Practical Issues When Calibrating Weights for Multiple Skewed Variables

Yan K. Liu, Kimberly Henry, Michael Strudler
Statistics of Income, Internal Revenue Service
77 K Street, NE, Washington, DC 20002.

Abstract
The Statistics of Income Division of the IRS started a panel sample of individual returns in 2007 for longitudinal analyses. This panel sample has also been used for cross-sectional estimations of multiple variables that are skewed and weakly correlated. Therefore, cross-sectional weights are needed to refer to the out-year population and multiple variables of interest. Calibration method is therefore applied to adjust weights such that sample estimates are close to population benchmarks. In this paper, we look at issues such as calibration cells, initial weights, prediction models and weight bounding. We share some experience in calibrating weights using R. We aim to produce final weights that are reasonable for multiple variables of interest.

Key words: calibration weighting, cross-sectional estimation, panel sample, R.

1. Introduction

The Statistics of Income Division (SOI) of the IRS started a new panel sample of individual returns in Tax Year 2007. This panel sample is intended to be used for longitudinal analyses and for cross-sectional estimations. Therefore, there are two sets of estimation weights: the longitudinal weights and the cross-sectional weights. The longitudinal weights refer to the population at the initial selection of the longitudinal sample. These weights are usually adjusted to take into account the attrition of the sample over time. The longitudinal weights are used when performing analyses of the longitudinal data through years. The cross-sectional weights are used to produce point estimates each year. Because of changes in the population through time, the cross-sectional weights are different from the longitudinal weights. A set of cross-sectional weights for each year should refer to the population of that year.

The first year of the panel is termed the *base-year*, while subsequent years are called *out-years*. The base-year panel sample was a stratified sample, where stratum boundaries were formed using the return's selection income. Selection Income is derived from the components of a taxpayer's adjusted gross income plus certain nontaxable items. Since different selection probabilities are used across the strata, varying from 0.34% to 100%, the base weights vary dramatically. This poses a particular problem for returns whose out-year income grows so dramatically that its associated base weight is no longer appropriate for cross-sectional estimations. In addition, a small refreshment sample is added each year for the purpose of cross-sectional estimations. How to incorporate the refreshment sample returns is another issue related to weighting the panel returns. Further, the base-year sample was stratified by one variable, while the out-year cross-sectional estimations are needed for multiple variables that are highly skewed and weakly correlated. Therefore, cross-sectional weights derived from base-year selection probabilities may need to be adjusted to refer to the out-year population and multiple variables of interest. In this paper, we develop cross-sectional weights for the sample of surviving panel returns and refreshment returns and look at weighting issues for the cross-sectional estimations. We consider some practical issues related to producing the out-year cross-sectional weights, including issues such as calibration cell definitions, initial weights, prediction models and weight bounding. We share some experience in calibrating weights using R. Our goal is to produce a set of final weights that produce reasonable cross-sectional estimates for multiple variables of interest.

2. Outline of the Base-Year Panel Sample (TY2007)

The TY2007 base-year panel sample was selected from the population of TY2007 returns. It was a stratified sample, where the stratum was defined by the selection income The sample included a random selection part and a secondary CWHS part. The random selection part was selected using permanent random number that was the transformed last-four digits of the return's primary SSN. The sampling rates were different across strata. The secondary CWHS part included returns having any of ten specific last four digits of the secondary SSN. Including the secondary CWHS sample incurred no additional cost since these returns would be processed for other purposes. Returns were retained in the sample in out-years if it was possible to link returns across years using the primary and secondary SSN's. A base-year panel return stays in the panel in the following years if either its primary SSN, secondary SSN, or both file tax returns in the out-year. These are referred as "surviving panel returns".

For the base-year panel return weights, notice that the secondary CWHS part represents only returns of married couples filing jointly, while the random part represents returns of all marital and filing statuses. To calculate return weights, we poststratify the random part into two groups within each stratum: returns that are married and filing jointly and all other returns. Within each poststratified cell, the base sampling weight is simply the population size divided by the sample size.

The base-year sampling rates across strata, along with the associated base weights, are shown in Table 1.

Table 1. Base-Year Return Weights (TY2007)

Stratum	Sampling Rates (%)	Sampling Weigh (N/n)	
		Married Filing Jointly	Other filing Statuses
1	100.00	1.0	1.0
2	100.00	1.0	1.0
3	50.00	2.1	2.0
4	50.00	2.0	2.0
5	22.51	4.4	4.4
6	3.38	30.0	25.9
7	2.00	47.7	55.4
8	1.40	480.0	968.2
9	1.40	502.5	998.7
10	0.34	252.7	331.3
11	1.86	51.5	52.9
12	2.44	39.2	39.3
13	12.18	8.1	7.9
14	28.60	3.5	3.6
15	50.00	2.0	2.1
16	100.00	1.0	1.0
17	100.00	1.0	1.0

3. Out-Year Cross-Sectional Sample and Initial Weights

To support the cross-sectional estimation, a small refreshment sample is selected every year. It is also a stratified sample that has the same stratum definition as the base-year panel sample. The selection income is adjusted for the inflation so that the out-year selection income is compatible to the base-year selection income. The refreshment sample adds some newly rich returns[1] and some new entrants. It helps to keep the out-year sample representative.

The surviving panel returns and refreshment returns together compose the cross-sectional sample that is used to make out-year estimations on Sales and Capital Assets (SOCA). Therefore, it is also referred as the SOCA cross-sectional sample. In this paper, we look at the weight development for one of the out-years, Tax Year 2010. The TY2010 SOCA cross-sectional sample includes 222,545 returns. Of those, 200,907 are surviving panel returns and 21,638 are refreshment sample returns. Our goal is to develop return weights so that the TY2010 SOCA sample can best represent the TY2010 population. In terms of the SOCA cross-sectional sample return weights, we start with 'probability-based' initial weights. Then we make adjustment by throwing in additional auxiliary information. The initial weights before adjustments are described here.

The initial return weight is based on the inverse of return's selection probabilities when the return is linked to base-year return/returns. For returns that did not file in the base-

[1] The newly rich filers are either new filers with high-income or stratum jumpers.

year, ad hoc weights are assigned. We first link TY2010 SOCA sample returns to the TY2007 population using the primary SSN (PSSN) and secondary SSN (SSSN). If there is a match, the primary filer's selection probability P_1 is assigned based on the stratum of the base-year return's. If the return's filing status is joint, we check to see if the SSSN is linked to the base-year PSSN or SSSN (sample data first, then population data). If there is a match, the secondary filer's selection probability P_2 is assigned based on the stratum of the base-year return. Otherwise, P_2=0. Out of the 222,545 TY2010 SOCA cross-sectional sample returns, 204,455 have matched returns in the base-year population and 18,090 returns have no match.

For the 204,455 returns that are linked to the base-year population, the initial weight of any return is calculated based on its primary filer's selection probability P_1 and secondary filer's selection probability P_2. If P_2=0, then the initial weight $d = 1/P_1$. If P_1 and P_2 are both from the same base-year return, then $d = 1/P_1$. If P_1 and P_2 are from two different base-year returns, then d is the inverse of the joint selection probability, i.e., $d = 1/ ((1/P_1) + (1/P_2) - (1/P_1)(1/P_2))$. Technically, some returns were subject to both base-year panel sample selection and the out-year refreshment sample selection, which creates a complicated selection probability. Therefore, we take this ad hoc approach, since the initial weights may be adjusted later through subsequent calibration and trimming adjustments.

The 18,090 unmatched returns are treated as new filers. Since sampling rates across strata for the refreshment are small, new filers are under-represented. Therefore, weights for unmatched returns should be large compared to the matched. We use an ad hoc approach for initial weights of unmatched returns, using the empirical 90th percentile of matched return weights in the corresponding poststratified cell. This ensures that the weights are relatively large but not influential.. Again, these adjustments are performed within each poststratified cell, i.e., married filing jointly vs. all other filing statuses within each selection income stratum.

4. Weight Adjusting Using Calibration Approach

With this SOCA cross-sectional sample, we are interested in the estimation for some key variables. Since these variables are not closely correlated with each others, a set of weights that works well for the estimation of one variable may not be good for other variables. In order to find a set of compromised weights that balance all the key variables, we make use of an available resource of known out-year population totals of those key variables. In the end, we hope to have a weighted sample that reflects the out-year population and provides reasonable estimates of multiple key variables.

SOI has rich information on important variables from the large yearly cross-sectional sample (at the return-level). We then use a calibration approach to adjust the initial weights to produce weighted estimates to match these known population totals (e.g., Särndal *et. al*. 1992; Kott 2009).

Let d_k be the initial weight of return k (after trimming) and w_k be the calibration weight of return k. The calculation weights w_k are calculated through an iterative process of adjustments. In other words, calibration is a weight-adjustment method that creates a set of weights, $\{w_k\}$, such that (1) they are close to the original design weights d_k (i.e., as the sample size grows arbitrarily large, w_k converges to d_k) and are therefore nearly design-unbiased; and (2) satisfy a set of calibration equations:

$$\begin{aligned} &\sum_S w_k = N \\ &\sum_S w_k \mathbf{x}_k = \sum_U \mathbf{x}_k \end{aligned} , \qquad (4.1)$$

where N and $\sum_U \mathbf{x}_k$ are the known control totals. There is one calibration equation for each auxiliary x-variable. Different distance functions specified for d_k and w_k produce different kinds of weights (Deville and Särndal 1992). For example, using a linear distance function along the linear prediction produces the generalized regression estimator. Several extensions, including bounds on the size of the final weights have been proposed in the literature (Rao and Singh 1997; Singh and Mohl 1996; etc.).

Table 2 gives the variables that we are interested. We would like to have the adjusted weights that give the estimates on these variables close to the population benchmarks. These variables have low correlation with each other. We may not be able to take care of all 16 variables since too many variables in the calibration model can cause convergence problem. So, we consider these variables in the order of the importance as shown in Table 2.

Table 2. Key Auxiliary Variables in the Order of Importance

Variable	Description
x_1	Net short-term capital gain/loss
x_2	Net long-term capital gain/loss
x_3	Net short-term gain/loss from Sales of Capital Assets
x_4	Net long-term gain/loss from Sales of Capital Assets
x_5	Short-term gain from Form 6252 and short-term gain/loss from Forms 4684, 6781, and 8824
x_6	Net short-term gain or (loss) from partnerships, S corporations, estates, and trusts from Schedule(s) K-1
x_7	Gain from Form 4797, Part I; long-term gain from Forms 2439 and 6252; and long-term gain/loss from Forms 4684, 6781, and 8824
x_8	Net long-term gain or (loss) from partnerships, S corporations, estates, and trusts from Schedule(s) K-1
x_9	Capital gain distributions reported on Schedule D
x_{10}	Total short-term sales of capital assets amount
x_{11}	Total long-term sales of capital assets amount
x_{12}	Short-term capital loss carryover
x_{13}	Long-term capital loss carryover
x_{14}	Net capital gain/loss(limited) reported on Schedule D
x_{15}	Adjusted Gross Income
x_{16}	Selection Income (Stratifying variable)

5. Application of Weight Calibration

In this section, we look at how calibration weighting is performed. We discuss some practical issues and solutions when creating calibration groups, assigning initial weights and dealing with outliers, putting bounds on final weights, and choosing a calibration distance function. We then analyze the calibration results and compare with the benchmarks. Finally, we summarize our weight adjustment strategy. Our data are highly dispersed. So we form calibration groups and apply calibration method within each group. In this paper, we look at two groups to discuss our weighting strategy and address our issues.

5.1 Calibration Data

Throughout the text, we use two example cases of data to illustrate the practical issues and our solutions. The Case 1 includes 574 returns whose selection incomes range from -$20 million to -$150 million and the second group includes 2,273 returns whose selection incomes range from -$1 million to -$5 million. The key variables are $x_1 - x_{16}$ in the order of their importance, as shown in Table 2. The selection income x_{16} is the least important in terms of estimation bias in its out-year total. Figure 1 is the scatterplot matrix of the first eight variables in Case 1. We can see that variables are not closely related to each other and there are extreme values in all variables (most of the correlation coefficients range from 0.01 – 0.3.0). Similar relationship holds for other variables and for Case 2 example.

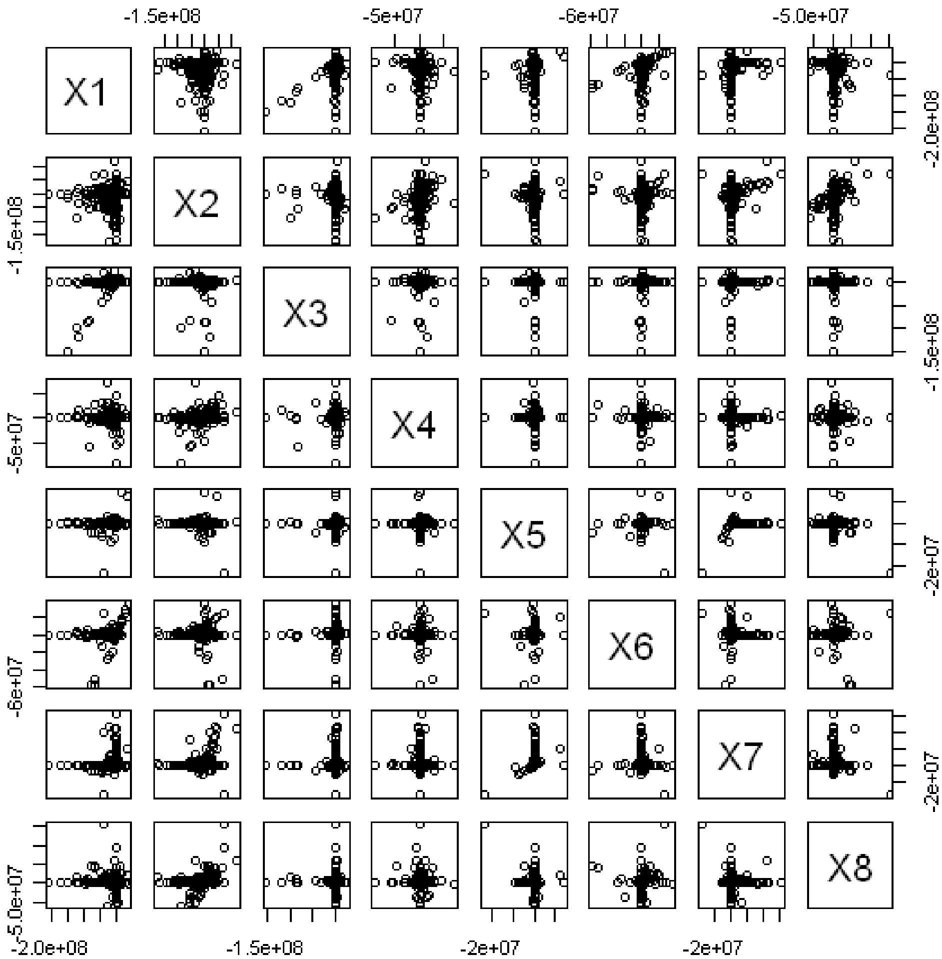

Figure 1. Scatterplot Matrix of Variables $x_1 - x_8$ (Case 1 Example)

5.2 Calibration Using R

We use the software R to produce the calibrated weights. R's survey package (Lumley 2008) includes the *calibrate()* function for producing various calibration weights. There are some particular choices to make when using this function, since there are options to vary the following: (1) the prediction model that relates the study variable *y* to the auxiliary *x*-variables, (2) the distance function, and (3) other parameters.

After the sample dataset is read into R, we must create a survey design object as input to the calibration function. This ensures that the calibration function uses the proper sample design and weighting information. We offer the following arbitrary code to achieve this (where terms within the [ ] brackets are user-specified):

```
sam.dsgn<-svydesign(id=~[ID variable name], weights=~D,
                    data=as.data.frame([sample dataset name]))
```

Here is the R code of our weight calibration:

```
W1 <- calibrate(design = sam.dsgn,
formula=function of x-variables,
 population = [vector of population control totals],
bounds = c(U, L),
calfun = c([distance function]),
maxit=[integer value],
epsilon=([decimal value])
```

Our choices for the calibrate function that determine the weights are explained in the following.

```
W1 <- calibrate(design = sam.dsgn,
formula=~X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11+X12+X13+X14+X15+X16,
population = poptot,
bounds = c(0, 8),
calfun = c("linear"),
maxit=2000,
epsilon=0.07)
```

- *Formula* = specifies the prediction model. Our variables of interest *Y* are the same as the auxiliary variables *X*. So we choose the linear model as our prediction model. We start by throwing all 16 variables in the model. If R's console shows a warning that the calibration does not converge due to model problem, then we need to adjust the prediction model for convergence.

- *population* = the known control totals for the population size (N) and totals of the variables x_1-x_{16}.

- *calfun=c("linear")* specifies that the distance function is linear. Our resulting calibration weights will be those associated with the generalized regression estimator (e.g., Sarndal *et. al*. 1992). R also supports other two distance functions: raking and logit. Empirically we found little differences in the weights when using different distance functions.

- *c(L, U)* is the argument that puts a bound on weight adjustment such that $L < w_k/d_k < U$. Since the linear distance function can produce negative weights, we set L=0 to force weights to be positive and U=8 to limit a calibrated weight not to exceed 8 times of its initial weight. R does not offer the option to specify different bounds for each sample unit *k*, which can be problematic when we want the calibrated weights w_k (not w_k/d_k) to be bounded. In particular, we want all w_k to be at least 1. We will deal with this subsequently using R's *trimWeight()* function.

- *maxit* is the parameter that specifies the maximum number of iterations in the calculation. We set *maxit=2,000* here to allow up to 2,000 iterations.

- *epsilon* is the tolerance parameter and is an extremely useful feature. It is a maximum allowable relative distance between the estimated total and the population total for calibration variables, including the intercept if specified in the model. We normally start with a small tolerance parameter. If the calibration fails to converge because the specified epsilon larger than allowed, then the calibrate function will produce an error message in the console that with the value of epsilon from the last specified iteration. We can then reset the correct value for *epislon*. In the above example, if we first set *epsilon=0.001*, the convergence could fail with the message: "*Failed to converge: eps=0.06985287 in 2,001 iterations,*" After resetting *epsilon=0.07,* the calibration will converge. Note that *W1* is the R object that is produced from the calibrate function. To obtain an output object that includes only the vector of calibrated weights, we need to use the *weights()* function in R.

5.3 Further Adjusting by Poststratification

When the calibration converges at a particular *epsilon* tolerance parameter, there can be a difference between the estimated total and the population total for the auxiliary variables and the population size. For example, if we set *epsilon=0.07*, then the relative difference should not be more than 7%. We actually do not require the estimates to match the benchmark totals exactly, as long as the difference is reasonable. However, we do need to have the sum of calibration weights match the post-strata population sizes exactly. For our Case 1 example, the above calibration with tolerance parameter *epsilon=0.07* results in calibration weights w_k such that $\sum w_k$ = 1,018, while the known population size *N*=1,095. In order to match the population size *N,* we make further adjustment through poststratification using R's function *postStratify() function:*

W2 <- postStratify(W1, [stratum variable], [population count by stratum])

Let w_k' denote the poststratified weights. Then w_k' satisfy that $\sum w_k' = N$. This adjustment may slightly increase the relative difference on calibration variables.

6. Issues and Solutions in the Weight Calibration

In this section, we discuss issues we have to consider when we develop weights through calibration. Specifically, we address issues, such as assigning initial weights and dealing with outlier weights, creating groups for calibration, putting bounds on final weights, choosing the prediction model. We look at these issues through some examples.

6.1 Initial Weights

The initial weights will be adjusted using calibration method. But before weight adjusting, it is important to realize that there may be outliers in initial weights. The initial weights are carried from the base-year selection. However, income changes over time due to economic success/failure or return composition change (e.g., marriage or divorce), which leads to units shifting to different strata where selection probabilities were different in the base-year sample. Returns with very low income in the base-year may end up with extremely high income in the out-year. A problem arises when returns shift strata due to dramatic changes in yearly income[2]. Some stratum jumpers that experience very large growth in income (the absolute value), along with their large weights, will have extremely large influence upon the estimates of income and tax variables at income levels where most panel members have much smaller weights. Those returns with both large weights and large incomes can inflate the variance and cause estimation bias. Therefore, appropriate weight trimming for the extreme stratum jumpers is considered here. Instead

[2] All incomes in out-years are adjusted so that they are comparable to the base-year 2007 income.

of customized trimming, we trim weights to the 90th percentile within each cell for matched returns. In other words, the ad hoc initial weight assigned to unmatched return is the cap for trimming initial weights.

Calibration is a reweighting procedure that adjusts initial weights so that equations in (4.1) are satisfied or approximately satisfied. As discussed in Section 3, it would be really complicated to calculate exact selection probability for some returns. Therefore, we had some ad hoc treatment for some returns to come up with initial weights. We want to know if this is a big deal. Another issue on initial weights is outlier weights from stratum jumpers. We want to see if we have to trim those extreme weights.

We still take the above group 1 data for example and look at three sets of initial weights:

d_k – initial weights derived from the selection probabilities, $1 <= d_k <= 1{,}007$
d_k^T – truncated weights where large initial weights d_k are truncated, $1 \le d_k^T <= 3.47$
d_k^C – a constant weight for all returns, i.e., $d_k^C = N/n = 1.9077$ for all k

The initial weights d_k are carried from the base-year selection probabilities with ad hoc treatments for some returns. There are outliers in d_k from stratum jumpers. 12 returns have an initial weight larger than 39 and rest 562 returns are all under 8.2. To avoid extreme initial weights, we truncate d_k to its 90th percentile within each weighting cell. The truncated weights are denoted as d_k^T. The truncated weights in this group are no larger than 3.47. The third set of initial weights, d_k^C, ignore the selection probabilities and simply assumes an ad hoc constant weight for all returns.

We apply the same calibration procedure described in Section 4 for each of three sets of initial weights. They all converge with a prediction model of 16 variables, but at different *epsilon* levels - 0.024 for d_k^T 0.077 for d_k^C and 0.293 for d_k. Then the calibrated weights are further adjusted by poststratification to match the population size and by trimming to force weights to fall in the range of 1 to 10. We look at the relative bias between the estimate and the population bench mark and compare them for the three sets of initial weights and for each of 16 variables. Figure 2 summarizes the relative biases. The probability-based initial weights without trimming outliers (d_k) result in much larger bias than truncated weights (d_k^T) and constant (d_k^C) for most variables. d_k^T and d_k^C are close in terms of the bias.

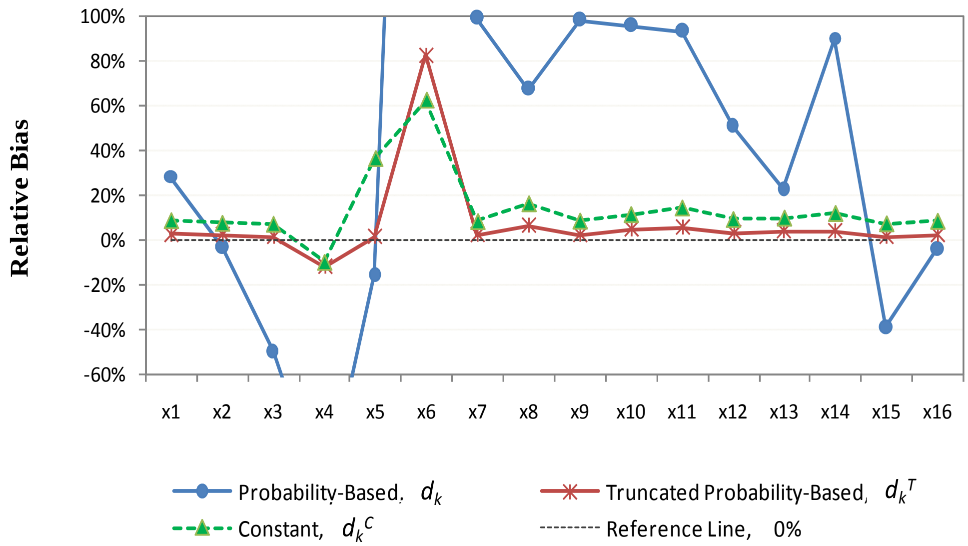

Figure 2. Relative Bias Comparison for Three Sets of Initial Weights

The above analysis shows that trimming outlier initial weights to 90th percentile within each weighting cell works well. For returns without base-year matches, the ad hoc treatment of assigning an initial weight of 90th percentile is also a reasonable choice since the constant weight weights d_k^C is not much different from d_k^T in terms of estimation bias. Therefore, we choose d_k^T , the trimmed probability weights with ad hoc treatment, as the initial weights in the calibration procedure for all groups. In summary, the ad hoc treatment to initial weights is fine, as long as those initial weights are in a reasonable range. This is because calibration adjustment can fine tune initial weights to a certain degree. Extreme initial weights can cause problem in calibration, either convergence problem (does not converge at all) or large bias problem (converge, but at a large tolerance level). So they should be trimmed to a reasonable level.

In the following analyses, only trimmed initial weight d_k^T is used. Other forms of initial weights (d_k and d_k^C) are not considered anymore.

6.2 Forming Calibration Cells

Intuitively, we want calibration cells to be homogeneous, while having a reasonably large enough cell sample size for each of them. We look at the example of Case 2. There are 2,273 returns with initial weight d_k^T ranging from 1 to 57. We still choose the linear prediction model and relax the bound option to be *c(0, 300)*. For any number of variables in the prediction model, the tolerance parameter *epsilon* has to be so large in order for the calibration convergence. To calibrate on all 16 variables, we need at least *epsilon* = 21.05 for the calibration to converge. That is, the maximum allowable relative bias for calibration variables and the intercept is 2105%. Even to calibrate on only two variables x_1 and x_2, we need at least *epsilon* = 4.41. This precision level is obviously not acceptable.

Then we divide the group into four cells by the combination of I_1 and I_2, where

$$I_1 = \begin{cases} 0, \text{ if } x_1 < 0 \\ 0, \text{ if } x_1 \geq 0 \end{cases} \quad \text{and} \quad I_2 = \begin{cases} 0, \text{ if } x_2 < 0 \\ 0, \text{ if } x_2 \geq 0 \end{cases}.$$

The cell sample sizes are 379, 322, 309 and 1,263 for cells 1-4. We apply calibration procedure within each cell. For a full prediction model of 16 variables, the minimum tolerance parameters *epsilon* needed for calibration convergence are 0.172, 0.137, 0.120 and 0.002 for four cells respectively. Other examples also show that homogeneity helps calibration convergence and the precision.

6.3 Choice of Variables in the Prediction Model

Our specified prediction model is a linear with a maximum of 16 variables. Sometimes, too many variables in the prediction model may not converge or converge at a large tolerance level. Take for example the Case 2 data. There are four calibration cells. As indicated in Section 5.2, with a full prediction model of x_1-x_{16}, calibration converges in each cell, but at high tolerance parameters. Another option is to see if the calibration may converge at a much smaller tolerance level after dropping out a few less important variables from the model. Table 3 gives two model options. Option 1 is to use the full model in each cell. Option 2 is to choose a model that can converge at a much smaller *Epsilon*. The variables dropped from the model are less important anyway. The estimation biases for two options are summarized in Figure 3. It obviously shows that option 2 outperforms option 1 in terms of bias.

Table 3. Model Choice and Epsilon for Convergence

Cell	Option 1		Option 2	
	Variables in the model	***Epsilon* for convergence**	**Variables in the Model**	***Epsilon* for Convergence**
1	Full	0.173	14 variables (without x_{10} and x_{16})	0.001
2	Full	0.137	13 variables (without x_{10}, x_{15} and x_{16})	0.001
3	Full	0.120	9 variables (x_1 - x_9)	0.001
4	Full	0.002	16 variables (full model)	0.002

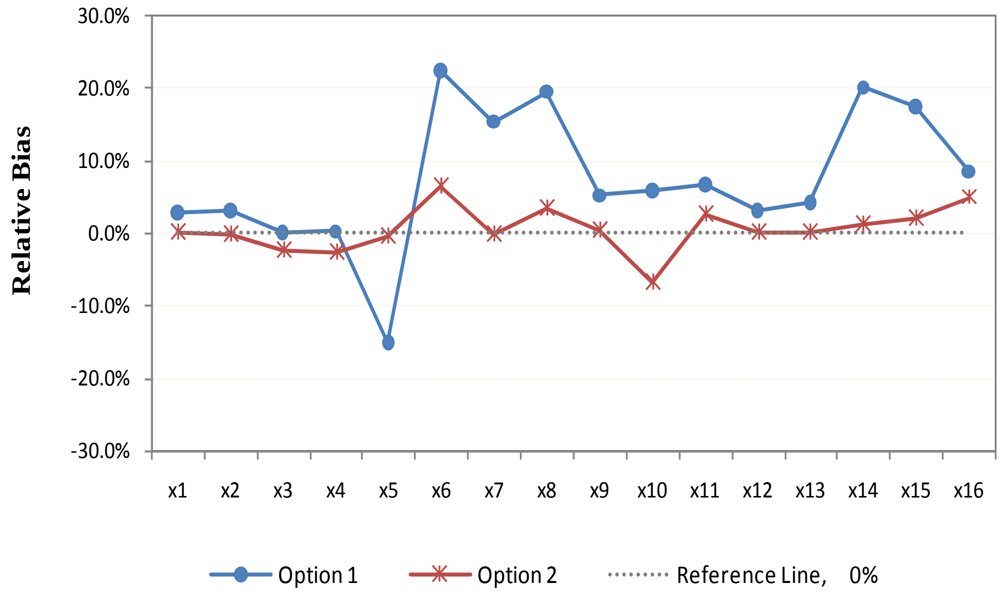

Figure 3. Relative Bias Comparison for Two Model Options

6.4 Final Adjustment to the Desired Weight Bounds

The weight bounding option in R's calibration function is only for w_k/d_k , not for w_k. Therefore, after the calibration and the poststratification adjustments, there is still a problem with the adjusted weights, $w_k^{'}$, in that some weights may be smaller than 1 and some weights may be outside the desired upper bound. We may force the final weights to fall in the desired range using R's function *trimWeight()*. For example, if our desired final weight bounds are [1, 10], then we can restrict the weights to be within these bounds by further adjust $w_k^{'}$ using R's *trimWeights()* function*:*

```
W3 <- trimWeights(W2, lower=1, upper=10, strict=TRUE)
```

where W_2 is the calibrated and poststratified weight and W_3 is the trimmed weight. Weights outside the bounds are trimmed to be equal to the boundary values and the total amount trimmed is redistributed among the weights for observations that were not trimmed. This ensures that the total of the weights before and after the trimming remains the same. The reapportionment of the 'excess weight' can push the non-trimmed weights over the boundary limits. If the option *strict=TRUE* is used, then the function calls itself recursively to prevent this. This adjustment may increase the relative bias on calibration variables. Let denote w_k^* the final weights after calibration, poststratification and weight trimming, then w_k^* satisfy that $\sum w_k^* = N$ and $1 \leq w_k^* \leq 10$.

7. Summary of Weight Adjustment Strategy

In our weight adjustment applications, we make use of the known population benchmarks and balance the estimation precision of multiple variables. The calibration approach using R works well for this purpose. This section summarizes our weight adjustment strategy.

1. Prepare data for calibration
   - Trim outlier initial weights
   - Form calibration cells – balance homogeneity and the cell sample size.
   - Sort out auxiliary variables in the order of their importance.

2. Adjust initial weights by calibration approach using R
   - Start with a small tolerance parameter, say ***epsilon***=0.001. Then adjust the value of *epsilon* as necessary. If the calibration needs a larger *epsilon* value to converge, R log gives this value.
   - Choose a prediction model. If calibration converges for the full model at a large value of the tolerance parameter *epsilon*, balance between the number of variables and the tolerance level, dropping some variables (less important ones) from the prediction model may achieve the convergence at a much smaller epsilon (better precision).
   - Choose reasonable bounds: *L*=0 to prevent negative weights, U too small may cause convergence problem.

3. Further adjust calibrated weights
   - Use function ***postStratify*** to adjust calibrated weights so that the total of calibrated weights matchs the known population size.
   - Trim poststratified weights that fall outside the bounds (L_0, U_0) using function ***trimWeight***

References

Deville, J.C. and Sarndal, C.E. (1992). "Calibration estimators in survey sampling. *JASA*, 87, 376-382.

Kott, P. (2009), "Calibration weighting: combining probability samples and linear prediction models," in D. Pfeffermann and C. R. Rao (Eds.), *Handbook of Statistics*, *Sample Surveys: Design, Methods and Application,* 29B, Amsterdam: Elsevier BV.

Lumley, T. (2008), "Post-stratification and calibration," Presentation slides.

Rao, J.N.K. and Singh, A.C. (1997), "A ridge-shrinkage method for range-restricted weight calibration in survey sampling," *Proceedings of the Section on Survey Research Methods,* American Statistical Association, Washington, D.C., 57-65.

Särndal, C.E., Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer: Berlin, New York.

Singh, A.C., and Mohl, C.A. (1996), "Understanding Calibration Estimators in Survey Sampling." *Survey Methodology*, 22, 107-115.