gives minimum result at m_5. Now as the sample size increases beyond m_5, i.e. $m_i \ \ (i=6,...11)$ the value of G^*also increases. This shows that in this dataset the best choice for *m* in *m-out-of-n* bootstrap is m_5.

- The results for G^* of *Boston* data are shown in the second column of *Table.5*. In this dataset we have 14 variables and each variable contain506 observations.For all choices of m, the value of G^* is computed. From the results again it is clear that the value of G^*decreases from $m_i \ \ (i=1,2,3,4,5)$ and gives minimum result at m_5 and as the value of m exceeds m_5, the value of G^*also increases. Here again m_5 seems the minimum value for all the 11 choices of mand is considered as the best choice for m in the rest of the $m_i \ (1,2,...4,6,...11)$ choices.
- Column 4 & 5 of *Table 5*, consists of dataset "*forest fires*" and "*Crime*". Each dataset having 517 observations with 12 & 23 variables respectively. In both the cases, we observe that the value of G^* has minimum result at m_5 . This shows that the choice of m in *m-out-of-n* bootstrap is robust to the number of variables.

- Similarly, column 6 & 7 of *Table .5 consists of* the data set "*College*" and "*RiceFarms*". Each the datasets has the same number of variables and different number of observations. But in both the cases, the value of G^* has a minimum value for m_5 as compared to the all the other choices of $m_i \ (1,2,...4,6,...11)$. This indicates that choice of m in *m-out-of-n* bootstrap is robust to the sample size.

- Four different data sets were analyzed in the last four column of *Table.5*. The value of G^* is computed for each of the 11 different choices of *m.* Each data set has different number of samples

and different number of variables. In each case the value of G^* has minimum value for m_5 as compared to all other values of m_i $(i = 1,2,...4,6,...11)$ .This means that m_5 is the best choice for m in *m-out-of-n* bootstrap.

Table 5: *Simulation results of G^* based on TEN real data sets, with B=1000.*

	fgl Data, n = 214 p=9	Boston Data, n = 506 p=13	forest fires n = 517 p=12	Crime Data, n = 517 p=23	College Data, n = 777 p=17	Rice Farms n = 1026 p=17	Fishing n = 1182 p=11	Patents HGH n = 1730 p=15	Weather n = 3655 p=14	Student n = 9679 p=12
m1	12.25851	19.08306	14.14871	30.1178	23.35234	28.28096	10.7003	17.3711	16.48415	12.39353
m2	10.20964	15.0214	13.94845	26.01766	21.02459	21.67415	10.30961	15.6005	16.20127	12.18855
m3	9.926174	13.91374	13.51994	24.32427	19.76231	19.92632	10.30924	15.44638	15.2795	12.09386
m4	9.712122	13.84631	12.1275	23.95359	17.82219	16.19464	10.18947	15.30109	15.11734	11.99153
m5	9.119285	12.83847	11.40994	22.51791	15.5592	16.89808	9.874201	14.74359	14.11572	11.18829
m6	9.351633	13.21453	11.74593	22.71786	15.89744	17.35493	10.04206	14.80735	14.26396	11.48057
m7	9.360707	13.21754	12.1832	23.02301	16.23878	17.67569	10.04947	14.90583	14.43688	11.2828
m8	9.389025	13.25694	12.2071	23.06259	16.93731	17.86537	10.06904	15.10413	14.54635	11.55546
m9	9.436968	13.34787	12.71816	23.22827	17.2592	18.43474	10.0855	15.14799	14.59379	12.03657
m10	9.466206	13.77501	12.73642	23.267	17.55316	19.09708	10.10739	15.22051	14.619	12.35747
m11	9.554179	13.93344	12.75143	23.62927	17.74638	19.62817	10.15719	15.25454	15.0557	12.77217

Conclusion

The basic theme of this study was to select the optimal value of m in *m-out-of-n* bootstrap. Extensive simulations studies have been conducted to estimate the optimal *m* in *m-out-of-n* bootstrap on the data set generated from the different distributions. In each study ELEVEN different choices of *m* were considered. The same study was carried out using TEN real data sets. From the results and analysis of the study, we observed that $m_5 = (1.73 / pi) * n$ was the best choice in all the given choices. Based on the findings of the study we conclude that if 55 % of the sample size is used, it will give the minimum value for G^* and consequently will result in the best value of *m*.Moreover, in selecting the optimal *m*in*m-out-of-n*, we can increase the number of choices of *m* for further investigating the selection of best choice of *m*.

References

1. Abadie, A. & Imbens, G. W. (2008). On the failure of the bootstrap for matching estimators. Econometrica, 76(6), 1537-1557.
2. Amatulli, G., Peréz-Cabello F., de la Riva, J. (2007) Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. Ecological Modelling 200, 321–333.
3. Baltagi B. H. (2006). "Estmating an economic model of crime using panel data from North Carolina." Journal of Applied Econometrics, 21(4).
4. Bickel, P., Götze, F. & van Z, W. (1997). Resampling fewer than n observations: gains, losses and remedies for losses. Statist. Sinica, Springer, 7, 1-31.
5. Efron, B. (1979). Bootstrap methods. another look at the jackknife. Ann. Statist, 7, 1-26.
6. Efron, B. & Tibshirani, R.J. (1993). An introduction to the Bootstrap. New York: Chapman and Hall.
7. Hall, P. (1992). The bootstrap and edgeworth expansion. N.Y: Springer Verlag.
8. Harrison, D., and Rubinfeld, D. L.(1978). "Hedonic Housing Prices and the Demand for Clean Air." Journal of Environmental Economics and Management 5 (1): 81–102.
9. Herriges, J. A. and Kling, C. L. (1999) "Nonlinear Income Effects in Random Utility Models", Review of Economics and Statistics, 81, 62-72.
10. Higham, N. J. (1990): "Analysis of the Cholesky decomposition of a semi-definite matrix." 161-185.
11. Mariyono, J. (2014): Pest Management Science: formerly Pesticide Science 64 (10), 1069-1073.
12. Peter, J. B. and Anat, S (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. Statistica Sinica 18, 967-985
13. Račkauskas, F. Götze & A. (2001). Adaptive choice of bootstrap sample sizes. In , . . State of the Art in Probability and Statistics, 286-309.

14. Silvia Goncalves and Timothy J. Vogelsang (2011). Block bootstrap HAC robust tests: The Sophistication of the naive bootstrap. Econometric Theory , Volume 27, Issue 4, pp. 745 – 791
15. Ingels, S. J. (2002)Education Longitudinal Study of Base Year Data File User's Manual. NCES 2004-405.
16. Venables, W. N. and Ripley, B. D. (2002). "Modern Applied Statistics with S. Fourth edition". Springer.