METHODS FOR EXTREME WEIGHTS IN SAMPLE SURVEYS

by

Francis J. Potter

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1875T

May 1990
Methods for Extreme Weights in Sample Surveys

by

Francis J. Potter

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill, 1990

Approved by:

[Signatures]

Advisor
Reader
Reader
FRANCIS J. POTTER. Methods for Extreme Weights in Sample Surveys
(under the direction of William D. Kalsbeek.)

ABSTRACT

In survey sampling practice, planned and unplanned variation in
the sampling weights can result in inflated sampling variances. As a
result, extreme sampling weights are sometimes trimmed to reduce the
sampling variance. However, when sampling weights are trimmed, a bias
can be introduced into the survey estimates. The goal of sampling
weight trimming is to reduce the sampling variance while avoiding the
introduction of substantial bias; thereby reducing the mean square
error of the point estimates. Because of the bias potential, the
coverage probability for interval estimates can also be affected.

The two currently used procedures are basically ad hoc, with
little theoretical basis and no empirical evaluations of the effects
of their weight trimming. This research summarizes the two currently
used procedures, provides an analytical framework for weight trimming,
and proposes three alternative procedures. In this research an
empirical investigation of the weight trimming procedures uses 200
replicated samples of 100 units from a data base for 2,989 counties in
the U.S. The effect of each procedure is measured in terms of the
sampling variance, the bias introduced, the estimated mean square
error, and the coverage probability for interval estimates.
Of the five weight trimming procedures investigated, three procedures use observed survey data and two use only the sampling weights to identify extreme sampling weights. Of the three procedures that use observed data, one of the proposed procedures results in less bias and equivalent variance reduction relative to a comparable currently used procedure based on minimizing an estimate of the mean square error. The proposed procedure utilizes a Taylor series linearized variate for a conditional estimate of a mean when trimmed weights are used and a measure of the bias introduced. The two procedures that use only the sampling weights include a new procedure that uses a distribution model for the sampling weights, and a currently used procedure that compares each weight to a multiple of the average squared weights. Of these two procedures, the new procedure is preferable because the criteria for the procedure can be determined for prespecified probability levels.
ACKNOWLEDGEMENTS

I wish to express appreciation to my advisor Bill Kalsbeek and to Ralph Folsom. Bill Kalsbeek has provided encouragement, direction, and support throughout the years and Ralph Folsom has provided many useful comments and insights into survey sampling design and estimation issues. I wish also to thank Dr. Norman L. Johnson for his assistance and the other members of my committee, Gerardo Heiss, Gary Koch, and B. V. Shah.

I wish to thank my wife, Linda, for her encouragement, understanding, and support, especially during the last two years of this effort. Without her encouragement, this effort would not been completed. I wish to thank my stepsons, Bill and Jay, whose understanding and own personal achievements has been a source of support for me.

Finally, I wish to thank the Research Triangle Institute (RTI) and my colleagues at RTI. Especially, I wish to thank B. J. York for her support and programming assistance, Rick Williams for programming assistance, and Trudi Averette and Pat Parker for typing assistance. I also want to express my appreciation of the RTI management which supported my graduate studies at UNC and which awarded me a Professional Development Award to provide financial support for the completion of this dissertation and to continue research in this area.
# TABLE OF CONTENTS

## I. INTRODUCTION

A. Overview .................................................. 1
B. Research Goals ........................................... 5
C. Sources of Unequal Sampling Weights .................. 7

## II. LITERATURE REVIEW

A. Overview .................................................. 11
B. Effects of Unequal Weights ............................... 11
C. Current Practices and Procedures ....................... 15

## III. ANALYTICAL DERIVATIONS

A. Overview .................................................. 29
B. Expectation and Variance of an Estimator of the Mean Using Trimmed Weights ......................... 31
C. A Generalized Design Effect Measure for Unequal Weights ..................................................... 58
D. A Distribution Model for Sampling Weights .......... 64

## IV. WEIGHT TRIMMING PROCEDURES

A. Overview .................................................. 76
B. Comparison of Proposed and Current Procedures ..... 78
C. Proposed Procedures ..................................... 79
D. Currently Used Procedures ............................... 96

## V. EMPIRICAL INVESTIGATION

A. Overview .................................................. 100
B. Summary of Results ..................................... 102
C. Design of the Empirical Study ......................... 104
D. The Effect of Trimming on the Weights ............... 107
E. Changes in Estimated Sampling Variances ............. 112
F. Change in Estimated Mean Square Error ............... 116
G. Coverage Probabilities .................................. 124

## VI. SUMMARY, CONCLUSION, AND FUTURE RESEARCH TOPICS

A. Summary .................................................. 130
B. Conclusion ............................................... 135
C. Future Research Topics ................................ 138
LIST OF TABLES

Table 1. Approximate Absolute Bias Bounds For Population Proportion........................................ 43

Table 2. Summary of Characteristics of Trimming Procedures...... 80

Table 3. Summary of Adjustments to Sampling Weights from 200 Replicated Samples.........................108

Table 4. Skewness of Untrimmed and Trimmed Sampling Weights Based on Weights from Individual Samples...............111

Table 5. Percentage Change in Variance From Weight Trimming Based on Estimates from 200 Replicated Samples.......113

Table 6. Summary of Sampling Variances from Trimming Procedures.............................................115

Table 7. Percentage Change In Mean Square Error From Weight Trimming Based on MSE Estimates from 200 Replicated Samples..............................................................118

Table 8. Summary of Mean Square Error from Weight Trimming Procedures..........................................120

Table 9. Comparison of Bias, Sampling Variance, and Mean Square Error Averages over 200 Replicated Samples....122

Table 10. Percentage Relative Standard Errors Based on Estimates from 200 Replicated Samples...............125

Table 11. Percentage of Confidence Intervals Covering True Value Based on 200 Replicated Samples.................126

Table 12. Comparison of Actual and Estimated Design Effects......129
Chapter I
INTRODUCTION

A. Overview

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 variance when the data and the selection probabilities are not positively correlated. A few extreme weights can offset the precision gained from an otherwise well-designed and executed survey design. However, if a positive correlation exists between the data and the selection probability for the sampling units, the extreme weights can result in reductions in the sampling variances, that is the extreme weights are beneficial. Unplanned extreme variation in the sampling weights may result from the sample selection procedure, inaccuracies or errors in frame data, the nonresponse compensation procedures, or other sources.

The purpose of this research effort is to explore the effect of weight trimming procedures both analytically and empirically and to propose alternative approaches that will utilize more information available to the research analyst. The underlying goal of this research is to provide a research statistician or analyst with a better understanding of the effect of trimming and improved tools to assess whether trimming is needed and how to evaluate which weights impact the precision.
The analytical investigation seeks to demonstrate the impact of weight trimming on the expectation of the estimate and its variance. The empirical investigation of the effect of weight trimming assesses the current procedures in a somewhat controlled setting and evaluates components of the proposed procedure. In this research, I also propose a systematic approach to weight trimming that utilizes more of the information that may be available to the statistician. This approach is designed:

1. to assist the statistician in determining if weight trimming is needed by assessing the relationship between the observed data and the sampling weights;

2. to provide a procedure by which a statistician can determine which observed data or auxiliary data should be used in investigating if weight trimming is needed; and

3. to provide a method for determining extreme values in the weighted data that can adversely affect the precision.

The empirical investigation utilizes the proposed approaches (as well as current procedures) to refine the overall strategy for weight trimming.

Unequal sampling weights can result from probability proportional to size sampling, from inaccurate or out-dated sampling information in multi-stage surveys, or from nonresponse compensation procedures. Various weight trimming procedures are currently used by government and private survey research organizations. While identifying procedures in current use (Potter 1989), I found that individual
survey groups tended to know of only the procedure used by their group and, for some national health related surveys, the sampling weights were trimmed using ad hoc trimming rules. Weight trimming and the potential effect of this trimming are rarely discussed in methodological reports (the notable exception is the Current Population Survey (Hanson 1978)) and exploratory investigations, either analytical or empirical, are not present in these reports.

In practice, several procedures are used to limit or reduce the number and size of extreme sampling weights. The practices and procedures fall into two categories that depend on whether the trimming is implemented during the weight computation process or after the weights are fully computed:

1. procedures used to avoid or minimize the number and size of extreme weights are implemented during the weight computation process; and

2. procedures used to identify, trim, and explicitly compensate for extreme sampling weights are implemented after the weights are fully computed.

The most notable use of procedures to avoid or minimize the number and size of extreme weights is the Census Bureau's Current Population Survey (CPS) and the Consumer Expenditure Survey (CES). In the CPS, the Census Bureau limits the size of the noninterview adjustment factor and the first-stage ratio adjustment factor so that extreme weights are less likely to occur (Hanson 1978, Bailar, et al. 1978). In the CES, the Census Bureau also sets a limit on an intermediate weighting factor (Alexander 1986). Any weight trimming here is compensated for implicitly by post-stratification.
In most survey situations, the final adjusted sampling weights are analyzed for extremely large sampling weights. In some of these situations, the survey statistician may impose a trimming strategy for excessively large weights. A trimming strategy will generally include 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 a bias in the estimate and a decrease in the sampling variance. In sum, 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.

One of the currently used weight trimming procedures implicitly utilizes the possible correlation between the data and the selection probability that may exist when probability proportional to size (pps) sampling is used. This potential correlation is incorporated through repeated calculation of the estimated sampling variance. However, this procedure does not use the potential relationship to identify extreme weights. The usual measure of the design effect (DEFF) attributable to unequal weights (based on Kish's (1965, Section 11.7) DEFF concept) is sometimes used to determine if extreme weights may exist, but this statistic assumes no correlation between the data and the selection probability.
Folsom (1984) described a use of the single draw variate (SDV) in survey sampling. The SDV is the observed data divided by the single draw selection probability. That is, the SDV $y_k$ for the $k$th unit in a sample of size $n$ is

$$y_k = \frac{Y_k}{p_k} = n \times w_k \times Y_k$$

where

- $Y_k$ = the observed data,
- $p_k$ = the single draw selection probability, and
- $w_k$ = the sampling weight

$$w_k = \frac{1}{n p_k}$$

Because the SDV is a direct function of both the data and the sampling weight, any beneficial correlation (that is, a negative correlation) between the data and the sampling weight will be accounted for in the SDV. Therefore, the SDV is a useful measure for investigating problematic extreme weights.

### B. Research Goals

In this dissertation, the proposed research goals can be categorized into three sets: analytical derivations, the proposed new approaches, and the empirical investigation. The analytical derivation goals are the following.

1. Develop an analytical form for the effect of weight trimming in terms of the bias and sampling variance.

2. Develop an estimator of the variance for a trimmed weight estimator of the population mean.
3. Develop a generalized measure \( (DEFF_{WG}) \) of the DEFF attributable to unequal weighting that incorporate the possible correlation between the data and the selection probability.

4. Develop the distribution for sampling weights which can be used to determine weight trimming criterion levels.

The goal of the new procedure development is to develop alternate extreme weight identification and trimming procedures that utilize more of the information potentially available to the researcher. The first procedure is based on the Taylor series approximation to the bias and the variance. I will call this the Taylor series procedure because it utilizes the Taylor series linearized variate. The second procedure utilizes the single draw variates (SDVs) and will be labeled as the SDV procedure. The third procedure will use a probability distribution derived for the sampling weights. This procedure (called the weight distribution procedure) is proposed for situations when weight trimming must be performed before any data are available for use in the other procedures. These procedures will be based on the sampling methods and methods for outlier detection.

The empirical study goal in this research is to provide a demonstration of the current and proposed procedures in a setting when the population can be fully enumerated. The specific empirical goals are the following:

1. Using a data base containing data that are correlated or uncorrelated to the sampling weights as a population, empirically evaluate the two current primary algorithms and the proposed procedures in terms of bias, sampling variance...
reduction, and mean square error as well as the consistency and variability of trimming levels.

2. Empirically investigate a generalized measure of the DEFF attributable to unequal weighting.

C. Sources of Unequal Sampling Weights

Unequal sampling weights may result from design features or from unplanned or unexpected occurrences experienced during the study. The unequal weights can be both beneficial or deleterious in the same survey or in repeated samples from the same population. Statistical surveys are often designed for one objective but are used to provide information to address multiple topics, sometimes not directly related to the original objective. In some samples from a given population, a sample design can result in samples with minor or negligible weight variation and in other samples from the same population, the weights may exhibit substantial variation. More specific examples of possible sources of unequal sampling weights are the following:

1. Probability proportional to size (pps) sampling;
2. Disproportionate stratified sampling;
3. Adjustments for differential nonresponse or poststratification;
4. Inaccurate frame information in multi-stage designs; and
5. Sampling weights for analysis units defined at intermediate stages in a multi-stage design.

More specifically, these five sources contribute to unequal sampling weights in the following ways.
1. **Probability proportional to size (pps) sampling:**

   In pps sampling, units are selected relative to a size measure. In finite population sampling, the Horvitz-Thompson estimator of a population total, \( \hat{Y} \), is of the form
   \[
   \hat{Y} = S \frac{Y_k}{\pi_k},
   \]
   where
   - \( Y_k \) = the observed value for the kth unit,
   - \( \pi_k \) = the probability of selection the kth unit in a sample of size \( n \), and
   - \( S \) = the sum over the \( n \) units in the sample.

   The inverse of the selection probability (\( 1/\pi_k \)) is the sampling weight (\( w_k \)) for the ith unit. If the selection probability for a sampling unit is directly proportional (and the sampling weight is inversely proportional) to the magnitude of the observed value (\( Y_k \)) for the unit, then the sampling variance for the estimator, \( \hat{Y} \), equals zero (Raj 1968). In surveys where a pps sampling strategy is used and the observed data are not positively correlated with the unit selection probabilities, the sampling variance will be larger than that expected for a simple random sample of the same size.

2. **Disproportionate stratified sampling:**

   Disproportionate stratified sampling can be viewed as a form of pps sampling that uses the square root of the within-stratum variance component as the size measure of units in the stratum. Assuming constant cost across all strata such a design attempts to be optimal in the sense of Neyman allocation by assigning a
relatively small sample (and low selection probabilities) to strata with small per-unit variance contributions and a relatively large sample (and high selection probability) in strata with large per-unit variances. For survey data with design variances that match the optimizing model, the stratification will result in increases in the precision of estimates. On the other hand, when the unit variance components are reasonably equal across strata and the equally weighted proportional allocation is near optimal, unequal weights resulting from the disproportionate allocation can substantially inflate the variance of estimates relative to an unstratified simple random sample.

3. **Adjustments for differential nonresponse or poststratification:**
   Ratio adjusting sampling weights to account for nonresponse generally entails the assignment of sampling units into classes of 'similar' units (Oh and Scheuren 1983). Nonresponse adjusted sampling weights are the product of the sampling weight and the inverse of the weighted or unweighted response rate within the class. Unequal weights will result from differences in response rates among classes. Similarly, poststratification adjustment of the sampling weight can also further contribute to increased variance arising from unequal weights because a beneficial relationship (that is, a positive correlation) between the selection probability and the data can be reduced.

4. **Inaccurate frame information in multi-stage designs:**
   In multi-stage sampling designs (such as area household surveys), when first stage units are selected pps, inaccurate or outdated
frame information can result in unequal weights for final stage units. For example, many multi-stage household survey designs are designed to give equal selection probabilities to all sample households. In a multi-stage household survey, the first stage unit may be an area segment (such as a census block) and the segments are selected with pps to the most recent Census count of households in the segment. A current count of households in the sampled segment is obtained and the households are listed to form a sampling frame. If the count of households in the segment is different from the count used to select the segment and the within-segment sample of households cannot be modified (increased or decreased) sufficiently to preserve equal weights, then the sampled households will have unequal weights. If the difference between the counts used to select the segments and the current counts are substantially different for large proportion of the segments then the large variation may be present in the sampling weights.

5. **Sampling weights for analysis units defined at intermediate stages in a multi-stage design:**

In multi-stage designs when pps selection is used in all but the last stage of sampling, the sampling weights for all but the last stage will generally be unequal by design. Data for a unit at an intermediate stage may not have a proportional relationship to the size measure and therefore, estimates using these data can suffer substantial loss of precision from the unequal weights.
Chapter II

LITERATURE REVIEW

A. Overview

The literature review presented in this chapter will address topics which in one way or another relate to the research objectives set forth in this research. The topics are on the following:

1. the design effect (DEFF) attributable to unequal weighting and Folsom's formulation of components of the design effect; and

2. a description of current procedures for controlling or identifying and trimming extreme sampling weights.

B. Effects of Unequal Weights

Kish (1965) introduces the concept of the design effect (DEFF) for a sample survey. The DEFF is defined as the ratio of the variance \( \text{Var}_d(\hat{Y}_d) \) of an estimator when the actual design is taken into account to the variance \( \text{Var}_s(\hat{Y}_s) \) of an estimator when a simple random sample of the same size is selected. That is,

\[
\text{DEFF} = \frac{\text{Var}_d(\hat{Y}_d)}{\text{Var}_s(\hat{Y}_s)}.
\]

The DEFF from unequal weighting alone can be described for unstratified and stratified designs using simplifying assumptions. In the sense of a superpopulation model (assuming a constant unit variance for all observations, \( \text{Var}(Y_k) = \sigma^2 \)), the DEFF from unequal weighting (DEFF\(_w\)) can be represented for the design-based
estimator (\( \hat{Y} \)) of the mean by

\[
\text{DEFF}_w = \frac{\text{Var}_d(\hat{Y}_d)}{\text{Var}_s(\hat{Y}_s)}
\]

where

\[
\hat{Y}_d = \frac{\sum w_k Y_k}{\sum w_k},
\]

\[
\hat{Y}_s = \frac{\sum Y_k}{n},
\]

and

\[
S = \text{the summation over the sample of } n \text{ units}.
\]

The usual formulation for the DEFF attributable to unequal weighting can be derived as follows.

Note that

\[
\text{Var}_d(\hat{Y}_d) = \text{Var}_d\left(\frac{\sum w_k Y_k}{\sum w_k}\right)
\]

\[
= \sum w_k^2 \frac{\text{Var}(Y_k)}{(\sum w_k)^2}
\]

\[
= \sum w_k^2 \frac{\sigma^2}{(\sum w_k)^2}
\]

\[
= \sigma^2 \left(\frac{\sum w_k^2}{(\sum w_k)^2}\right).
\]

For the denominator of the DEFF

\[
\text{Var}_s(\hat{Y}_s) = \frac{\sigma^2}{n},
\]

where

\[
n \text{ is the common sample size}.
\]

The DEFF for unequal weighting (\( \text{DEFF}_w \)) can be represented as

\[
\text{DEFF}_w = \frac{\text{Var}_d(\hat{Y}_d)}{\text{Var}_s(\hat{Y}_s)}
\]

\[
= \left[ \frac{\sigma^2}{\sum w_k^2 / (\sum w_k)^2} \right] / \left[ \frac{\sigma^2}{n} \right]
\]

\[
= n \sum w_k^2 / (\sum w_k)^2.
\]

Analogously, for stratified designs with \( n_h \) sample members in stratum \( h \) and equal weights in each stratum, then the DEFF for the estimator \( \hat{Y}_{ds} \) can be represented as (assuming again constant unit variance)

\[
\text{DEFF}_{ws} = \frac{\text{Var}_d(\hat{Y}_{ds})}{\text{Var}_s(\hat{Y}_s)}
\]
\[ n \left[ \frac{\sum_{k} w_k^2}{\sum_{h} n_h} \right] = \frac{n \sum_{h} w_h^2}{\left( \sum_{h} n_h w_h \right)^2} \]

where

\( w_h \) = the sampling weight in stratum \( h \).

These representations of the design effect attributable to unequal weighting can be useful in assessing extreme weights. However, these representations do not incorporate the potential positive effect of the correlation between the data and the sampling weight.

In Folsom and Williams (1981) and Williams, Folsom, and LaVange (1983), the DEFF was partitioned into components representing design features such as multi-stage clustering, stratification, without replacement sampling, and unequal weights. The research was conducted as part of the National Assessment of Educational Progress (NAEP), a national study of the knowledge, skill, understanding, and attitudes of young Americans, ages 9, 13, or 17. NAEP employed a three stage unequal probability stratified design. In investigating methods for the analysis of the NAEP data and evaluating procedures proposed by Fellegi (1980) and Rao and Scott (1981) for adjusting test statistics from standard statistical software packages for survey design effects, Folsom demonstrated in the context of this design (2 primary sampling units selected without replacement in each primary stratum) that the overall DEFF effect can be partitioned into the effects of the without-replacement primary unit selection and the effects of the primary stratification. When estimating and analyzing proportions, Folsom also derived a component of the design effect that represents
the effect of unequal weighting. That is, for a categorical survey outcome variable with R response levels, the generalized design effect \( \Delta \) is described in matrix terms (R-1 x R-1) as

\[
\Delta = W \{ I + \bar{m} \Delta [- \Gamma_S^2 + (I - \Gamma_S^2) \Gamma_B^2] + (\bar{m} - 1) [\Delta + (I - \Delta) R_w^2] \}
\]

where

\( \bar{m} \) = average number of second stage units in each primary sampling unit (PSU);

\( I \) = the identity matrix (R-1 x R-1);

\( \Delta \) = the fraction of the total variation that is accounted for by the between PSU covariance component matrix;

\( \Gamma_S^2 \) = the PSU-level stratification effect matrix;

\( \Gamma_B^2 \) = the average (over the strata) of the between PSU cross-correlation effect matrix resulting from paired without replacement selection within primary strata;

\( R_w \) = the negative cross-correlation between second stage sampling units induced by without replacement selection; and

\( W \) = the effect of unequal weighting.

The form for the unequal weighting DEFF incorporated a component that is comparable to the usual estimator of the DEFF for unequal weighting and a component that he used to demonstrated (in the situation of a dichotomous response distribution) the effect of selecting population members with the response of interest with higher probability.
C. Current Practices and Procedures

In recent years, a number of articles and books (especially the three-volume *Incomplete Data in Sample Surveys* (Madow, et al. 1983)) have addressed the issues of missing data, response errors, weight adjustment, and imputation. However, a literature search and personal contact with various major survey research organizations have identified little documentation of weight trimming procedures currently in use. In 1988, I prepared a paper (Potter 1989) that included descriptions of the currently used procedures. The general trimming strategies identified are as follows.

1. Procedures to Minimize Number and Size of Extreme Sampling Weights

The Bureau of the Census uses procedures to reduce the variation of sampling weights in the Current Population Survey (CPS) and the Consumer Expenditure Survey (CES) (CPS: Hanson 1978, Bailar et al. 1978, Little 1986b, Scheuren 1986, CES: Alexander 1986). In the CPS methodology report (Hanson 1978) and in Alexander's (1986) discussion of the weighting methods used for Consumer Expenditure Surveys (CES), the sampling weights for the CPS and the CES are controlled by limiting the size of a components of the sampling weight. In the CES, at one stage of the sampling weight computation, the sampling weights are composed of the product of the base weight and a weighting control factor. To summarize the Census Bureau weighting procedure, the base weight is a first-stage weight for a area unit. The weighting control factor takes into account changes to the sampling rates that result
(a) from changes in the size measure of the second stage unit since
the last census (for example, substantial growth in a second stage
unit) and (b) from deviations from an overall sampling rate. In these
areas where increases in the size measure of the second stage unit
since the last census, subsampling may be employed to maintain the
desired workload. The weighting control factor takes this subsampling
into account. For the CES, Alexander reports that the weighting
control factor was arbitrarily limited to a magnitude of 8 until 1984
and, since 1985, limited to a value of 4. (The reason for the use of
a value of 4, or the previous value of 8, is not discussed by
Alexander.) The excess weight that was 'lost' was accounted for by
poststratification in a later step of the weight computation
procedure. Alexander states that, in other Census Bureau surveys,
similar limitations are set on intermediate weighting factors.

The Census Bureau also limits the size of some noninterview
adjustments in some surveys. For the CPS (Hanson 1978), if
noninterview weighting classes contain either less than 20 cases or
the noninterview adjustment factor is greater than 2.0, then a
restricted form of weighting class collapsing is used until these
requirements are achieved. The restriction allows for the collapsing
across race (white and not-white) but not across area of residence
categories. In some instances, these requirements can not be achieved
and the Census Bureau limits the weighting class noninterview
adjustment to a value of 3.0. In the CPS, the Census Bureau also
limits the first-stage ratio adjustment factor to a value of 1.3.
The use of such limits sometimes introduces some debate (Little 1986b, Scheuren 1986). Little (1986b) questions the use of ad hoc cut-off values. Scheuren (1986), in discussing Little's paper, indicates that cut-off values are usually developed by considering the cost/benefit tradeoffs between reduced bias for some data items and increased variances for other items. Because of periodic nature of the CPS, CES and other Census Bureau surveys, the Census Bureau can develop procedures using a historical perspective generally unavailable to survey statisticians conducting one-time surveys.

To provide insight into the effect of these procedures, let us consider only trimming of the weighting control factor in a simplified situation with complete response. For the kth sampled unit, let

\[ w_{kb} = \text{the base weight}, \]
\[ f_{kc} = \text{the untrimmed weighting control factor}, \]
\[ f'_{kc} = \text{the trimmed weighting control factor} \]

(that is, \( f'_{kc} = \begin{cases} f_{kc} & \text{if } f_{kc} \text{ less than 4, or} \\ 4 & \text{if } f_{kc} \text{ greater than 4} \end{cases} \)),

\[ r = \text{the ratio adjustment factor when no trimming is imposed}, \]

and

\[ r' = \text{the ratio adjustment factor when trimming is used}. \]

Then for the fully trimmed weight, \( w_{kt} \),

\[ w_{kt} = w_{kb} \times f'_{kc} \times r' \]

and the untrimmed weight, \( w_{ku} \),

\[ w_{ku} = w_{kb} \times f_{kc} \times r. \]

To estimate a total, \( Y \), from a sample of sample of size \( n \), \( Y_k \), \( k=1,2,.., n \), the Horvitz-Thompson estimator, \( \hat{Y}_t \), using the trimmed
weights is

\[ \hat{Y}_t = S w_{kt} \ast Y_k \]
\[ = S w_{kb} \ast f_{kc} \ast r' \ast Y_k \]

and using the untrimmed weights is

\[ \hat{Y}_u = S w_{ku} \ast Y_k \]
\[ = S w_{kb} \ast f_{kc} \ast r \ast Y_k. \]

The difference between \( \hat{Y}_t \) and \( \hat{Y}_u \) is

\[ \hat{Y}_t - \hat{Y}_u = S(w_{kt} - w_{ku}) Y_k \]
\[ = S w_{kb} Y_k (f_{kc} r' - f_{kc} r). \]

Define \( A = \{k: f_{kc} \text{ is not trimmed}\} \) and

\( B = \{k: f_{kc} \text{ is trimmed}\}. \)

Then

\[ \hat{Y}_t - \hat{Y}_u = S \sum_{k \in A} w_{kb} Y_k f_{kc} (r' - r) + S \sum_{k \in B} w_{kb} Y_k (4r' - f_{kc} r). \]

From this representation, the following observations can be made.

a. The \( f_{kc} \) will be greater than 4 in areas where the value of the size measure is a substantial underestimate of the size of the area (that is, the area experienced high growth since the last census). In areas of very high growth, the \( f_{kc} \) may be substantially larger than 4.

b. The ratio adjustment factors \( r \) and \( r' \) will compensate for a portion of the trimmed value. However, the ratio adjustment strata are generally broad classes of the population so the trimmed excess of \( w_{kc} \) will be distributed across a broad portion of the sample.
c. By viewing the persons in these high growth areas as a domain, this domain will be underestimated as the result of these procedures.

d. These procedures require the availability of external data for developing the ratio adjustments.

In CPS Design and Methodology Report (Hanson 1978), the Census Bureau describes the potential for bias from some of these procedures.

In summary, the procedures used by the Census Bureau provide various examples of how and where extreme weights can be avoided by limiting factors included in the weights. The Census Bureau sets limits on subsampling weights in areas experiencing growth in number of housing units, redefines weighting classes to limit adjustment factor, sets limits on nonresponse adjustment factors, and sets limits of ratio adjustment factors. Some possible effects of these procedures include those discussed above and the following:

a. groups with substantially different response experience would be combined by redefining and collapsing of weighting classes to reduce adjustment factors;

b. persons with relatively low response propensity are not being represented appropriately when the nonresponse adjustment factors are limited; and

c. persons and households in areas of high response propensity may be over-represented because poststratification is used to account for the weight and weight adjustment factor limitation.
Some of these issues related to the nonresponse adjustments may result in negligible bias effect because of the high response rates generally achieved in Census Bureau surveys.

2. Procedures to Trim and Compensate for Extreme Sampling Weights

Procedures for weight trimming and subsequent compensation differ on the basis of the amount and type of information used to determine a level of trimming. Three specific strategies were identified: (1) an inspection strategy, (2) a strategy involving the computation of an estimated mean square error for selected items, and (3) a strategy involving the relative contribution of extreme weights to the overall variance.

a. Inspection Strategy

Some form of inspection of the weight distribution is generally conducted regardless of whether trimming procedures are planned. At the Research Triangle Institute, the sum, mean, variance, coefficient of variation and selected percentiles are usually computed to describe the weight distribution. The coefficient of variation (CV) among the sampling weights is a useful descriptive measure because of its relationship to the constant unit variance model for the design effect for unequal weighting (DEFF_w). That is, for a sample of n cases

$$DEFF_w = 1 + \left( \frac{n - 1}{n} \right) \times (CV)^2.$$  

$$= 1 + (CV)^2 \text{ for large } n.$$  

In addition to these descriptive statistics, the 25 largest and 25 smallest weights are listed along with components to the weight (such
as the initial or prior stage weight, the nonresponse adjustment factor, and post-stratification adjustment factors). This listing generally identifies a sufficient portion of the tails to ascertain the essential characteristics of the weight distribution. The largest 25 weights will include all weights that are likely to be trimmed and the 25 smallest weights can provide an indication of the skewness of the distribution.

In most cases, the large weights that can affect (increase) the sampling variance can be easily identified because these weights will differ from the other weights by a substantial amount. In some cases, inspection of the listing of the largest 25 weights can identify logical trimming limits. Assuming that an adjustment is made to the untrimmed weights so that the original weight sum is preserved, the effect of the trimming on $\text{DEFF}_W$ can be determined by multiplying $\text{DEFF}_W$ by the adjustment to the untrimmed weights (denoted by $A_S$) and subtracting a value based on the trimmed weights. This value is the sum of the square of the trimmed weight(s) adjusted by $A_S$ and the square of the trimming limit divided by the squared weight sum. If only the $n_t$ largest weights in a sample of size $n$ are trimmed, then the adjustment factor $A_S$ is given by

$$A_S = \left[ \frac{\sum_{k=1}^{n} w_k - (n_t w_0)}{n} \right] / \left[ \frac{\sum_{k=1}^{n} w_k - S w_k}{n_t} \right]$$

where

- $w_k = \text{original weight for the trimmed case(s)}$ and
- $w_0 = \text{the trimming weight limit}$.

That is, the revised design effect, $\text{DEFF}'_W$, is
\[
\text{DEFF}_W^t = A_s^2 \text{DEFF}_W - \{ n \left[ A_s^2 \sum w_k^2 - (n_t w_o^2) \right] \} / (W_s)^2
\]

where

\[
W_s = \text{the sum of the weights.}
\]

The problems associated with the inspection strategy are easily apparent. First, the relationship between key data items and the weights are not used. Second, the procedure is subjective and the choice of a trimming limit is arbitrary. Third, the effects on the sampling variance and bias of the estimates are unknown. However, this procedure can be implemented both inexpensively and quickly. When data for computing the effect of the trimming may not be available and because of time or funding constraints exist, this procedure may be justified.

b. **Estimated Mean Square Error (MSE) Trimming**

A method used by some of the major survey research organizations is the evaluation of an estimate of the mean square error for selected data items at various trimming levels to empirically determine the trimming level (Cox and McGrath 1981, Cox 1988, Heeringa 1988). In this procedure for determining cut-off values for weights, the statistician conducts a visual inspection of the distribution of the sampling weights. A set of key data items are identified and an estimate of the mean square error is calculated for the key data items at different candidate cut-off values. The values of the estimated mean square error can be plotted versus the cut-off values to determine a cut-off value that achieves adequate reductions in the estimated mean square error for all or most of the data items.
In this procedure, the trimmed excess is distributed across the untrimmed weights to reproduce the original weight sum.

The assumption underlying this procedure is that for a set of weights and data, a point exists at which the reduction in the sampling variance resulting from the trimming is offset by the increase in the square of the bias introduced into the estimate. The mean square error (MSE for an estimator, \( \hat{\gamma} \), before any trimming strategy is used) is

\[
\text{MSE}(\hat{\gamma}) = \text{Var}(\hat{\gamma}) + (\text{Bias}(\hat{\gamma}))^2.
\]

where

\[
\text{Var}(\hat{\gamma}) = \text{the sampling variance of the estimator } \hat{\gamma}
\]

and

\[
\text{Bias}(\hat{\gamma}) = \text{the difference between } E(\hat{\gamma}) \text{ and the true value, } \gamma.
\]

In the implementation of this procedure, an estimate of the mean square error, MSE(\( \hat{\gamma}_c \)) is computed for each data variable at each cut-off value \( c \). Assuming that \( \hat{\gamma} \) is an unbiased estimate of \( \gamma \), the MSE(\( \hat{\gamma}_c \)) can be estimated by noting that

\[
E(\hat{\gamma}_c - \hat{\gamma})^2 = \text{Var}(\hat{\gamma}_c) + \text{Var}(\hat{\gamma}) - 2 \text{Cov}(\hat{\gamma}_c, \hat{\gamma}) + (E(\hat{\gamma}_c) - E(\hat{\gamma}))^2
\]

\[
= (\text{BIAS}(\hat{\gamma}_c))^2 + \text{Var}(\hat{\gamma}_c) + \text{Var}(\hat{\gamma}) - 2 \text{Cov}(\hat{\gamma}_c, \hat{\gamma}).
\]

where

\[
\text{Cov}(\hat{\gamma}_c, \hat{\gamma}) = \text{the covariance between the two estimates.}
\]

Therefore, an unbiased estimator of the MSE(\( \hat{\gamma}_c \)) is given by

\[
\text{MSE}(\hat{\gamma}_c) = (\hat{\gamma}_c - \hat{\gamma})^2 - \text{Var}(\hat{\gamma}) + 2 \text{Cov}(\hat{\gamma}_c, \hat{\gamma}).
\]

The covariance of \( \hat{\gamma}_c \) and \( \hat{\gamma} \) can be written as
\[ \text{Cov}(\hat{Y}_C, \hat{Y}) = (\text{Var}(\hat{Y}_C) \text{ Var}(\hat{Y}))^{1/2} \rho(\hat{Y}_C, \hat{Y}) \]

where

\[ \rho(\hat{Y}_C, \hat{Y}) = \text{the correlation between } \hat{Y}_C \text{ and } \hat{Y}. \]

Assuming that \( \rho(\hat{Y}_C, \hat{Y}) \) is close to 1.0, then

\[ \text{Cov}(\hat{Y}_C, \hat{Y}) \approx [\text{Var}(\hat{Y}_C) \text{ Var}(\hat{Y})]^{1/2} \]

Therefore, an alternative estimator of the MSE(\( Y_C \)) is

\[ \hat{\text{MSE}}(\hat{Y}_C) \approx (\hat{Y}_C - \hat{Y})^2 - \text{Var}(\hat{Y}) + 2 [\text{Var}(\hat{Y}_C) \text{ Var}(\hat{Y})]^{1/2} \]

The estimated MSEs are plotted relative to the cut-off values (for example, the percentage of truncation) to determine reasonable truncation level. Under the assumption of the offsetting effects of the reductions in the sampling variance and increases in the bias, the plots are expected to be U-shaped. The 'optimal' level of truncation is the point that minimizes estimated MSE (i.e., minimizes sampling variance and estimated square bias) for the set of key data items.

The advantages of this procedure are that the effect of trimming on the estimates and the sampling variances is utilized to determine the extent of trimming and the selection of the trimming value. In a sense, any correlation that exists between the observed data and the sample weights is used as the result of computing the estimated sampling variance. In addition, the actual design effect can be computed for each data variable at each trimming level. However, the disadvantages of this approach are as follows:

1. This approach is computer intensive; requiring repeated computation of estimates and sampling variances for each data variable at each trimming level.
(2) Key data items must be determined and the data available.

(3) The trimming levels are determined by trial and error.

(4) The plots may not show a single 'best' truncation level for all data items and a single truncation level must be selected.

If the statistician has the available resources (time, computer resources, and data), this procedure is superior to the inspection strategy.

c. **Comparison of the square weight to the mean square weight**

This procedure uses the comparison of the contribution of each weight to the sampling variance of an estimate by systematically comparing all weights to a value computed from the sum of the square 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 sum of the square 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 square weights. The procedure has been reported in conjunction with the National Assessment of Educational Progress (NAEP) and, for brevity sake, I will refer to this procedure as the NAEP procedure.

The basis for the NAEP procedure can be described as evaluating the contribution of each weighted observation to the overall variance of the weighted estimate. For a sample of size n, let
\( Y_k \) = the observed for the kth unit and \\
\( w_k \) = the original weight for the kth unit.

Also assume that the unit variance (under a superpopulation model) for 
\( Y_k \) is \( \sigma^2 \). Under a superpopulation model,

\[
\text{Var}(w_k Y_k) = w_k^2 \text{Var}(Y_k) = w_k^2 \sigma^2.
\]

For an estimate of a total,

\[ \hat{Y} = \sum w_k Y_k, \]

a similar representation of the variance of \( \hat{Y} \) under a superpopulation model is

\[
\text{Var}(\hat{Y}) = \text{Var}(\sum w_k Y_k) = s \sum w_k^2 \text{Var}(Y_k) = s \sum w_k^2 \sigma^2.
\]

The relative contribution of the variance associated to the kth unit, 
\( \text{Var}(w_k Y_k) \), to \( \text{Var}(\hat{Y}) \) is

\[
\frac{\text{Var}(w_k Y_k)}{\text{Var}(\hat{Y})} = \frac{w_k^2 \sigma^2}{s \sum w_k^2 \sigma^2} = \frac{w_k^2}{s \sum w_k^2}.
\]

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 square weights. That is,

\[
w_k^2 < K \left( s \sum w_k^2 \right).
\]

To take into account the sample size, \( K \) is set as a function of \( n \),

\[ K = c / n. \]

The algorithm is then

\[
w_k^2 \leq c s \sum w_k^2 / n \text{ or } w_k \leq K_n.
\]

(1)
where

$$K_n = (c \, S \, w_k^2 / n)^{1/2}$$

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

$$n \, w_k^2 / S \, w_k^2.$$ 

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 NAEP trimming procedure facilitates the identification of
trimming levels through a systematic and relatively objective
algorithm. This procedure also provides an indication of the effect
of the trimming on the sampling variance because the average of the
square weights is the numerator of the design effect attributable to
unequal weighting. That is,

$$\text{DEFF}_w = \left( S \, w_k^2 / n \right) / \left( S \, w_k / n \right)^2
= \left( K_n^2 / c \right) / \left( S \, w_k / n \right)^2.$$ 

The primary disadvantage of this procedure is that the survey outcome
data are not utilized. The actual effects of the trimming on the
sampling variance and bias for the estimates are unknown. Complete
reliance on the algorithm to select the trimming value may impose more
extensive trimming then can be justified.

As mentioned previously, this procedure has been used for the
sampling weights of the National Assessment of Educational Progress
(NAEP) for over 10 years. The procedure was alluded to in an RTI NAEP
methodology report (Benrud, et al. 1978). A variation of this procedure that used some external data has been reported in a more recent NAEP methodology report (Johnson, et al. 1987). The initial version of this procedure is attributed to John Tukey but no specific reference was found.

The next chapter includes the analytical derivations which are the basis for the trimming procedures described in Chapter IV. Chapter V contains the description of the design and the results of the empirical study.
Chapter III
ANALYTICAL DERIVATIONS

A. Overview

In this chapter, I will describe various analytical formulations and derivations related to assessing extreme weights and weight trimming strategies. In Chapter IV, these results will be used as the basis for the proposed weight trimming strategies.

Because no known analytical assessment of the effect of weight trimming has been performed and the procedures currently used are somewhat ad hoc, the purpose of the analytical derivations are as follows:

1. to evaluate the potential effect of the trimming on the bias and the variance of the estimator, and

2. to develop more theoretically sound procedures to evaluate the need for trimming and to identify sampling weights to trim.

Trimming sampling weights using the current weight trimming procedures has taken on the form of an art. The derivations described in the following sections seek to introduce a theoretically based structure to weight trimming.

Because a primary issue in weight trimming is the potential effect of the trimming on the bias and the variance of the estimator, I will
derive the conditional expectation and variance for an estimator of a population mean when a weight trimming strategy is used and a fixed trimming level \( w_0 \) is assumed for all samples. This formulation for the expected value (and bias) and for the variance provides insight into the effect of weight trimming. An estimator of the variance is developed using the Taylor series linearization approach to derive a variate accounting for the weight trimming.

The first weight trimming procedure (the Taylor series procedure, described in Chapter IV) is developed using the derived forms for the bias and Taylor series linearized variate. Therefore, this procedure permits the assessment of both the potential for bias and the variance reduction introduced by the weight trimming.

Because the Taylor series procedure and the SDV procedure (described in Chapter IV) utilize data in the determination of the weights for trimming, a generalized unequal weighting design effect measure (DEFF\(_{WG}\)) is developed to assist in identifying the key data items for use in the weight trimming. The DEFF\(_{WG}\) is developed using a simple linear model that represents the relationship between the observed data and the selection probabilities and sampling weights. By using this model, the generalized design effect attributable to unequal weighting is derived that is analogous to Kish's DEFF for unequal weighting. The DEFF\(_{WG}\) is an improvement over Kish's DEFF because it incorporates the relationship between the data and the sampling weights. This measure is also potentially useful to determine if weight trimming is necessary.

Because some estimates resulting from sample surveys are based on
dichotomous or polytomous response variables or because sometimes data are not available when weight trimming is performed, a new weight trimming procedure is proposed in Chapter IV that utilizes an assumed distribution for the sampling weights. In this chapter, I will derive the theoretical distribution for 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. A weight trimming level can be determined based on the distribution using the average weight and an assumed value for the DEFF.

The next section derives the expectation of the trimmed mean estimator and the variance estimator and the estimator of the variance. Following these derivations is the derivation of the generalized DEFF measure for unequal weighting. The last part of this chapter will describe the derivation of the distribution for sampling weights.

B. Expectation and Variance of an Estimator of the Mean Using Trimmed Weights

In this section, I will derive the expected value and the bias of an estimator of the population mean and the variance of the estimator when trimmed weights are used. An estimator of the variance of the trimmed weight estimator is also derived. For the variance estimator, a Taylor series linearized variate is derived that can be used to identify extreme weights for trimming.
While weight trimming can conceivably be used to deal with both large weights and small weights the following derivation assumes only the more common scenario when large weights are trimmed. The derivation for two-sided trimming would be a natural extension of this approach.

Assume a sampling frame of \( N \) units. Define the following:

\[
Y_k = \text{the observed data for the } k\text{th unit.}
\]

\[
P_k = \text{the single draw selection probability for the } k\text{th unit.}
\]

\[
\pi_k = \text{the selection probability (or the expected number of selections) for the } k\text{th unit when a sample of size } n \text{ is selected and assume } \pi_k \text{ is less than } 1 \text{ for all } k; \text{ that is,}
\]

\[
\pi_k = n \cdot p_k.
\]

\[
w_k = \text{the untrimmed sampling weight for the } k\text{th unit in a sample of size } n; \text{ that is,}
\]

\[
w_k = 1 / \pi_k.
\]

\[
w_{kt} = \text{the sampling weight for the } k\text{th unit when a weight trimming strategy is used.}
\]

\[
\bar{Y} = \text{the population mean of the } Y_k; \text{ that is,}
\]

\[
\bar{Y} = \frac{\sum Y_k}{N}.
\]

\[
\hat{Y} = \text{the usual expansion estimator for a population mean when the untrimmed sampling weights are used; that is,}
\]

\[
\hat{Y} = \frac{S w_k Y_k}{S w_k}
\]

where \( S \) denotes the summation over the sample of \( n \) units.

\[
\hat{Y}_t = \text{the usual expansion estimator for a population mean when a weight trimming strategy is used on the sampling weights; that is,}
\]
\[ \hat{Y}_t = S w_{kt} Y_k / S w_{kt}. \]

1. **Expectation of An Estimator of the Mean With Trimmed Weights**

Assume a sample of size \( n \) is selected with unequal probability and with replacement. The following derivations are conditional on a fixed weight trimming value of \( w_o \) for all possible samples of size \( n \). All weights below this value are adjusted by a factor \( A_s \) so that the original weight sum \( (W_s) \) for the sample is preserved. That is, the trimmed weight \( w_{kt} \) is defined as

\[
\begin{align*}
  w_{kt} &= w_o & \text{if } w_k \geq w_o, \\
  &= A_s w_k & \text{if } w_k < w_o
\end{align*}
\]

such that

\[
S w_{kt} = S w_k = W_s.
\]

The assumption of a fixed \( w_o \) for all samples can be based on the concept that \( w_o \) represents a percentile of the distribution of the population of sampling weights for samples of size \( n \). That is, if \( F_w(w) \) is the cumulative distribution function for the population of sampling weights, then \( w_o \) is the value such that \( F_w(w_o) = \delta \), where \( \delta \) may be a value such as .99.

In the following, I show that the expected value of the estimator of this mean with trimmed weights is

\[
E(\hat{Y}_t) = \bar{y}_{NT} + w_o \Pi_t (\bar{y}_t - \bar{y}_{NT})
\]

where

33
\( \overline{y}_{NT} \) = the population mean of the observations with untrimmed weights;

\[ \Pi_t = E_t/N, \]

where

\[ E_t = E(n_t) \]

is the expected number of units in samples of size \( n \) that will have weights greater than or equal to \( w_0 \) (the trimming levels) and \( N \) is the frame size; and

\( \overline{y}_{rt} \) = weighted population average of data for units with a sampling weight greater than or equal to \( w_0 \), where the weights are the inclusion probabilities.

The bias of \( \hat{y} \) is then shown to be

\[
\text{Bias} (\hat{y}_t) = (\theta_t - w_0 \, \Pi_t) \, \overline{y}_{NT} \\
- (\theta_t \, \overline{y}_T - w_0 \, \Pi_t \, \overline{y}_{rt})
\]

where

\[ \theta_t = \text{the population proportion of units with trimmed weights;} \]

and

\( \overline{y}_T = \text{the population mean for units with trimmed weight.} \)

For the subsequent estimation of the bias, the bias of \( \hat{y}_t \) is also shown in terms of individual observations. That is

\[
\text{Bias} (\hat{y}_t) = - \Sigma \, \tau_k \, (1 - w_0 \, \pi_k) \, (Y_k - \overline{y}_{NT})/N.
\]

In this form, \( \overline{y}_{NT} \) is clearly shown as a pivotal quantity for the bias. The following describes the detailed derivations.

Because the sampling weights \( w_k \) that are less than \( w_0 \) are adjusted
to compensate for the trimming of the sampling weights \( w_k \) that are
greater than or equal to \( w_0 \), the adjustment factor can be denoted as
\( A_s \) and
\[
A_s = \frac{(S w_k - n_t w_0)}{(S w_k - S_t w_k)} \frac{(W_s - n_t w_0)}{(W_s - S_t w_k)} 
\tag{2}
\]
where
\( n_t \) = the number of weights that are trimmed in the sample
(that is, \( n_t \) is the number of units with sampling
weights in excess of \( w_0 \)), and
\( S_t \) = the summation over the \( n_t \) units for which the sampling
weights were trimmed.

If an indicator function, \( \tau_k \), is defined such that
\[
\tau_k = 1 \quad \text{if } w_k \geq w_0,
= 0 \quad \text{if } w_k < w_0,
\]
then, from equation (1), the trimmed weight for the \( k \)th unit can be
written as
\[
w_{kt} = \tau_k w_0 + (1 - \tau_k) A_s w_k. \tag{3}
\]
Also, the adjustment factor \( A_s \) can be written as
\[
A_s = \frac{(S w_k - S \tau_k w_0)}{(S w_k - S \tau_k w_k)} = \frac{(S (w_k - \tau_k w_0))}{(S (1 - \tau_k) w_k)}. \tag{4}
\]
Since we defined \( \pi_k \) as the expected number of hits (or selections) for
the \( k \)th unit in a sample of size \( n \), that is,
\[
\pi_k = 1 / w_k,
\]
then equation (3) can be rewritten as
\[
w_{kt} = \tau_k w_0 \pi_k w_k + (1 - \tau_k) A_s w_k \tag{5}
= [\tau_k w_0 \pi_k + (1 - \tau_k) A_s] w_k. \tag{6}
\]
Also, the adjustment factor $A_s$ can be written as

\[
A_s = \frac{S \left( w_k - \tau_k w_0 \pi_k w_k \right)}{S \left( 1 - \tau_k \right) w_k} = \frac{S \left( 1 - \tau_k w_0 \pi_k \right) w_k}{S \left( 1 - \tau_k \right) w_k}.
\]  
(7)

Alternatively, we can write $A_s$ as the ratio of two totals, $A_{s1}$ and $A_{s2}$. That is,

\[
A_s = \frac{A_{s1}}{A_{s2}}
\]

where

\[
A_{s1} = S \left( 1 - \tau_k w_0 \pi_k \right) w_k \quad \text{(8)}
\]

\[
A_{s2} = S \left( 1 - \tau_k \right) w_k \quad \text{(9)}
\]

For an estimator of the mean of observed data $Y_k$ using the trimmed weights, the usual estimator is given by

\[
\hat{Y}_t = S w_{kt} Y_k / S w_{kt}.
\]  
(10)

Because the adjustment factor $A_s$ was developed so that sum of the weights is preserved, we have that

\[
S w_{kt} = W_s.
\]

Using the definition of the trimmed weights in equation (5), then equation (10) can be written as a function of weighted totals. That is,

\[
\hat{Y}_t = \left\{ S \left[ \tau_k w_0 \pi_k w_k + \left( 1 - \tau_k \right) A_s w_k \right] Y_k \right\} / W_s
\]

\[
= \left[ S \tau_k w_0 \pi_k w_k Y_k + A_s S \left( 1 - \tau_k \right) w_k Y_k \right] / W_s
\]

\[
= S \tau_k w_0 \pi_k w_k Y_k / W_s + A_s S \left( 1 - \tau_k \right) w_k Y_k / W_s
\]

\[
= C_s / W_s + \left( A_{s1} / A_{s2} \right) (B_s / W_s)
\]  
(11)

where

\[
B_s = S \left( 1 - \tau_k \right) w_k Y_k
\]  
(12)

\[
C_s = S \tau_k w_0 \pi_k w_k Y_k
\]  
(13)
To derive the expectation of \( \hat{y}_t \), I will use the Taylor series linearization method to develop an expression for the expected value. This method is based on a Taylor series expansion of the function. For example, consider a function \( F(X,Y) \), where \( X \) and \( Y \) are linear sample statistics. Let \( \mu_X \) and \( \mu_Y \) denote the expected value of \( X \) and \( Y \), respectively. The function \( F(X,Y) \) can be expanded in a Taylor series about \( \mu_X \) and \( \mu_Y \), such that

\[
F(X,Y) = F(\mu_X, \mu_Y) + \delta F_x(\mu_X, \mu_Y)(X - \mu_X) + \delta F_y(\mu_X, \mu_Y)(Y - \mu_Y)
\]

+ higher order terms,

where \( \delta F_x(\mu_X, \mu_Y) \) and \( \delta F_y(\mu_X, \mu_Y) \) are the first-order partial derivatives of the function \( F \) with respect to \( X \) and \( Y \) and evaluated at the expectations \( \mu_X \) and \( \mu_Y \).

Assuming that the sampling variances of \( X \) and \( Y \) and the \((X,Y)\) covariance are of order no larger than \((1/n)\) and that the second order partial derivatives of the function \( F \) are bounded in the neighborhood of the point \((\mu_X, \mu_Y)\), the expected value of the higher order terms is of an order no larger than \(1/n\). Therefore, for sufficiently large sample sizes the expected value of \( F(X,Y) \) is \( F(\mu_X, \mu_Y) \) because the expected value of the first order terms are zero.

In this situation, the function of interest is

\[
F(A_{s1}, A_{s2}, B_s, C_s, W_s) = C_s / W_s + (A_{s1} / A_{s2}) \ (B_s / W_s).
\] (14)
The expected value of $\hat{Y}_t$ is obtained as

$$E(\hat{Y}_t) = E(F(A_{s1}, A_{s2}, B_s, C_S, W_s))$$

$$= F(E(A_{s1}), E(A_{s2}), E(B_s), E(C_s), E(W_s))$$

for large sample sizes.

In order to obtain the expected value of $\hat{Y}_t$, we need the expected value of each of the component totals.

1. For $A_{s1}$,

$$E(A_{s1}) = E(\sum (1 - \tau_k w_o \pi_k) w_k)$$

$$= \sum (1 - \tau_k w_o \pi_k)$$

$$= N - w_o \sum \tau_k \pi_k$$

where $N$ is the population size. Because $\pi_k$ is the expected number of selections for the $k$th unit,

$$\sum \pi_k = n$$

and, therefore,

$$\sum \tau_k \pi_k = E(n_t),$$

the expected number of units that will have a weight in excess of $w_o$ in a sample of size $n$. Let $E_t$ denote $E(n_t)$, then

$$E(A_{s1}) = N - w_o E_t$$

$$= (1 - w_o \Pi_t) N$$

(15)

where

$$\Pi_t = E_t / N.$$

The term ($w_o E_t$) is the product of the weight trimming level and the expected number of units in a sample of size $n$ that would have a sampling weights greater than or equal to $w_o$. The product ($w_o E_t$) corresponds, therefore, to the expected value of the estimated population count associated with the trimmed weight units. The term
(\(w_o \Pi_t\)) corresponds to the reduced proportion of the population count represented by the trimmed weight units whereas

\[ \theta_t = \frac{N_t}{N} \]

corresponds to the true proportion of the population units with trimmed weights.

2. For \(A_{S2}\),

\[
E(A_{S2}) = E \left( \sum (1 - \tau_k) w_k \right) \\
= \sum (1 - \tau_k) \\
= N - N_T \\
= (1 - \theta_t) N
\]

where \(N_T\) and \(\theta_t\) is the population count and proportion of the units with a sampling weight in greater than or equal to \(w_o\).

3. For \(B_s\),

\[
E(B_s) = E \left( \sum (1 - \tau_k) w_k Y_k \right) \\
= \sum (1 - \tau_k) Y_k \\
= (N - N_T) \bar{y}_{NT}
\]

where \(\bar{y}_{NT}\) is the population mean for units with a weight less than \(w_o\),

\[ = (1 - \theta_t) N \bar{y}_{NT}. \]  

4. For \(C_s\),

\[
E(C_s) = E \left( \sum \tau_k w_o \pi_k w_k Y_k \right) \\
= \sum \tau_k w_o \pi_k Y_k \\
= w_o E_t \left( \sum \tau_k \pi_k Y_k / E_t \right) \\
= w_o E_t \bar{y}_{\pi t}
\]

where

\[ \bar{y}_{\pi t} = \sum \tau_k \pi_k Y_k / E_t. \]
In equation (18), \( \hat{\bar{y}}_{\pi t} \) can be interpreted as a weighted population average of the observed data for the units with a sampling weight greater than or equal to \( w_0 \). The weights in this average are the expected number of selections for the kth unit.

5. For \( W_s \),

\[
E(W_s) = E(\sum w_k)
\]

\[
= N.
\]

For large sample sizes, we can use the Taylor series linearization to arrive at an expression for the expected value of \( \hat{\bar{y}}_t \).

\[
E(\hat{\bar{y}}_t) = \frac{E(A_{s1})}{E(A_{s2})} \left[ \frac{E(B_s)}{E(W_s)} \right] E(C_s) / E(W_s)
\]

\[
= \left[ \frac{(1 - w_0 \Pi_t)N}{(1 - \theta_t) N} \right] (1 - \theta_t) N \bar{y}_{NT} / N
\]

\[
+ w_0 E_t \hat{\bar{y}}_{\pi t} / N
\]

\[
= (1 - w_0 \Pi_t) \bar{y}_{NT} + w_0 \Pi_t \bar{y}_{\pi t}
\]

\[
= \bar{y}_{NT} + w_0 \Pi_t (\bar{y}_{\pi t} - \bar{y}_{NT}).
\]

(19)

The bias of \( \hat{\bar{y}}_t \) is obtained by first noting that \( \bar{y} \) can be written as

\[
\bar{y} = [\theta_t \bar{y}_T + (1 - \theta_t)\bar{y}_{NT}],
\]

where

\[
\theta_t = N_t / N,
\]

\( \bar{y}_T = \) the population mean for the units with trimmed weights, and

\( \bar{y}_{NT} = \) the population mean for the units with untrimmed weights.

The bias of \( \hat{\bar{y}}_t \) is given as
$$\text{Bias}(\hat{\bar{Y}}_t) = E(\hat{\bar{Y}}_t) - \bar{Y}$$

$$= \bar{Y}_{NT} + w_o \Pi_t \left[ \bar{Y}_{xt} - \bar{Y}_{NT} \right]$$
$$- \left[ \Theta_t \bar{Y}_T + (1 - \Theta_t) \bar{Y}_{NT} \right]$$
$$= (\Theta_t - w_o \Pi_t) \bar{Y}_{NT} - (\Theta_t \bar{Y}_T - w_o \Pi_t \bar{Y}_{xt}) \quad (20)$$

where

$$w_o \Pi_t = w_o \frac{E_t}{N}$$

is the proportional reduction of population represented by units with a sampling weight greater than or equal to \(w_o\) caused by the assignment of \(w_o\) as the unit's weight.

In this form, the bias can be easily seen as a direct function of the difference between the proportion of units with large weights and the expected value of the proportion of the estimated population count associated with the trimmed weights (\(w_o \Pi_t\)).

For example, if we can assume that \(\bar{Y}_T\) and \(\bar{Y}_{xt}\) are approximately equal then equation (20) can be written as

$$\text{Bias}(\hat{\bar{Y}}_t) \approx (\Theta_t - w_o \Pi_t) (\bar{Y}_{NT} - \bar{Y}_T). \quad (21)$$

From this representation of the bias, we can see the three primary factors contributing to the bias:

1. \(\Theta_t\) = the proportion of units with a sampling weight greater than or equal to \(w_o\);

2. \(w_o \Pi_t\) = the proportional reduction of population represented by units with a sampling weight greater than or equal to \(w_o\) caused by the assignment of \(w_o\) as the unit's weight; and
3. \((\bar{\gamma}_{NT} - \bar{\gamma}_{T})\) = the difference between the population mean value for the units with weights less than \(w_o\) and the population mean value for units with weights greater than or equal to \(w_o\).

If \(\bar{\gamma}_{NT}\) and \(\bar{\gamma}_{T}\) are approximately equal, then the bias is negligible.

If \(\bar{\gamma}_{NT}\) and \(\bar{\gamma}_{T}\) are not equal, then the bias may not be negligible.

However, if the \(\theta_t\) is small (say, less than 1 percent) then the approximate bias will be less than 1 percent of the difference between \(\bar{\gamma}_{NT}\) and \(\bar{\gamma}_{T}\) because \(\Pi_t\) will always be less than \(\theta_t\).

If the observed data represents a dichotomous (0-1) response variable, then the \(\bar{\gamma}_{NT}\) and \(\bar{\gamma}_{T}\) proportions and the absolute value of the difference between \(\bar{\gamma}_{NT}\) and \(\bar{\gamma}_{T}\) will be less than 1.0. Therefore, an upper bound can be developed for the absolute value of the bias. That is, if the observations are a 0-1 response variables, then

\[
\text{Bias}(\bar{\gamma}_t) \leq (\theta_t - w_o \Pi_t) (\bar{\gamma}_{NT} - \bar{\gamma}_{T}) \leq (\theta_t - w_o \Pi_t).
\]

Table 1 below provides approximate bounds for the absolute value of the bias of proportions for selected values of \(\theta_t\) and \(w_o \Pi_t\). As can be seen from this table, if weight trimming is used extensively substantial bias can be introduced. If approximately 20 percent of the population are subject to weight trimming, then the absolute value of the bias will be up to 10 percent when the proportional reduction of coverage is around 10 percent. On the other hand, if a relatively small proportion of the population is subject to trimming (that is, \(\theta_t\)
Table 1. Approximate Absolute Bias Bounds For Population Proportion

<table>
<thead>
<tr>
<th>$w_o$ II_t</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>$\delta_t$</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.009</td>
<td>0.049</td>
<td>0.099</td>
<td>0.149</td>
<td>0.199</td>
<td>0.249</td>
<td></td>
</tr>
<tr>
<td>0.005</td>
<td>0.005</td>
<td>0.045</td>
<td>0.095</td>
<td>0.145</td>
<td>0.195</td>
<td>0.245</td>
<td></td>
</tr>
<tr>
<td>0.010</td>
<td>0.0</td>
<td>0.04</td>
<td>0.09</td>
<td>0.14</td>
<td>0.19</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>-</td>
<td>0.0</td>
<td>0.05</td>
<td>0.10</td>
<td>0.15</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>-</td>
<td>-</td>
<td>0.0</td>
<td>0.05</td>
<td>0.10</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0</td>
<td>0.05</td>
<td>0.10</td>
<td></td>
</tr>
</tbody>
</table>
is small) then substantial trimming of the weights will introduce relatively minor bias.

Lower and upper bounds for the bias can be developed by assuming that the observed data for the units with weights greater or equal to w₀ are all equal 1 or are all equal to 0. From equation (20),

\[ \text{Bias}(\hat{\Theta}_t) = (\Theta_t - w_0 \Pi_t) \bar{\gamma}_{NT} - (\Theta_t \bar{\gamma}_T - w_0 \Pi_t \bar{\gamma}_{\pi T}). \]

If all Y_k's equal 0 for units with sampling weight ≥ w₀ then an upper bound for the bias can be given as

\[ \text{Bias}(\hat{\Theta}_t)_u = (\Theta_t - w_0 \Pi_t) \bar{\gamma}_{NT}. \tag{23} \]

If all Y_k's equal 1 for units with sampling weight ≥ w₀ then a lower bound for the bias will be

\[ \text{Bias}(\hat{\Theta}_t)_l = (\Theta_t - w_0 \Pi_t) \bar{\gamma}_{NT} - (\Theta_t - w_0 \Pi_t) \]
\[ = (\Theta_t - w_0 \Pi_t) (\bar{\gamma}_{NT} - 1). \tag{24} \]

The impact of individual observation values can be assessed by an alternative formulation of the bias. This formulation can be developed by noting that

\[ (\Theta_t - w_0 \Pi_t) = [(N_t/N) - (w_0 E_t / N)] \]
\[ = \Sigma \tau_k (1 - w_0 \pi_k) / N, \]

and

\[ \Theta_t \bar{\gamma}_T - w_0 \Pi_t \bar{\gamma}_{\pi T} = (N_t/N) \bar{\gamma}_T - (w_0 E_t / N) \bar{\gamma}_{\pi T} \]
\[ = \Sigma \tau_k (1 - w_0 \pi_k) y_k / N. \]

The bias of \( \hat{\Theta}_t \) (equation (20)) can be written as

\[ \text{Bias}(\hat{\Theta}_t) = \left[ (N_t/N) - (w_0 E_t / N) \right] \bar{\gamma}_{NT} \]
\[ - \{ \left[ N_t/N \right] \bar{\gamma}_T - (w_0 E_t / N) \bar{\gamma}_{\pi T} \} \]
\[ = - \Sigma \tau_k (1 - w_0 \pi_k) (y_k - \bar{\gamma}_{NT}) / N, \text{ or} \tag{25} \]
\[ = - \Theta_t \Sigma \tau_k (1 - w_0 \pi_k) (y_k - \bar{\gamma}_{NT}) / N_T. \tag{26} \]
Because \( \tau_k \) equal to 1 for \( w_k \) greater than or equal to \( w_o \) and \( w_k \) equals the inverse of \( \pi_k \), the factor \( (1 - w_o \pi_k) \) is always positive when \( \tau_k \) equals 1.

By letting

\[
P_{0k} = \frac{w_o}{w_k}
\]
denote the proportional decrease in the weight for the \( k \)th unit, then the bias of \( \hat{y}_t \) can be written as

\[
\text{Bias}(\hat{y}_t) = - \beta_t \sum \tau_k (1 - P_{0k}) (Y_k - \bar{Y}_{NT})/N_T. \tag{27}
\]

Therefore, if all units with sampling weights, \( w_k \), greater than or equal to \( w_o \) have \( Y_k \) greater than \( \bar{Y}_{NT} \), then the bias is negative. Conversely, if all units with sampling weights greater than or equal to \( w_o \) have observed data, \( Y_k \), less than \( \bar{Y}_{NT} \), then the bias is positive. If the sample design uses probability proportional to size selection and the observed data \( Y_k \) are positively correlated with the size measure, then the larger weights will be associated with the smaller values of \( Y_k \). In this case, based on equation (27), a positive bias would be expected to result, that is the expected value of the estimator would be larger than the true value. Also, the bias is zero if \( Y_k \) equals \( \bar{Y}_{NT} \) for all units with a sampling weight greater than or equal to \( w_o \).

If the observations are 0-1 response variables, then a second bias bound can be formulated. A lower bound \((\text{Bias}(\hat{y}_t))_l\) can be approximated by assuming that all units with weights greater than or equal to \( w_o \) have an observed value of 1 and an upper bound \((\text{Bias}(\hat{y}_t))_u\) if these units are assumed to have an observed value of 0. That is,
\[ \text{Bias}(\hat{\gamma}_t)_1 = -\theta_t \left( \sum_k (1 - \rho_{ok}) / N_t \right) (1 - \bar{\gamma}_{NT}), \]

and

\[ \text{Bias}(\hat{\gamma}_t)_u = \theta_t \left( \sum_k (1 - \rho_{ok}) / N_t \right) \bar{\gamma}_{NT}. \]

Once again the bias can be minimized if the units with trimmed weights represents a small proportion of the population and the extent of trimming is minimized.

Using the form for the bias given in equation (27), the expected value of \( \hat{\gamma}_t \) can be written as

\[ E(\hat{\gamma}_t) = \bar{\gamma} - \theta_T \sum_k (1 - \rho_{ok})(\gamma_k - \bar{\gamma}_{NT}) / N_T. \]  

(28)

Alternatively, using equation (20), the expected value of \( \hat{\gamma}_t \) can be represented as

\[ E(\hat{\gamma}_t) = \bar{\gamma} + (\theta_t - \omega_t \Pi_t) \bar{\gamma}_{NT} \]

\[ - (\theta_T \bar{\gamma}_T - \omega_t \bar{\gamma}_t). \]  

(29)

These representations of the expected value of \( \hat{\gamma}_t \) exhibit more clearly the effect of weight trimming on the estimator of the mean.

2. **The Variance of the Trimmed Weight Estimator**

In this section, the variance of the estimator of the population mean with trimmed weights is developed using the Taylor series approximation. A linearized variate is developed for the variance of the estimator. An estimator of the variance is derived then based on the sample based estimate of the linearized variate.

Because the estimator \( \hat{\gamma}_t \) is a nonlinear function, the variance of the estimator can be approximated using the Taylor series linearization method. As defined previously, the function of interest is
\[ F(A_{s1}, A_{s2}, B_s, C_s, W_s) = C_s / W_s + (A_{s1} / A_{s2}) \ (B_s / W_s). \tag{30} \]

For simplicity, let
\[
F_T = F(A_{s1}, A_{s2}, B_s, C_s, W_s), \\
\mu_1 = E(A_{s1}), \\
\mu_2 = E(A_{s2}), \\
\mu_B = E(B_s), \\
\mu_C = E(C_s), \\
\mu_W = E(W_s),
\]
and
\[
F_\mu = F(\mu_1, \mu_2, \mu_B, \mu_C, \mu_W).
\]

For the partial derivatives of \( F_T \) evaluated at \( \mu_1, \mu_2, \mu_B, \mu_C, \mu_W \), let
\[
\delta F_1 = \delta F_{A1}(\mu_1, \mu_2, \mu_B, \mu_C, \mu_W), \\
\delta F_2 = \delta F_{A2}(\mu_1, \mu_2, \mu_B, \mu_C, \mu_W), \\
\delta F_B = \delta F_B(\mu_1, \mu_2, \mu_B, \mu_C, \mu_W), \\
\delta F_C = \delta F_C(\mu_1, \mu_2, \mu_B, \mu_C, \mu_W),
\]
and
\[
\delta F_W = \delta F_W(\mu_1, \mu_2, \mu_B, \mu_C, \mu_W).
\]

The Taylor series expansion can be represented as
\[
F_T = F_\mu + \delta F_1(A_{s1} - \mu_1) + \delta F_2(A_{s2} - \mu_2) + \delta F_B(B_s - \mu_B) \\
+ \delta F_C(C_s - \mu_C) + \delta F_W(W_s - \mu_W) + \text{higher order terms.}
\]

If the higher order terms are negligible, then the variance of \( \hat{Y}_t \) can be approximated by
\[ \text{Var}(\hat{\gamma}_t) = \text{Var}(F_T) = E[F_T - F_\mu]^2 \]
\[ = \delta F_1^2 \text{Var}(A_{s1}) + \delta F_2^2 \text{Var}(A_{s2}) + \delta F_B^2 \text{Var}(B_s) + \delta F_C^2 \text{Var}(C_s) + \delta F_W^2 \text{Var}(W_s) + 2 \delta F_1 \delta F_2 \text{Cov}(A_{s1}, A_{s2}) + 2 \delta F_1 \delta F_B \text{Cov}(A_{s1}, B_s) + 2 \delta F_1 \delta F_C \text{Cov}(A_{s1}, C_s) + 2 \delta F_1 \delta F_W \text{Cov}(A_{s1}, W_s) + 2 \delta F_2 \delta F_B \text{Cov}(A_{s2}, B_s) + 2 \delta F_2 \delta F_C \text{Cov}(A_{s2}, C_s) + 2 \delta F_2 \delta F_W \text{Cov}(A_{s2}, W_s) + 2 \delta F_B \delta F_C \text{Cov}(B_s, C_s) + 2 \delta F_B \delta F_W \text{Cov}(B_s, W_s) + 2 \delta F_C \delta F_W \text{Cov}(C_s, W_s). \]

The partial derivatives of \( F_T \) evaluated at \( \mu_1, \mu_2, \mu_B, \mu_C, \) and \( \mu_W \) are the following:

1. \( \delta F_1 = (1 / \mu_2) (\mu_B / \mu_W) \)
   \[ = \bar{\gamma}_{NT} / N. \]  
   \[ \text{(32)} \]

2. \( \delta F_2 = - (\mu_1 / \mu_2^2) (\mu_B / \mu_W) \)
   \[ = - [(N - w_0 E_t)/(N - N_T)] (\bar{\gamma}_{NT} / N) \]
   \[ = - [(1 - w_0 \Pi_t)/(N - N_T)] \bar{\gamma}_{NT} \]  
   \[ \text{(33)} \]
   where
   \[ \Pi_t = E_t / N. \]

3. \( \delta F_B = \mu_1 / (\mu_2 \mu_W) \)
   \[ = (N - w_0 E_t)/[(N - N_T) N] \]
   \[ = (1 - w_0 \Pi_t) / (N - N_T). \]
   \[ \text{(34)} \]

4. \( \delta F_C = (1 / \mu_W) \)
   \[ = 1 / N. \]
   \[ \text{(35)} \]
5. \[ \delta F_W = - \left( \frac{\mu_C}{\mu_W^2} + \frac{\mu_1}{\mu_2} \right) \left( \frac{\mu_B}{\mu_W^2} \right) \]
\[ = - \bar{\gamma}_t / N. \]
\[ = - \left\{ \left( \frac{(N - w_o E_t)}{N} \right) \bar{\gamma}_{NT} / N + w_o \bar{\gamma}_{\pi t} / N \right\} \]
\[ = - \left\{ \left( 1 - w_o \Pi_t \right) \bar{\gamma}_{NT} / N + w_o \bar{\gamma}_{\pi t} / N \right\} \]
\[ (36) \]

where
\[ \bar{\gamma}_{\pi t} = \Sigma \tau_k \pi_k \gamma_k / N. \]

Note that the divisor for \( \bar{\gamma}_{\pi t} \) is \( N \) and not \( E_t \) or \( N_T \). This divisor is used to simplify the development of variance expression.

With these items defined, in lieu of working with equation (31) the Taylor series approximation for the variance can be derived by developing a Taylor series linearized variate, \( z \). That is,
\[ z_k = \delta F_1 A_{s1k} + \delta F_2 A_{s2k} + \delta F_B B_{sk} \]
\[ + \delta F_C C_{sk} + \delta F_W w_k. \]

where
\[ A_{s1k} = (w_k - \tau_k w_o) \]
\[ = w_k \left( 1 - \tau_k w_o \pi_k \right), \]
\[ A_{s2k} = (w_k - \tau_k w_k) \]
\[ = w_k \left( 1 - \tau_k \right), \]
\[ B_{sk} = (w_k - \tau_k w_k) \gamma_k \]
\[ = w_k \left( 1 - \tau_k \right) \gamma_k, \]

and
\[ C_{sk} = \tau_k w_o \gamma_k \]
\[ = w_k \tau_k w_o \pi_k \gamma_k. \]

The linearized variate is given by
\[ z_k = \delta F_1 w_k \left( 1 - \tau_k w_o \pi_k \right) + \delta F_2 w_k \left( 1 - \tau_k \right) + \delta F_B w_k \left( 1 - \tau_k \right) \gamma_k \]
\[ + \delta F_C w_k \tau_k w_o \pi_k \gamma_k + \delta F_W w_k. \]
\[= (\bar{v}_{NT} / N) \ w_k (1 - \tau_k \omega \pi_k) \]
\[- [(1 - w_o \Pi_t) / (N - N_T)] \ \bar{v}_{NT} \ w_k (1 - \tau_k) \]
\[+ [(1 - w_o \Pi_t) / (N - N_T)] \ w_k (1 - \tau_k) \ y_k \]
\[+ (1/N) \ w_k \ \tau_k \ w_o \ \pi_k \ y_k \]
\[- w_k [(1 - w_o \Pi_t) \ \bar{v}_{NT} / N + w_o \ \bar{v}_{\pi t} / N]. \]

\[= w_k \{ \bar{v}_{NT} / N - \tau_k \omega \pi_k \bar{v}_{NT} / N \]
\[+ [(1 - w_o \Pi_t) / (N - N_T)] (1 - \tau_k) (y_k - \bar{v}_{NT}) \]
\[+ (1/N) \ \tau_k \ w_o \ \pi_k \ y_k \]
\[= \bar{v}_{NT}/N + w_o \Pi_t \ \bar{v}_{NT}/N - w_o \ \bar{v}_{\pi t}/N \}. \]

\[= w_k\{[(1 - w_o \Pi_t) / (N - N_T)] (1 - \tau_k) (y_k - \bar{v}_{NT}) \]
\[+ (w_o / N) (\tau_k \ \pi_k \ y_k - \bar{v}_{\pi t}) \]
\[- (w_o \ \bar{v}_{NT} / N) (\tau_k \ \pi_k - \Pi_t) \}. \]

Let
\[v_k = (1 - \tau_k) (y_k - \bar{v}_{NT}), \quad (37)\]
\[u_k = (\tau_k \ \pi_k \ y_k - \bar{v}_{\pi t}), \quad (38)\]
and
\[g_k = (\tau_k \ \pi_k - \Pi_t), \quad (39)\]
then
\[z_k = w_k\{[(1 - w_o \Pi_t) / (N - N_T)] \ v_k \]
\[+ (w_o / N) \ u_k - (w_o \ \bar{v}_{NT} / N) \ g_k \}. \]

Because
\[N - N_T = (1 - \theta_t) \ N, \]

50
we have
\[ z_k = w_k \left\{ [(1 - w_o \Pi_t)/(1 - \Theta_t)] v_k / N \\
+ (w_o / N) u_k - (w_o \bar{\gamma}_{NT} / N) g_k \right\}. \] (40)

The variance of \( \hat{\gamma}_t \) can be approximated by
\[
\text{Var}(\hat{\gamma}_t) = \text{Var}(F_T) \\
\approx \text{Var}(\Sigma z) \\
= N [E(z^2) - (Ez)^2].
\]

Because \( Ez \) equals zero,
\[
\text{Var}(\hat{\gamma}_t) \approx \text{Var}(\Sigma z) \\
= N E z^2 \\
= N E(w_k \left\{ [(1 - w_o \Pi_t)/(1 - \Theta_t)] v_k / N \\
+ (w_o / N) u_k - (w_o \bar{\gamma}_{NT} / N) g_k \right\})^2.
\]
\[
= N E(w_k \left\{ [(1 - w_o \Pi_t)/(1 - \Theta_t)]^2 v_k^2 / N^2 \\
+ (w_o / N)^2 u_k^2 + (w_o \bar{\gamma}_{NT} / N)^2 g_k^2 \\
+ 2 [(1 - w_o \Pi_t) w_o / ((N - N_T) N)] v_k u_k \\
- 2 [(1 - w_o \Pi_t) w_o \bar{\gamma}_{NT} / ((N - N_T) N)] v_k g_k \\
- 2 (w_o^2 \bar{\gamma}_{NT} / N^2) u_k g_k \right\}).
\]

Because, when \( \tau_k \) equals to 1, \( v_k \) equals 0, and when \( \tau_k \) equals to 0,
\[ v_k = (Y_k - \bar{\gamma}_{NT}), \]
we have
\[ v_k u_k = (1 - \tau_k) (Y_k - \bar{\gamma}_{NT}) (\bar{\gamma}_{NT} - \bar{\gamma}_t), \]
\[ v_k g_k = (1 - \tau_k) (Y_k - \bar{\gamma}_{NT}) (-\Pi_t), \]
\[ E(v_k u_k) = 0, \]

51
and
\[ E(v_k g_k) = 0. \]

Thus,
\[
\text{Var}(\hat{\gamma}_t) = N \left\{ \left[ (1 - w_0 \Pi_t)/(1 - \theta_T) \right]^2 E(v_k^2) + w_0^2 E(u_k^2)/N^2 + (w_0 \bar{v}_{NT})^2 E(g_k^2)/N^2 \right. \\
- 2 w_0 \bar{v}_{NT} E(u_k g_k)/N^2 \right\}.
\]

\[
= \left\{ \left[ (1 - w_0 \Pi_t)/(1 - \theta_T) \right]^2 E(v_k^2)/N \\
+ w_0^2 E(u_k^2)/N + (w_0 \bar{v}_{NT})^2 E(g_k^2)/N \\
- 2 (w_0^2 \bar{v}_{NT}) E(u_k g_k)/N \right\}.
\]

\[
= (1/n) \left\{ \left[ (1 - w_0 \Pi_t)/(1 - \theta_T) \right]^2 \text{Var}(\bar{v}_{NT}) \\
+ w_0^2 \text{Var}(\bar{v}_{\pi t}) + (w_0 \bar{v}_{NT})^2 \text{Var}(\Pi_t) \\
- 2 w_0^2 \bar{v}_{NT} \text{Cov}(\bar{v}_{\pi t}, \Pi_t) \right\}. \tag{41}
\]

The components of the variance in equation (41) can be interpreted in the following fashion:

1. \[ [ (1 - w_0 \Pi_t)/(1 - \theta_T) ]^2 \text{Var}(\bar{v}_{NT}) \]

   consists of two parts: \((1 - w_0 \Pi_t)/(1 - \theta_T)\), and \(\text{Var}(\bar{v}_{NT})\).

   a. \((1 - w_0 \Pi_t)/(1 - \theta_T)\) = the population level value of the adjustment factor that is used to inflate the weights of the units with sampling weights less than \(w_0\). The term, \(w_0 \Pi_t\), is the reduced proportion of the population accounted for by the units with trimmed weights caused by the trimming and the term, \(\theta_T\), is the true proportion of the population accounted for by units with the trimmed weights.
b. \( \text{Var}(\bar{y}_{NT}) \) = the variance of the mean of the units with weights less than \( w_0 \).

In combination,
\[
[(1 - w_0 \Pi_t)/(1 - \beta_T)]^2 \text{Var}(\bar{y}_{NT}) = \text{the inflated variance of the population of the units with untrimmed weights.}
\]

2. \( w_0^2 \text{Var}(\bar{y}_T') = \text{the variance of the units with weights greater than or equal to } w_0 \text{ multiplied by the square of } w_0. \text{ This is the variance of } \bar{y}_T' = \sum \tau_k y_k / N. \text{ Because this mean is based on all population units and if few units are subject to trimming in a sample of size } n, \text{ the mean and the variance can be relatively small.} \]

3. In the component \( (w_0 \bar{y}_{NT})^2 \text{Var}(\Pi_t) \),
\( \text{Var}(\Pi_t) = \text{the variance of the proportion of units that are trimmed in a sample of size } n. \text{ If few units are subject to trimming and the sampling weight distribution is relatively smooth at the higher end (i.e., the weights gradually get larger), the mean and the variance are also likely to be small. Because the variance is multiplied by the square of } w_0 \bar{y}_{NT}, \text{ if, over repeated sampling, the proportion of units with a sampling weight } \geq w_0 \text{ varies substantially, this factor can contribute substantially to the variance of } \hat{y}_T. \text{ On the other hand, if the weight distribution has a cluster of weights substantially higher than the remainder of the distribution, the proportion of units with large weights may vary substantially among repeated samples.} \)
4. \( w_0^2 \bar{y}_{NT} \text{Cov}(\bar{y}_{x_t}, \Pi_t) \) is the covariance of the mean of the data for the units with weight greater than or equal to \( w_0 \) and the proportion of units with weight greater than or equal to \( w_0 \).

The variance can be also written as

\[
\text{Var}(\hat{\bar{y}}_t) = \left\{ \left[1 - w_0, \Pi_t \right]/(1 - \theta_T) \right\}^2 \text{Var}((1 - \tau_k)Y_k) \\
+ w_0^2 \text{Var}(u_k - \bar{y}_{NT}g_k) \}. \\
= \left\{ \left[1 - w_0, \Pi_t \right]/(1 - \theta_T) \right\}^2 \text{Var}((1 - \tau_k)Y_k) \\
+ w_0^2 \text{Var}(\tau_k \bar{y}_k - \bar{y}_{NT} - \bar{y}_{NT} \tau_k - \Pi_t \bar{y}_{NT}) \}. \\
= \left\{ \left[1 - w_0, \Pi_t \right]/(1 - \theta_T) \right\}^2 \text{Var}((1 - \tau_k)Y_k) \\
+ w_0^2 \text{Var}(\tau_k \bar{y}_k - \bar{y}_{NT} - (\bar{y}_{xt} - \Pi_t \bar{y}_{NT})) \}. \\
\] (42)

The second variance component,

\[
\text{Var}(\tau_k \bar{y}_k - \bar{y}_{NT} - (\bar{y}_{xt} - \Pi_t \bar{y}_{NT})) \\
\]
indicates that if the data for the units with a weight greater than or equal to \( w_0 \) are similar to the mean of the untrimmed weights \( \bar{y}_{NT} \), both individually and in mean, this component will be small. On the other hand, if the data for the units with a weight greater than or equal to \( w_0 \) are substantially different from the data for the units with smaller weights, this variance can contribute substantially to the overall variance of the estimator.

The increase in this component of the variance is analogous to the effect shown in the bias of \( \hat{\bar{y}}_t \). That is, if the data for the units
with a trimmed weight are similar to the mean for the units with an untrimmed weight, the bias is less than if the data differ substantially.

3. The Taylor Series Estimator of the Variance

Based on the Taylor series approximation to the variance, a variance estimator can be developed. To estimate the variance of \( \hat{Y}_t \), the sample estimates are substituted for the population values in \( z \).

That is,

\[
\hat{z}_k = w_k \{(1 - w_o \hat{\Pi}_t)/(1 - \hat{\Theta}_t)\} \hat{v}_k / \hat{N} \\
+ (w_o / \hat{N}) \hat{u}_k - (w_o \hat{\gamma}_{NT} / \hat{N}) \hat{g}_k \}
\]

\[
= (1/\hat{N}) \{ w_k (1 - w_o \hat{\Pi}_t)/(1 - \hat{\Theta}_t) \} \hat{v}_k \\
+ w_k w_o \hat{u}_k - \hat{\gamma}_{NT} w_k w_o \hat{g}_k \}
\]

(43)

Now

\( \hat{\gamma}_{NT} = S(1 - \tau_k) w_k Y_k / (\hat{N} - \hat{N}_T) \),

\( \hat{\gamma}_{\pi t}' = S \tau_k \pi_k w_k Y_k / \hat{N} \)

\( = S \tau_k \hat{Y}_k / \hat{N} \),

\( \hat{\Pi}_t = S \tau_k \pi_k w_k / \hat{N} \)

\( = S \tau_k / \hat{N} \)

\( = n_t / \hat{N} \),

and

\( \hat{\Theta}_t = \hat{N}_T / \hat{N} \).
Also,
\[
(1 - w_o \tilde{\Pi}_t)/(1 - \Theta_t) = (\hat{N} - w_o n_t) / (\hat{N} - \hat{N}_t) = A_s,
\]
as defined in equation (2).

Now, the linearized components are as follows:
\[
\begin{align*}
\hat{v}_k &= (1 - \tau_k) \left( \gamma_k - \hat{\gamma}_{NT} \right) \\
w_k w_o \hat{u}_k &= w_k w_o \left( \tau_k \pi_k \gamma_k - \hat{\gamma}_{\pi_t} \right) \\
 &= (\tau_k w_o \gamma_k - w_k w_o \hat{\gamma}_{\pi_t}) \\
w_k w_o \hat{g}_k &= w_k w_o \left( \tau_k \pi_k - \tilde{\Pi}_t \right) \\
 &= (\tau_k w_o - w_k w_o n_t / \hat{N})
\end{align*}
\]

The linearized variate \( z \) in equation (43) can be estimated as
\[
\hat{z}_k = (1/\hat{N}) \{ A_s w_k (1 - \tau_k) \left( \gamma_k - \hat{\gamma}_{NT} \right) \\
+ (\tau_k w_o \gamma_k - w_k w_o S \tau_k \gamma_k / \hat{N}) \\
- \hat{\gamma}_{NT} (\tau_k w_o - w_k w_o n_t / \hat{N}) \}
\]

\[
= (1/\hat{N}) \{ [A_s w_k (1 - \tau_k) + \tau_k w_o] \left( \gamma_k - \hat{\gamma}_{NT} \right) \\
+ w_k \hat{\gamma}_{NT} w_o (n_t / \hat{N}) - w_k w_o S \tau_k \gamma_k / \hat{N} \}
\]

\[
= (1/\hat{N}) \{ w_k \tau_k (\gamma_k - \hat{\gamma}_{NT}) \\
+ w_k w_o \left[ (n_t / \hat{N}) \hat{\gamma}_{NT} - S \tau_k \gamma_k / \hat{N} \right] \}
\]

where
\[
w_{kt} = A_s w_k (1 - \tau_k) + \tau_k w_o.
\]
Thus,
\[
\hat{z}_k = \left\{ \frac{1}{N} \right\} \left\{ w_{kt} (y_k - \hat{y}_{NT}) + w_k \left[ \hat{y}_{NT} - \hat{y}_{NT} + \left( w_o nt / \hat{N} \right) \hat{y}_{NT} - w_o s \tau_k y_k / \hat{N} \right] \right\}
\]

\[
= \left\{ \frac{1}{N} \right\} \left\{ w_{kt} (y_k - \hat{y}_{NT}) + w_k \left[ \hat{y}_{NT} - \left( \hat{N} / \hat{N} - w_o nt / \hat{N} \right) \hat{y}_{NT} - w_o s \tau_k y_k / \hat{N} \right] \right\}
\]

\[
= \left\{ \frac{1}{N} \right\} \left\{ w_{kt} (y_k - \hat{y}_{NT}) + w_k \left[ \hat{y}_{NT} - \left( \hat{N} - w_o nt / \hat{N} \right) / \left( \hat{N} - \hat{N}_T \right) \hat{y}_{NT}' - w_o s \tau_k y_k / \hat{N} \right] \right\}
\]

where
\[
\hat{y}_{NT}' = S (1 - \tau_k) w_k y_k / \hat{N}.
\]

Therefore,
\[
\hat{z}_k = \left\{ \frac{1}{N} \right\} \left\{ w_{kt} (y_k - \hat{y}_{NT}) + w_k \left[ \hat{y}_{NT} - \hat{y}_t \right] \right\}
\]

\[
\text{where}
\]
\[
\hat{y}_t = \left( \hat{N} - w_o nt \right) / \left( \hat{N} - \hat{N}_T \right) \hat{y}_{NT}' - w_o s \tau_k y_k / \hat{N}
\]
\[
= S w_{kt} y_k / \hat{N}
\]

The variance is estimated using the usual variance estimator for an unequal probability with replacement sample design when a Taylor series linearized variate is being used. That is,
\[
s^2 = S n z_k^2 / (n - 1)
\]
because
\[ \hat{z} = S \frac{z_k}{n} = 0. \]
The computational form will be relatively easy to use because the only factor added to the computation is \( \hat{y}_{NT} \).

An alternative form for \( \hat{z}_k \) can be developed from equation (44) by noting that
\[ n_t = S \tau_k. \]
That is,
\[ \hat{z}_k = \frac{1}{N} \left\{ \frac{w_{kt}}{S} (Y_k - \hat{y}_{NT}) \right\} \\
+ \frac{w_k}{N} \left\{ \left( S \tau_k w_o \hat{y}_{NT} - S \tau_k w_o Y_k \right) / N \right\} \]

\[ = \frac{1}{N} \left\{ \frac{w_{kt}}{S} (Y_k - \hat{y}_{NT}) \right\} \\
- \frac{w_k}{N} \left\{ \left( S \tau_k w_o (Y_k - \hat{y}_{NT}) / N \right) \right\}. \] (46)

In this equation, we can see that \( \hat{y}_{NT} \) becomes the pivotal quantity for variance estimate. In the first part of equation (46), each observation is compared to \( \hat{y}_{NT} \) and, in the second term, the summation inside the brackets is one of the components of the bias.

C. A Generalized Design Effect Measure for Unequal Weights

The procedures to identify extreme weights are based on the assumption that extreme weights do adversely affect the precision of the estimate. If the selection probabilities are positively correlated with the observed data then the extreme weights may be beneficial. As a quick evaluation of whether the extreme weights adversely effect the precision, a measure of the dispersion of the weights is needed that incorporates the potential correlation between
the selection probabilities (and, therefore, the sampling weights) and the data.

In addition, both the Taylor series procedure and the single draw variate (SDV) procedure (described in Chapter IV) utilize data in the weight trimming process. In situations when multiple data items are available for the weight trimming, selecting from among these data items may be necessary to efficiently use resources. The generalized measure of the design effect of unequal weighting (DEFF\textsubscript{WG}) described in this section can be used both as a tool to assess the need for weight trimming and to assist the analyst in selecting the data items for the trimming procedure. DEFF\textsubscript{WG} is found to be a function of the usual design effect measure (DEFF\textsubscript{W}) and the proportion of the total variation in the data that is not explained by the sampling weights.

Kish (1965) introduced the concept of the design effect (DEFF) for a sample survey. In the literature review (Chapter II), the usual estimator (DEFF\textsubscript{W}) of the design effect of unequal weighting was described. The DEFF for unequal weighting (DEFF\textsubscript{W}) can be represented as

\[
\text{DEFF}_{\text{W}} = \frac{\text{Var}_d(\hat{Y})}{\text{Var}_s(\hat{Y})} = \left[ \frac{\sigma^2 \sum w_k^2 / (\sum w_k)^2}{\sigma^2 / n} \right] = n \frac{\sum w_k^2 / (\sum w_k)^2}. \tag{47}
\]

When the selection probabilities are determined based on an assumed relationship between the observed data and a size measure (as in the case of probability proportional to size sampling), then the observed data are expect to be related to the sampling weights. The
usual form for assessing the design effect attributable to unequal weighting does not incorporate this relationship.

Let the observed data $Y_k$ be represented in the form of the linear model such that

$$Y_k = a + b \ p_k + \epsilon_k'$$

where

$a$ and $b$ = unknown constants
$p_k$ = the single draw selection probability, and
$\epsilon_k'$ = the unit error term for the model.

Assume the expected value of $\epsilon_k'$ equals zero and the variance of $\epsilon_k'$ equals $\sigma_m^2$ for all units, that is,

$$E(\epsilon_k') = 0$$

and

$$V(\epsilon_k') = \sigma_m^2$$

where

$$\sigma_m^2 = \text{the unit variance under the model (b not equal to 0).}$$

The model can be rewritten in terms of the sampling weight $w_k$, where

$$w_k = 1 / n \ p_k$$

as

$$Y_k = a + b / (n \ w_k) + \epsilon_k'.$$

Then

$$w_k \ Y_k = b / n + w_k \ (a + \epsilon_k')$$

$$= b / n + a \ w_k + \epsilon_k' \ w_k.$$ 

The variance of $w_k \ Y_k$, assuming the model, is

$$\text{Var}(w_k \ Y_k) = \text{Var}(b / n + a \ w_k + \epsilon_k' \ w_k)$$

$$= w_k^2 \ \text{Var}(\epsilon_k')$$

$$= w_k^2 \ \sigma_m^2.$$
Because the estimator for a total $Y$ can be written as
\[
\hat{Y} = S w_k Y_k,
\]
the variance (assuming this model) can be written as
\[
\text{Var}_d(\hat{Y}_d) = \text{Var}(S w_k Y_k) = \sigma_m^2 S w_k^2.
\]
(49)

Alternatively, an estimator of a total, assuming a simple random sample (srs), can be written as
\[
\hat{Y}_s = N S Y_k / n
\]
where
\[
N = \text{the count of frame units.}
\]

By assuming the model-based unit variance (assuming no relationship between the data and the selection probability, i.e., $b=0$) is $\sigma_e^2$, that is
\[
\text{Var}(Y_k) = \text{Var}(a + \epsilon_k) = \text{Var}(\epsilon_k) = \sigma_e^2.
\]

The variance of the srs estimator of a total can be written as
\[
\text{Var}_s(\hat{Y}_s) = \text{Var}(N S Y_k / n) = N^2 \sigma_e^2 / n.
\]
(50)

By taking the ratio of equation (49) and equation (50), the DEFF attributable to unequal weighting is then
\[
\text{DEFF}_{wg} = \frac{\text{Var}_d(\hat{Y})}{\text{Var}_s(\hat{Y}_s)} = \frac{\left( n S w_k^2 / N^2 \right)}{\left( \sigma_m^2 / \sigma_e^2 \right)}.
\]
(51)

The first term of equation (51) is nearly equivalent to the usual form of the DEFF attributable to unequal weighting (equation (47)) because the sum of the weights is an estimate of $N$. 

61
The second term of equation (51) is the ratio of the unit variance assuming a relationship (b not equal to 0) to the unit variance assuming no relationship (b equal to 0). In terms of simple linear regression, this is equivalent to the ratio of the residual error to the total variation in the data. The residual error can be written as the difference between the total variation and the variation explained by the model, that is

$$
\sigma_m^2 = \sigma_e^2 - \sigma^2(\text{model}).
$$

Therefore, the second part of equation (51) can be written as

$$
\frac{\sigma_m^2}{\sigma_e^2} = \frac{\sigma_e^2 - \sigma^2(\text{model})}{\sigma_e^2} = 1 - \left(\frac{\sigma^2(\text{model})}{\sigma_e^2}\right) = 1 - R^2. \tag{52}
$$

The term $R^2$ in equation (52) denotes the proportion of the total variation explained by the model when $b$ is not equal to zero. More simply, this is the square of the correlation coefficient between the selection probabilities and the data. Using equation (51) and equation (52), the generalized DEFF attributable to unequal weighting can be written as

$$
\text{DEFF}_{WG} = \left( n \sum w_k^2 / N^2 \right) \left( 1 - R^2 \right) \equiv \text{DEFF}_w (1 - R^2). \tag{53}
$$

The generalized DEFF measure can be also derived for stratified designs with different linear models for each stratum.

This formulation of the design effect can be used to better assess the loss of precision attributable to unequal weighting and to determine if weight trimming is needed. For example, if for some key data items, the correlation between the data and the sampling weights
is highly negative (close to -1.0), the DEFF_{WG} measure may indicate that trimming is unnecessary even if the usual DEFF_{W} indicates a large unequal weighting effect. The negative correlation between the data and the sampling weights imply a positive correlation between the data and the selection probability as was desired if pps selection was used. If the correlation of 1.0 exists between the data and the selection probabilities, and the intercept (the term a) in the model equals 0.0, then sampling theory indicates that the sampling variance will be zero. The generalized DEFF may indicate that for some strata the weights do not need trimming and for other strata the weights do need to be trimmed, although DEFF_{W} is the same for all strata.

If a weight trimming strategy is used when a high negative correlation existed initially between the weights and the data, the form demonstrates (equation 53) the relationship between the effect of weight trimming or smoothing and reductions in the correlation between the weights and the data. DEFF_{WG} can be used to determine if the weight trimming is reducing the correlation between the data and the weights more than reducing the variation in the weights. A substantial reduction in the correlation may offset the reduction imposed by trimming the weights and, thereby, affect the overall reduction in the design effect.

The assessment of the correlation between the weights and the data can also be used to identify candidate data items to assess the effect of trimming. For example if, among a set of key data items, some items are highly negatively correlated to the sampling weights and others are not, then selecting some items from each group would
provide a better set of key data items for use in a trimming procedures such as the estimated mean squared error procedure.

D. A Distribution Model for Sampling Weights

Some weight trimming procedures assume that data are available when the weight trimming is performed. However, in some instances, the weights need to be ready prior to having the data available. For example, in some epidemiological research surveys, the data collected need to be combined to form composite indexes with score weights developed based on factor analysis. The sampling weights are needed to perform the factor analyses.

The NAEP procedure can be performed when no data are available. However, based on personal experience with the NAEP procedure, the procedure can lead to excessively trimmed weights.

The following derivations show a distributional model for the sampling weights based on somewhat general assumptions about the selection probabilities. Because the sampling weight is the inverse of the selection probability, a distribution model for the sampling weights can be derived assuming a beta distribution for the selection probability. The distribution model for sampling weights is essentially a form of the beta distribution with parameters $\alpha$ and $\beta$. The expected value and variance are derived and estimators are developed for $\alpha$ and $\beta$. It is also shown that the cumulative distribution function (CDF) can be evaluated from the beta distribution using the estimated values for $\alpha$ and $\beta$. 

64
The key results are that, when a standard beta distribution is assumed for the single draw selection probabilities, the distribution model for the sampling weights is in a form of a beta distribution. The rth moment of the weights distribution is shown to be

\[ E(w^r) = \frac{1}{n^r} \frac{\Gamma(a + \beta) \Gamma(a - r)}{\Gamma(a + \beta - r) \Gamma(a)} \]

The mean and variance of the weight distribution is then

\[ E(w) = \frac{a + \beta - 1}{n (a - 1)} \]

\[ \text{Var}(w) = \frac{\beta (a + \beta - 1)}{n^2 (a - 1)^2 (a - 2)} . \]

Using the method of moments, estimates of \( a \) and \( \beta \) are derived. The estimators for \( a \) and \( \beta \) are

\[ \hat{a} = \left[ \bar{w} (n \bar{w} - 1) / n \right] s_{w}^2 + 2 \]

\[ \beta = (n \bar{w} - 1) \left[ \bar{w} (n \bar{w} - 1) / n \right] s_{w}^2 + 1 \]

where

\[ \bar{w} = \text{sample mean of the weights}; \text{ and} \]

\[ s_{w}^2 = \frac{S (w_k - \bar{w})^2}{n} . \]

It is also shown that the cumulative distribution function (CDF) for the weight distribution is a simple function of the CDF for the beta distribution.

Let \( z_k \) denote a random variable representing the selection probability \( p_k \) then
\[ f_{z_k}(z) = z^{a-1}(1-z)^{\beta-1} / B(a, \beta) \]  

where

\[ B(a, \beta) = \Gamma(a) \cdot \Gamma(\beta) / \Gamma(a + \beta) \]

and, for integer values of \( a \) and \( \beta \),

\[ \Gamma(a) = (a-1)!, \]
\[ \Gamma(\beta) = (\beta-1)!, \]

and

\[ \Gamma(a + \beta) = (a + \beta-1)! \]

The use of the beta distribution is proposed for two reasons. First, the beta distribution is very flexible and can be used to model various distributional patterns. For example, if both \( a \) and \( \beta \) equal 3, the beta distribution is similar to the normal distribution in the range 0 to 1. The beta distribution can also be skewed right or left depending on the values of \( a \) and \( \beta \). The expected value of a beta random variable is

\[ E(z_k) = a/(a + \beta) \]

The variance is

\[ V(z_k) = a \beta / (a + \beta)^2 (a + \beta + 1) \]

The second reason is that, in the sense that the selection probabilities can be represented as a function of random variables (for example, size measures with a standard gamma distribution), the selection probabilities can be shown to have a beta distribution. Let \( X_k (k=1, \ldots, N) \) be random variables with the standard gamma distribution. That is, with parameter \( a_k \)
\[ f_{X_k}(x) = x^{a_k-1} e^{-x} / \Gamma(a_k) \]

From basic distribution theory, we know that a random variable \( z_k \), defined as

\[ z_k = X_k / (X_k + X_j) \]

has the beta distribution with parameters \( a_k \) and \( a_j \).

In terms of the selection probabilities, a random variable \( z_k \) can be defined as

\[ z_k = X_k / (\Sigma X_k) \]

\[ = X_k / (X_k + \Sigma_{j \neq k} X_j) \]

Since the sum of standard gamma random variables with parameters \( a_k \) has a gamma distribution with parameter \( \Sigma a_k \) then \( z_k \) has a beta distribution with parameters \( a_k \) and \( \Sigma_{j \neq k} a_j \).

Thus, in probability proportional to size sampling, if the size measures \( X_k \) follow the standard gamma distribution with parameter \( a_k \) then the single draw selection probability \( p_k \) will have the beta distribution \( a_k \) and \( \Sigma_{j \neq k} a_j \).

Let \( z \) be a random variable distributed with beta distribution with parameters \( \alpha \) and \( \beta \). Then

\[ f_z(z) = z^{\alpha-1} (1-z)^{\beta-1} / B(\alpha, \beta) \]

for \( 0 \leq z < 1 \)
If $z = p_k$, then for the sampling weight $w_k$, that is
\[ w_k = \frac{1}{n} p_k, \]
and the comparable random variable
\[ w = \frac{1}{n} z. \]
The range of $w$ is from $1/n$ to infinity. The distribution for $w$ can be derived by noting that for
\[ w = \frac{1}{n} z, \text{ and} \]
\[ z = \frac{1}{n} w. \]
The distribution for $w$ can be derived using the usual transformation techniques for a univariate distribution. The derivative of $z$ with respect to $w$ is
\[ \frac{dz}{dw} = \frac{-1}{nw^2}. \]
Then
\[ f_w(w) = f_z(z) \frac{dz}{dw}. \]
\[ f_w(w) = \frac{(1/nw)^{a-1}(1 - 1/nw)\beta^{-1}(1/nw^2)}{B(a, \beta)} \]
\[ = n \frac{(1/nw)^{a+1} (1 - 1/nw)\beta^{-1}}{B(a, \beta)} \]
for $1/n \leq w \leq \infty$. (55)
The expected value and variance of $w$ can be derived as follows by first deriving the expectation of $w^r$.
\[ E(w^r) = \int_{1/n}^{\infty} w^r \left[ n \frac{(1/nw)^{a+1} (1 - 1/nw)\beta^{-1}}{B(a, \beta)} \right] dw \]
\[ = \frac{1}{n^r} \int_{1/n}^{\infty} \left[ n \frac{(1/nw)^{a-r+1} (1 - 1/nw)\beta^{-1}}{B(a, \beta)} \right] dw \]
Let $u = 1 / nw$, then

$$w = 1 / nu$$

and

$$dw = 1 / nu^2 du.$$ 

Note also $0 \leq u \leq 1$.

$$E(w^n) = \frac{1}{n^r} \int_1^{\frac{1}{nu} \Gamma(a+r)} \left[ u^{a-r+1} (1 - u)^{\beta-1} \right] \frac{\Gamma(a+\beta)}{\Gamma(a) \Gamma(\beta)} du$$

$$= \frac{1}{n^r} \frac{\Gamma(a+\beta)}{\Gamma(a+\beta-r) \Gamma(a)} \int_0^1 u^{a-r-1} (1 - u)^{\beta-1} \frac{\Gamma(a+\beta-r)}{\Gamma(a-r) \Gamma(\beta)} du$$

Therefore, $E(w^n)$ is

$$E(w^n) = \frac{1}{n^r} \frac{\Gamma(a+\beta)}{\Gamma(a+\beta-r) \Gamma(a)}$$

Using this form for $E(w^n)$, we can evaluate for $E(w)$ and $E(w^2)$.

$$E(w) = \frac{a + \beta - 1}{n (a - 1)}$$

$$E(w^2) = \left( \frac{1}{n} \right)^2 \frac{(a + \beta - 1)(a + \beta - 2)}{(a - 1)(a - 2)}$$

The variance of $w$ is given by

$$\text{Var}(w) = \frac{1}{n^2} \left[ \frac{(a+\beta-1)(a+\beta-2)}{(a-1)(a-2)} \right] - \frac{1}{n^2} \frac{(a+\beta-1)^2}{(a-1)^2}$$

$$= \frac{1}{n^2} \frac{\beta (a + \beta - 1)}{(a - 1)^2(a - 2)}$$

Therefore the mean and variance of $w_k$ are

$$E(w) = (a + \beta - 1) / n (a - 1)$$

$$\text{(56)}$$

$$\text{(57)}$$

$$\text{(58)}$$
\[ \text{Var}(w) = \left[ \frac{\beta (\alpha + \beta - 1)}{n^2 (\alpha - 1)^2 (\alpha - 2)} \right] \]  

(59)

The distribution has a positive finite mean if \( \alpha \) is greater than 1, and a positive finite variance if \( \alpha \) is greater than 2.

Using the method of moments, estimates for \( \alpha \) and \( \beta \) can be derived assuming an unequal probability with replacement sample design. That is, estimates for \( \alpha \) and \( \beta \) can be determined by setting

\[ \bar{w} = \frac{(\alpha + \beta - 1)}{n (\alpha - 1)}, \]  

(60)

and

\[ s_w^2 = \frac{\beta (\alpha + \beta - 1)}{n^2 (\alpha - 1)^2 (\alpha - 2)} \]  

(61)

where

\[ \bar{w} = \frac{\sum w_k}{n} \]

\[ s_w^2 = \frac{\sum (w_k - \bar{w})^2}{n}. \]

The estimators for \( \alpha \) and \( \beta \) are the following.

\[ \hat{\alpha} = \left[ w \left( \frac{n \bar{w} - 1}{n} \right) s_w^2 \right] + 2 \]

\[ \hat{\beta} = (n \bar{w} - 1) \left[ \bar{w} \left( \frac{n \bar{w} - 1}{n} \right) s_w^2 + 1 \right] \]

and are derived as follows.

By assigning

\[ \bar{w} = \frac{(\alpha + \beta - 1)}{n (\alpha - 1)}, \]

then

\[ \alpha = \left[ \frac{\beta}{(n \bar{w} - 1)} \right] + 1 \]

(62)

Note that

\[ \alpha - 1 = \frac{\beta}{(n \bar{w} - 1)} \]
and

\[ a - 2 = (\alpha - 1) - 1 \]

\[ = \left[ \beta / (n\bar{w} - 1) \right] - 1 \]

\[ = (\beta - n\bar{w} + 1) / (n\bar{w} - 1) \]

From equation (61)

\[ s_w^2 = \frac{\beta (\alpha + \beta - 1)}{n^2 (\alpha - 1)^2 (\alpha - 2)} \]

Then

\[ s_w^2 n^2 (\alpha - 1)^2 (\alpha - 2) = \beta (\beta + \alpha - 1) \]

and

\[ s_w^2 \left[ n^2 \frac{\beta^2}{(n\bar{w} - 1)^2} \cdot \frac{(\beta - n\bar{w} + 1)}{(n\bar{w} - 1)} \right] = \beta \left[ \beta + \frac{\beta}{n\bar{w} - 1} \right] \]

\[ \beta - (n\bar{w} - 1) = \bar{w} (n\bar{w} - 1)^2 / n \ s_w^2 \]

Therefore,

\[ \beta = \left[ \bar{w} (n\bar{w} - 1)^2 / n \ s_w^2 \right] + (n\bar{w} - 1) \]

\[ = (n\bar{w} - 1) [\bar{w} (n\bar{w} - 1)/n \ s_w^2 + 1] . \]

That is, the estimator for \( \beta \) is

\[ \hat{\beta} = (n\bar{w} - 1) [\bar{w} (n\bar{w} - 1)/n \ s_w^2 + 1] . \]  

(63)

Because from equation (62)

\[ a = \left[ \beta / (n\bar{w} - 1) \right] + 1 \]

the estimator for \( \alpha \) is given by

\[ \hat{\alpha} = \left[ \bar{w} (n\bar{w} - 1) / n \ s_w^2 + 1 \right] + 1 \]

\[ = \left[ \bar{w} (n\bar{w} - 1) / n \ s_w^2 \right] + 2 \]  

(64)
In summary, when the selection probabilities are assumed to follow the beta distribution with parameters \( \alpha \) and \( \beta \), then the distribution for the sampling weights, \( w \), follow a distribution in the form of a beta distribution with

\[
E(w) = \frac{[\alpha + \beta - 1]}{[n(\alpha - 1)]}
\]

\[
Var(w) = \frac{[\beta(\alpha + \beta - 1)]/[n^2 (\alpha - 1)^2 (\alpha - 2)]}{n}
\]

The estimators for \( \alpha \) and \( \beta \) are

\[
\hat{\alpha} = \frac{\bar{w}(n\bar{w} - 1)}{n s_w^2} + 2
\]

\[
\hat{\beta} = \frac{(n\bar{w} - 1) [\bar{w}(n\bar{w} - 1)]/n s_w^2 + 1}{n}
\]

where

\[
\bar{w} = \frac{\sum w_i}{n}
\]

\[
s_w^2 = \frac{\sum (w_i - \bar{w})^2}{n - 1}
\]

The estimators for \( \alpha \) and \( \beta \) can be rewritten as a function of \( \text{DEFF}_w \) (see equation (47)). Note that

\[
s_w^2 = \frac{\sum (w_i - \bar{w})^2}{n} = \bar{w}^2 [\text{DEFF}_w - 1].
\]

Let \( D = \text{DEFF}_w \), then using equation (67) in equation (65)

\[
\hat{\alpha} = \frac{\bar{w}(n\bar{w} - 1)}{[n \bar{w}^2 (D - 1)]} + 2
\]

\[
= \frac{[n \bar{w} (2 D - 1) - 1]}{[n \bar{w} (D - 1)]}.
\]

Similarly for \( \hat{\beta} \), using equation (67) in equation (66)

\[
\hat{\beta} = \frac{(n \bar{w} - 1) [\bar{w}(n\bar{w} - 1)]/n \bar{w}^2(D - 1) + 1}{n}
\]
Therefore, the estimators for $\alpha$ and $\beta$ can both be considered as a function of the design effect $D$ and the sum of the weights $(\bar{w})$. That is

$$\hat{\alpha} = \frac{[n \bar{w} (2D - 1) - 1]}{[n \bar{w} (D - 1)]},$$

(68)

$$\hat{\beta} = \frac{(n \bar{w} - 1)(n \bar{w} D - 1)}{(n \bar{w} (D - 1))}$$

(69)

As $n\bar{w}$ gets large, then the estimator for $\hat{\alpha}$ can be approximated by

$$\hat{\alpha} = \frac{(2D - 1)}{(D - 1)}.$$

(70)

However, the estimator for $\hat{\beta}$ will increase as $n\bar{w}$ increases.

To develop a simple form for the cumulative distribution function (CDF), we note that the density function for the weight distribution is given as

$$f_{w_k}(w) = n \frac{1}{nw}^{\alpha+1} (1 - \frac{1}{nw})^{\beta-1} / B(\alpha, \beta)$$

for $1/n \leq w \leq \infty$

where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + \beta)}$$

That is, the density function is given by

$$f_{w_k}(w) = \left( \frac{1}{n w} \right)^{\alpha+1} \left( 1 - \frac{1}{n w} \right)^{\beta-1} \frac{n}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\beta)}.$$

Because this function is based on the beta distribution, we can use the beta distribution to determine approximate criterion levels for assessing extreme weights.
The cumulative distribution function (CDF) for this weight distribution \( F_w \) is

\[
F_w(w_0) = \int_{1/n}^{w_0} f_w(w) \, dw.
\]

Let \( u = 1 / nw \), then

\[
F_w(w_0) = P(w \leq w_0) = P(nw \leq nw_0) = P(1/nw > 1/nw_0) = 1 - P(u \leq 1/nw_0).
\]

The CDF for the weight distribution can be written as

\[
F_w(w_0) = \int_{1/n}^{w_0} n (1 - 1/nw)^{\beta-1} (1/nw)^{\alpha+1} / B(\alpha, \beta).
\]

By using \( u = 1/nw \), the limits of integration for \( u \) become 0 and 1. The change of variable transformation is

\[
F_w(w_0) = \int_{1}^{1/nw_0} n (1 - u)^{\beta-1} u^{\alpha+1} \left( \frac{1}{n u^2} \right) / B(\alpha, \beta)
\]

\[
F_w(w_0) = 1 - \int_{0}^{1/nw_0} (1 - u)^{\beta-1} u^{\alpha-1} / B(\alpha, \beta)
\] (71)

Because the integral in equation (71) is the CDF for the beta distribution with parameters \( \alpha \) and \( \beta \) evaluated from 0 to 1/nw_0, the criterion levels for the weight distribution can be determined from the beta distribution. The values for \( \alpha \) and \( \beta \) can be estimated from
the sample weights (equations (65) and (66)) and critical values for the weight distribution can be estimated using the beta CDF function in computer software such as SAS (SAS Institute Inc. 1985).

Alternatively, if an upper bound for $DEFF_w$ is set then equations (68) and (69) can be used to generate estimates of $a$ and $\beta$.

In Chapter IV, I will describe a weight trimming procedure based on this distribution using criterion values generated from the beta CDF.
Chapter IV

EMPIRICAL INVESTIGATION: TRIMMING PROCEDURES

A. Overview

In this chapter, three proposed weight trimming procedures will be described and compared to the estimated MSE and the NAEP procedure described in Chapter II. I will also describe the how these five procedures are implemented in the empirical study.

The three proposed weight trimming procedures are the following:

(1) Taylor Series Procedure

The Taylor series procedure uses the estimated MSE and the estimated relative bias computed for each data item at multiple candidate trimming levels. The "optimal" trimming level is the trimming level (from among the candidate trimming levels) that results in a minimum composite score for the estimated MSE and relative bias across the data items. The estimated MSE is computed using the derived forms for the bias and Taylor series linearized variate. Therefore, this procedure permits the assessment of both the potential for bias and the variance reduction introduced by the weight trimming.

(2) The SDV Procedure

The second weight trimming procedure is based on single draw variates (SDVs). As described later in this chapter, a SDV is a function of both the observed data and the sampling weights. Therefore, the SDV for a sample member naturally incorporates the
relationship between the sampling weight and the observed data and an extreme SDV may be caused by excessively large sampling weights.

The weight trimming strategy is developed using studentized residuals of the SDVs. The studentized residuals for the sampling weights and the SDVs are compared to a criterion value in a systematic fashion to identify extreme SDVs that can be attributed to excessively large sampling weights. Weights identified as excessively large are trimmed and the excess is distributed among the other weights. Revised SDVs and studentized residuals are computed using these trimmed adjusted weights and are evaluated for extreme SDVs again. The studentized residuals of the SDVs are compared to the criterion value a total of 10 times to identify the "optimal" trimming level.

(3) The Weight Distribution Procedure

Because some estimates resulting from sample survey are based on dichotomous or polychotomous response variables or sometimes data are not available when weight trimming is performed, the third weight trimming procedure utilizes 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 prespecified 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 prespecified 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. The comparison of the sampling weights to the trimming level is performed 10 times. This weight trimming procedure identifies and trims sampling weights with a small probability of occurrence, based on the model.

The two currently used procedures also will be briefly reviewed in this chapter. These procedures are: (1) the estimated mean square error (estimated MSE procedure) and (2) the NAEP procedure. The estimated MSE procedure uses data in the trimming procedure; the NAEP procedure, as described, does not use data. The full descriptions of these procedures are in Chapter II.

B. Comparison of Proposed and Current Procedures

The five procedures can be classified as two pairs of direct competitors and the fifth procedure which is a competitor to both pairs of direct competitors. The Taylor series procedure and the estimated MSE procedure are direct competitors because both use the survey data and the estimated MSE for determining the trimming necessary. The Taylor series procedure represents an improvement over the estimated MSE procedure because it is based on a representation of the trimmed weight variance estimate that takes into account the
weight trimming. This procedure also provides an assessment of the potential bias introduced when a weight is trimmed. The Taylor series procedure is an improvement over the NAEP procedure because it takes into account the observed data, the potential bias and the variance effect.

The weight distribution and the NAEP procedures also are direct competitors because both use statistics based on only the sampling weights. The weight distribution procedure can be viewed as an improvement over the NAEP procedure because the trimming criterion can be stated in terms of a probability statement. The only documentation of the NAEP procedure suggests that the trimming criterion be established by an empirical assessment of the sampling weights.

The SDV procedure incorporates some of the characteristics of both pairs of direct competitors. Because the SDV uses the observed data, it is a competitor to both the Taylor series and estimated MSE procedures. Also because the SDV procedure directly incorporates the sampling weights, it is a competitor to the weight distribution and the NAEP procedure. A disadvantage to the SDV procedure is that a distribution model is assumed for assessing the studentized residuals of the SDVs. The primary characteristics of the five procedures are summarized in Table 2.

C. The Proposed Procedures

Two of the three proposed procedures (the Taylor series procedure and the SDV procedure) utilize survey data. For these two procedures, the analyst needs to select the data items of key importance and
Table 2. Summary Of Characteristics of Trimming Procedures

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Uses Data</th>
<th>Uses Weights</th>
<th>Trimming Criterion Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Taylor series</td>
<td>Yes</td>
<td>No</td>
<td>Minimize MSE and relative bias</td>
</tr>
<tr>
<td>2. Weight Distribution</td>
<td>No</td>
<td>Yes</td>
<td>Distribution model</td>
</tr>
<tr>
<td>3. SDV Procedure</td>
<td>Yes</td>
<td>Yes</td>
<td>Assume distribution model</td>
</tr>
<tr>
<td>4. Estimated MSE</td>
<td>Yes</td>
<td>No</td>
<td>Minimize MSE</td>
</tr>
<tr>
<td>5. NAEP</td>
<td>No</td>
<td>Yes</td>
<td>Empirical evaluation of weights</td>
</tr>
</tbody>
</table>
represent the possible relationships between the weights and the data expected in the full set of data items. The generalized DEFF measure can be used to determine what data items should be used in the weight trimming. Most or all of these key data items should be non-zero because a zero value may mask extreme weights. The proposed procedures are described below.

(1) The Taylor Series Procedure

(a) Procedure Description

This extreme weight identification procedure is based on the derivation for the Taylor series approximation to the expectation, bias, and variance of the trimmed weight estimator and the estimate of the variance. This procedure utilizes the data and permits an assessment of the impact of the weight trimming on the potential bias and the variance.

For this procedure, we assume that a set of key data items are available for all responding units. That is, for each unit we have data \( Y_{k1} \) for \( l=1,\ldots,m \).

From the derivation of the bias of the trimmed mean estimator, we have from equation (25) of Chapter III.

\[
\text{Bias}(\hat{\tau}_{1}) = - \sum \tau_k \left( 1 - w_0 \pi_k \right) \left( Y_{k1} - \bar{Y} \right) / N
\]

where

\[
w_0 = \text{the trimming level for the weights},
\]

\[
\tau_k = 1 \text{ if } w_k > w_0
\]

\[
= 0 \text{ otherwise},
\]

\[
\pi_k = \text{the selection probability for the kth unit in a sample of size n}; \text{ that is},
\]
\[ \pi_k = 1 / w_k, \]

and

\[ \hat{y}_{NT1} = \Sigma (1 - \tau_k) y_{k1} / (\Sigma (1 - \tau_k)). \]

An estimator of the bias is

\[
\text{Bias}(\hat{y}_{t1}) = -S \tau_k w_k (1 - w_0 \pi_k) (y_{k1} - \hat{y}_{NT1}) / \hat{N} = -S \tau_k (w_k - w_0) (y_{k1} - \hat{y}_{NT1}) / \hat{N} \tag{1}
\]

where

\[ \hat{y}_{NT1} = S (1 - \tau_k) w_k y_{k1} / (\hat{N} - \hat{N}_T). \]

Note that, if all units with sampling weights greater than or equal to \( w_0 \) have observed data, \( y_k \), greater than \( \hat{y}_{NT} \), then the bias is negative. Conversely, if all units with sampling weights greater than or equal to \( w_0 \) have \( y_k \) less than \( \hat{y}_{NT} \), then the bias is positive.

From equation (46) in Chapter III, a form of the Taylor series linearized variate when the weight trimming is accounted for is given for the \( k \)th unit by

\[
\hat{z}_{k1} = (1/\hat{N}) \{w_{kt} (y_{k1} - \hat{y}_{NT1}) - w_k [S \tau_k w_0 (y_{k1} - \hat{y}_{NT1}) / \hat{N}] \} \tag{2}
\]

where

\[ w_{kt} = A_s w_k (1 - \tau_k) + \tau_k w_0. \]

and

\[ A_s = \text{the adjustment factor for the untrimmed weights.} \]

Note that the second component to equation (2) incorporates a component of the estimated bias in equation (1).

In this procedure, I will also use the Taylor series linearized variate when no weight trimming is performed. I will define \( \hat{z}_{kt1} \) as the Taylor series linearized variate for the \( l \)th data item when...
trimming is accounted for and \( \hat{z}_{kull} \) as the Taylor series linearized variate for the lth data item when no trimming is performed (that is, the linearized variate for untrimmed weights).

The Taylor series linearized variate \( (\hat{z}_{kull}) \) when no trimming is performed is of the form
\[
\hat{z}_{kull} = \left( \frac{1}{\hat{N}} \right) w_k (Y_{kl} - \hat{\gamma}_l) 
\]
where
\[
\hat{\gamma} = S w_k Y_k / \hat{N}
\]
It can easily be shown that
\[
\hat{z}_{kull} = \left( \frac{1}{\hat{N}} \right) \left\{ w_k (Y_{kl} - \hat{\gamma}_{NTl}) - w_k s \tau_k w_k (Y_{kl} - \hat{\gamma}_{NTl}) / \hat{N} \right\}
\]
Therefore, the Taylor series linearized variate \( (\hat{z}_{kull}) \) when no trimming is performed can be written in a form that is directly comparable to the Taylor series linearized variate \( (\hat{z}_{kltl}) \) when weight trimming is incorporated.

Using equation (2) and equation (4), the linearized variate \( (\hat{z}_{kltl}) \) can be written as a function of \( \hat{z}_{kull} \). That is,
\[
\hat{z}_{kltl} = \hat{z}_{kull} - \{(w_k - w_{kt})(Y_{kl} - \hat{\gamma}_{NTl})
\]
\[
+ w_k \text{ Bias}(\hat{\gamma}_{tll})\} / \hat{N}.
\]
In this form, the effect of the change in the weights and the effect of the bias on the linearized variate \( (\hat{z}_{kltl}) \) when weight trimming is incorporated can be easily seen. That is, when \( w_k \) is greater than \( w_o \) then the difference \( (w_k - w_{kt}) \) is positive and (if the estimated bias term is ignored) the absolute value of \( \hat{z}_{kltl} \) is reduced. Conversely, when \( w_k \) is less than \( w_o \), the difference \( (w_k - w_{kt}) \) is negative and (again ignoring the estimated bias term) the absolute value of \( \hat{z}_{kltl} \) is
increased. When the bias term is considered, we can see that a large estimated bias can offset a portion of the reductions sought by the weight trimming.

(b) The Measures For Trimming

Because the purpose of the weight trimming is to reduce the estimated variance of the estimate, the estimated variances are considered as key measures for this weight trimming procedure. The estimated variance of \( \hat{\gamma}_1 \) when no trimming is performed is

\[
\hat{\text{Var}}(\hat{\gamma}_1) = S_n \sum z_{k1}^2 / (n - 1).
\]

The estimated variance of \( \hat{\gamma}_{t1} \) when trimming is performed is

\[
\hat{\text{Var}}(\hat{\gamma}_{t1}) = S n \sum z_{k1}^2 / (n - 1).
\]

For the Taylor series trimming procedure, two key measures will be used that incorporate both the variance reduction and the potential bias effects. These measures as follows (computed for each key data item 1):

1. an estimated mean square error measure:

\[
\text{MSE}_1 = \text{Var}(\hat{\gamma}_{t1}) + \text{Bias}(\hat{\gamma}_{t1})^2,
\]

and

2. the relative bias:

\[
\text{RelBias}_1 = \text{Bias}(\hat{\gamma}_{t1}) / \hat{\gamma}_{t1}.
\]

These two measures summarize the primary information necessary for weight trimming. The approximate MSE provides a measure of the mean square error resulting from the variance reduction and the potential bias caused by weight trimming. The relative bias (RelBias_1) provides information on the estimated magnitude of the bias.
(c) The Empirical Study Procedure

To implement this procedure, I used these two measures to identify a trimming level among a set of candidate trimming levels that, jointly for multiple data items, has the smallest estimated MSE and absolute value of the relative bias. Because I used multiple data items, the identified trimming level may not be the smallest estimated MSE and bias for all or any of the data items.

In the empirical study, I identified 20 candidate trimming levels and computed 20 sets of weights for each sample. Using these candidate weights, the procedure implemented is the following:

1. For each data item, the estimated MSE and relative bias are computed for each set of weights.
2. The estimated MSE and the square of the relative bias are assigned a rank (1 to smallest value and 20 to the largest value) for each data item.
3. An average rank is computed for the estimated MSEs across the data items for each set of weights; similarly an average rank is computed for the square of the relative biases for each set of weights.
4. The two average ranks are then averaged to result in a single score for each set of weights.
5. The identified trimming level corresponds to the set of weights with the lowest average rank.

This procedure seeks the joint minimum of the estimated MSE and absolute value of the relative bias.
(2) **Single Draw Variate Based Trimming Procedure**

(a) **Single Draw Variates**

In this section, I will describe single draw variates (SDV) and, in the following sections, the SDV weight trimming procedure will be described. To discuss SDVs, assume a sample of size \( n \) from sampling frame of \( N \) units and define the following:

- \( Y_k \) = the observed data for the \( k \)th unit.
- \( p_k \) = the single draw selection probability for the \( k \)th unit.
- \( w_k \) = the untrimmed sampling weight for the \( k \)th unit.
- \( w_{kt} \) = the sampling weight for the \( k \)th unit when a weight trimming strategy is used.

\[ Y = \text{the total of the } Y_k, \text{ that is,} \]
\[ Y = \sum Y_k, \]

where

\( \Sigma \) denotes the summation over the \( N \) units.

- \( \hat{Y} \) = the usual expansion estimator for a total when the untrimmed sampling weights are used, that is,

\[ \hat{Y} = S w_k Y_k \]

where

\( S \) denotes the summation over the sample of \( n \) units.

- \( \hat{Y}_t \) = the usual expansion estimator for a total when a weight trimming strategy is used on the the sampling weights, that is

\[ \hat{Y}_t = S w_{kt} Y_k. \]

The SDV for the \( k \)th unit is defined as

\[ y_k = y_k / p_k \]
\[ = n \ w_k \ Y_k. \]  

(6)

The SDV is, therefore, a direct function of the sampling weight \( w_k \) and the observed data \( Y_k \).

In a more general sense, if propensity to response is assumed to be a random variable for each unit with expected value \( p_{rk} \), a single draw variate \( y_{rk} \) can be written that incorporates this response propensity. That is,

\[
y_{rk} = Y_k / p_k \ p_{rk}
\]

\[ = n_r \ w_{rk} \ Y_k\]

where

\[
n_r = \text{the number of respondents}
\]

\[
w_{rk} = \text{the response adjusted sampling weight}.
\]

Therefore, SDVs can be developed using the response adjusted weights. Essentially, SDVs can be developed using the final analysis weights, where these weights may incorporate response, poststratification, and the other adjustments.

The usual expansion estimator for estimating a total \( Y \) is

\[
\hat{Y} = S \ w_k \ Y_k
\]

\[ = S \ n \ w_k \ Y_k / n
\]

\[ = S \ y_k / n
\]

\[ = \bar{y}.\]

It can easily be shown that the expected value of \( y_k \) and \( \hat{Y} \) is \( Y \). That is,

\[
E(y_k) = E(\sum \lambda_k \ y_k)
\]

where

\[
\lambda_k = 1 \text{ if the kth unit is selected in a sample of size 1}
\]

\[ = 0 \text{ otherwise.}\]
By definition,

\[ E(\lambda_k) = p_k, \]

and

\[
E(y_k) = \Sigma p_k \frac{Y_k}{p_k} \\
= \Sigma p_k Y_k/p_k \\
= Y.
\]

Using this result, it can be shown that

\[ E(\hat{Y}) = E(\bar{y}) = Y. \]

When a with-replacement sampling design is used, the \( y_k \) are independent and identically distributed (iid) random variables with expectation of \( Y \) and the same variance \( V(y_k) \) where (Raj 1968)

\[
V(y_k) = \Sigma p_k \left[ \frac{Y_k}{p_k} - Y \right]^2 \\
= \Sigma p_k (y_k - \bar{y})^2.
\]

The estimator of the variance of \( y_k \) is

\[ V(\hat{y}_k) = S(y_k - \bar{y})^2/(n-1). \]  \( (7) \)

The variance of \( \hat{Y} \) is given by

\[ V(\hat{Y}) = \frac{1}{n} \Sigma p_k (y_k - \bar{y})^2 \]

The estimator of the variance of \( \hat{Y} \) is given as

\[ \hat{V}(\hat{Y}) = S(y_k - \bar{y})^2 / n(n-1) \]

Because the SDVs are iid, a distributional model can be assumed for SDVs.
(b) Procedure Description

The goal of the Single Draw Variate (SDV) based trimming procedure is to provide a tool to determine which units have excessive sampling weights and to provide candidate values to assign to the excessive weights. Using a set of key data items and the variance estimators, the procedure identifies units with relatively large SDVs. Because of the direct relationship between this SDV and the sampling weights, this procedure can identify the units for which large sampling weights may affect the precision of the estimate. The proposed procedure utilizes the ratio of the deviation of the SDV from its mean, $\bar{y}$, to an estimate of the variance of the SDV. This procedure is analogous to the current procedure that assesses the relative contribution of each weight to the variance (called the NAEP procedure in the literature review). However, because the SDV procedure utilizes data, the SDV procedure is not a direct competitor to the NAEP procedure.

The form of this procedure is to compute for each unit $k$ a measure,

$$ R_{SDV}(k) = \frac{(y_k - \bar{y})}{\sqrt{\frac{S(y_1 - \bar{y})^2}{(n-1)}}} \quad (8) $$

This ratio is essentially a studentized residual for the $y_k$ because

$$ E(y_k) = E(\bar{y}) = y, $$

and

$$ \text{Var}(y_k) = \frac{S(y_1 - \bar{y})^2}{(n-1)}. $$

An excessively large SDV can be defined as that unit for which $R_{SDV}(k)$ exceeds such a specific level $c$. The testing algorithm can also be
written as

$$R_{SDV}(k) > c(a) \text{ for a prespecified } a.$$  

The SDV procedure is developed using the usual estimators of the mean and the variance of the SDVs. These estimates can be adversely affected by extreme values. For example, the variance estimate can be inflated by extreme values and, thereby, possible mask extreme values. More robust estimators of the mean and variance can be used. Some possible estimators include Winsorized estimates or M-estimators of the mean and variance (see Barnett and Lewis 1984). The approach taken in this research is to test the basic concept of using SDVs for weight trimming. Future research efforts can build on the results of the current research to refine the basic strategy.

When the ratio ($R_{SDV}$) is computed for a single data item, it may be unclear whether the sampling weight or the observed data is affecting the SDV. Therefore, a second set of ratios (one for each unit) should be computed using only the weights, that is, with $Y_k = 1$ for all $k$. The second ratio would be computed similar to equation (6) with

$$y_k = n_k w_k$$

$$\bar{y} = S w_k$$

Therefore

$$R_{SDV}(k, w) = \frac{(n w_k - S w_i)}{[S (n w_i - S w_i)^2 / (n-1)]^{1/2}}$$

$$= (w_k - \bar{w}) / [S (w_i - \bar{w})^2 / (n - 1)]^{1/2}$$  \hspace{1cm} (9)$$

where

$$\bar{w} = S w_k / n.$$
For the weights, $R_{SDV}(k,w)$ is also compared to a prespecified value $c$. A bivariate plot of $R_{SDW}(k)$ and $R_{SDW}(k,w)$ can be used to demonstrate the effect of unequal weights. These ratios and criteria are investigated in the empirical study.

(c) **Multiple Key Data Items**

The SDV procedure can easily be generalized to a multivariate situation. The measure $R_{SDV}(k)$ can be computed for several key data items (say, $p$ items) as well as for the weights alone (by assigning $Y_k = 1$ for all $k$). A $p+1$ component vector can be formed for each unit $k$

\[
\{R_{SDV}(k)\} = \\
\{R_{SDV}(k,1), \ldots, R_{SDV}(k,1), \ldots, R_{SDV}(k,p), R_{SDV}(k,w)\}
\]

where

$R_{SDV}(k,1) = \text{the } R_{SDV}(k) \text{ value for the data item } l, l=1,\ldots,p,$

and

$R_{SDV}(k,w) = \text{the } R_{SDV}(k) \text{ value for the weights.}$

By using this vector, extreme values can be identified that adversely affect the sampling variances. This is, for units with both $R_{SDV}(k,w)$ large and $\{R_{SDV}(1), \ldots, R_{SDV}(p)\}$ large, the extreme weights can affect the precision.

(d) **The Empirical Study Procedure**

In the empirical study, multiple data items and the following rule were used to identify extreme weights. Also, instead of a $p+1$-dimensional vector, a composite measure was used. The procedure is the following:
1. The sample members are ordered in descending weight order (that is from the largest to the smallest weight).

2. Each studentized SDV ratio \( \{(R_{SDV}(i), i=1, \ldots, p), R_{SDV}(k,w)\} \) is compared to the criterion \( c \).

   a. For the largest weight,
      If any one of the \( R_{SDV} \)'s exceed \( c \) then the weight is classified as an extreme value, otherwise no weights are classified as extreme.

   b. For other than the largest weight,
      1. If any one of the \( R_{SDV} \)'s exceed \( c \) and the prior sampling weight was classified as extreme then the weight is classified as an extreme value.
      2. If all of the \( R_{SDV} \)'s are less than or equal to \( c \) and the prior sampling weight was classified as extreme then the weight is not classified as an extreme value and the checking for extreme studentized SDV ratios is stopped.
      3. If any one of the \( R_{SDV} \)'s exceed \( c \) and the prior sampling weight was not classified as extreme then the weight is also not classified as an extreme value.

The rationale for this comparison is that the value for an \( R_{SDV} \) may exceed \( c \) for any sample member. However, to focus the procedure on the weights (as opposed to the data) resulting in an extreme studentized SDV ratio, I limited the identification of extreme values to those sample members with the largest weights. After placing the
file with the weights in descending order, the checks for extreme SDV's is stopped at the first occurrence of all of the $R_{SDV}$'s being less than or equal to $c$.

(e) Assignment of Values to Trimmed Weights

When a unit has an extreme SDV or weight, a value $w_0$ is assigned to the trimmed weights and the weights for the other units can be adjusted to reproduce the original weight sum. For the SDV-based procedure, various candidate values for $w_0$ are possible. For the first iteration in the empirical study, I assigned the average of the weights deemed extreme (excluding the largest weight). The largest weight was excluded in this average because, when the largest weight was included, the trimming level was inflated by this value. For the second to the tenth iteration, the trimming level was assigned the value of the largest adjusted weight with all of the $R_{SDV}$'s being less than or equal $c$.

This procedure and the assignment of maximum allowable weight are done repeatedly until all weights are within the desired bounds. That is, after the first iteration, a revised set of SDVs are computed and the mean and variance of the SDVs are also recomputed.

(f) Distribution Model for SDVs

The SDVs are independent and identically distributed when a with-replacement sample design is used. Therefore, the criterion value, $c$, can be assigned by assuming a specific distribution for the SDVs.

If the SDVs are assumed to be normally distributed, then a criterion value can be determined to identify extreme values. In the
previous section,

\[ R_{SDV}(k) = \frac{(y_k - \bar{y})}{\sqrt{\frac{S (y_i - \bar{y})^2}{(n - 1)}}} \]

The \( R_{SDV}(k) \) are the studentized residual of the SDV. \( R_{SDV}(k) \) has zero mean and unit variance for all \( k \). For large \( n \) and assuming a normal distribution for the SDVs, the criterion level \( c \) can be assigned from the normal distribution or, for small \( n \), the \( t \)-distribution with \( n-1 \) degrees of freedom can be used to determine the criterion level. In the empirical study, the normal distribution is assumed with a criterion level of \( c = 2.33 \). This value corresponds to a one-tailed probability of 0.01 and was selected to match the probability value used in the weight distribution procedure, described in the next section.

(3) Weight Distribution

(a) Procedure Description

In Chapter III, a theoretical distribution was developed for sampling weights assuming an unequal probability with replacement sampling design with all selection probabilities less than 1.0. The distribution utilizes a Beta distribution for the selection probabilities. The density function for the distribution is

\[ f_{W_k}(w) = n \frac{(1/nw)^{a+1} (1 - 1/nw)^{\beta-1}}{B(a,\beta)} \]

for \( 1/n \leq w \leq \infty \)

where

\[ B(a,\beta) = \frac{\Gamma(a) \Gamma(\beta)}{\Gamma(a + \beta)} \]

It was also shown that estimates for alpha and beta can be computed
from the sample size, the mean weight, and the variance of the weights. That is, (from Chapter III, equations (65) and (66))

\[ \hat{a} = \frac{\tilde{w} (n\tilde{w} - 1)}{n s_w^2} + 2 \]  

\[ \hat{\beta} = (n\tilde{w} - 1) \frac{\tilde{w} (n\tilde{w} - 1)}{n s_w^2} + 1 \]  

where

\[ \tilde{w} = \frac{\sum w_i}{n} \]

\[ s_w^2 = \frac{\sum (w_i - \tilde{w})^2}{n} \]

The percentiles for the cumulative distribution function \( F_w(w) \) for the distribution can be computed using the complete Beta distribution \( \text{Beta}(x, \alpha, \beta) \) 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

\[ F_w(w_0) = 1 - \text{Beta}(1 / nw_0, \alpha, \beta) \]

That is,

\[ F_w(w_0) = 1 - \int_0^{1/nw_0} (1 - u)^{\beta-1} u^{\alpha-1} / B(\alpha, \beta) \]

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

\[ 1 - F_w(w_k) = \text{Beta} \left( \frac{1}{n w_k}, \alpha, \beta \right). \]

A weight value with an extremely low probability of occurring can be trimmed to a specific probability of occurrence. For example, an observed weight with a probability of occurrence of less than, say, 0.01 can be trimmed to a value with a larger probability (say, 0.01).

(b) The Empirical Study Procedure

For the empirical study, I set the probability of occurrence criterion 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 \( w_{op} \). For the first of 10 iterations, the original weights were used to estimate \( \alpha \) and \( \beta \) using equations (8) and (9), respectively. For the second to the tenth iteration, \( \alpha \) and \( \beta \) was estimated using the weights from the prior iteration.

D. Currently Used Procedures

(1) Estimated Mean Square Error (MSE)Trimming

(a) Procedure Description

A method used by some of the major survey research organizations is the evaluation of an estimate of the mean square error for selected data items at various trimming levels to empirically determine the trimming level (Cox & McGrath 1981, Cox 1988, Heeringa 1988). This procedure is described in detail in Chapter II.

In brief, the assumption underlying this procedure is that, for a
set of weights and data, a point exists at which the reduction in the sampling variance resulting from the trimming is offset by the increase in the square of the bias introduced into the estimate. In this procedure, the MSE($\hat{y}_t$) is estimated by

$$\text{MSE}(\hat{y}_t) = (\hat{y}_t - \hat{y})^2 - \hat{\text{Var}}(\hat{y}) + 2 [\hat{\text{Var}}(\hat{y}_t) \hat{\text{Var}}(\hat{y})]^{1/2}$$  \hspace{1cm} (12)

where

$\hat{y} =$ the estimate of the mean using the untrimmed weights;

$\hat{y}_t =$ the estimate of the mean using trimmed weights;

$\hat{\text{Var}}(\hat{y}) =$ the estimated variance of $\hat{y}$; and

$\hat{\text{Var}}(\hat{y}_t) =$ the estimated variance of $\hat{y}_t$.

The procedure is implemented by repeatedly computing the estimate of the MSE for selected set of data items at differing levels of weight truncation. The 'optimal' level of truncation is the point that minimizes estimated MSE (i.e., minimizes sampling variance and estimated squared bias) for the set of key data items.

(b) The Empirical Study Procedure

To implement this procedure, the estimated MSE is used to identify a trimming level among a set of candidate trimming levels that, jointly for multiple data items, has the smallest estimated MSE. Because multiple data items were used, the identified trimming level may not be the smallest estimated MSE for all or any of the data items.

In the empirical study, 20 candidate trimming levels were used and 20 sets of weights were computed for each sample. Using these
candidate weights, the procedure implemented is the following:

1. For each data item, the estimated MSE is computed for each set of weights.
2. The estimated MSE is assigned a rank (1 to smallest value and 20 to the largest value) for each data item.
3. An average rank is computed for the estimated MSEs across the data items for each set of weights.
4. The identified trimming level corresponds to the set of weights with the lowest average rank.

This procedure seeks the joint minimum of the estimated MSE across the data item.

(2) The NAEP Procedure

(a) Procedure Description

In Chapter II, the NAEP procedure is described in detail. In brief, the NAEP procedure entails the comparison of each weight to a value computed from the sum of the squared weights for the sample, $K_n$. That is,

$$w_k \leq K_n$$

where

$$K_n = \left( c \Sigma w_k^2 / n \right)^{1/2}.$$  

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 sum 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.

(b) The Empirical Study Procedure

For the empirical study, I will allow the NAEP procedure to go through 10 iterations. The only documentation of the use of the NAEP procedure is in the methodological reports of the National Assessment of Education Progress (NAEP) study and, in these reports, \( c \) is assigned a value of 10. The reports indicate no specific rationale for the use of \( c = 10 \) but suggest that a \( c \) value can be selected by inspection of the distribution of \( w_k^2 / (S w_k^2 / n) \). Smaller or larger values of \( c \) will generate different trimming levels. Future investigation should evaluate the effect of different values of \( c \).
Chapter V
EMPIRICAL INVESTIGATION: FINDINGS

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 specific goals of the empirical study are:

1. to investigate the currently used procedures (the NAEP procedure and the estimated MSE procedure) to identify extreme weights;

2. to evaluate the Taylor Series, weight distribution, and the SDV based procedures relative to the currently used procedures in terms of the variance change; and

3. to evaluate the bias and mean square error that results from each of the weight trimming procedures.

The performance measures used in the empirical study include the change in the estimated variance of the estimate (that is, how much variance reduction is achieved), the extent of bias introduced and the change in the mean square error of the estimate (that is, whether the bias introduced by these procedures offsets the variance reduction), and the average and variance of the trimming levels (that is, whether these procedures result in consistent trimming levels over repeated
samples).

For the empirical study, I used county-level data on medical resources and demographic characteristics of the county population from the Area Resource File (ARF) data base developed by the Health Resources and Services Administration of the U.S. Department of Health and Human Services. For each county, the number of households was used as a size measure (for probability proportional to size (pps) sample selection) and four data items (both correlated and uncorrelated to the size measure) to assess the impact of the weight trimming. A total of 2,989 of the 3,080 county units in the ARF data base was used. County units excluded either had a very large or a very small count of households, or were likely to have a zero value for one or more of the data items. I selected 200 samples of 100 units each using the probability minimal replacement sampling procedure developed by Chromy (1979) for the pps selection.

The four variables used were median family income, birth rate among teenagers, percentage of 5 to 17 year old population that are white, and the average temperature in June. These four variables were chosen because of the correlation between the data items and the sampling weight across the 200 samples. More specifically, these four data items were selected for the following reasons:

1. The average correlation between the sampling weights and median family income over the 200 samples is negative; therefore the weight trimming is expected to introduce a negative bias (the expected value of the estimator is less
than the true value).

2. The average correlation between the sampling weights and birth rate among teenagers over the 200 samples is positive; therefore the weight trimming is expected to introduce a positive bias (the expected value of the estimator is greater than the true value).

3. The average correlation between the sampling weights and both percentage of 5 to 17 years old population that are white with the average temperature in June over the 200 samples is essentially zero; therefore the weight trimming is expected to introduce no bias.

For each of the 200 samples, I implemented the five weight trimming procedures. The following is a brief summary of the primary results. After this summary, the design of the empirical study and the finding are presented in more detail.

B. Summary of Results

Of the five procedures, the Taylor series and the estimated mean square error procedure tended to perform similarly. Also, the NAEP procedure and the weight distribution procedure operated almost identically. The SDV based procedure tended to have some of the characteristics of both pairs of procedures.

The average reduction in the variance over the 200 replicated
samples and averaged over the four variables (see Table 5) was the largest for the SDV procedure (34 percent reduction), followed by the estimated MSE procedure and the Taylor series procedure (25 and 22 percent reduction, respectively), then the NAEP procedure and the weight distribution procedure (17 and 16 percent reduction, respectively). The average trimming level resulting from the five procedures (see Table 3) exhibited the same relationship (that is, the average trimming level was the lowest for the SDV procedure and highest for the NAEP procedure and the weight distribution procedure).

When the bias was taken into consideration by evaluating the mean square error (see Table 7), The average reduction in the MSE over the 200 replicated samples averaged over the four variables was approximately the same for the Taylor series procedure, the NAEP procedure and the weight distribution procedure (3.4, 3.4, and 3.5 percent reduction, respectively), followed by the estimated MSE procedure (2.8 percent reduction) whereas the SDV procedure resulted in an increase in the average MSE (a 17.7 percent increase).

In summary, although the SDV procedure resulted in the greatest variance reduction, the bias introduced (especially for the median family income) offset the variance reduction when the MSE was estimated. The estimated MSE procedure and the Taylor series procedure resulted in greater average variance reduction than the NAEP procedure and the weight distribution procedure, but the average MSE for these four procedures were similar. Although the NAEP procedure and the weight distribution procedure (as implemented in this empirical study) resulted in the average MSE estimates equivalent to
the estimated MSE procedure and the Taylor series procedure and introduced the least bias, the NAEP procedure and the weight distribution procedure do not use the data.

C. Design of the Empirical Study

The county-level data from the Area Resource File formed the population for the empirical study. Of the 3,080 county units in the ARF data base, I included 2,989 county units. The counties were limited to those with more than 250 households and less than 200,000 households. The lower limit was set to avoid zero values among the variables used because a zero value would negate any potential biasing effect. The upper limit was established to avoid certainty selections. The counties in Hawaii and Alaska were also excluded because some variables were not reported in the ARF for the county units in these states.

For each county, the number of households was used as the size measure to select the unit with probability strictly proportional to this size measure, using Chromy's probability minimal replacement (PMR) sequential sampling procedure (Chromy 1979). The size measure varied sufficiently to insure large and small sampling weights. I selected 200 samples, each containing 100 units. The samples were selected by first ordering the data base by the size measure and then using Chromy's PMR sampling procedure to select the samples. By selecting the sample from the ordered frame, I ensured a wide variation in the sampling weights and introduced an implicit stratification by size.
A set of four variables were chosen to include some data items that are correlated to the size measure and others that are not correlated: median family income; birth rate among teenagers; percentage of 5 to 17 years old population that are white; and the average temperature in June. As described previously, these four variables were chosen because of the correlation between the data item and the sampling weights. More specifically, the correlation between the data items and the sampling weights are as follows:

1. The average correlation between the sampling weights and median family income over the 200 samples is negative (an average of $-0.441$ with a minimum of $-0.608$ and a maximum of $-0.185$ over the 200 samples).

2. The average correlation between the sampling weights and birth rate among teenagers over the 200 samples is positive (an average of $0.221$ with a minimum of $-0.124$ and a maximum of $0.533$ over the 200 samples).

3. The average correlation between the sampling weights and both percentage of 5 to 17 year old population that are white and the average temperature in June over the 200 samples are essentially zero (an average of $0.054$ with a minimum of $-0.458$ and a maximum of $0.301$ for the percentage of 5 to 17 year old population that are white and an average of $0.030$ with a minimum of $-0.358$ and a maximum of $0.250$ for the average temperature in June).

Therefore the weight trimming was expected to introduce a negative bias for the median family income estimates (the expected value of the estimator is less than the true value) and a positive bias for the
birth rate among teenagers. For the percentage of 5 to 17 years old population that are white and the average temperature in June, the weight trimming was expected to introduce little bias.

The Taylor series procedure and the estimated MSE procedure evaluate statistics for predetermined candidate trimming levels. For the empirical study, 20 candidate trimming levels were computed for each sample. The candidate trimming levels were computed as follows:

a. Trimming level 1 is the next to largest weight;
b. Trimming level 2 is the sum of the next to largest and the third largest weight, divided by 2;
c. Trimming level 3 is the sum of the next to largest, the third largest, and the fourth largest weight, divided by 3;
d. Trimming level 4 to 20 were computed as similar averages of the largest weights (excluding the largest weight).

These candidate trimming levels represented up to a 66 percent reduction relative to the largest sampling weight. For each trimming candidate level, a set of trimmed adjusted weights were computed. That is, all weights in excess of the trimming level was assigned the value of the candidate trimming level and the trimmed excess was distributed among the untrimmed weights. Therefore, 20 separate sets of trimmed adjusted weights were computed for use with the Taylor series procedure and the estimated MSE procedure for each sample.

For the other procedures (as described in Chapter IV), the trimming levels were generated within the procedure. For these procedures, 10 iterations were performed within each sample to identify the final trimming level.
D. The Effect of Trimming on The Weights

On each of the 200 replicated samples, the following weight trimming procedures were implemented:

1. the estimated mean square error (MSE) procedure;
2. the Taylor series procedure;
3. the NAEP procedure;
4. the weight distribution procedure; and
5. the SDV procedure.

To assess the effect of the trimming on the sampling weights, I looked at the maximum weight, the design effect attributable to unequal weighting, and the standard deviation of the weights before and after trimming and the adjustment made to the untrimmed weights. These measures are summarized on Table 3.

(1) Average Maximum Sampling Weight

The average maximum sampling weight before trimming across the 200 replicated samples was 346.3. The average trimming level was less than 50 percent of the maximum weight for all procedures. The average reduction in the maximum weight was the greatest for the SDV procedure (73 percent, an average trimming level of 83.6), followed by the estimated MSE procedure (58.8 percent, an average trimming level of 125.4). The Taylor series procedure, the NAEP procedure, and the weight distribution procedure resulted in approximately the same reduction in the maximum weight. The other measures (the adjustment factor to the untrimmed weights, the design effect attributable to unequal weights, and the standard deviation of the weights) follow this pattern.
Table 3. Summary of Adjustments to Sampling Weights from 200 Replicated Samples

<table>
<thead>
<tr>
<th>Item</th>
<th>Minimum Value</th>
<th>Maximum Value</th>
<th>Standard Deviation of Mean</th>
<th>Std. Error C.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untrimmed Weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum Weight</td>
<td>346.3</td>
<td>222.0</td>
<td>1,375.1</td>
<td>165.0</td>
</tr>
<tr>
<td>Design Effect</td>
<td>3.57</td>
<td>2.81</td>
<td>12.79</td>
<td>1.26</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>47.57</td>
<td>38.1</td>
<td>139.5</td>
<td>13.37</td>
</tr>
<tr>
<td>Maximum Trimmed Weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MSE</td>
<td>125.4</td>
<td>79.6</td>
<td>222.0</td>
<td>42.0</td>
</tr>
<tr>
<td>Taylor Series</td>
<td>142.2</td>
<td>79.6</td>
<td>222.0</td>
<td>42.2</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>147.3</td>
<td>138.7</td>
<td>199.6</td>
<td>8.3</td>
</tr>
<tr>
<td>Weight Dist'n</td>
<td>151.0</td>
<td>142.3</td>
<td>204.7</td>
<td>8.5</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>83.6</td>
<td>69.6</td>
<td>116.4</td>
<td>7.1</td>
</tr>
<tr>
<td>Proportional Reduction to Maximum Weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MSE</td>
<td>0.588</td>
<td>0.932</td>
<td>0.051</td>
<td>0.184</td>
</tr>
<tr>
<td>Taylor Series</td>
<td>0.538</td>
<td>0.895</td>
<td>0.051</td>
<td>0.185</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>0.524</td>
<td>0.855</td>
<td>0.358</td>
<td>0.121</td>
</tr>
<tr>
<td>Weight Dist'n</td>
<td>0.512</td>
<td>0.851</td>
<td>0.343</td>
<td>0.124</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>0.730</td>
<td>0.917</td>
<td>0.591</td>
<td>0.072</td>
</tr>
<tr>
<td>Adjustment to Untrimmed Weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MSE</td>
<td>1.21</td>
<td>1.01</td>
<td>2.07</td>
<td>0.164</td>
</tr>
<tr>
<td>Taylor Series</td>
<td>1.16</td>
<td>1.01</td>
<td>1.68</td>
<td>0.131</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>1.11</td>
<td>1.05</td>
<td>1.51</td>
<td>0.062</td>
</tr>
<tr>
<td>Weight Dist'n</td>
<td>1.10</td>
<td>1.05</td>
<td>1.50</td>
<td>0.062</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>1.36</td>
<td>1.19</td>
<td>1.82</td>
<td>0.093</td>
</tr>
<tr>
<td>Design Effect Due to Unequal Weights After Trimming</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MSE</td>
<td>2.23</td>
<td>1.67</td>
<td>2.95</td>
<td>0.341</td>
</tr>
<tr>
<td>Taylor Series</td>
<td>2.37</td>
<td>1.72</td>
<td>2.95</td>
<td>0.321</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>2.46</td>
<td>2.42</td>
<td>2.50</td>
<td>0.015</td>
</tr>
<tr>
<td>Weight Dist'n</td>
<td>2.48</td>
<td>2.44</td>
<td>2.53</td>
<td>0.017</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>1.87</td>
<td>1.71</td>
<td>2.04</td>
<td>0.069</td>
</tr>
<tr>
<td>Standard Deviation of Trimmed Weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MSE</td>
<td>32.8</td>
<td>26.7</td>
<td>50.4</td>
<td>4.680</td>
</tr>
<tr>
<td>Taylor Series</td>
<td>34.7</td>
<td>26.7</td>
<td>50.4</td>
<td>4.575</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>36.0</td>
<td>33.8</td>
<td>48.8</td>
<td>2.048</td>
</tr>
<tr>
<td>Weight Dist'n</td>
<td>36.4</td>
<td>34.1</td>
<td>49.1</td>
<td>2.068</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>27.8</td>
<td>24.3</td>
<td>38.1</td>
<td>1.888</td>
</tr>
</tbody>
</table>
(2) Variation OfTrimming Levels Among Samples

The NAEP procedure, the weight distribution procedure, and the SDV procedure exhibit substantially less variation in the trimming level over the samples than either the estimated MSE procedure or the Taylor series procedure. The standard deviation of the maximum trimmed weight for the NAEP procedure, the weight distribution procedure, and the SDV procedure is between 7.1 and 8.5 whereas the standard deviation of the maximum trimmed weight for the estimated MSE procedure or the Taylor series procedure is approximately 42.0.

(3) Correlation Between Trimming Level and Maximum Weight

The correlation between the trimming level and the maximum untrimmed weight over the 200 samples was significantly different from zero for the NAEP procedure, the weight distribution procedure, and the SDV procedure (0.988, 0.989, and 0.626, respectively), but was not significantly different from zero for either the estimated MSE procedure or the Taylor series procedure (0.008 and 0.109, respectively). The high correlation for the NAEP procedure and the weight distribution procedure is expected because these procedures use only the sampling weights. For the SDV procedure, the significant correlation is also expected because this procedure incorporates the weight as well as the data in the procedure. The insignificant correlation for the estimated MSE procedure or the Taylor series procedure is not surprising because the MSE (and, for the Taylor series procedure, the relative bias) is the basis for these procedures. However, the small size of these correlations (especially for estimated MSE procedure) is somewhat surprising. The dependence
or lack of dependence of the trimming level on the maximum untrimmed weight over the 200 samples is a key difference among these procedures.

(4) Skewness of Weight Distributions

To assess the shape of the sampling weight distribution, the skewness measure was computed for the untrimmed weights and for the trimmed weights for each of the 200 samples. A symmetric distribution, like the normal distribution, has a skewness of zero. The average skewness of the untrimmed sampling weights over the 200 samples (see Table 4) is 3.96 (ranging from 2.69 to 9.07). For the SDV procedure the average skewness was reduced to 0.96 (ranging from 0.69 to 1.24 over the 200 samples). The average skewness for the estimated MSE procedure was 1.58 (ranging from 0.57 to 2.89 over the 200 samples), varying substantially more than the SDV procedure. For the Taylor series procedure, the NAEP procedure, and the weight distribution procedure, the average skewness was 1.81, 1.93 and 1.98, respectively. Although the average skewness for these three procedures was similar, the NAEP procedure and the weight distribution procedure exhibited substantially less variation among the samples than did the Taylor series procedure. The standard deviation of the skewness was 0.02 for both the NAEP procedure and the weight distribution procedure, and 0.56 for the Taylor series procedure.

In summary, over the 200 samples the SDV procedure resulted in the most trimming; and, among the samples, resulted in the least variation in the trimming levels. The estimated MSE procedure resulted in the next lowest average trimming level; and, among the samples, the
Table 4. Skewness of Untrimmed and Trimmed Sampling Weights  
Based on Weights from Individual Samples  
From 200 Replicated Samples

<table>
<thead>
<tr>
<th>Minimum Average Value</th>
<th>Maximum Average Value</th>
<th>Standard Deviation</th>
<th>Std. Error Of Mean</th>
<th>C.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untrimmed Weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>2,972.2</td>
<td>2,812.2</td>
<td>4,041.1</td>
<td>167.5</td>
</tr>
<tr>
<td>Mean Weight</td>
<td>29.72</td>
<td>28.12</td>
<td>40.41</td>
<td>1.67</td>
</tr>
<tr>
<td>Maximum Weight</td>
<td>346.3</td>
<td>222.0</td>
<td>1,375.1</td>
<td>165.0</td>
</tr>
<tr>
<td>Skewness</td>
<td>3.96</td>
<td>2.69</td>
<td>9.07</td>
<td>1.38</td>
</tr>
<tr>
<td>Skewness After Weight Trimming</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MSE</td>
<td>1.58</td>
<td>0.57</td>
<td>2.89</td>
<td>0.59</td>
</tr>
<tr>
<td>Taylor Series</td>
<td>1.81</td>
<td>0.71</td>
<td>2.89</td>
<td>0.56</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>1.93</td>
<td>1.89</td>
<td>1.98</td>
<td>0.02</td>
</tr>
<tr>
<td>Weight Dist'n</td>
<td>1.98</td>
<td>1.93</td>
<td>2.03</td>
<td>0.02</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>0.96</td>
<td>0.69</td>
<td>1.24</td>
<td>0.11</td>
</tr>
</tbody>
</table>
estimated MSE procedure resulted in substantial variation in the trimming levels, with the standard deviation of the trimming levels being 42.0 for the estimated MSE procedure versus 7.1 for the SDV procedure. The Taylor series procedure, the NAEP procedure, and the weight distribution procedure resulted in similar average trimming levels; but variation in the trimming levels was substantially higher for the Taylor series procedure than for the NAEP procedure or the weight distribution procedure.

E. Change in Estimated Sampling Variance

The change in the estimated sampling variance was computed as the estimated sampling variance using the trimmed weights minus the estimated sampling variance using the untrimmed weights divided by the estimated sampling variance using the untrimmed weights. In Table 5, the change is summarized as percentages; a negative change implies that the trimmed weights resulted in a reduction in the estimated sampling variance and a positive value implies that the estimated sampling variance was larger using the trimmed weights.

All procedures resulted in an average reduction in estimated sampling variance (relative to the estimated sampling variance using the untrimmed weights) over the 200 samples. However, all procedures resulted in a larger estimated sampling variance for at least one sample among the 200 samples. That is, in the column shown as the maximum, all procedures had a positive value (an increase in the estimate sampling variance).
Table 5. Percentage Change in Variance From Weight Trimming Based on Variances Estimates from 200 Replicated Samples

<table>
<thead>
<tr>
<th>Procedure/Variable</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated MSE Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>-21.7</td>
<td>-86.5</td>
<td>45.7</td>
<td>132.1</td>
<td>24.8</td>
</tr>
<tr>
<td>2</td>
<td>-31.7</td>
<td>-87.4</td>
<td>25.2</td>
<td>112.6</td>
<td>25.4</td>
</tr>
<tr>
<td>3</td>
<td>-22.8</td>
<td>-94.1</td>
<td>28.7</td>
<td>122.8</td>
<td>26.2</td>
</tr>
<tr>
<td>4</td>
<td>-25.1</td>
<td>-90.9</td>
<td>25.3</td>
<td>116.3</td>
<td>25.3</td>
</tr>
<tr>
<td>Average</td>
<td>-25.3</td>
<td>-89.7</td>
<td>31.2</td>
<td>121.0</td>
<td>25.4</td>
</tr>
<tr>
<td>Taylor Series Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-19.9</td>
<td>-89.9</td>
<td>49.6</td>
<td>139.4</td>
<td>24.9</td>
</tr>
<tr>
<td>2</td>
<td>-28.0</td>
<td>-88.8</td>
<td>25.2</td>
<td>113.9</td>
<td>24.0</td>
</tr>
<tr>
<td>3</td>
<td>-19.1</td>
<td>-93.7</td>
<td>27.6</td>
<td>121.4</td>
<td>23.7</td>
</tr>
<tr>
<td>4</td>
<td>-21.7</td>
<td>-81.7</td>
<td>37.5</td>
<td>119.1</td>
<td>24.3</td>
</tr>
<tr>
<td>Average</td>
<td>-22.2</td>
<td>-88.5</td>
<td>35.0</td>
<td>123.5</td>
<td>24.2</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-15.0</td>
<td>-85.5</td>
<td>36.6</td>
<td>122.2</td>
<td>23.3</td>
</tr>
<tr>
<td>2</td>
<td>-22.1</td>
<td>-86.5</td>
<td>28.0</td>
<td>114.5</td>
<td>24.2</td>
</tr>
<tr>
<td>3</td>
<td>-12.4</td>
<td>-89.8</td>
<td>31.0</td>
<td>120.8</td>
<td>22.1</td>
</tr>
<tr>
<td>4</td>
<td>-17.2</td>
<td>-83.1</td>
<td>45.3</td>
<td>128.4</td>
<td>23.1</td>
</tr>
<tr>
<td>Average</td>
<td>-16.7</td>
<td>-86.2</td>
<td>35.2</td>
<td>121.5</td>
<td>23.2</td>
</tr>
<tr>
<td>Weight Distribution Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-14.5</td>
<td>-85.3</td>
<td>35.8</td>
<td>121.1</td>
<td>23.1</td>
</tr>
<tr>
<td>2</td>
<td>-21.4</td>
<td>-86.1</td>
<td>27.7</td>
<td>113.8</td>
<td>24.0</td>
</tr>
<tr>
<td>3</td>
<td>-11.8</td>
<td>-89.4</td>
<td>30.3</td>
<td>119.7</td>
<td>21.8</td>
</tr>
<tr>
<td>4</td>
<td>-16.5</td>
<td>-82.6</td>
<td>44.7</td>
<td>127.3</td>
<td>22.8</td>
</tr>
<tr>
<td>Average</td>
<td>-16.0</td>
<td>-85.9</td>
<td>34.6</td>
<td>120.5</td>
<td>22.9</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-31.3</td>
<td>-87.7</td>
<td>45.7</td>
<td>133.3</td>
<td>23.9</td>
</tr>
<tr>
<td>2</td>
<td>-41.1</td>
<td>-92.9</td>
<td>16.4</td>
<td>109.3</td>
<td>23.9</td>
</tr>
<tr>
<td>3</td>
<td>-31.4</td>
<td>-95.4</td>
<td>30.4</td>
<td>125.8</td>
<td>27.6</td>
</tr>
<tr>
<td>4</td>
<td>-31.9</td>
<td>-90.2</td>
<td>36.9</td>
<td>127.2</td>
<td>25.9</td>
</tr>
<tr>
<td>Average</td>
<td>-33.9</td>
<td>-91.5</td>
<td>32.4</td>
<td>123.9</td>
<td>25.3</td>
</tr>
</tbody>
</table>

Percentage Change = 100 * (Trimmed - Untrimmed) / Untrimmed.

* Variables: 1 = median family income; 2 = birth rate among teenagers; 3 = percentage white age 10-17; 4 = temperature in June.
In terms of the average reduction in estimated sampling variance (relative to the estimated sampling variance using the untrimmed weights) over the 200 samples, the SDV procedure resulted in the greatest average change (-33.9 percent), followed by the estimated MSE procedure and the Taylor series procedure (-25.3 percent and -22.2 percent, respectively). The NAEP procedure and the weight distribution procedure exhibited approximately the same percentage reduction (-16.7 percent and -16.0 percent, respectively). A key finding is that, although the relative reduction is approximately the same for the estimated MSE procedure and the Taylor series procedure, the average trimming level was higher (implying less trimming) for the Taylor series procedure than for the estimated MSE procedure. This implies that the estimated MSE procedure perhaps trimmed more than necessary.

The estimates of the sampling variances are summarized on Table 6. In the first column, the variance among the samples, is computed as the variance of the estimated means across the 200 samples. The second to fifth columns are the average, minimum, maximum, range and standard deviation of the estimated variances from the individual samples. For the untrimmed weights, the variance among the samples is greater than the average estimated variance from the individual samples for all but the first variable (median family income). For the estimated MSE procedure and the Taylor series procedure, the variance among the samples is greater than the average estimated variance from the individual samples for all variables. For the NAEP and the weight distribution procedure, the pattern shown in the
Table 6. Summary of Sampling Variances from Weight Trimming Procedures

<table>
<thead>
<tr>
<th>Procedure/Variables</th>
<th>Variance Among Samples (^1)</th>
<th>Based on Estimates from Individual Samples</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Variance</td>
<td>Average</td>
<td>Minimum</td>
</tr>
<tr>
<td>Untrimmed Weights</td>
<td>311,799</td>
<td>329,258</td>
<td>114,086</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7.21</td>
<td>6.10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>18.79</td>
<td>15.04</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.003</td>
<td>0.853</td>
</tr>
<tr>
<td>Estimated MSE Procedure</td>
<td>243,745</td>
<td>220,025</td>
<td>111,075</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.81</td>
<td>3.33</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>9.12</td>
<td>7.58</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.693</td>
<td>0.540</td>
</tr>
<tr>
<td>Taylor Series Procedure</td>
<td>244,692</td>
<td>226,783</td>
<td>106,550</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.26</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10.08</td>
<td>8.42</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.727</td>
<td>0.576</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>209,568</td>
<td>243,237</td>
<td>116,486</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.07</td>
<td>3.92</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10.34</td>
<td>9.13</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.728</td>
<td>0.615</td>
</tr>
<tr>
<td>Weight Distribution Procedure</td>
<td>212,015</td>
<td>245,477</td>
<td>118,433</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.14</td>
<td>3.97</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10.50</td>
<td>9.27</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.737</td>
<td>0.622</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>151,055</td>
<td>188,315</td>
<td>102,683</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.63</td>
<td>2.77</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.74</td>
<td>6.08</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.546</td>
<td>0.475</td>
</tr>
</tbody>
</table>

\(^1\) Variance among samples = \(\Sigma (\hat{\sigma}_s - \hat{\sigma})^2/199\), where \(\hat{\sigma}_s\) is the estimate for sample \(s\) and \(\hat{\sigma}\) is average of \(\hat{\sigma}_s\).

*Variables: 1 = median family income; 2 = birth rate among teenagers; 3 = percentage white age 10-17; 4 = temperature in June.
estimated variance using the untrimmed weights is also present; that is, the variance among the samples is greater than the average estimated variance from the individual samples for all but the first variable (median family income). For the SDV procedure, the variance among the samples is greater than the average estimated variance from the individual samples for two variables (median family income and the birth rate among teenagers) but the reverse is true for the other two variables.

The fact that the average estimated variances computed from the individual samples for the estimated MSE procedure and the Taylor series procedure is less than the variance among the samples is consistent with the procedure because these procedures seek the minimum estimated MSE for each sample. Therefore, the average estimated sampling variance from individual samples will tend to underestimate the true sampling variance. Because the NAEP and the weight distribution procedure use only the sampling weights, the similarity of the pattern between the results of these procedures and the results using the untrimmed weights could be expected. The SDV procedure resulted in the smallest average variances and the smallest range of estimated sampling variances of the five procedures.

F. Change in Estimated Mean Square Error

The change in the estimated mean square error (MSE) was computed as the estimated mean square error using the trimmed weights minus the estimated sampling variance using the untrimmed weights divided by the estimated sampling variance using the untrimmed weights. For the estimated MSE procedure, the NAEP procedure, the weight distribution
procedure, and the SDV procedure, the estimated MSE for the individual samples using the form

\[
\text{MSE}(\hat{\gamma}) = (\hat{\gamma}_t - \hat{\gamma}_u)^2 - \text{Var}(\hat{\gamma}_u) + 2 \times (\text{Var}(\hat{\gamma}_u) \times \text{Var}(\hat{\gamma}_t))^{1/2}
\]

where

\( \hat{\gamma}_t \) = the estimated mean using the trimmed weights;
\( \hat{\gamma}_u \) = the estimated mean using the untrimmed weights;
\( \text{Var}(\hat{\gamma}_t) \) = the estimated sampling variance for \( \hat{\gamma}_t \); and
\( \text{Var}(\hat{\gamma}_u) \) = the estimated sampling variance for \( \hat{\gamma}_u \).

For the Taylor series procedure, an estimate of the bias is generated. For this procedure, the estimated mean square error was computed as follows:

\[
\text{MSE}(\hat{\gamma}_t) = \text{Var}(\hat{\gamma}_t) + (\text{Bias}(\hat{\gamma}_t))^2
\]

where

\( \text{Bias}(\hat{\gamma}_t) \) = the estimated bias of \( \hat{\gamma}_t \),
\( \text{Bias}(\hat{\gamma}_t) = \hat{\gamma}_t - \hat{\gamma}_u \).

(1) **Comparison of Estimated MSE from Five Procedures**

In Table 7, the change in the estimated MSE is summarized as percentages; a negative change implies that the trimmed weights resulted in a reduction in the estimated MSE and a positive value implies that the estimated MSE was larger using the trimmed weights.
Table 7. Percentage Change In Mean Square Error From Weight Trimming Based On MSE Estimates From 200 Replicated Samples

<table>
<thead>
<tr>
<th>Procedure/Variables</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated MSE Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>32.7</td>
<td>-85.2</td>
<td>428.2</td>
<td>513.4</td>
<td>74.1</td>
</tr>
<tr>
<td>2</td>
<td>-19.8</td>
<td>-105.5</td>
<td>229.2</td>
<td>334.7</td>
<td>32.8</td>
</tr>
<tr>
<td>3</td>
<td>-10.7</td>
<td>-96.5</td>
<td>395.8</td>
<td>492.4</td>
<td>42.8</td>
</tr>
<tr>
<td>4</td>
<td>-13.5</td>
<td>-86.5</td>
<td>68.9</td>
<td>155.4</td>
<td>26.4</td>
</tr>
<tr>
<td>Average</td>
<td>-2.8</td>
<td>-93.4</td>
<td>280.5</td>
<td>374.0</td>
<td>44.0</td>
</tr>
<tr>
<td>Taylor Series Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>15.3</td>
<td>-63.5</td>
<td>241.4</td>
<td>304.9</td>
<td>46.5</td>
</tr>
<tr>
<td>2</td>
<td>-14.4</td>
<td>-68.3</td>
<td>61.2</td>
<td>129.4</td>
<td>20.3</td>
</tr>
<tr>
<td>3</td>
<td>-6.0</td>
<td>-70.3</td>
<td>173.2</td>
<td>243.5</td>
<td>27.4</td>
</tr>
<tr>
<td>4</td>
<td>-8.5</td>
<td>-58.0</td>
<td>68.5</td>
<td>126.5</td>
<td>20.7</td>
</tr>
<tr>
<td>Average</td>
<td>-3.4</td>
<td>-65.0</td>
<td>136.1</td>
<td>201.1</td>
<td>28.7</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td>-51.8</td>
<td>79.2</td>
<td>131.0</td>
<td>19.1</td>
</tr>
<tr>
<td>2</td>
<td>-9.2</td>
<td>-62.8</td>
<td>76.7</td>
<td>139.6</td>
<td>21.4</td>
</tr>
<tr>
<td>3</td>
<td>-1.4</td>
<td>-68.6</td>
<td>101.1</td>
<td>169.8</td>
<td>23.1</td>
</tr>
<tr>
<td>4</td>
<td>-5.6</td>
<td>-57.9</td>
<td>72.1</td>
<td>130.0</td>
<td>19.7</td>
</tr>
<tr>
<td>Average</td>
<td>-3.4</td>
<td>-60.3</td>
<td>82.3</td>
<td>142.6</td>
<td>20.8</td>
</tr>
<tr>
<td>Weight Distribution Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.8</td>
<td>-51.4</td>
<td>72.9</td>
<td>124.4</td>
<td>18.4</td>
</tr>
<tr>
<td>2</td>
<td>-9.0</td>
<td>-60.7</td>
<td>75.6</td>
<td>136.3</td>
<td>20.8</td>
</tr>
<tr>
<td>3</td>
<td>-1.2</td>
<td>-67.9</td>
<td>96.4</td>
<td>164.4</td>
<td>22.3</td>
</tr>
<tr>
<td>4</td>
<td>-5.5</td>
<td>-57.2</td>
<td>71.3</td>
<td>128.5</td>
<td>19.2</td>
</tr>
<tr>
<td>Average</td>
<td>-3.5</td>
<td>-59.3</td>
<td>79.1</td>
<td>138.4</td>
<td>20.2</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>87.1</td>
<td>-79.1</td>
<td>446.1</td>
<td>525.3</td>
<td>83.8</td>
</tr>
<tr>
<td>2</td>
<td>-11.3</td>
<td>-107.0</td>
<td>201.2</td>
<td>308.2</td>
<td>45.8</td>
</tr>
<tr>
<td>3</td>
<td>-3.2</td>
<td>-95.4</td>
<td>227.5</td>
<td>322.9</td>
<td>56.5</td>
</tr>
<tr>
<td>4</td>
<td>-1.5</td>
<td>-86.6</td>
<td>141.9</td>
<td>228.5</td>
<td>46.6</td>
</tr>
<tr>
<td>Average</td>
<td>17.7</td>
<td>-92.0</td>
<td>254.2</td>
<td>346.2</td>
<td>58.2</td>
</tr>
</tbody>
</table>

Percentage Change = 100 * (Trimmed - Untrimmed) / Untrimmed.
*Variables: 1 = median family income; 2 = birth rate among teenagers; 3 = percentage white age 10-17; 4 = temperature in June.
For four of the five procedures (excluding the SDV procedure), the average change in the estimated MSE across the four variables was similar. For the SDV procedure, the average change was positive; this resulted from the average estimated MSE for the first variable (median family income). The bias introduced overwhelmed the variance reduction. Over the four variables, the reduction in the sampling variance shown in Table 5 were mostly negated by the biases introduced.

In comparing the estimated MSE procedure and the Taylor series procedure for the first variable, although the estimated MSE procedure resulted in slightly greater variance reduction, the variance reduction seemed to be offset by the increased bias. For the other three variable, the estimated MSE procedure resulted in greater estimated MSE reduction. The higher estimated MSE occurring for the Taylor series procedure could have been caused by the estimate of the bias used for this procedure. The bias estimator does not take into account the variance of the bias. However, when averaged over the four variables, the Taylor series resulted in a slightly greater estimated MSE reduction.

Once again the NAEP procedure and the weight distribution procedure resulted in almost the same average changes for each of the four variables.

The mean square error among the samples and the average, minimum, maximum, range and the standard deviation of the estimated MSE from the individual samples are given in Table 8. The mean square error among the samples is computed as variance of the estimated means
Table 8. Summary of Mean Square Error from Weight Trimming Procedures From 200 Replicated Samples

<table>
<thead>
<tr>
<th>Procedure/Variable</th>
<th>MSE Among Samples</th>
<th>Based on Estimates from Individual Samples</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>Estimated MSE Procedure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>376,138</td>
<td>425,245</td>
<td>100,565</td>
</tr>
<tr>
<td>2</td>
<td>3.83</td>
<td>4.30</td>
<td>-0.70</td>
</tr>
<tr>
<td>3</td>
<td>9.44</td>
<td>10.34</td>
<td>0.82</td>
</tr>
<tr>
<td>4</td>
<td>0.700</td>
<td>0.659</td>
<td>0.211</td>
</tr>
<tr>
<td>Taylor Series Procedure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>326,468</td>
<td>375,736</td>
<td>126,924</td>
</tr>
<tr>
<td>2</td>
<td>4.26</td>
<td>4.89</td>
<td>1.51</td>
</tr>
<tr>
<td>3</td>
<td>10.32</td>
<td>12.67</td>
<td>1.45</td>
</tr>
<tr>
<td>4</td>
<td>0.731</td>
<td>0.738</td>
<td>0.228</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>261,580</td>
<td>324,446</td>
<td>126,183</td>
</tr>
<tr>
<td>2</td>
<td>4.07</td>
<td>5.01</td>
<td>1.56</td>
</tr>
<tr>
<td>3</td>
<td>10.47</td>
<td>11.79</td>
<td>1.41</td>
</tr>
<tr>
<td>4</td>
<td>0.735</td>
<td>0.743</td>
<td>0.235</td>
</tr>
<tr>
<td>Weight Distribution Procedure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>259,774</td>
<td>321,666</td>
<td>127,147</td>
</tr>
<tr>
<td>2</td>
<td>4.14</td>
<td>5.03</td>
<td>1.56</td>
</tr>
<tr>
<td>3</td>
<td>10.63</td>
<td>11.86</td>
<td>1.41</td>
</tr>
<tr>
<td>4</td>
<td>0.744</td>
<td>0.745</td>
<td>0.234</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>508,230</td>
<td>592,852</td>
<td>123,968</td>
</tr>
<tr>
<td>2</td>
<td>2.83</td>
<td>4.86</td>
<td>-0.89</td>
</tr>
<tr>
<td>3</td>
<td>7.02</td>
<td>10.38</td>
<td>1.10</td>
</tr>
<tr>
<td>4</td>
<td>0.548</td>
<td>0.727</td>
<td>0.189</td>
</tr>
</tbody>
</table>

MSE Among Samples = \( \sum (\hat{\theta}_s - \theta)^2 / 199 \), where \( \hat{\theta}_s \) is estimate for sample s.

*Variables: 1 = median family income; 2 = birth rate among teenagers; 3 = percentage white age 10-17; 4 = temperature in June.
around the true value for the mean. For each procedure and for all but the fourth variable, the mean square error among the samples is smaller than the average estimated MSE from the individual samples. For the fourth variable (temperature in June), the mean square error among the samples is almost identical to the average from the individual samples for all but the SDV procedure.

(2) Bias Introduced By the Five Procedures

To assess the bias introduced by the trimming procedures, I compared the average estimated mean over the 200 replicated samples to the true value (see Table 9) to compute a bias and a relative bias estimate. Except for the first variable (median family income) and the SDV procedure, the absolute value of the relative bias (bias/true value) is between 0.09 and 0.66 percent. For the median family income variable, the relative bias ranged from 3.6 percent for the SDV procedure to 1.32 percent for the weight distribution procedure. Except for the temperature in June (variable 4), the SDV procedure introduced the largest biases among the five procedures. The estimated MSE procedure and the Taylor series procedure (both of which tended to trim more weights than the NAEP procedure or the weight distribution procedure) tended to have larger absolute value of the relative bias for the four variables.

The variances among the samples for all but the SDV procedure (see column 6 in Table 9) show a marked similarity in the average change from the variance among samples computed using the untrimmed weights. However, the MSE among the samples, which incorporates the bias, (column 8) show that the NAEP procedure and the weight distribution
<table>
<thead>
<tr>
<th>Var</th>
<th>True Value</th>
<th>Average Over Samples</th>
<th>Relative Variance (%)</th>
<th>MSE Among Samples</th>
<th>Change %</th>
<th>MSE Among Samples</th>
<th>Change %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untrimmed Weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16,592.2</td>
<td>16,618.0</td>
<td>25.8</td>
<td>0.16</td>
<td>311,779</td>
<td>312,464</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>33.70</td>
<td>33.73</td>
<td>0.04</td>
<td>0.10</td>
<td>7.21</td>
<td>7.21</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>85.64</td>
<td>85.97</td>
<td>0.33</td>
<td>0.39</td>
<td>18.79</td>
<td>18.90</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>75.88</td>
<td>75.86</td>
<td>-0.02</td>
<td>-0.02</td>
<td>1.003</td>
<td>1.003</td>
<td></td>
</tr>
<tr>
<td>Estimated MSE Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16,592.2</td>
<td>16,956.1</td>
<td>363.9</td>
<td>2.19</td>
<td>243,745</td>
<td>-21.8</td>
<td>376,138</td>
</tr>
<tr>
<td>2</td>
<td>33.70</td>
<td>33.55</td>
<td>-0.15</td>
<td>-0.43</td>
<td>3.81</td>
<td>-47.2</td>
<td>3.83</td>
</tr>
<tr>
<td>3</td>
<td>85.64</td>
<td>86.21</td>
<td>0.57</td>
<td>0.66</td>
<td>9.12</td>
<td>-51.5</td>
<td>9.44</td>
</tr>
<tr>
<td>4</td>
<td>75.88</td>
<td>75.96</td>
<td>0.08</td>
<td>0.10</td>
<td>0.693</td>
<td>-30.8</td>
<td>0.700</td>
</tr>
<tr>
<td>Taylor Series Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16,592.2</td>
<td>16,878.2</td>
<td>286.0</td>
<td>1.72</td>
<td>244,692</td>
<td>-21.5</td>
<td>326,468</td>
</tr>
<tr>
<td>2</td>
<td>33.70</td>
<td>33.63</td>
<td>-0.06</td>
<td>-0.18</td>
<td>4.26</td>
<td>-41.0</td>
<td>4.26</td>
</tr>
<tr>
<td>3</td>
<td>85.64</td>
<td>86.13</td>
<td>0.49</td>
<td>0.57</td>
<td>10.08</td>
<td>-46.4</td>
<td>10.32</td>
</tr>
<tr>
<td>4</td>
<td>75.88</td>
<td>75.95</td>
<td>0.07</td>
<td>0.09</td>
<td>0.73</td>
<td>-27.6</td>
<td>0.731</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16,592.2</td>
<td>16,820.3</td>
<td>228.1</td>
<td>1.37</td>
<td>209,568</td>
<td>-32.8</td>
<td>261,580</td>
</tr>
<tr>
<td>2</td>
<td>33.70</td>
<td>33.75</td>
<td>0.06</td>
<td>0.16</td>
<td>4.07</td>
<td>-43.5</td>
<td>4.07</td>
</tr>
<tr>
<td>3</td>
<td>85.64</td>
<td>86.01</td>
<td>0.36</td>
<td>0.43</td>
<td>10.34</td>
<td>-45.0</td>
<td>10.47</td>
</tr>
<tr>
<td>4</td>
<td>75.88</td>
<td>75.96</td>
<td>0.09</td>
<td>0.11</td>
<td>0.73</td>
<td>-27.4</td>
<td>0.735</td>
</tr>
<tr>
<td>Weight Distribution Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16,592.2</td>
<td>16,810.8</td>
<td>218.5</td>
<td>1.32</td>
<td>212,015</td>
<td>-32.0</td>
<td>259,774</td>
</tr>
<tr>
<td>2</td>
<td>33.70</td>
<td>33.76</td>
<td>0.06</td>
<td>0.19</td>
<td>4.14</td>
<td>-42.6</td>
<td>4.14</td>
</tr>
<tr>
<td>3</td>
<td>85.64</td>
<td>86.00</td>
<td>0.36</td>
<td>0.42</td>
<td>10.50</td>
<td>-44.1</td>
<td>10.63</td>
</tr>
<tr>
<td>4</td>
<td>75.88</td>
<td>75.96</td>
<td>0.08</td>
<td>0.11</td>
<td>0.74</td>
<td>-26.5</td>
<td>0.744</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16,592.2</td>
<td>17,189.9</td>
<td>597.6</td>
<td>3.60</td>
<td>151,055</td>
<td>-51.6</td>
<td>508,230</td>
</tr>
<tr>
<td>2</td>
<td>33.70</td>
<td>33.25</td>
<td>-0.44</td>
<td>-1.31</td>
<td>2.63</td>
<td>-63.5</td>
<td>2.83</td>
</tr>
<tr>
<td>3</td>
<td>85.64</td>
<td>86.18</td>
<td>0.53</td>
<td>0.62</td>
<td>6.74</td>
<td>-64.2</td>
<td>7.02</td>
</tr>
<tr>
<td>4</td>
<td>75.88</td>
<td>75.92</td>
<td>0.05</td>
<td>0.06</td>
<td>0.55</td>
<td>-45.5</td>
<td>0.548</td>
</tr>
</tbody>
</table>

Variance Among Sample = Σ(̂θs - ̄θ)²/199; where ̂θs is sample estimate and ̄θ is average of ̂θ; MSE Among Samples = Σ(̂θs - θ)²/199.

Percentage Change = 100 x (Trimmed - Untrimmed)/Untrimmed Variables.

Variables: See Table 8.
procedure tended to have greater average reduction from the MSE among samples computed using the untrimmed weights. This finding is basically the result of the larger biases introduced by the estimated MSE procedure and the Taylor series procedure for the median family income variable.

In the Taylor series procedure, the bias was estimated for each sample. The average of the bias from the 200 individual sample estimates of the bias is 260.2 (with a standard error of 20.2) for variable 1, -0.097 (with a standard error of 0.080) for variable 2, 0.159 (with a standard error of 0.146) for variable 3, and 0.086 (with a standard error of 0.028) for variable 4. On average, these estimates are relatively close to the bias estimate computed from the average of the estimated sample means (see Table 9). Except for variable 3, the bias estimate computed from the average of the estimated sample means, which is an asymptotically unbiased estimate of the bias, is inside the 95 percent confidence interval for the average bias computed from the individual sample bias estimates.

In summary, when considering the mean square error, the biasing effect of weight trimming can offset the variance reduction. Therefore, in this empirical study there is, on average, less of a distinction among the results of the estimated MSE procedure, the Taylor series procedure, the NAEP procedure, and the weight distribution procedure. The estimated MSE procedure and the Taylor series procedure can produce better precision than the NAEP procedure, and the weight distribution procedure for some variables but poorer precision for other variables.
G. Coverage Probabilities

The previous analyses in this chapter assessed the impact of weight trimming on measures related to point estimates. Because the weight trimming can reduce the estimated sampling variance and introduce a bias in the estimated means, the effect of the trimming on interval estimates are also of interest. For each variable, I investigated interval estimates computing the relative standard errors (Table 10) and by assessing the proportion of 90, 95, and 99 percent confidence intervals computed using the individual sample estimates that contained the true value (see Table 11).

(1) Relative Standard Errors

The relative standard errors are of interest because these show the relative half-width of the confidence intervals. As seen in Table 10, the relative standard errors exhibit a pattern similar to the estimated sampling variances. That is, when compared to the relative standard errors using the untrimmed weights, the SDV procedure shows the greatest reduction in the relative standard errors. Following this procedure, the estimated MSE procedure and the Taylor series procedure have the next greatest reduction.

(2) Interval Estimates

In terms of the interval estimates, the proportion of intervals containing the true value using the untrimmed sampling weights was on or near the nominal level for the first variable for all three interval estimates. For the 99 percent confidence interval estimates, the true value for the fourth variable was in the estimated interval in 98 percent of the 200 samples. For the other two
<table>
<thead>
<tr>
<th>Variable</th>
<th>Average</th>
<th>Relative Change(%)</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untrimmed Weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.35</td>
<td>1.93</td>
<td>7.57</td>
<td>5.64</td>
<td>2.12</td>
<td>0.975</td>
</tr>
<tr>
<td>2</td>
<td>6.97</td>
<td>3.70</td>
<td>19.15</td>
<td>15.45</td>
<td>2.306</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.04</td>
<td>1.19</td>
<td>24.82</td>
<td>23.63</td>
<td>2.776</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.17</td>
<td>0.58</td>
<td>2.91</td>
<td>2.32</td>
<td>0.344</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>3.88</td>
<td>1.85</td>
<td>13.61</td>
<td>11.76</td>
<td>1.600</td>
<td></td>
</tr>
<tr>
<td>Estimated MSE</td>
<td>Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.74</td>
<td>-18.1%</td>
<td>1.95</td>
<td>4.07</td>
<td>2.12</td>
<td>0.455</td>
</tr>
<tr>
<td>2</td>
<td>5.36</td>
<td>-23.1%</td>
<td>3.50</td>
<td>7.92</td>
<td>4.42</td>
<td>0.844</td>
</tr>
<tr>
<td>3</td>
<td>3.10</td>
<td>-23.3%</td>
<td>1.30</td>
<td>7.23</td>
<td>5.93</td>
<td>0.994</td>
</tr>
<tr>
<td>4</td>
<td>0.96</td>
<td>-18.6%</td>
<td>0.62</td>
<td>1.54</td>
<td>0.93</td>
<td>0.163</td>
</tr>
<tr>
<td>Average</td>
<td>3.04</td>
<td>-20.8%</td>
<td>1.84</td>
<td>5.19</td>
<td>3.35</td>
<td>0.614</td>
</tr>
<tr>
<td>Taylor Series</td>
<td>Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.79</td>
<td>-16.6%</td>
<td>1.86</td>
<td>4.39</td>
<td>2.53</td>
<td>0.481</td>
</tr>
<tr>
<td>2</td>
<td>5.55</td>
<td>-20.4%</td>
<td>3.45</td>
<td>8.83</td>
<td>5.38</td>
<td>0.962</td>
</tr>
<tr>
<td>3</td>
<td>3.25</td>
<td>-19.6%</td>
<td>1.31</td>
<td>7.35</td>
<td>6.04</td>
<td>1.139</td>
</tr>
<tr>
<td>4</td>
<td>0.98</td>
<td>-16.0%</td>
<td>0.62</td>
<td>1.61</td>
<td>0.99</td>
<td>0.181</td>
</tr>
<tr>
<td>Average</td>
<td>3.14</td>
<td>-18.2%</td>
<td>1.81</td>
<td>5.54</td>
<td>3.74</td>
<td>0.691</td>
</tr>
<tr>
<td>NAEP</td>
<td>Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.90</td>
<td>-13.3%</td>
<td>1.99</td>
<td>4.29</td>
<td>2.30</td>
<td>0.463</td>
</tr>
<tr>
<td>2</td>
<td>5.77</td>
<td>-17.1%</td>
<td>3.48</td>
<td>8.62</td>
<td>5.14</td>
<td>0.952</td>
</tr>
<tr>
<td>3</td>
<td>3.40</td>
<td>-15.8%</td>
<td>1.30</td>
<td>6.71</td>
<td>5.41</td>
<td>1.139</td>
</tr>
<tr>
<td>4</td>
<td>1.02</td>
<td>-13.2%</td>
<td>0.62</td>
<td>1.67</td>
<td>1.04</td>
<td>0.185</td>
</tr>
<tr>
<td>Average</td>
<td>3.27</td>
<td>-14.9%</td>
<td>1.85</td>
<td>5.32</td>
<td>3.47</td>
<td>0.685</td>
</tr>
<tr>
<td>Weight Distribution</td>
<td>Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.92</td>
<td>-12.9%</td>
<td>2.01</td>
<td>4.34</td>
<td>2.34</td>
<td>0.473</td>
</tr>
<tr>
<td>2</td>
<td>5.81</td>
<td>-16.7%</td>
<td>3.48</td>
<td>8.77</td>
<td>5.29</td>
<td>0.973</td>
</tr>
<tr>
<td>3</td>
<td>3.42</td>
<td>-15.2%</td>
<td>1.29</td>
<td>6.85</td>
<td>5.56</td>
<td>1.166</td>
</tr>
<tr>
<td>4</td>
<td>1.02</td>
<td>-12.8%</td>
<td>0.62</td>
<td>1.70</td>
<td>1.07</td>
<td>0.189</td>
</tr>
<tr>
<td>Average</td>
<td>3.29</td>
<td>-14.4%</td>
<td>1.85</td>
<td>5.41</td>
<td>3.56</td>
<td>0.700</td>
</tr>
<tr>
<td>SDV</td>
<td>Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.51</td>
<td>-25.05%</td>
<td>1.84</td>
<td>3.40</td>
<td>1.56</td>
<td>0.304</td>
</tr>
<tr>
<td>2</td>
<td>4.96</td>
<td>-28.78%</td>
<td>3.28</td>
<td>6.58</td>
<td>3.30</td>
<td>0.604</td>
</tr>
<tr>
<td>3</td>
<td>2.82</td>
<td>-30.12%</td>
<td>1.37</td>
<td>4.44</td>
<td>3.07</td>
<td>0.641</td>
</tr>
<tr>
<td>4</td>
<td>0.90</td>
<td>-23.15%</td>
<td>0.60</td>
<td>1.22</td>
<td>0.63</td>
<td>0.119</td>
</tr>
<tr>
<td>Average</td>
<td>2.80</td>
<td>-26.78%</td>
<td>1.77</td>
<td>3.91</td>
<td>2.14</td>
<td>0.417</td>
</tr>
</tbody>
</table>

Percentage Relative Error = 100 * standard error / estimate.
Relative Change = (Trimmed - Untrimmed) / Untrimmed.
Variables: See Table 8.
### Table 11. Percentage of Confidence Intervals Covering True Value Based on 200 Replicated Samples

<table>
<thead>
<tr>
<th></th>
<th>Variable 1</th>
<th></th>
<th>Variable 2</th>
<th></th>
<th></th>
<th>Variable 3</th>
<th></th>
<th></th>
<th>Variable 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>90 Percent Confidence Interval</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untrimmed Weights</td>
<td>90.0</td>
<td>87.0</td>
<td>77.5</td>
<td>82.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MSE Procedure</td>
<td>76.0</td>
<td>87.5</td>
<td>79.5</td>
<td>84.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor Series Procedure</td>
<td>81.5</td>
<td>85.0</td>
<td>77.5</td>
<td>82.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>87.0</td>
<td>87.5</td>
<td>81.5</td>
<td>82.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight Distribution Procedure</td>
<td>87.0</td>
<td>87.5</td>
<td>82.0</td>
<td>82.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>61.5</td>
<td>86.5</td>
<td>83.0</td>
<td>87.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>95 Percent Confidence Interval</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untrimmed Weights</td>
<td>94.5</td>
<td>90.5</td>
<td>84.5</td>
<td>90.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MSE Procedure</td>
<td>83.5</td>
<td>91.0</td>
<td>85.5</td>
<td>90.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor Series Procedure</td>
<td>88.0</td>
<td>90.5</td>
<td>85.5</td>
<td>89.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>91.0</td>
<td>91.5</td>
<td>88.0</td>
<td>92.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight Distribution Procedure</td>
<td>91.5</td>
<td>91.5</td>
<td>87.5</td>
<td>92.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>74.5</td>
<td>92.0</td>
<td>86.0</td>
<td>91.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>99 Percent Confidence Interval</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untrimmed Weights</td>
<td>96.5</td>
<td>95.5</td>
<td>91.0</td>
<td>98.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MSE Procedure</td>
<td>93.5</td>
<td>97.0</td>
<td>90.5</td>
<td>98.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor Series Procedure</td>
<td>95.0</td>
<td>97.5</td>
<td>91.5</td>
<td>98.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td>97.5</td>
<td>97.5</td>
<td>92.0</td>
<td>98.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight Distribution Procedure</td>
<td>97.5</td>
<td>97.5</td>
<td>92.0</td>
<td>98.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDV Procedure</td>
<td>88.5</td>
<td>96.0</td>
<td>92.0</td>
<td>98.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Variables: 1 = median family income; 2 = birth rate among teenagers; 3 = percentage white age 10-17; 4 = temperature in June.

2 $90 \text{ Percent Confidence Interval} = \hat{\theta}_s \pm 1.645 \times \text{SQRT(Var(\hat{\theta}_s))}$

3 $95 \text{ Percent Confidence Interval} = \hat{\theta}_s \pm 1.96 \times \text{SQRT(Var(\hat{\theta}_s))}$

4 $99 \text{ Percent Confidence Interval} = \hat{\theta}_s \pm 2.58 \times \text{SQRT(Var(\hat{\theta}_s))}$

126
variables, the coverage probabilities were sometimes substantially lower than the nominal levels. It should be noted that, in the population of counties, the median family income has a distribution that is close to the normal distribution whereas the distributions of the other variables are asymmetric.

When considering the trimmed weights for the first variable (median family income) that was expected to have a negative bias from weight trimming, the proportions of intervals containing the true value were higher for the NAEP and the weight distribution procedure than the proportions for the other three procedures and, for the 99 percent confidence interval, larger than the proportion for the untrimmed weights.

For the other three variables, the proportion of intervals containing the true value was generally larger for the trimmed weights than for the untrimmed weights. The reason for this improvement is that the extreme weights in the samples increased the variation of the estimated means over the repeated sample for these three variables. The weight trimming introduced relatively little variation in the estimated means for these variables in the repeated samples and reduced the variation of the means over the repeated samples, thus resulting in improved interval estimates.

H. Design Effects

In Chapter III, a generalized design effect (DEFFWG) measure of unequal weighting was derived that incorporates the correlation between the sampling weights and the observed data. The DEFFWG is
based on a simple variance model for the data similar to the model for Kish's DEFF<sub>W</sub> measure (1965). In Table 12, the actual DEFF measure (computed as the ratio of the estimated variance to the variance assuming simple random sampling) and DEFF<sub>WG</sub> for estimates is summarized using the untrimmed weights and using the trimmed weights.

For estimates using the untrimmed weights, the generalized DEFF measure tends to underestimate the actual DEFF for two variables (variable 2, the birth rate among teenagers, and variable 3, the percent of 5 to 17 year old population that is white) and overestimates the actual DEFF for the other two variables. When the trimmed weights are used, the generalized DEFF measure tends to underestimate the actual DEFF for all but the fourth variable (the temperature in June). The generalized DEFF measure does provide an indication of the relative magnitude of the design effect.

The tendency of the DEFF<sub>WG</sub> to underestimate the actual DEFF may be caused by the fact that the DEFF<sub>WG</sub> does not account for the direction of the correlation between the weights and the data. A negative correlation between the data and sampling weights can result in a smaller estimated sampling variance (relative to an estimated variance computed with equal weights) whereas a positive correlation can result in a larger estimated sampling variance. The DEFF<sub>WG</sub> utilizes the square of the correlation coefficient. Therefore, a positive correlation and a negative correlation are treated equivalently, although only a negative correlation is expected to result in a decrease in the sampling variance.
Table 12. Comparison of Actual and Estimated Design Effects  
Average over 200 Replicated Samples

<table>
<thead>
<tr>
<th>Procedure/Variables</th>
<th>Actual Design Effect</th>
<th>Generalized DEFF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Standard Error</td>
</tr>
<tr>
<td>Untrimmed Weights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>2.58</td>
<td>0.110</td>
</tr>
<tr>
<td>2</td>
<td>4.78</td>
<td>0.256</td>
</tr>
<tr>
<td>3</td>
<td>6.10</td>
<td>0.662</td>
</tr>
<tr>
<td>4</td>
<td>3.15</td>
<td>0.135</td>
</tr>
<tr>
<td>Estimated MSE Procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.72</td>
<td>0.036</td>
</tr>
<tr>
<td>2</td>
<td>2.62</td>
<td>0.057</td>
</tr>
<tr>
<td>3</td>
<td>3.22</td>
<td>0.112</td>
</tr>
<tr>
<td>4</td>
<td>2.00</td>
<td>0.044</td>
</tr>
<tr>
<td>Taylor Series Procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.78</td>
<td>0.040</td>
</tr>
<tr>
<td>2</td>
<td>2.84</td>
<td>0.071</td>
</tr>
<tr>
<td>3</td>
<td>3.56</td>
<td>0.138</td>
</tr>
<tr>
<td>4</td>
<td>2.13</td>
<td>0.050</td>
</tr>
<tr>
<td>NAEP Procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.91</td>
<td>0.041</td>
</tr>
<tr>
<td>2</td>
<td>3.08</td>
<td>0.070</td>
</tr>
<tr>
<td>3</td>
<td>3.67</td>
<td>0.132</td>
</tr>
<tr>
<td>4</td>
<td>2.28</td>
<td>0.053</td>
</tr>
<tr>
<td>Weight Distribution Procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.93</td>
<td>0.042</td>
</tr>
<tr>
<td>2</td>
<td>3.12</td>
<td>0.073</td>
</tr>
<tr>
<td>3</td>
<td>3.92</td>
<td>0.138</td>
</tr>
<tr>
<td>4</td>
<td>2.30</td>
<td>0.054</td>
</tr>
<tr>
<td>SDV Procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.48</td>
<td>0.024</td>
</tr>
<tr>
<td>2</td>
<td>2.18</td>
<td>0.031</td>
</tr>
<tr>
<td>3</td>
<td>2.62</td>
<td>0.051</td>
</tr>
<tr>
<td>4</td>
<td>1.76</td>
<td>0.027</td>
</tr>
</tbody>
</table>

*Variables: 1 = median family income; 2 = birth rate among teenagers; 3 = percentage white age 10-17; 4 = temperature in June.
Chapter VI

SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH TOPICS

A. Summary

In this chapter, I will summarize the key results of this research and provide potential topics for future research.

The analytical derivations in Chapter III were the basis for developing the three new weight trimming procedures. More specifically, trimming sampling weights using the two current weight trimming procedures has taken on the form of an art. The derivations sought to introduce a theoretically based structure to weight trimming.

Because a primary issue in weight trimming is the potential effect of the trimming on the bias and the variance of the estimator, the expectation and variance was derived for an estimator of a population mean using a weight trimming strategy. An estimator of the variance was also developed using the Taylor series linearization approach to derive a variate accounting for the weight trimming.

The first proposed weight trimming procedure, the Taylor series procedure (described in Chapter IV), was developed using the Taylor series linearized variate and an estimator of the bias. This procedure permits the assessment of both the potential for bias and the variance reduction introduced by the weight trimming. The second proposed weight trimming procedure, the SDV procedure, is based on
single draw variates (SDVs) and uses the studentized residuals of the SDVs to identify units with extreme weights.

Because the Taylor series procedure and the single draw variate procedure utilize data in the determination of the weights for trimming, a generalized unequal weighting design effect measure (DEFFWG) was developed to assist in identifying the key data items for use in the weight trimming. The DEFFWG was developed using a simple linear model that represents the relationship between the observed data and the selection probabilities and sampling weights.

For situations in which weight trimming is performed either with dichotomous or polychotomous survey response variables and with no data available, a third weight trimming procedure, the weight distribution procedure, was proposed that utilizes an assumed distribution for the sampling weights. The theoretical distribution for sampling weights was derived based on the assumption that the selection probabilities follow a beta distribution.

In the empirical study two currently used methods, the estimated MSE procedure and the NAEP procedure, and the three proposed procedures were evaluated using a population constructed from the 1986 Area Resource File (ARF) of county-level data. Two hundred (200) samples of 100 counties each were selected. Each sample was selected independently using Chromy's probability minimal replacement (PMR) sequential selection algorithm (1979). The sampled units were selected with probability proportional to size using the count of households as the size measure.
For the empirical study, four variables were used from the ARF data base: median family income, birth rate among teenagers, percentage of 5 to 17 year old population that are white, and the average temperature in June. For each of the 200 samples, I implemented the five weight trimming procedures and measured the statistical effect of weight trimming on estimates using these four variables.

The findings of the empirical study show that, of the five procedures, the Taylor series and the estimated mean square error procedure tended to perform similarly. Also, the NAEP procedure and the weight distribution procedure operated almost identically. The SDV based procedure tended to have some of the characteristics of both pairs of procedures.

In comparing the Taylor series and the estimated MSE procedure, the estimated MSE procedure resulted in the larger average reduction (25.3 percent reduction) in the variance over the 200 replicated samples than the Taylor series procedure (22.2 percent) for the four variables. However, for the one variable (median family income) where a negative bias was expected, the estimated MSE procedure resulted in a larger average MSE than the Taylor series procedure (32.7 percent average increase (relative to the variance of the untrimmed weights) for the estimated MSE procedure versus a 15.3 percent average increase for the Taylor series procedure). Therefore, although the average variance reduction was larger for the estimated MSE procedure, the average reduction in the MSE was slightly larger for the Taylor series procedure (3.4 percent reduction for the Taylor series procedure versus 2.8 percent for the estimated MSE procedure).
Both the NAEP procedure and the weight distribution procedure produced almost identical trimming levels over the 200 samples. The NAEP procedure and the weight distribution procedure resulted in an average variance reduction of 16.7 and 16.0 percent and average MSE reduction of 3.4 and 3.5 percent, respectively. Although the NAEP procedure and the weight distribution procedure provide the same results, the trimming level established for the weight distribution procedure can be specified by a probability statement (for example, extremely large weights with probability of less than 1 percent are trimmed) whereas the trimming level for the NAEP procedure is established empirically for each sample.

When considering the bias through the estimates of the MSE for these procedures, both the NAEP procedure and the weight distribution procedure resulted in approximately the same average reduction in the MSE as the Taylor series and the estimated MSE procedures. However, the Taylor series and the estimated MSE procedures produced the greater reduction in the estimated sampling variance. It should be noted that the NAEP and the weight distribution procedures trimmed a maximum of three weights, whereas the estimated MSE and the Taylor series procedures trimmed between one and nine weights.

The SDV based procedure resulted in substantially more trimming than the other procedures. This procedure resulted in the greatest variance reduction (33.9 percent reduction in the estimated variance) but also exhibited more bias than the other procedures for two of the four variables and an average proportional increase in MSE of 17.7 percent over the 200 samples.
For each variable, interval estimates were investigated by assessing the proportion of 90, 95, and 99 percent confidence intervals (computed using the individual sample estimates) that contained the true value. For the first variable (median family income) that was expected to have a negative bias from weight trimming, the proportions of intervals containing the true value were higher for the NAEP and the weight distribution procedure than the proportions for the other three procedures and, for the 99 percent confidence interval, larger than the proportion for the untrimmed weights.

For the other three variables, the proportion of intervals containing the true value was generally larger for the trimmed weights than for the untrimmed weights. The reason for this improvement is that the extreme weights in the samples increased the variation of the estimated means over the repeated sample for these three variables. Because the weight trimming introduced relatively little bias in the estimated means for these variables in the repeated samples and reduced the variation of the means over the repeated samples, this resulted in improved interval estimates.

In summary, the NAEP procedure and the weight distribution procedure are direct competitors in the sense that both procedures use only the sampling weights for weight trimming. These two procedures resulted in the least trimming and almost the same average variance and average MSE reductions. However, the weight distribution procedure permits a statistical basis for the choice of a trimming level whereas the criterion for the NAEP procedure is empirically
based. The Taylor series and the estimated MSE procedures are also
direct competitors in the sense that both use the observed survey data
and an estimator of the mean square error in the trimming procedure.
The Taylor series procedure also incorporates an estimator of the bias
introduced by the trimming into the weight trimming procedure. The
SDV procedure incorporated some aspects of both pairs of the other
procedures. It resulted in the largest reduction in the estimated
sampling variances but introduced the largest biases.

B. 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 this research, the five weight
trimming procedures investigated include three procedures that use the
observed survey data and two that use only the sampling weights.
In terms of the five weight trimming procedure as evaluated in the empirical study, the estimated MSE procedure and the Taylor series procedure utilize the data and an estimate of the MSE, and these procedures are preferable to the other three procedures. However, since the MSE estimate can mask the true extent of the bias introduced by weight trimming, the Taylor series procedure is preferable to the estimated MSE procedure because the Taylor series procedure directly incorporates the relative bias into the trimming procedure. The Taylor series procedure also utilizes a more correct variance estimator because the linearized variate took into account the weight trimming.

When no data or only categorical data are available, the weight distribution procedure is preferred over the NAEP procedure because the trimming level for the weight distribution procedure can be established on a theoretical basis. The trimming level for the NAEP procedure is based on an empirically evaluated subjective criterion.

As a general protocol for weight trimming, it is recommended that, if data are available for weight trimming, the analyst should evaluate the data used in the weight trimming procedure to ensure that they are representative of the data to be analyzed. That is, if some of the data to be analyzed are expected to be correlated (either positively or negatively) with the sampling weights, data with similar correlations should be used in the weight trimming.

If no data are available for weight trimming, the analyst should evaluate the weight trimming to assess the bias that may be introduced by evaluating the bias after data are available. In the empirical
study, the NAEP procedure trimmed fewer weights than did the estimated MSE procedure or the Taylor series procedure; however, in prior uses of the NAEP procedure, I found that the NAEP procedure sometimes indicated substantial trimming. Since the NAEP procedure does not utilize data, the effect of the trimming (in terms of the actual variance reduction and the bias introduced) is not readily apparent.

The five weight trimming procedures were evaluated in the empirical study as separate procedures. In practice, components of these procedures can be combined to form a single approach. For example, the sampling weight distribution can be used to determine candidate trimming levels for the Taylor series procedure or the SDV procedure. Also, the SDV procedure can incorporate a bias factor to limit the estimated bias introduced.

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.
C. **Future Research Topics**

Three areas for future research are recommended. These are: (a) to evaluate the effect of the weight trimming for other nonlinear statistics such as regression coefficients and correlation coefficients; (b) to further refine the SDV based procedure by developing a distribution model for SDVs and by using more robust estimators of the mean and variance of SDVs; and (c) to evaluate the weight trimming strategies on additional data bases. The following is a discussion of the recommended areas for future research:

1. In the analytic derivations and the empirical study, I used a simple ratio-type statistic (the estimated mean). The weight trimming reduced the estimated sampling variance and, in some cases, introduced bias. The effect of the bias may become more pronounced when a more complex nonlinear statistic (such as a regression coefficient) is estimated. An estimator of the bias for the estimated mean was described in Chapter III. Future research can develop a similar estimator for more complex nonlinear statistics.

2. For the SDV procedure, I assumed a normal distribution for developing a trimming criterion. Future research into other distributions for weight trimming criteria should improve this procedure. If the weight distribution model derived in Chapter III is used, the distribution for the SDVs can be derived as the product of the weight distribution random variable and the random variable for the observed data.
The SDV procedure in this research effort utilized the usual estimators for the mean and variance. These estimators can be adversely affected by extreme values. Barnett and Lewis (1984) describe various more robust estimators that can be used to detect extreme values, or outliers.

3. I attempted to enhance the realism of the empirical study by using the Area Resource File data base. However, the sample design needed to be simplified to allow evaluation of these procedures. The true test of these procedures is their application and evaluation on actual survey data bases. Because most major surveys include multiple stages, these procedures should be evaluated on real multi-stage survey data bases.

These procedures can be evaluated using a real multi-stage survey in the following fashion. The evaluation would use a large survey data base (such as the longitudinal National Health Interview Survey conducted by the National Center for Health Statistics). A set of data items would be identified that include data that are correlated and not correlated with the sampling weights as well as data for which independent external values are available. The five weight trimming procedures would be implemented and the statistical effect of the weight trimming would be measured using sample-based
estimates of the sampling variance, the estimated bias introduced, and the mean square error computed for each procedure. For data with external values available, a true measure of the bias could be computed. For other data, the use of a longitudinal survey data base would allow for the estimation of the potential bias introduced by the weight trimming by the comparison of the estimates using trimmed weights to estimates computed or projected from prior years.
REFERENCES


141


Kish, L., (1965), Survey Sampling, New York: John Wiley.


