Survey Methodology, June 1998
Vol. 24, No. 1, pp. 3-9
Statistics Canada

Sampling and Estimation From Multiple List Frames

PHILLIP S. KOTT, JOHN F. AMRHEIN and SUSAN D. HICKS[1]

ABSTRACT

Many economic and agricultural surveys are multi-purpose. It would be convenient if one could stratify the target population of such a survey in a number of different ways to satisfy a number of different purposes and then combine the samples for enumeration. We explore four different sampling methods that select similar samples across all stratifications thereby reducing the overall sample size. Data from an agriculture survey is used to evaluate the effectiveness of these alternative sampling strategies. We then show how a calibration (*i.e.*, reweighted) estimator can increase statistical efficiency by capturing what is known about the original stratum sizes in the estimation. Raking, which has been suggested in the literature for this purpose, is simply one method of calibration.

KEY WORDS: Calibration; Collocated sampling; Permanent random numbers; Poisson sampling; Systematic probability proportional to size sampling.

1. INTRODUCTION

Many of the list frame surveys conducted by the National Agricultural Statistics Service (NASS) are integrated in the sense that data on a range of heterogenous items, such as planted crop acres and grain stock inventories, are collected in a single survey rather than through a number of independent surveys. Bankier (1986), Skinner (1991), and Skinner, Holmes and Holt (1994) have shown how an old method of combining independently drawn stratified simple random samples – where each sample comes from a (list) frame with a different stratification scheme – can be made more efficient; that is, the variances resulting from such a combined estimation strategy would not be as large as those from the independent surveys summarized by themselves.

Even more appealing for many applications would be a sampling design that tends to select the same units from every frame, thereby reducing both the cost and respondent burden of an integrated survey. This paper explores several such designs. Three make use of permanent random numbers. The fourth uses a variation of systematic probability proportional to size sampling. The goal for each is to meet or exceed – at least on average – a particular set of sample size targets.

The paper shows how a calibration (*i.e.*, reweighted) estimator can provide relative efficiency by capturing what we know about the original stratum sizes in the estimation. A final section points out that the use of a calibration technique can do more than simply reflect original stratum sizes.

An alternative strategy for burden reduction is to use separate instruments for different survey targets and to select distinct samples for each instrument. This increases the number of units selected over all, but reduces the burden per selected unit. NASS is using that approach in its Agricultural Resources Management Study (see Kott and Fetter 1997), but it is *not* the approach to be discussed here.

2. INDEPENDENT SAMPLING AND UNBIASED ESTIMATION

Suppose we have F independent frames; for example, a sorghum frame, an oats frame, and a general grain stocks frame. Each frame is stratified independently, and without replacement simple random samples are drawn from each stratum of every frame. Frame f (say, the oats frame) contains H_f strata; stratum h (large oats operations) in frame f has N_{fh} population units, out of which n_{fh} units are selected. The union of the F frames must cover the entire (list) population, but no single frame need be complete. The frames may overlap.

One unbiased estimator for a population total $T = \sum_{i \in P} y_i$ is the simple multiplicity estimator suggested by Skinner (1991):

$$t_M = \sum_{i \in P} y_i n_{(i)} / E[n_{(i)}], \qquad (1)$$

here P denotes the entire population, and $n_{(i)}$ is the number of times unit i is selected for the sample from any frame. Observe that $n_{(i)} = 0$ for the population units not in the sample. In the great majority of applications, $n_{(i)}$ will be one for most sampled units, but $n_{(i)} > 1$ is a possibility with this design.

The expected number of times unit i will be selected for the sample is $E[n_{(i)}] = \sum^F p_{if}$, where p_{if} is the probability of selecting unit i in the stratified simple random sample from frame F; that is, $p_{if} = n_{fh}/N_{fh}$, where unit i is in stratum h of frame f.

There is also a Horvitz-Thompson estimator for T under the design, namely $t_{HT} = \sum_{i \in S} y_i / \pi_i$, where S denotes the sample and $\pi_i = 1 - (1 - p_{i1})(1 - p_{i2}) \cdots (1 - p_{iF})$. See Bankier (1986) for further discussion of this approach.

[1] Phillip S. Kott, Research Division; John F. Amrhein, Survey Sampling Branch; and Susan D. Hicks, Estimates Division, National Agricultural Statistics Service, USDA.

3. SAMPLING STRATEGIES USING PERMANENT RANDOM NUMBERS

The sampling design discussed above is independent across frames. For many surveys, however, it would be convenient if the design were *not* independent across frames. This is because all units in the combined sample are given the same survey instrument, and many units are in a number of frames. Therefore, given frame/stratum sample-size targets, a design with a tendency towards selecting the same unit in every frame should result in a smaller overall number of contacts (and consequently survey costs) than independent sampling across frames.

To this end, suppose each unit has been given a target p_{if} in each frame to meet or exceed. This target value is constant for all units in stratum h of frame f. We will withhold judgement on the policy of focussing on target p_{if} values – or equivalently on target n_{if} values – until the concluding section. Suffice it to say that many statistical agencies, including NASS, have such a policy.

One potential sampling design assigns each unit in the population a *permanent random number* (PRN) drawn from the uniform distribution on the interval [0, 1). Unit i is selected for the frame f sample when its PRN is less than p_{if}.

The result is a Poisson sample where the probability of selecting unit i for the sample is $\pi_i = \max_f\{p_{if}\}$, which is clearly at least as large as each individual p_{if} for a given unit. An unbiased Horvitz-Thompson estimator for T under this design is $t_p = \sum_{i \in S} y_i / \max_f\{p_{if}\}$.

Under Poisson sampling, sample size is random. One way to reduce the variance of the sample size is with a variant of this sample design. In *collocated* PRN sampling, each population unit is assigned a unique PRN from among the members of the set $\{e/N, (1+e)/N, (2+e)/N, ..., (N-1+e)/N\}$, where e is a uniform random variable drawn from the interval [0, 1). To this end, one can first draw provisional PRN's for each unit followed by a value for e. The unit with the smallest provisional PRN is assigned a collocated PRN of e/N, the units with the second smallest provisional PRN is assigned $(1+e)/N$, and so on until $(N-1+e)/N$ is assigned to the unit with the largest provisional PRN. The estimator t_p remains unbiased under collocated sampling.

Due to random nature of the sample sizes resulting from Poisson and collocated sampling, frame/stratum sample size targets may not be met when a particular sample is drawn. A third PRN design begins with target n_{fh} values and removes this possibility. In this design, the units in stratum h of frame f with the n_{fh} smallest PRN's are selected for the sample (this is very similar to sequential Poisson sampling in Ohlsson 1995). A Horvitz-Thompson estimator under this *fixed-sample-size* PRN design requires one to compute the selection probabilities of the sampled units – a difficult task which may have to be approximated by simulation.

4. A SYSTEMATIC PROBABILITY PROPORTIONAL TO SIZE DESIGN

Another sampling design with the same selection probabilities as the Poisson (and collocated) sampling scheme described in the last section consists of the following steps:

0) When necessary, create an additional "stratum" for each frame consisting of those units not in any design stratum.

1) Divide up the population into mutually exclusive cells by cross-classifying the strata from the various frames. A pair of units in a particular cell will then be in the same stratum of each frame (*e.g.*, the large oats stratum, the medium grain stocks stratum, and the no sorghum stratum).

2) Randomly order the population units in each cell and then sort the cells themselves in any order. This results in a list of all population units.

3) Draw a systematic probability proportional to "size" (PPS) sample from this list using the π_i described in the discussion of Poisson sampling as the measures of size (the word "size" is in quotes because the π_i are not really size measures in a conventional sense). This ensures that a unit's selection probability equals π_i.

The systematic PPS sampling design introduced above will always result in a sample of size close to $\sum_{i \in P} \pi_i$. In fact, if $\sum_{i \in P} \pi_i$ is an integer, then the sample size will exactly equal that sum. Otherwise, the sample size will be one of the two integers closest to $\sum_{i \in P} \pi_i$. Similarly, the expected number of sampled units in a cell, C, will be $\sum_{i \in C} \pi_i$, while the actual sample size will either be $\sum_{i \in C} \pi_i$ or one of the two integers closest to it.

Consider now a particular stratum h in a particular frame f with target sample size n_{fh}. For a unit i in this stratum, $\pi_i \geq n_{fh}/N_{fh}$ by design. Let $P(fh)$ denote the set of population units in stratum fh. The expected number of sampled units in fh is $\sum_{i \in P(fh)} \pi_i \geq n_{fh}$. There is no *guarantee* that the realized sample size in the stratum will be greater than or equal to n_{fh}. Nevertheless, given the above inequality and the lower bounds on the sample sizes of the cells within fh, the sample size in stratum fh will never be far below n_{fh}.

The advantages of this design over Poisson and collocated sampling is that it produces a more stable size and a greater likelihood of meeting frame/stratum requirements. Fixed-sample-size PRN, by contrast, will always meet frame/stratum requirements, but it does so at a cost: the design has a less stable overall sample size, and selection probabilities can be very difficult to determine.

5. EVALUATION OF THE ALTERNATIVE SAMPLING TECHNIQUES

To evaluate these sampling techniques empirically, we selected three states that conduct NASS's Vegetable Chemical Use Survey and replicated the three PRN techniques, the systematic PPS method, and independent sampling across frames 100 times. The assigned PRN's were maintained across the three PRN techniques within each replicate. A separate frame was constructed for each commodity of interest within a state (the number of frames ranged from two in Minnesota to 23 in California). Population units were allocated to one of four strata in each frame; two probability strata, one take-all stratum, and one zero stratum were used in each frame. Stratum boundaries were determined using a modified Lavallée and Hidiroglou (1988) method, and units were assigned to strata based on a $\text{cum}^3\sqrt{f(x)}$ rule (Sweet and Sigman 1995). This stratification was chosen to mimic what might be a reasonable or reasonably common univariate sample design.

A target sample size of one-third the population was selected from each of the probability strata. Table 1 compares the overall sample sizes realized from each of the sampling techniques. As expected, the independent frame approach realized the largest sample sizes. The three PRN techniques realized sample sizes of similar size with the Poisson method experiencing the highest standard deviations in each of 3 trials (states). The PPS method appears to be the most stable.

Table 1
Mean realized sample sizes (over 100 replications)

State	Independent Frame Method	Fixed Sample Size Method	Poisson PRN Method	Collocated PRN Method	Systematic PPS Method
CA	496 (8.8)	388 (9.6)	375 (11.1)	374 (5.6)	373 (.14)
MI	658 (9.3)	513 (9.2)	504 (13.6)	501 (6.0)	502 (.48)
NJ	563 (8.1)	359 (8.6)	343 (13.8)	344 (4.6)	343 (.17)

Population sizes are: CA-775; MI-1041; NJ-785.
Standard deviations are in parentheses.

Table 2 shows the percentage of strata-level Poisson and PPS samples that fell short of their target sample sizes. One reason more shortfalls were not observed in the Poisson methods' realized sample sizes is the occurrence of what we call "visitors". A visitor is a sample unit that was not chosen within a specific commodity's frame, but ends up in the sample because it was selected in another commodity's frame. The existence of visitors tend to cause frame-level sample sizes to be larger, on average, than the targeted sizes.

Figure 1 shows cumulative distributions of differences between realized and desired sample sizes as percents of the desired sample sizes for the sampled strata. That is, the cumulative distribution of (realized – desired)/desired at the probability stratum level. For example, Michigan had 13 commodity frames each with two probability strata. Sampling from these frames was replicated 100 times so that the cumulative distribution function (CDF) for each technique utilized 2600 points. The two Poisson methods are shown as a single line since they coincide. The Poisson methods do not over-sample as much as the fixed-sample-size and independent frame methods, but at the risk of under-sampling as we saw in Table 2. The fixed-sample-size techniques (with dependent and independent frames) do not experience under-sampling, but do experience more over-sampling than the Poisson and PPS methods. The PPS method experiences some under-sampling but not to the extent of the Poisson methods. The PPS design also shows the steepest gradient of all the CDF's, indicating that it realizes less over-sampling.

Table 2
Percentage of probability strata for which the realized sample size fell short of the target (in 100 replications)

State	Poisson PRN Method	Collocated PRN Method	Systematic PPS Method
CA	11%	11%	6.3%
MI	12%	12%	6.3%
NJ	11%	8%	1.4%

Under the Poisson and collocated techniques, the probability of selection for unit i is $\pi_i = \max_f(p_{fh})$ where h corresponds to the stratum in which i belongs for frame f. The same probability of selection is used for the PPS technique. By contrast, the probabilities of selection under the fixed-sample-size PRN method are difficult to determine and may need to be simulated.

Such a simulation was conducted using the California data. The fixed-sample-size technique was run 10,000 times. Since all probability strata were sampled at a rate of 1/3, the simulated probabilities (*i.e.*, relative frequencies) can be compared to 1/3. The mean simulated probabilities of selection over the 10,000 trials are shown in Figure 2 as a function of the number of frames in which the unit is contained within a probability stratum. There were 19 commodities of interest in this state, but no units existed in probability strata in exactly 16 or 19 frames. A unit's probability of selection tends to increase with the number probability strata containing it. This selection probability is 1/3 only when the unit is in exactly one such stratum.

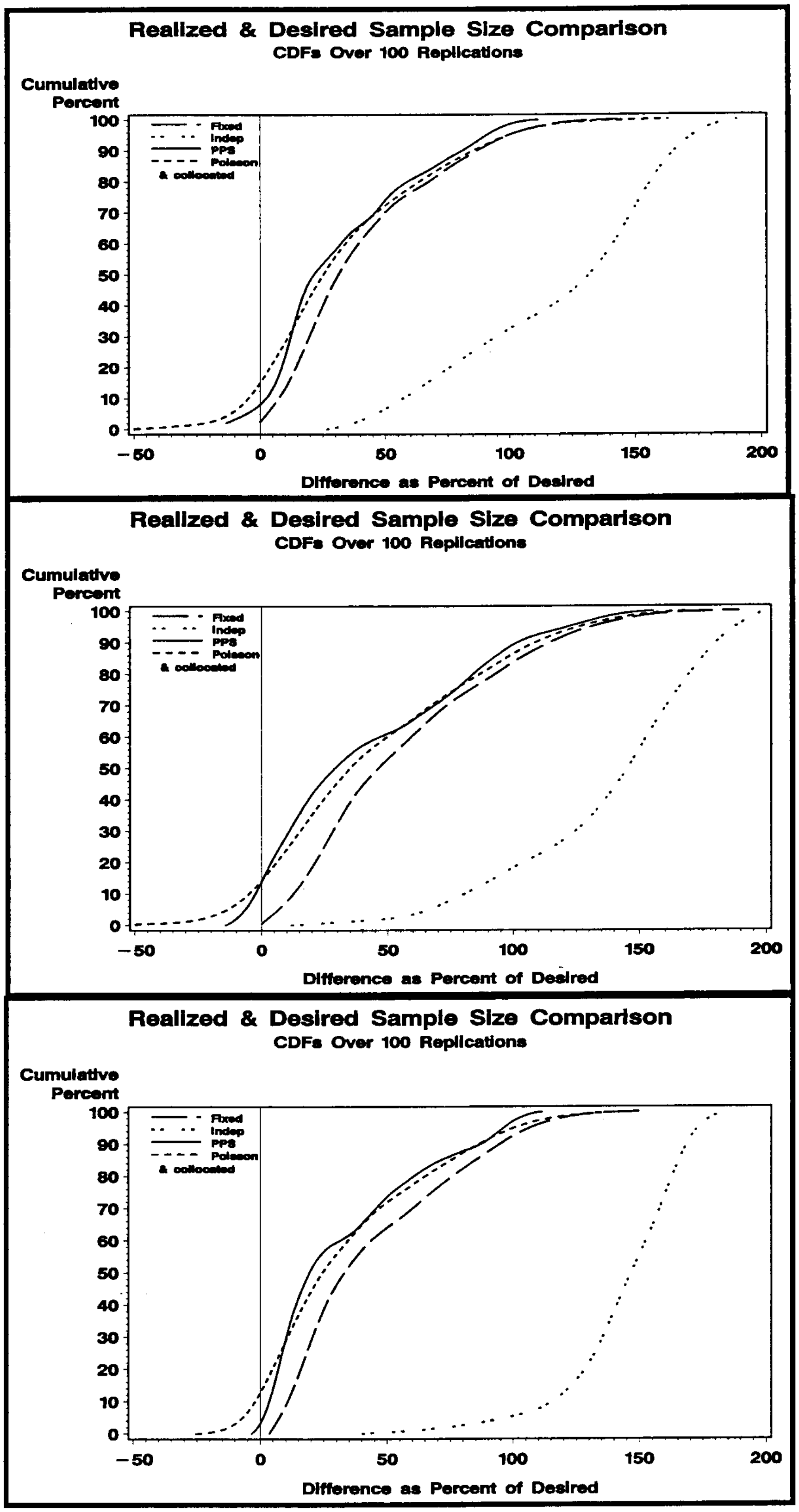

Figure 1. Comparison of realized and desried sample sizes for sampled strata. Top - MI; middle - CA; bottom - NJ.

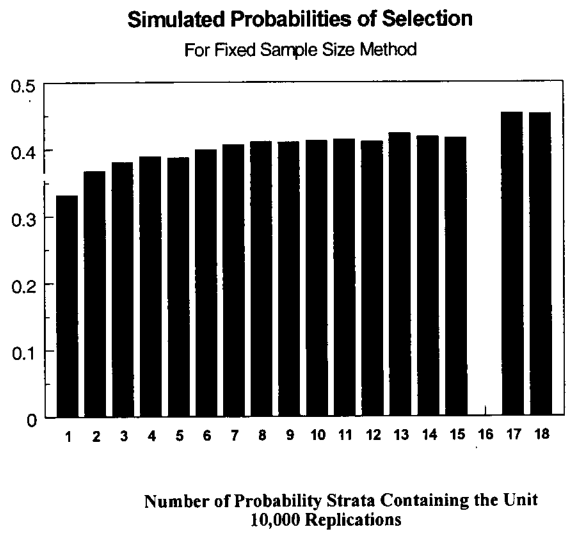

Figure 2. Simulated probabilities of selection for the fixed-sample-size method-California

6. CALIBRATION

The problem with both t_M and t_P (or t_{HT}) is that they are often not very good estimators for T in term of precision (variance). One of the properties of single-frame, stratified simple random sampling is that the conventional expansion estimator estimates the stratum population size perfectly (*i.e.*, with zero variance). In our multiple frame set up, however, neither t_M nor t_P will estimate the N_{fh} perfectly in most applications.

Let us define $w_i^0 = n_{(i)}/E[n_{(i)}]$ as the *original sampling weight* of unit i in t_M. Similarly, $w_i^0 = 1/\max_f \{p_{if}\}$ in t_P and $1/\pi_i$ more generally for a Horvitz-Thompson estimator. Bankier (1986) proposed raking to create a set of adjusted weights such that

$$\sum_{i \in S_{fh}} w_i^C = N_{fh} \tag{2}$$

for each stratum h in every frame f, where S_{fh} is that part of the sample that is in stratum h of frame f regardless of the frame(s) from which the units were selected.

Deville and Särndal (1992) call (2) a *calibration equation*. They point out that there are a number of ways to compute the *calibration weights*, the w_i^C, so that equation (2) is satisfied and w_i^C/w_i^0 is in some sense close to 1 for all i. One method is raking as suggested by Bankier (1986). Another method, discussed at length by Deville and Särndal (1992), uses least squares. Either way, the resulting estimator

$$t_C = \sum\nolimits_{i \in S} w_i^C y_i,$$

where S denotes the entire sample, will be nearly design unbiased because w_i^C/w_i^0 is close to 1 for all i.

The estimator t_C is also unbiased under the model:

$$y_i = \beta_0 + \sum_{f=1}^{F} \sum_{h=2}^{H_f} d_{ifh} \beta_{fh} + \epsilon_i, \tag{3}$$

where the dummy variable, d_{ifh}, is 1 when unit i is in stratum h of frame f (sampled or not) and zero otherwise, while ϵ_i is a random variable with a mean of zero. The β_0 and the β_{fh} are unknown constants (β_0 represents the mean y-value for a unit in the first stratum of every frame; that is why the second sum excludes $h = 1$). The same d_{ifh} values apply to every survey item (y) of interest, while the β values change with the survey item. For many survey items, β_{fh} values will be zero when frame f (say, grain stocks) is irrelevant to the item (say, planted oat acres).

Isaki and Fuller (1982) call the model expectation of the design mean squared error of t_C the "anticipated mean squared error" of the estimator. This value is of most use at the planning stage of a sample survey.

If the model in equation (3) holds, and the ϵ_i are uncorrelated, then the anticipated mean squared error of t_C is

$$\begin{aligned}
E_\epsilon[\mathrm{MSE}_D(t_C)] &= E_\epsilon\{E_D[\textstyle\sum_s w_i^C y_i - \sum_P y_i)^2]\} \\
&= E_D\{E_\epsilon[(\textstyle\sum_s w_i^C y_i - \sum_P y_i)^2]\} \\
&= E_D\{E_\epsilon[(\textstyle\sum_s w_i^C \epsilon_i - \sum_P \epsilon_i)^2]\} \\
&= E_D\{\textstyle\sum_s [(w_i^C)^2 - 2w_i^C] E_\epsilon(\epsilon_i^2)\} + \sum_P E_\epsilon(\epsilon_i^2) \\
&\approx E_D\{\textstyle\sum_s [(1/\pi_i)^2 - 2/\pi_i] E_\epsilon(\epsilon_i^2)\} + \sum_P E_\epsilon(\epsilon_i^2) \\
&= \textstyle\sum_P (1/\pi_i - 1) E_\epsilon(\epsilon_i^2),
\end{aligned} \tag{4}$$

since $w_i^C \approx 1/\pi_i$. It is of some interest to note that using Poisson, collocated, and systematic PPS sampling result in estimators with approximately equal anticipated mean squared errors asymptotically. This surprising result is in part due to the nature of a calibrated estimator, but it is also a repercussion of the fact that when we take the design expectation of the approximate model variance in the last line of equation (4), we average over all possible samples and remove the biggest source of variation among the three sampling designs.

Now suppose we had used stratified simple random sampling and selected unit i with probability $p_{if} \le \pi_i$, where f is the frame relevant to y. It is not hard to show that the anticipated variance of the simple expansion estimator would have been $\sum_P (1/p_{if} - 1) E_\epsilon(\epsilon_i^2)$, which is at least as large as the right hand side of equation (4). Thus, there are gains – in large samples, at least – from "integrating" the samples from various frames as we have effectively done. How large the samples must be in practice for the asymptotic results to be relevant is unclear. At the very least, the sample size must be many times the number of model parameters in equation (3).

A few words on mean squared error estimation for t_C are in order. The mean squared error estimator advocated by Deville and Särndal (1992) – an estimator with both good design and model-based properties – can not be implemented unless the joint selection probability (π_{ij}) for every pair of sample units (i and j) is known. Among the designs we have discussed, these probabilities are easily calculated only for the Poisson variant of PRN (where $\pi_{ij} = \pi_i\pi_j$).

As we have observed in equation (4), the anticipated mean squared error of the calibration estimator is the same under Poisson PRN, collocated PRN, and systematic PPS sampling. This suggests that the Poisson mean squared error estimator may be reasonable under each of the three designs. A stronger model-driven argument exists for this contention, but will not be made here.

7. DISCUSSION

In the last section, it was pointed out that if calibration weights were designed to satisfy equation (2), the resulting estimator would be unbiased under the model in equation (3). In many applications, there may be a more appropriate model on which to base calibration than the one in equation (3). For example, if there was a continuous control variable used to stratify a particular frame, it makes more sense to use that variable directly in the model rather than indirectly through frame/stratum identifiers.

Raking is a form of calibration under a particular model. With that in mind, it makes sense to use the most reasonable model available. Least squares has the advantage over raking that it can easily be applied to continuous control variables. Singh and Mohl (1996) provide an extensive review of alternative calibration algorithms including an extension of raking to continuous variables. An intriguing least-squares variant missed by Singh and Mohl (1996) can be found in Brewer (1994).

Many economic and agricultural surveys employ rotating sample designs. This has proved an effective way to balance cost and burden considerations. Although our empirical findings demonstrated an advantage of the systematic PPS methodology in terms of meeting target sample sizes, the three PRN designs are much more conducive to sample rotation. See, for example, Ohlsson (1995) on this topic. Moreover, with the PRN methods, one can integrate different frames at different times of the year (with systematic PPS there is no easy way to allocate the sample back to the frame of origin). This is a particularly useful property for agricultural surveys because different crops have different growing seasons.

In summary, the fixed-sample-size PRN sample design is excellent for meeting target sample sizes but is hard to use in practice because selection probabilities are usually unknown and must be simulated. The systematic PPS design is very good at meeting target sample sizes but is difficult to incorporate into a sample rotation scheme. Moreover, mean squared error estimation requires invocation of model assumptions. Our empirical example shows that collocated sampling may only be slightly better than Poisson at meeting target sample sizes. It should be recognized, however, that other configurations of the frames, strata, and sampling fractions may produce different results. Moreover, collocated sampling is conducive to rotation schemes, like Poisson sampling. On the other hand, like PPS sampling, it requires the assumption of a model to estimate mean squared error.

Finally, setting p_{if} or n_{if} targets is a popular, but indirect, means of controlling the variance of the estimator t_C associated with each frame. These targets lead to our *ad hoc* decision to set π_i equal to $\max_f\{p_{if}\}$. A more direct strategy would be to set (asymptotic) anticipated variance targets for each frame estimator using equation (4) and postulated values for the $E_\epsilon(\epsilon_i^2)$. One could then choose, say, the set of π_i that minimizes the expected sample size yet satisfy these variance targets. A similar approach is taken by Amrhein, Fleming, and Bailey (1997) who use Chromy's algorithm in a manner analogous to Sigman and Monsour (1995). Poisson PRN, collocated PRN, and systematic PPS sampling remain three viable alternatives for selecting the sample once optimal π_i are determined.

REFERENCES

AMRHEIN, J.F., FLEMING, C.M., and BAILEY, J.T. (1997). Determining the probabilities of selection in a multivariate probability proportional to size sample design. In *Proceedings: Symposium 97: New Directions in Surveys and Censuses.* Statistics Canada. To appear.

BANKIER, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. *Journal of the American Statistical Association*, 81, 1074-1079.

BREWER, K.R.W. (1994). Survey sampling inference: some past perspectives and present prospects. *Pakistan Journal of Statistics,* 10(1)A, 213-233.

DEVILLE, J-C., and SÄRNDAL, C.-E. (1992). Calibration estimator in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

ISAKI, C.T., and FULLER, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

KOTT, P.S., and FETTER, M.J. (1997). A multi-phase sample design to co-ordinate surveys and limit response burden. *Proceedings of the Section on Survey Research Methods, American Statistical Association.* To appear.

LAVALLÉE, P., and HIDIROGLOU, M. (1988). On the Stratification of Skewed Populations. *Survey Methodology*, 14, 33-43.

OHLSSON, E. (1995). Coordination of samples using permanent random numbers. In *Business Survey Methods* (Eds. B.G. Cox, D.A. Binder, N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott). New York: Wiley, 153-169.

SINGH, A.C., and MOHL, C.A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, 22, 107-115.

SIGMAN, R.S., and MONSOUR, N.J. (1995). Selecting samples from list frames of businesses. In *Business Survey Methods* (Eds. B.G. Cox, D.A. Binder, N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott). New York: Wiley, 133-152.

SKINNER, C.J. (1991). On the efficiency of raking ratio estimation for multiple frame surveys. *Journal of the American Statistical Association*, 86, 779-784.

SKINNER, C.J., HOLMES, D.J., and HOLT, D. (1994). Multiple frame sampling for multivariate stratification. *International Statistical Review*, 62, 3, 333-347.

SWEET, E., and SIGMAN, R.S. (1995). *User Guide for the Generalized* SAS *Univariate Stratification Program*, Economical Statistical Methods and Programming Division, Bureau of the Census, U.S. Department of Commerce, Report number ESM-9504.