Comparing Alternate Designs For A Multi-Domain Cluster Sample

Pedro J. Saavedra, Mareena McKinley Wright and Joseph P. Riley
Mareena McKinley Wright, ORC Macro, 11785 Beltsville Dr., Calverton, MD 20705

Abstract

Key words: PPS sampling, simulations, design effect, alternate designs

The Assisted Housing Income (Re)certification Study calls for a national sample of housing projects and tenants from three assisted housing programs. Unfortunately, different programs predominate in different localities. In the past a composite measure of size was used to select clusters, and allocations were made for each PSU so as to insure sampling the same total number of tenants for each program. For the current cycle, a two-phase sample was used to select projects from each program so that probabilities within programs remain proportional to size. While this method leads to a lower design effect due to weighting, the work load is slightly uneven across PSUs. An alternate design begins by selecting many samples without constraining the total number of projects selected for each program. Then the samples that yield the desired numbers are identified and one of them is chosen. The sample is used only to determine the number of projects to be selected from each cell (PSU/program). Two weighting schemes are considered for this method and are compared for the design effect due to weighting.

Introduction

This paper describes the examination of alternate designs for the sample of the *Quality Control for Rental Assistance Subsidies Study.* The sample provides nationally representative rent error rates for HUD's assisted housing programs, and is drawn on a yearly basis. Rent subsidy errors are estimated for the Public and Indian Housing (PIH)-administered Public Housing, PIH-administered Section 8, and Office of Housing owner-administered programs, separately and in combination. The sample consists of 2,400 tenants, from 60 PSUs and 600 projects, with 200 projects drawn from each of the three programs described above, and separate estimates required for each program.

The primary purpose of the *Quality Control for Rental Assistance Subsidies: FY2005 Study* is to provide national estimates of improper rent payments for subsidized housing in the United States. The data are also used to investigate the causes of such rent errors or improper payments, and try to understand how the rent errors are associated with tenant characteristics and the project staff administration of the housing programs and rent determination process. The universe includes all projects and tenants located in the continental United States, Alaska, Hawaii, and Puerto Rico.

The sample is designed to obtain a 95 percent likelihood that estimated aggregate national rent errors for all programs are within two percentage points of the true population rent calculation error, assuming an error of ten percent of the total rents (based on OMB criteria). Based on previous studies, it has been determined that a tenant sample size 2,400 would yield an acceptable precision for estimates of the total average error.

Two levels of clustering are used in this study, namely:

- Tenants clustered within projects
- Projects clustered within PSUs (generally groups of counties)

Optimizing costs and variances, four tenants per project, ten projects per cluster, and sixty clusters, for a total of 2,400 tenants, 800 for each program type would be the ideal size. This approach has created problems in past years. The difficulty lies in the fact that the distribution of programs by PSU is not uniform, and hence there is no measure of size that fits all three programs and no clear-cut way of allocating the units by program after selection of PSUs. In fact there are cases where a PSU may not have all three programs present, or where if it having only one or two projects in one of the three programs, those programs have had to be replaced, creating the choice between going to a different PSU or altering the allocations per PSU in other PSUs than the one where the problem existed.

One issue that has presented logistical difficulties has been that of unequal allocations per PSU. The ideal design would sample 200 projects (without replacement for Public Housing and Owner-Administered, with minimum replacement for Voucher). But the distribution by PSU should be such that every tenant would have approximately the same probability of selection.

Thus the design is confronted with the following desirable properties:

1) Exactly 200 projects in each program
2) Exactly 10 projects in each PSU
3) The same probability of selection for each tenant
4) Absence of bias
5) Sampling of projects without replacement in two of the programs and with minimum replacement in the third
6) PSUs sufficiently small geographically to permit data collectors to cover the PSU from one place
7) PSUs sufficiently large so that a small PHA or owner will not be burdened with the inclusion of too many projects in the sample.

1. Definition, Allocation and Sampling of Clusters for the 2005 Study

A sample of 60 PSUs was designed with ten projects per PSU and four tenants per project (allowing PSUs and projects to be selected more than once if sufficiently large). Size measures for PSUs and projects were inflated to add to the same amount. While sampling variance could have resulted in differences in the number of units sampled from each of the three programs, this precluded forcing the number of tenants from each of the programs to be the same. The design called for equal allocation of the three HUD programs: Public Housing, Voucher, and Owner-Administered projects. Public Housing and Owner-Administered projects were usually actual buildings, and there was no point in sampling more than four tenants per building.

The clustering procedure began by using counties as the initial cluster. A restriction was placed that a cluster had to include a minimum of 1,000 tenants, 30 projects and two PSUs. When a county did not meet the criterion, the clustering program identified the nearest county and merged the two. At total of 531 PSUs were created. The clustering program was effective, except that from time to time the resulting PSUs were unnecessarily large. This was resolved by a manual revision of PSUs after selection.

PSUs were selected with probabilities proportional to size (PPS), using the systematic PPS approach. However, the study called for an equal number of tenants to be selected from each of the three program types. In order to accomplish this, we selected PSUs with a size measure calculated as the average of the proportions of tenants from each of the three programs found in the PSU. The number of tenants in each program within a PSU was then divided by the number nationwide. The three values were averaged to create a measure of size that summed to one.

The size measure was then multiplied by 60—the number of PSUs to be selected—to obtain the expectation of selection for each PSU. If this expectation was less than one it was interpreted as the probability of selection of the PSU. If it was greater than one, the PSU was selected with certainty. The integer part of the expectation indicated the minimum number of times the PSU could be selected and the fractional part indicating the probability that the PSU be selected one additional time.

The PSUs were grouped within states and then within HUD-defined regions. States were sorted in a random order within regions, and PSUs were randomly sorted within states. As the frame was prepared for the selection of PSUs, PSUs were arranged in order and each was assigned expectation values. A random number was generated as a starting point to select the PSUs. A cumulative distribution of the expectations was calculated by adding the expectation of a PSU to the cumulative expectation of the previous one (starting with the random number). Thus the real numbers between 0 and 60 were divided into segments where each PSU is represented by the segment between the cumulative expectation of the previous PSU (or 0 for the first PSU) and its cumulative expectation. A random number x between 0 and 1 was selected, and the integers from 0 to 59 were added to the random number. The numbers x, 1+x, 2+x ... 59+x defined the selected PSUs and a PSU was selected as many times as one of these numbers falls into its corresponding segment.

This is essentially the Goodman-Kish approach (1950) but using sampling with minimal replacement (Chromy, 1979). This procedure results in sample sizes proportional to the number of tenants in each region. Rather than allocate a number of clusters to this region, this method implicitly stratifies the sample and essentially allows a fractional allocation. In other words, if the expectation for a region should be 4.6 PSUs, it would have a 40 percent chance of getting 4 and a 60 percent chance of getting 5.

In addition, once the PSUs were selected, the larger PSUs were divided and one of the parts was selected with PPS. The decision to divide or not was implemented subjectively, using a map to determine data collection burden. Once a division was made, one of the parts was selected with PPS using the same combined size measure used in selecting the PSUs.

2. Allocation and Sampling of Projects for the 2005 Study

Unlike previous years' second stage sampling, in 2005 the selection of projects was done as a second phase. This means that the PSUs were pooled and the projects selected from the combined set of counties. This was done so as to make the expected number of projects to be approximately ten per county, but the actual number was allowed to vary. It permitted sampling the three HUD programs independently, insuring both equal initial weights and the ability to replace a project with minimum disruption.

Let us begin with a description of the design for Owner-Administered projects. Let t_{ijk} be the number of tenants in PSU k, program j and project i. Let p_k be the probability of selection of PSU k and $s_{k.}$ be the number of times PSU k was selected. The size measure assigned to the project for the second phase was $q_{ijk} = t_{ijk}\, s_k / p_k$. Now the size measures are added for the whole program, and probability of selection for the project will be defined as $p_{ijk} = 200\, q_{ijk}/q_{.j.}$ where $q_{.j.}$ means the sum of the q_{ijk} for program j over all PSUs and projects. The actual formula used 201 as opposed to 200, as it was considered desirable to have an extra project in the event that a project proved to present difficulties at the last minute, but this refinement will not be considered in this study.

Note that the p_{ijk} could be greater than one. This did not happen with Owner-Administered projects, but it did happen with some of the larger Public Housing projects and in Voucher projects. These two programs were treated differently in that respect.

For Public Housing projects, the probabilities greater than one were set to one, the units were removed from the list, the number 200 (representing the number of projects to be sampled) was reduced accordingly and the process was repeated. Additional iterations were used until no units had p values greater than one.

For Voucher projects, the process was closer to that of the selection of PSUs. The sampling was done with minimal replacement, and the probabilities were treated as expectations, allowing them to be selected more than once.

3. Difficulties with the 2005 Method and Alternatives Used in Previous Studies

The major difficulty with the 2005 method is that the number of projects selected in each PSU varied. It did have the advantage that three independent samples were drawn and the design effect due to weighting was small using unconditional weights. It presented logistic problems, however, as data collectors are hired with the intent that each have an equal work load but this approach did not permit controlling for the number in each PSU. As a result, this method was not a favorite of the data collection team.

In previous years' studies, alternate approaches were used with greater or lesser success. In every case the selection of PSUs was made using PPS with minimum replacement and using a composite measure of size (the average of the proportion of tenants in each of the three programs). Clusters were larger in some years (resulting in larger travel budgets) and smaller in other years (resulting in excessive burdens on some Public Housing Authorities).

In early iterations of the study, a simpler design was used. It simply inflated the sizes of the two smaller programs so that the expected sample would represent the three programs approximately equally. A second re-adjustment of the sizes then took place at the second stage, but even this would not guarantee the equal representation of the three programs.

Another approach first calculated the probabilities needed for a sample of 200. The number of expected hits in each PSU were then computed and adjusted to add to 10 in each case (or a multiple of 10 if the PSU was selected more than once). Next, randomized rounding was used at the PSU level, forcing the PSU totals to add to 10 or a multiple of 10. The process did not guarantee that the sum of the allocations for each program would add to 200, but it was repeated until it did. The program/PSU combination became a cell for the second stage, and the probability of selection was based on using the rounded number as an allocation.

When this approach was first used, two other problematic features occurred, which were changed in subsequent iterations that used the approach. First, the PSUs were smaller. Second, minimum replacement was required for all three programs for at least one of the years when it was in place, resulting in excessive burden. Replacements also created a problem because if a project turned out to be out-of-scope, the same cell might not have a replacement.

An approach considered but never implemented was to select two of the programs independently and allocate the difference between ten and the sum of the two to the third. Unfortunately, as will be seen, there is no guarantee that the units sampled for the first two programs by themselves will not exceed ten units by themselves.

4. Simulations of Alternate Approaches

In order to explore possibilities for obtaining a sample design having the desirable properties listed above, several simulations were conducted. At this point it is not possible to evaluate the alternative approaches, for two reasons. In the first place, we do not have frame values that could be used to simulate the efficiency and degree of bias of the sampling methodology. Second, the advantages and disadvantages of each approach are not commensurate. How much potential bias is justified to obtain an equal number of projects in each PSU? The objective of these simulations is merely to obtain a better picture of the alternatives available.

All the simulations used the same frame and clustering structure used in the 2005 study. An initial size measure was the average of the proportions of each of the three programs found in the PSU. The Chromy sampling algorithm was used to select 60 PSUs and 1,000 sets of PSUs were selected.

The first set of simulations used essentially the same sampling approach as used in the 2005 study. Each project was given the size measure equal to the number of tenants times the number of times the PSU was selected divided by the probability of selection of the PSU. The allocation was made at the program level (200 per program). These sizes yielded a probability of selection for the PSU, and this was adjusted to be no greater than one for the Public Housing and Owner-Administered projects. The projects were then sampled by region, PSU and county within PSU. A total of 1,000 selections (one for each PSU sample selected at first stage) were drawn, using first the Chromy sampling approach and then the systematic PPS approach in PROC SURVEYSELECT in SAS.

With this sampling approach, the variation of total projects selected per PSU was much larger using the Chromy procedure than the systematic PPS procedure. Therefore, all subsequent analysis used the systematic PPS procedure at the second stage. The disappointment in this approach was that in every single one of the 1,000 simulations with either method, there were PSUs where two of the three programs added to more than 10 units (20 if the PSU was sampled twice). Thus, the approach of selecting only two programs through this method and assigning the right number to the third program will not be feasible in general.

The second set of simulations calculated the probabilities the same way, but this time controlled first by PSU and then by program. In this case all the PSUs had exactly 10 projects (20 for those sampled twice) but only 28 out of 1000 samples yielded exactly 200 projects. A single PSU structure was selected and 1,000 samples with just that set of PSUs was selected using the same approach. The result was that 49 of 1,000 samples yielded exactly 200 for each program.

The temptation here is to use some sort of rejective sampling. In other words we select a large number of samples and then choose one which meets the desired characteristics, In such a case, however, the probabilities of selection would not hold. However, these are unconditional probabilities of selection. They reflect the probabilities of selection from the beginning of the sampling process. It is entirely possible to obtain probabilities of selections and weights at the second stage that are conditional on the results of the first stage.

The design in this instance would work as follows. The PSUs are selected as in the current study, using either PPS systematic or PPS sequential algorithms. At the second stage the projects are selected controlling for PSU, thus guaranteeing ten projects for each PSU (twenty if selected twice) using the second approach described above. One thousand samples will be drawn, and the samples that include 200 for each program will be separated. One will be selected (and this need not be a random choice, other characteristics of the sample may be taken into account).

Now the program/PSU combinations may be defined as strata, and the number of hits in those cells will become allocations for those cells. The second stage sample may now be defined as a stratified sample with fixed allocations for each cell. The origin of those allocations need not enter into the weighting. A new sample would now be drawn using the allocations resulting from the arbitrarily selected sample, and this will be a probability sample.

It may be argued that using this approach it is possible that there could be a small cell with zero allocations, and this would create a bias. Here we can return to the practice in the current sample of drawing at least one additional project to cover any cases discovered to be unusable at the last minute. Of the 49 samples with exactly 200 projects per program, seven had exactly one program/PSU cell with no sampled cases. One such sample could be selected and an extra project assigned to the zero allocation cell.

5. Evaluation of Effective Sample Size Due to Weighting and Practical Considerations

In theory one wishes to select a sample with equal probability for every unit. In this instance the equal probability objective applies to the individual programs rather than the sample as a whole. Given the minimum replacement procedure for the selection of PSUs there are two ways in which one can assign weights (Saavedra, 2005). One can calculate the probability of selection of each unit from the start, and one can calculate it at each stage starting from the results of the previous stage. The first method has a lower design effect due to weighting, but in cases where the intra-class correlation is higher the same method can be more efficient. In particular, if one has to make some logistically required decisions at the second stage, the second approach may be necessary.

The second approach is almost a requirement for some of the various designs discussed here, so it will be used uniformly. Unequal weights will be the result of minimum replacement at stages 1 and 2 (for Voucher projects). It will be the result of adjusting for the number of tenants in the PSUs selected at stage 1. And it will be the result of using integer allocations at the cell level with the final method.

Finally, there is a way of smoothing the ways so as to reduce the design effect due to weighting. One can first do 10,000 simulations and then separate those with the desired total units per program. One simulation can be randomly selected. But the probability of selection of the units would be based on the average probability across the sets of allocations that yield the 200 projects per program. This approach yielded a significantly higher effective sample size than using a single allocation to establish the probability of selection.

Effective sample size due to weighting can be obtained by dividing the sum of the squares of the weights into the square of the sum of the weights.

Domain	Controlling for Program	Controlling for PSU	Controlling for cell (single allocation)	Controlling for cell (allocation average)
Vouchers	769	764	742	749
Public Housing	786	762	759	774
Owner-Administered	794	791	772	786

The results should not be surprising. The more one controls, the more one sacrifices equal weights and the smaller the effective sample size. The situation is then as follows. If one controls for PSU one may have fewer than exactly 800 tenants (4 from each of 200 projects) per program, but the effective sample size will still be larger than if one controlled by cell and guaranteed the 800 tenants per program.

Controlling for cell has one additional drawback. The design calls for replacing a project which is found to have closed, dropped out of the program, or be participating in an experimental program. Controlling for cell makes the replacement problem more problematic due to the presence of small cells or cells consisting of a PHA that has joined an experimental program.

Naturally, controlling for program (the method used in the 2005 study) seems to be the method of choice from the perspective of effective sample size as well as from the perspective of replacement. However, it is the most problematic approach from the perspective of data collection management and cost.

References

Chromy JR. Sequential sample selection methods. In *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp 401-406, 1979.

Saavedra, P.J. (2205) Comparison of Two Weighting Schemes for Sampling with Minimal Replacement. Presented at the Joint Statistical Meetings, Minneapolis, MN.

Goodman R. and Kish, L. (1950) Controlled Selection - a Technique in Probability Sampling J. Americ. Statist. Assoc. 45, 350-372.