Methods and Issues in Trimming Extreme Weights in Sample Surveys

Frank Potter and Yuhong Zheng
Mathematica Policy Research, P.O. Box 2393, Princeton, NJ 08543

Abstract

In survey sampling practice, unequal sampling weights (the inverse of the selection probabilities) can be both beneficial and deleterious. Extreme variation in the sampling weights can result in excessively large sampling variances when the data and the selection probabilities are not positively correlated. In addition, extreme variation in the weights can result from unplanned subsampling, nonresponse adjustments, or post-stratification. In some survey situations, the survey statistician may impose a trimming strategy for excessively large weights. Because of the weight trimming, the survey statistician will usually expect an increased potential for a bias in the estimate and a decrease in the sampling variance. The ultimate goal of weight trimming is to reduce the sampling variance more than enough to compensate for the possible increase in bias and, thereby, reduce the mean square error. In this paper, we discuss some current methods used to identify the appropriate trimming values and provide guidance on selecting the final trimming level, which may be different from the values suggested by the algorithms.

1. Introduction

In survey sampling practice, unequal sampling weights can be developed by design through over-sampling and under-sampling of specific subpopulations in a sampling frame and reflect the original selection probabilities. Unequal sampling weights can also arise from operational issues (what may require subsampling) or from nonresponse adjustment and post-stratification procedures used to produce the final analytic weight. The unequal weights can be both beneficial and deleterious. Extreme variation in the sampling weights can result in excessively large sampling variances, especially when the data and the weights are positively correlated. In such situations, a few extreme weights can offset the precision gained from an otherwise well-designed and executed survey. On the other hand, if a negative correlation exists between the data and the weights for the sampling units, the variation in the weights can result in reductions in the sampling variances, so that this variation will be beneficial.

In practice, several post-design procedures are used to limit or reduce the number and size of extreme sampling weights (Potter 1990). A summary of weight trimming procedures used in international assessments of adult competencies is given in a paper presented at 2014 JSM (Van de Kerckhove et al. 2014). The practices and procedures fall into three categories:

1. Procedures used to avoid or minimize the number and size of extreme weights by trimming or limiting components of the weights during the weight computation process; and
2. Procedures in which the size of the weights are controlled as part of the nonresponse adjustment procedure.
3. Procedures used to identify, trim, and explicitly compensate for extreme sampling weights are implemented after the weights are fully computed.

In the first type of procedures, while established criteria are used to limit the number of extreme weights, some extreme values may still arise. In the second category of procedures, the nonresponse adjustments are developed using a restriction on the weights so that, in principle, few, if any, weights will have an extreme value (Deville and Särndal 1992). In the third category of procedures, a trimming strategy generally includes a procedure to determine excessive weights and a method to distribute the trimmed portion of the weights among the untrimmed weights. Because of the weight trimming, the survey statistician will usually expect an increased potential for bias in the estimate and a decrease in the sampling variance. Hence, a trimming strategy may reduce the sampling variance for an estimate but increase the mean square error (the sampling variance plus the bias squared). The ultimate goal of weight trimming is to reduce the sampling variance more than enough to compensate for the possible increase in bias and, thereby, reduce the mean square error (MSE).

In this research, we investigated some current procedures that utilize the final adjusted sampling weights. In the empirical study, the procedures are demonstrated in a setting where the population can be fully enumerated. The specific empirical goal is to evaluate the procedures in terms of bias, sampling variance reduction, and mean square error as well as the consistency and variability of trimming levels using a data base containing data that are correlated or uncorrelated to the sampling weights. The simulation study contains 1,000 samples of 1,000 cases selected with probability proportional to a size measure from the tract-level file of the 2014 Planning Database contains selected 2010 Census and selected 2008-2012 5-year American Community Survey (ACS) estimates.[1] The tract-level file contained information on 72,890 tracts. The final count of tracts used in the simulation was 72,516 tracts with 374 tracts with populations counts of zero or one excluded.

2. Methods

A. Overview

Four weight trimming procedures are discussed in this paper. The procedures are (1) the contribution to entropy procedure, (2) the procedure based on a distribution model for the weights, and (3) the median weight plus 4 or 5 times interquartile range of the weights. The first two procedures are described in Potter (1990) and the median weight plus 4 or 5 times interquartile range of the weights is discussed in the context of the National Immunization Survey (Chowdhury et al. 2007 and Van de Kerckhove et al. 2014). None of the procedures use survey data in determining the trimming level. Two weight trimming procedures that utilizes the survey data (The Taylor series procedure and the estimated MSE procedure) are described in Potter 1990. An alternative version of the contribution to entropy procedure used with the 1983-1984 National Assessment of Educational Progress (NAEP) and uses data, is described by Johnson et al. (1987). The current weight trimming procedure used in NAEP for the student weight uses a multiple of the median of the weights.[2]

[1] See https://www.census.gov/research/data/planning_database/2014/

[2] https://nces.ed.gov/nationsreportcard/tdw/weighting/2008/ltt_weighting_2008_trimming_adjustments.aspx .

B. Trimming Procedures

1. The Contribution to Entropy Procedure

The contribution to entropy procedure uses the comparison of the contribution of each weight to the sampling variance of an estimate by systematically comparing the individual weights to a value computed from the average of the squared weights for the sample. If a weight is above the computed value, the weight is assigned this value and the other weights are adjusted to have the new weights sum to the original weight total. The average of the squared adjusted weights is computed again and used in a second comparison of each individual adjusted weight. The procedure is repeated until all adjusted weights are below or equal the value based on the sum of the adjusted squared weights. A variation of this procedure has been reported in conjunction with the NAEP and, for simplicity, it is referred to as the NAEP procedure in this paper.

In the NAEP procedure, the relative contribution is limited to a specific value by comparing the square of each weight to a multiple of the sum of the squared weights. That is,

$$w_k^2 \le c \sum w_k^2 / n \quad \text{or}$$

$$w_k \le K_n . \qquad (1)$$

where $K_n = (c \sum w_k^2 / n)^{1/2}$.

The value for c is arbitrary and can be chosen empirically by looking at the distribution of the square root of the values of

$$n\, w_k^2 \le \sum w_k^2 .$$

In the NAEP algorithm, each weight in excess of K_n is given this value and the other weights are adjusted to reproduce the original weight sum. The sum of square adjusted weights is computed and each weight is again compared using equation (1). The procedure is performed repeatedly until none of the weights exceed this criterion.

The use of this procedure is documented in a methodological report of the NAEP study (Benrud et al. 1978) and, in this report, c is assigned a value of 10. That is the square of any weight is less than 10 times the average of the squared weights. In the NAEP 1983-84 Technical Report (Johnson et al. 1987), an analogous weight trimming procedure is described that uses data (estimated student counts). In this report, an empirical method is described to determine a value for c; c was assigned a value of 10. Smaller or larger values of c will generate different trimming levels.

2. Weight Distribution

This trimming procedure is based on an assumed distribution for the sampling weights. If the selection probabilities are assumed to follow a Beta distribution, the sampling weight distribution can be shown to be of a form that is essentially an inverse of a beta variate.

In this procedure, the parameters for the sampling weight distribution are estimated using the sampling weights and a trimming level is computed that has a pre-specified probability of occurrence, based on the distribution model. Sampling weights in excess of this trimming level are trimmed to this level and the excess is distributed among the untrimmed weights. The parameters for the sampling weight distribution are then estimated using the trimmed adjusted sampling weights and a revised trimming level is computed that has the pre-specified probability of occurrence. The trimmed adjusted sampling weights are then compared to the revised trimming levels. If any weights are in excess of this trimming level, they are trimmed to this level and the excess is distributed among the untrimmed weights. This weight trimming procedure identifies and trims sampling weights with a small probability of occurrence, based on the model.

The key result is that, when a standard beta distribution is assumed for the single draw selection probabilities, the distribution model for the sampling weights (w) is in a form of a beta distribution. The density function for the distribution is

$$fw_k(w) = n\,(1/n\,w)^{\alpha+1}\,(1 - 1/n\,w)^{\beta-1} / B(\alpha,\beta)$$

for $1/n < w < \infty$

where

$$B(\alpha,\beta) = \Gamma(\alpha)\,\Gamma(\beta) / \Gamma(\alpha + \beta)$$

Estimates for alpha and beta can be computed from the sample size, the mean weight, and the variance of the weights. That is,

$$\hat{\alpha} = [\overline{w}\,(n\overline{w} - 1) / n\,s_w^2] + 2$$

$$\hat{\beta} = (n\overline{w} - 1)\,[\overline{w}\,(n\overline{w} - 1) / n\,s_w^2 + 1]$$

Where

$$\overline{w} = \sum w_i / n$$
$$s_w^2 = \sum (w_i - n\overline{w})^2 / n$$

The percentiles for the cumulative distribution function ($F_w(w)$) for the distribution can be computed using the standard Beta distribution (Beta(x,α,β)) where

$$\text{Beta}(x,\alpha,\beta) = \int_0^x (1 - u)^{\beta-1}\, u^{\alpha-1}\, du / B(\alpha,\beta)$$

The values for the cumulative distribution function of the weight distribution $F_w(W)$ is

$$F_w(x_o) = 1 - \int_0^{1/nx} (1 - u)^{\beta-1}\, u^{\alpha-1}\, du / B(\alpha, \beta)$$

The weight distribution trimming procedure compares the distribution of the weights relative to the theoretical distribution. The probability of weights as large or larger than an observed weight (w_k) is given by

$$I - F_w(w_k) = \text{Beta}(1 / n\, w_k, \hat{\alpha}, \hat{\beta}).$$

A weight value with an extremely low probability of occurring can be trimmed to a specific probability of occurrence.

For the empirical study, the probability of occurrence criterion was set at 0.01; that is, a weight with a value in excess of w_{op} where $1 - F(w_{op}) = 0.01$ was trimmed to w_{op}.

3. Median Plus Multiple of the Interquartile Range

By definition, the distribution of the weights with some extreme values are skewed so it is natural to consider a measure of dispersion that is expected to less affected by extreme values. Various authors have discussed the use of a multiple of the interquartile range (IQR) as a measure of dispersion for sampling weights with some extreme values (Chowdhury et al. 2007 and Van de Kerckhove 2014). For this study, we will used the median plus either 4 or 5 times the IQR.

3. Simulations

A. Overview

The goals of the empirical study are to investigate and evaluate weight trimming procedures using multiple data items from a population that can be fully enumerated. The performance measures used in the empirical study include

1. the change in the estimated variance of the estimate (that is, how much variance reduction is achieved),
2. the extent of bias introduced
3. the change in the mean square error of the estimate (that is, whether the bias introduced by these procedures offsets the variance reduction), and
4. the average and variance of the trimming levels (that is, whether these procedures result in consistent trimming levels over repeated samples) and
5. the accuracy of coverage probabilities for a 95 percent confidence interval.

These performance measures will be described below.

B. Empirical Study Design

For the empirical study, we used the 2014 Planning Database (PDB) from the US Census Bureau.[3] The PDB includes a range of housing, demographic, socioeconomic, and census operational data that was compiled for survey and census planning. Data includes selected Census and selected American Community Survey (ACS) estimates. Data are provided at both the census block group and the tract levels of geography; only tract-level data were used for the empirical study. As noted previously, the tract-level file contained information

[3] See https://www.census.gov/research/data/planning_database/2014/

on more than 72,890 tracts. The final count of tracts used in the simulation was 72,516 tracts with only tracts with populations counts of zero or one excluded.

We selected 1,000 samples of 1,000 tracts using the probability proportional to the total counts of persons in the tract based on the 2010 Census. The Chromy sequential selection procedure was used with implicit stratification based on the total counts of persons in the tract based on the 2010 Census (Chromy 1979). As shown in the top row of Table 1, the average of the weights across the 1,000 sample of a 1,000 tracts is 526 with a median of 377. Three samples had a weight of more than 10,000 with the largest weight of 12,864. The design effect from unequal weights (see the top row of Table 2) also shows this variation with an average design effect of 1.35 and a maximum value of 23.8.

For estimating totals and means, the five variables were chosen from the PDB because of the varying levels of estimated correlation between the data items and the sampling weight across the 1,000 samples.

1. Count of persons between the ages 45-64 in 2010 (2010 Census)
2. Count of persons 65 or older in the tract in 2010 (2010 Census)
3. Median household income for the tract in 2010 (2010 Census)
4. Median value of house in the tract in 2010 (2010 Census)
5. Count of unemployed civilians among persons 16 or older in the tract (American Community Survey)

We expected the tract-level count variables to be highly correlated to the size measure and negatively correlated with the weights (that is larger counts were associated with smaller weights). The correlations between the data and the weights ranged from -0.67 for the count of persons between the ages 45-64 in 2010 and 0.12 for the count of unemployed civilians among persons 16 or older.

Although some of the trimming procedures can be used iteratively to identify a final trimming level, a single trimming level was computed for each of the 1,000 samples for each procedure.

C. Summary of Results

The findings of the empirical study show that, of the four procedures, the two procedures using the median plus a multiple of the interquartile range performed very consistently and produced almost the same trimming value for all samples. The NAEP procedure and the weight distribution procedure were affected by the value of the largest weight in the sample.

In Figure 1, we show the plot of the trimming levels relative to the largest weight in a sample and the average, minimum and maximum trimming level is shown in Table 1. For both the NAEP procedure and the weight distribution procedure, the trimming level computed is affected by the size of the largest weight. For the NAEP procedure the largest trimming level was 1,312 and this occurred with the sample that had the largest weight. Similarly the largest trimming level for the weight distribution procedure was with the sample with the largest weight. The NAEP procedure resulted in trimming between 1 and 4 weights, whereas the weight distribution procedure resulted in trimming 1 to 12 weights.

For the two procedures using the median and the IQR, the trimming level showed almost no variation regardless of the size of the largest weight in a sample. When the median plus 4 IQR is used, the trimming level was approximately 206 and for the median plus 5 IQR, the trimming levels varied between 240 to 242. This finding can be expected because the samples are selected from the sample population with relatively large sample sizes and the median and IQR in each sample will reflect the median and IQR in the population. The consistency of the trimming levels affected the number of weights trimmed. The median and 4 IQR procedure resulted in trimming between 8 and 11 weights, whereas the median and 5 IQR procedure resulted in trimming 3 and 5 weights.

In table 2, we show the design effects from unequal weighting. All four trimming procedures resulted in an average design effect of around 1.22 with the NAEP procedure and the weight distribution procedure showing some variation, but the median plus the IQR showed again almost no variation.

The effect of the trimming on the point estimates, the variances and the mean square error is shown in tables 3 and 4 and figure 2. In table 3, the estimated relative bias was computed as

Relative Bias = 100 * (True Value – Estimated Value) / True Value

The relative bias is consistently less than 1 percent with some effects shown when there is a positive or negative correlation between the weights and the data.

The relative variance (RelVar) estimates in table 3 were computed as

RelVar = 100 * (Trimmed Estimated – Untrimmed Estimate) / Untrimmed Estimate

where the trimmed estimate is the variance using the trimmed weights and untrimmed estimate is the variance using the untrimmed weights. The purpose of the weight trimming is to reduce the variance while avoiding the introduction of bias in the estimates. For 3 of the 5 variables (those with low correlations between the data and the weights), the weight trimming does decrease the relative variance. However, when a stronger correlation exists between the data and the weights, the benefits of trimming appears to be negated and an increase in the relative variances is shown.

In table 4, we show the percentage of the 95 percent confidence intervals that included the true value. Except for the variable with the strong negative correlation between the data and the weights, the coverage probabilities are very good, essentially rounding to 95 percent in most cases.

For the relative mean square error, the effect of the correlation between the data and the weights show a greater impact. Again for 3 of the 5 variables with low correlations, the relative mean square error is decreased, but when a stronger correlation exists the trimming may adversely affect the mean square error. In figure 2, we show the impact of the correlation and the relative bias, the relative variance, and the relative mean square error.

4. Conclusions

In survey sampling practice, an analyst may encounter a survey with substantial variation in the sampling weights and a few extreme weights. Before implementing a weight trimming procedure, the analyst should evaluate whether the sampling weight variation has beneficial or deleterious effects on the sampling variances. When observed survey data are negatively correlated with the sampling weights and extremely large weights are associated with very small data values, the sampling variances computed using the original weights can be smaller than the sampling variances computed using equal or trimmed sampling weights. Therefore, weight trimming is not needed and may result in increased sampling variance as well as biased estimates. However, if the extremely large weights are determined to have adverse effects, weight trimming is a reasonable strategy to reduce the estimated sampling variances.

In terms of the four weight trimming procedures as evaluated in the empirical study, the NAEP and the weight trimming procedure produce trimming levels that are affected by the size of the largest weight(s) in the sample survey. It is possible that through an iterative protocol, trimming levels would be lowered for the samples with extremely large weights. The procedures using the median and a multiple of the IQR appear to be unaffected by the size of the largest weight and a specific trimming level is computed immediately. A paper by Van de Kerckhore et al. (2014) provides a summary of procedures using the IQR and Chowdhury et al. (2007) discusses procedures for developing trimming levels using the IQR (as well as a procedure using an exponential distribution model for the tail of the weight distribution).

The analysis of the data from the PDB clearly pointed out the effect of a correlation between the data and the weights. In surveys with the potential strong correlations between the data and the weights (such as in business surveys), weight trimming should be evaluated with available data or historical data to avoid increasing the sampling variance and potentially also introducing bias in the survey estimates.

The primary conclusion based on the empirical study results is that weight trimming can have both positive and negative effects. The positive effects (for example, the improvement in the interval estimates) occurred for some variables when little or no bias is introduced. However, for some data, the estimates using trimmed weights may be biased and the weight trimming can result in misleading point and interval estimates. All of the procedures resulted in reductions in the estimated sampling variance. However, all procedures also resulted in an increase in the estimated sampling variance for at least some of the 200 replicated samples. Therefore, the survey analyst needs to be cautious when trimming sampling weights because, unless weight trimming is conducted carefully and evaluated for various data items, larger sampling variance or substantial bias can result for some survey estimates.

References

Benrud, C.H., et al. (1978). "Final Report on National Assessment of Educational Progress: Sampling and Weighting Activities for Assessment Year 08." prepared for National Assessment of Educational Progress, Research Triangle Park, N.C.: Research Triangle Institute.

Chowdhury S., M. Khare, K. Wolter (2007). "Weight Trimming in the National Immunization Survey." *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 2651-2658
Chromy, J. R. (1979). "Sequential Sample Selection Methods." *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 401-406
Deville, J.C. and C.E. Särndal. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376–382.
Johnson, E.G. et al. (1987). "Weighting Procedures in Implementing the New Design." in The NAEP 1983-1984 Technical Report, ed. A. E. Beaton, Princeton, N.J.: National Assessment of Educational Progress, 493-504.
Potter, F. (1990), "A Study of Procedures to Identify and Trim Extreme Sampling Weights." *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 225-230.
Van de Kerckhove, W., L. Mohadjer, T Krenzke. (2014), "A Weight Trimming Approach to Achieve a Comparable Increase to Bias across Countries in the Programme for the International Assessment of Adult Competencies." *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 655-666.

Table 1. Largest value weights before and after trimming

	Average	**Minimum**	**Maximum**	**Standard Deviation**
Untrimmed	526	282	12,864	795.5
NAEP	262	251	1,312	61.8
Weight Dist'n	207	193	561	33.3
Median + 4 IQR	206	205	206	0.2
Median + 5 IQR	241	240	242	0.2

Table 2. Design effect from unequal weights before and after trimming

	Average	**Minimum**	**Maximum**	**Standard Deviation**
Untrimmed	1.35	1.22	23.75	1.203
NAEP	1.23	1.22	1.45	0.012
Weight Dist'n	1.21	1.20	1.28	0.007
Median + 4 IQR	1.21	1.20	1.21	0.001
Median + 5 IQR	1.22	1.21	1.22	0.001

Table 3. Average percentage change in relative bias, relative variance and relative mean square error (MSE) for selected data items and trimming methods

	Correlation[a]	NAEP	Weight Dist'n	Median + 4 IQR	Median + 5 IQR
Relative Bias					
Variable 1	-0.67	0.40	0.71	0.70	0.46
Variable 2	-0.41	0.39	0.69	0.67	0.45
Variable 3	-0.14	0.02	0.09	0.08	0.03
Variable 4	-0.05	0.01	0.03	0.03	0.01
Variable 5	0.12	-0.22	-0.37	-0.37	-0.25
Relative Variance					
Variable 1	-0.67	0.88	1.92	1.85	1.08
Variable 2	-0.41	0.66	1.06	1.06	0.76
Variable 3	-0.14	-3.34	-4.97	-4.86	-3.70
Variable 4	-0.05	-2.08	-3.80	-3.70	-2.44
Variable 5	0.12	-7.33	-11.79	-11.42	-8.26
Relative MSE					
Variable 1	-0.67	203.2	256.7	298.9	252.9
Variable 2	-0.41	34.7	43.8	50.8	43.3
Variable 3	-0.14	-3.1	-4.3	-4.2	-3.4
Variable 4	-0.05	-1.9	-3.4	-3.3	-2.2
Variable 5	0.12	-6.8	-10.0	-9.7	-7.5

[a] The correlation is between the data and the size of the sampling weight.

Note: Variables are:

1 Persons between the ages 45-64 in 2010 (2010 Census)
2 Persons 65 or older in 2010 (2010 Census)
3 Median household income in 2010 (2010 Census)
4 Median value of house in 2010 (2010 Census)
5 Unemployed civilians among persons 16 or older (American Community Survey)

Table 4. Percentage of 95 percent confidence intervals covering the true value for selected data items and trimming method[a]

	Untrimmed Weights	NAEP	Weight Dist'n	Median + 4 IQR	Median + 5 IQR
Variable 1	94.7	90.2	83.7	82.6	88.5
Variable 2	95.8	94.4	93.7	93.4	94.0
Variable 3	94.5	94.6	94.5	94.5	94.7
Variable 4	94.8	95.1	94.7	94.7	95.0
Variable 5	95.4	95.4	95.4	95.4	95.0

[a] 95 percent confidence interval is $\hat{\theta}\ + 1.96\ Standard\ Error\ (\hat{\theta}\ )$

Note: variables are:

1 Persons between the ages 45-64 in 2010 (2010 Census)

2 Persons 65 or older in 2010 (2010 Census)

3 Median household income in 2010 (2010 Census)

4 Median value of house in 2010 (2010 Census)

5 Unemployed civilians among persons 16 or older (American Community Survey)

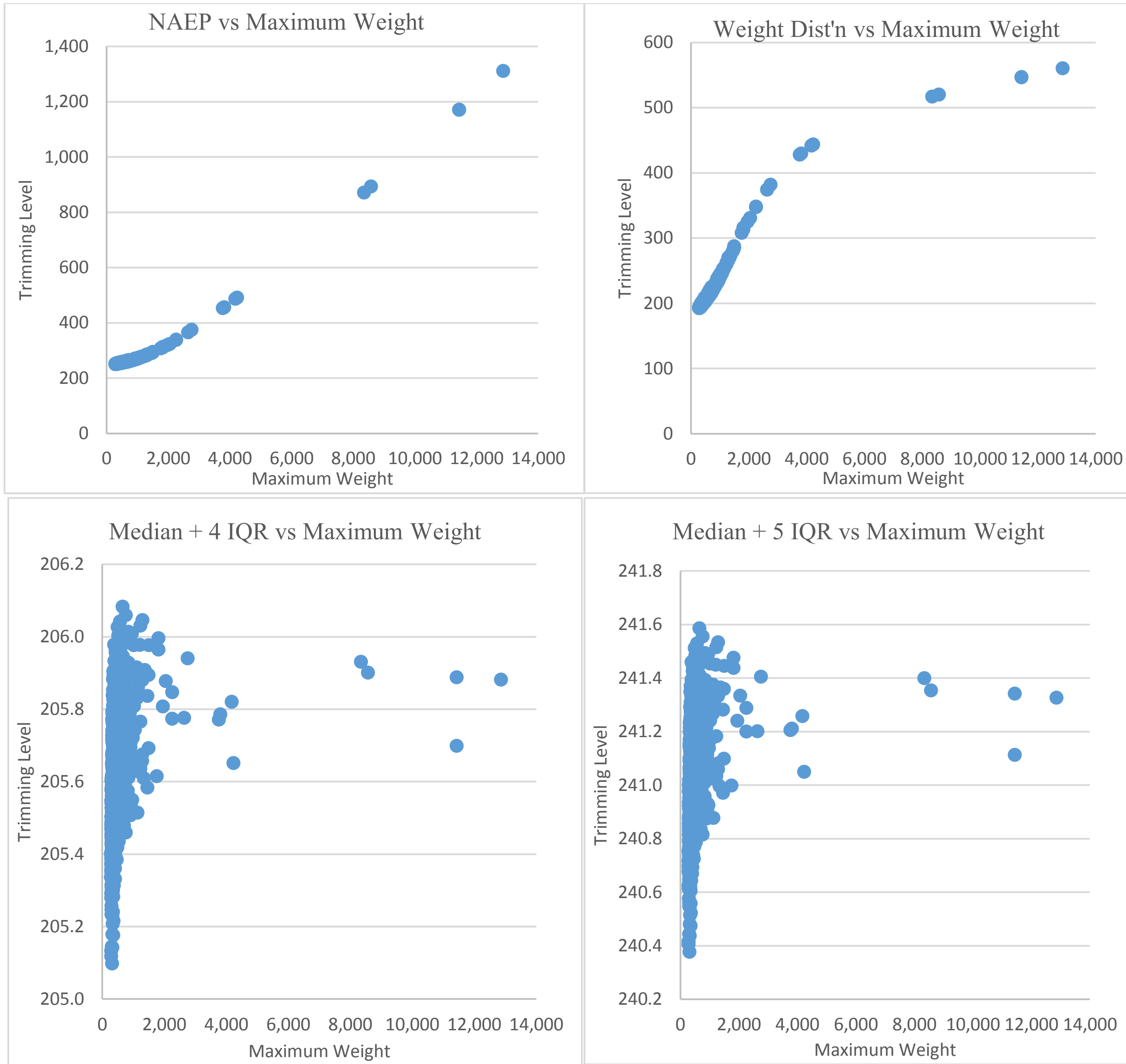

Figure 1. Maximum Weight and Trimming Levels

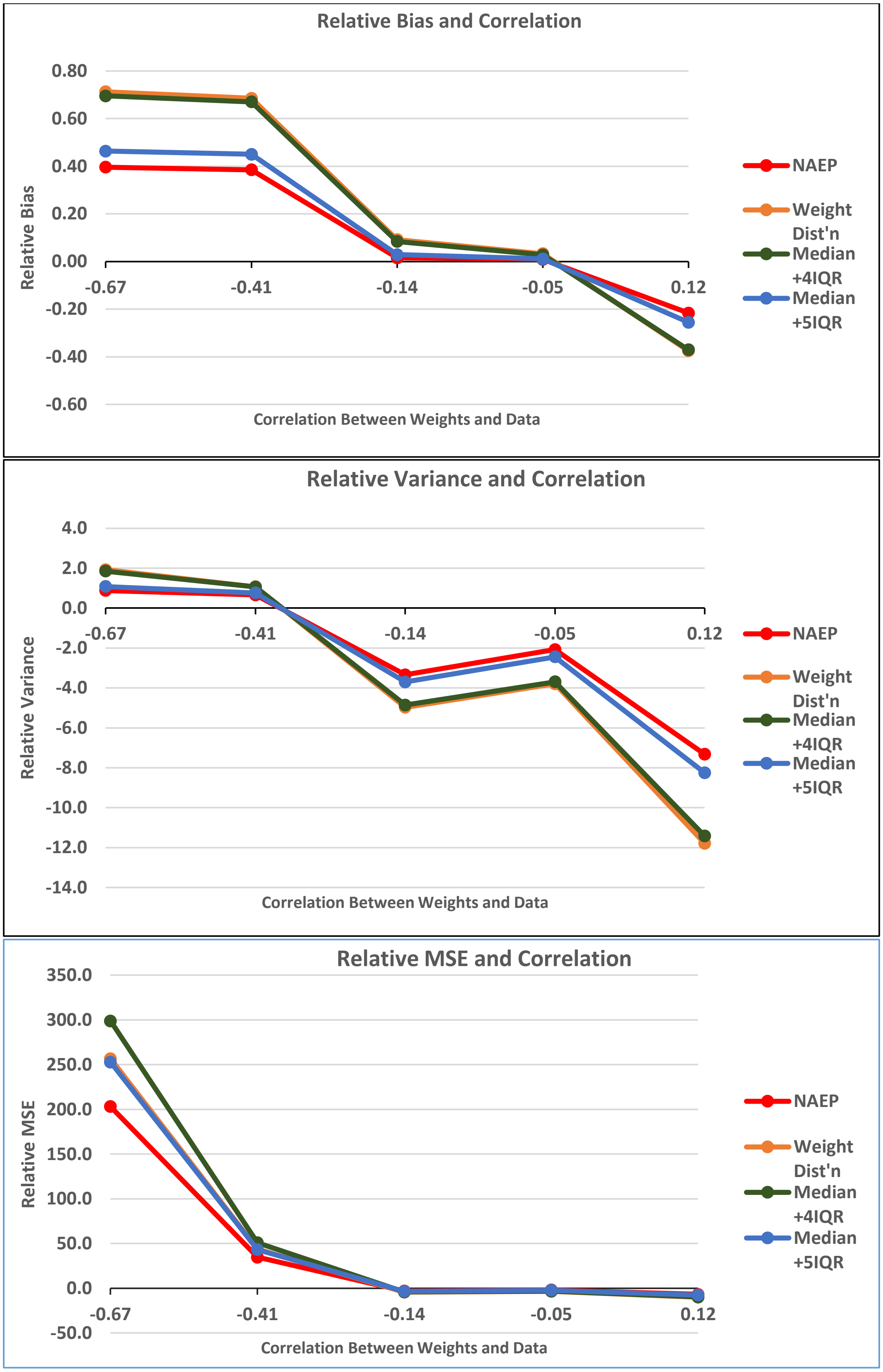

Figure 2. Relative bias, variance and mean square error by trimming method and correlation