

# Markov Chain Designs for One-Per-Stratum Sampling

F. JAY BREIDT<sup>1</sup>

## ABSTRACT

Classical results in finite population sampling tell us that systematic sampling is the most efficient equal-probability one-per-stratum design for certain kinds of autocorrelated superpopulations, but stratified simple random sampling may be much better than systematic sampling if the superpopulation is a trend with uncorrelated errors. What if the superpopulation consists of a trend plus autocorrelated errors? Intuitively, some sort of “compromise” between the two designs might be better than either. Such compromise designs are constructed in this paper and are shown to be examples of Markov chain designs, a wide class of methods for one-per-stratum selection from a finite population. These designs include as special cases systematic sampling, balanced systematic sampling and stratified simple random sampling with one sampling unit per stratum. First and second-order inclusion probabilities are derived for Markov chain designs, yielding the Horvitz-Thompson estimator and its variance. Efficiency of the Horvitz-Thompson estimator is evaluated using superpopulation models. Numerical examples show that new designs considered here can be more efficient than standard designs for superpopulations consisting of trend plus autocorrelated errors. An example of the implementation of Markov chain designs for the 1992 National Resources Inventory in Alaska is given.

KEY WORDS: Balanced systematic sampling; National Resources Inventory; Systematic sampling.

## 1. INTRODUCTION

A stratified sampling design, in which a finite population is divided into non-overlapping strata and samples are drawn from each stratum, is a common and effective technique for reducing sampling error. In practice, stratified sampling designs with only one sampling unit per stratum are widely used. Examples include stratified simple random sampling and systematic sampling with its variants (*e.g.*, Murthy and Rao 1988).

Systematic samples are susceptible to systematic errors. In large-scale spatial samples, for example, sources of systematic error could include roads, powerlines, irrigation systems, and so forth. A favorite example is the system of “section roads” in areas of the United States covered by the public land survey. This grid-based system is built up from square tracts of land called sections, each one mile on a side, which are often bounded by roads in midwestern agricultural regions. A systematic sampler with a one-mile sampling interval and an unlucky random start might conclude that Iowa is covered by gravel roads!

Systematic sampling does have the advantage of efficiency when the sampled population is positively autocorrelated, as is often the case in temporal and spatial sampling problems, since it forces observations to be as distant and hence as uncorrelated as possible.

Both autocorrelation and systematic error are of concern in the National Resources Inventory (NRI), an area sample of the nonfederal lands in the United States conducted every five years by the Soil Conservation Service of the

United States Department of Agriculture. NRI data items, collected by a combination of remote sensing and ground observation, include soil characteristics, land use, agricultural practices, erosion measures, and so on.

The 1992 NRI sample design for the northwestern region of the state of Alaska is a controlled version of one-per-stratum sampling. The region was divided into twenty-minute bands of latitude. Each band was divided into 500,000-acre strata. Each stratum was divided into a  $10 \times 10$  grid of cells indexed by latitude and longitude, and one cell per stratum was selected. Selection moved from east to west across the strata within a particular twenty-minute band. The random numbers which determined the longitude cells of the selected units and the random numbers which determined the latitude cells evolved as two independent Markov chains. (Basic results on Markov chains used in this paper can be found in an introductory text on stochastic processes such as Taylor and Karlin 1984). Details of the design are given in Section 2.

How does this *ad hoc* design compare to more standard one-per-stratum designs? It turns out, as shown in Section 2, that simple Markov chain techniques can describe a broad class of equal-probability designs for one-per-stratum selection from a finite population. This class includes standard techniques such as stratified simple random sampling, systematic sampling and balanced systematic sampling, as well as the Alaska designs described above. It is also easy to generate new designs within this class. This unified treatment of one-per-stratum designs allows for comparisons of efficiency.

<sup>1</sup> F. Jay Breidt, Iowa State University, Department of Statistics, Ames, IA 50011-1210, U.S.A.

First and second-order inclusion probabilities for all of these designs are derived in Section 3, yielding the Horvitz-Thompson estimator and its variance. As in much of the relevant literature (Madow and Madow 1944; Cochran 1946; Sedransk 1969; Bellhouse and Rao 1975; Wolter 1985; Bellhouse 1988; *etc.*) the average design variance of the Horvitz-Thompson estimator is evaluated under a variety of superpopulation models. Compact expressions for model-averaged design variances are obtained. Numerical examples in Section 4 show that designs introduced in this paper can be more efficient than standard one-per-stratum designs for superpopulations consisting of trend plus autocorrelated errors. Discussion follows in Section 5.

Though our motivating example is two-dimensional, one-dimensional designs will be considered throughout. Most proofs and derivations are straightforward and are omitted for brevity.

## 2. MARKOV CHAIN DESIGNS

Consider the problem of sampling from a finite population of  $N = na$  labeled units, denoted by

$$\begin{aligned} U &= \{1, \dots, N\} \\ &= \{1, \dots, a, a + 1, \dots, 2a, \dots, \\ &\quad (n - 1)a + 1, \dots, na\}. \end{aligned}$$

The value of a study variable  $y_k = y_{(i-1)a+j} = y_{ij}$  is associated with each label  $k$ ; the notation  $y_k$  or  $y_{ij}$  will be used for both random variables and realizations of random variables.

Here  $n$  is the sample size and  $a$  is the *sampling interval*. The  $n$  subsets

$$\{(i - 1)a + 1, \dots, (i - 1)a + a\} \quad (i = 1, \dots, n)$$

will be referred to as *strata*. The goal is to select one unit per stratum. Often, a stratified sampling design is defined to be one in which independent probability samples are selected in each stratum, but the restriction to independence is not used here.

Given a doubly stochastic transition probability matrix  $P$ , a *Markov chain sample* is given by

$$s = \{R_1, a + R_2, \dots, (n - 1)a + R_n\},$$

where  $R_1, \dots, R_n$  is the Markov chain defined by  $P$  and  $R_1 \sim \text{uniform}(1, \dots, a)$ . Formally, then, a *Markov chain design* (MC) is a function  $p(\cdot; P)$  such that

$$\begin{aligned} p(s; P) &= \Pr\{s = \{r_1, a + r_2, \dots, (n - 1)a + r_n\}\} \\ &= \Pr\{R_1 = r_1, R_2 = r_2, \dots, R_n = r_n\} \end{aligned}$$

$$= \begin{cases} P_{r_{n-1}, r_n} P_{r_{n-2}, r_{n-1}} \cdots P_{r_1, r_2} / a, \\ \quad \text{for } r_1, \dots, r_n \in \{1, \dots, a\}, \\ 0, \quad \text{otherwise.} \end{cases}$$

MC designs as defined in this paper are related to the designs given in Chandra, Sampath and Balasubramani (1992), in which a  $1 \times N$  vector of initial selection probabilities and a  $N \times N$  transition probability matrix of periodicity  $n$  determine a without-replacement sampling scheme. Chandra *et al.* focus on producing designs with strictly positive second-order inclusion probabilities. They do not explicitly consider the one-per-stratum designs of this paper, which can be imbedded in their structure in a straightforward way by constructing the appropriate initial probability vector and transition probability matrix.

The following result is useful in deriving the probabilistic features of MC designs.

**Result 1** Consider a Markov chain for which the transition probability matrix  $P$  is doubly stochastic (*i.e.*, all row sums and all column sums equal one) and  $R_1$  has a discrete uniform distribution, with mass  $1/a$  on each of the states  $1, \dots, a$ . Then  $R_i$  has a discrete uniform distribution on the states  $1, \dots, a$  for all  $i$ . In particular,  $R_i$  has mean  $(a + 1)/2$  and variance  $V(R_i) = (a^2 - 1)/12$ .

Some special cases of MC designs are of interest.

**Stratified simple random sampling.** If the transition probability matrix is

$$H = [1/a]_{j,j'=1}^a,$$

then

$$\begin{aligned} \Pr\{R_{i'} = j' \mid R_i = j\} &= 1/a = \Pr\{R_{i'} = j'\} \\ &\quad (j, j' = 1, \dots, a; i < i'), \end{aligned}$$

which, together with the Markov property, implies that  $R_1, \dots, R_n$  are probabilistically independent. In this case, the MC design is stratified simple random sampling with one unit per stratum (ST).

**Systematic sampling.** If the transition probability matrix is  $I$ , the  $a \times a$  identity matrix, then

$$\Pr\{R_{i'} = j' \mid R_i = j\} = \begin{cases} 1, & j = j', \\ 0, & j \neq j', \end{cases}$$

so that  $R_1, \dots, R_n$  are deterministically related. Thus,

$$s = \{R_1, a + R_1, \dots, (n - 1)a + R_1\},$$

and so the MC design is systematic sampling (SY).

**Compromise designs.** Intuitively, ST and SY are at opposite “extremes” in some sense. If  $\rho \in [0,1]$ , then

$$G_\rho = \rho H + (1 - \rho)I$$

is doubly stochastic. If  $\rho = 0$ , the design is SY and if  $\rho = 1$ , the design is ST. Any other choice of  $\rho$  will yield a sequence consisting of “runs” of SY samples. Thus, the class  $G_\rho$  includes ST and SY, as well as a continuum of “compromise” MC designs.

Other convex combinations of doubly stochastic matrices could be considered. The class of doubly stochastic matrices is also closed under matrix multiplication, transposition, and row and column permutation, so there are many ways to create MC designs.

**Balanced systematic sampling.** Murthy (1967, §5.9d) describes a one-per-stratum selection method which he calls *balanced systematic sampling* (BA). This method gives samples

$$s = \{R_1, a + (a + 1 - R_1), \dots, (n - 2)a + R_1, (n - 1)a + (a + 1 - R_1)\}$$

for  $n$  even and

$$s = \{R_1, a + (a + 1 - R_1), \dots, (n - 2)a + (a + 1 - R_1), (n - 1)a + R_1\}$$

for  $n$  odd. An interesting feature of this design is that if  $n$  is even and the population is perfectly linear ( $y_{ij} = \beta_0 + \beta_1[(i - 1)a + j]$ ), then the sample mean equals the population mean for any sample. With the transition probability matrix,

$$J = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix} a \times a$$

BA is a MC design.

**Alaska NRI design.** As described in Section 1, the 1992 NRI sample design for the northwestern region of the state of Alaska used two independent Markov chains in the controlled selection of latitude and longitude cells. The transition probability matrix for longitude cells,  $P_{\text{long}}$ , is given in Table 1. This design, henceforth denoted AK, is a MC design since  $P_{\text{long}}$  is doubly stochastic. Most of the transition probabilities are close to 0.10, so most “step sizes”

are approximately equally likely. Note, however, that mass has been removed from on and near the back diagonal and placed in the upper left and lower right corners, so that  $P_{\text{long}}$  discourages large east to west steps, such as from cell one to cell ten, and discourages small steps, such as from cell ten to cell one. On the other hand,  $P_{\text{long}}$  encourages steps of around length ten, such as from cell two to cell one, two or three. The realized sample of longitude cells is thus well-dispersed east to west, like a systematic sample would be, but its additional randomness guards against systematic error. Similarly, the Markov chain for latitude cells was set up to give good spatial dispersion north to south.

**Table 1**

Transition probability matrix for Markov chain sample of longitude cells, 1992 National Resources Inventory, Alaska. Entries are the conditional probabilities of selecting cell  $j'$  of stratum  $i + 1$  given that cell  $j$  of stratum  $i$  was selected.

Cell $j$ of stratum $i$	Cell $j'$ of stratum $i + 1$									
	1	2	3	4	5	6	7	8	9	10
1	0.05	0.15	0.15	0.15	0.15	0.15	0.10	0.10	0	0
2	0.15	0.15	0.15	0.10	0.10	0.10	0.10	0.10	0.05	0
3	0.15	0.15	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.10
4	0.15	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.10	0.10
5	0.15	0.10	0.10	0.10	0.05	0.05	0.10	0.10	0.10	0.15
6	0.15	0.10	0.10	0.10	0.05	0.05	0.10	0.10	0.10	0.15
7	0.10	0.10	0.05	0.10	0.10	0.10	0.10	0.10	0.10	0.15
8	0.10	0.10	0.05	0.05	0.10	0.10	0.10	0.10	0.15	0.15
9	0	0.05	0.10	0.10	0.10	0.10	0.10	0.15	0.15	0.15
10	0	0	0.10	0.10	0.15	0.15	0.15	0.15	0.15	0.05

### 3. HORVITZ-THOMPSON ESTIMATION UNDER MC

Write the population total as

$$t = \sum_U y_k = \sum_{i=1}^n \sum_{j=1}^a y_{(i-1)a+j} = \sum_{i=1}^n \sum_{j=1}^a y_{ij}$$

For all  $k$ , the first-order inclusion probabilities of a MC design are given by

$$\pi_k = \Pr\{k \in s\} = \Pr\{R_i = j\} = 1/a$$

and for  $k \leq l$ , the second-order inclusion probabilities are given by

$$\pi_{kl} = \begin{cases} 1/a, & \text{for } i = i', j = j', \\ 0, & \text{for } i = i', j \neq j', \\ P_{jj'}^{(i'-i)}/a, & \text{for } i < i'. \end{cases}$$

The design-unbiased Horvitz-Thompson estimator (Horvitz and Thompson 1952) for the population total is then

$$\hat{t}_\pi = \sum_s y_k / \pi_k = \sum_{i=1}^n \frac{y_i R_i}{1/a} = a \sum_{i=1}^n \sum_{j=1}^a y_{ij} I_{\{R_i=j\}},$$

where

$$I_{\{R_i=j\}} = \begin{cases} 1, & \text{if } R_i = j, \\ 0, & \text{if } R_i \neq j. \end{cases}$$

The design covariances of the indicators  $I_{\{R_i=j\}}$  are given by

$$\begin{aligned} C_{MC}(I_{\{R_i=j\}}, I_{\{R_{i'}=j'\}}) &= E_{MC}[I_{\{R_i=j\}} I_{\{R_{i'}=j'\}}] - \\ &E_{MC}[I_{\{R_i=j\}}] E_{MC}[I_{\{R_{i'}=j'\}}] \\ &= \pi_{(i-1)a+j, (i'-1)a+j'} - \\ &\pi_{(i-1)a+j} \pi_{(i'-1)a+j'}, \end{aligned}$$

and so the design variance of  $\hat{t}_\pi$  is

$$\begin{aligned} V_{MC}(\hat{t}_\pi) &= a^2 \sum_{i=1}^n \sum_{j=1}^a \left( \frac{1}{a} - \frac{1}{a^2} \right) y_{ij} y_{ij} \quad (1) \\ &+ a^2 \sum_{i=1}^n \sum_{j=1}^a \sum_{j' \neq j} \left[ 0 - \frac{1}{a^2} \right] y_{ij} y_{ij'} \\ &+ 2a^2 \sum_{i=1}^n \sum_{i' > i} \sum_{j=1}^a \left[ \frac{P_{jj'}^{(i'-i)}}{a} - \frac{1}{a^2} \right] y_{ij} y_{i'j} \\ &+ 2a^2 \sum_{i=1}^n \sum_{i' > i} \sum_{j=1}^a \sum_{j' \neq j} \left[ \frac{P_{jj'}^{(i'-i)}}{a} - \frac{1}{a^2} \right] y_{ij} y_{i'j'}. \end{aligned}$$

Since the design variance depends on all the values of the study variable in the finite population, (1) is not easily used for comparing designs. Following Cochran (1946), assume that the values of the study variable are generated from the superpopulation model

$$\xi : y_{ij} = \mu_{ij} + e_{ij},$$

where the  $\mu_{ij}$  are fixed and the  $e_{ij}$  are random variables with  $E_\xi[e_{ij}] = 0$ ,  $V_\xi(e_{ij}) = \sigma_{ij}^2$  and  $C_\xi(e_{ij}, e_{i'j'}) = \sigma_{ij, i'j'}$ . Then designs can be compared on the basis of model-averaged design variance.

**Proposition 1** Under the superpopulation model  $\xi$ , the average design variance of the Horvitz-Thompson estimator is

$$\begin{aligned} E_\xi[V_{MC}(\hat{t}_\pi)] &= a^2 V_{MC} \left[ \sum_{i=1}^n \mu_{iR_i} \right] + \\ &(a-1) \sum_{i=1}^n \sum_{j=1}^a \sigma_{ij}^2 - \sum_{i=1}^n \sum_{j=1}^a \sum_{j' \neq j} \sigma_{ij, i'j'} \\ &+ 2a \sum_{i=1}^n \sum_{i' > i} \sum_{j=1}^a \sum_{j'=1}^a \sigma_{ij, i'j'} \left[ \frac{P_{jj'}^{(i'-i)}}{a} - \frac{1}{a} \right] \end{aligned}$$

for any MC design. Note that if  $\mu_{ij}$  is independent of  $j$ , then  $V_{MC}[\sum_{i=1}^n \mu_{iR_i}] = 0$ .

The following proposition gives a sufficient condition under which no MC design has worse average design variance than SY.

**Proposition 2** Consider an uncorrelated additive model,

$$\xi : y_{ij} = \mu_{ij} + e_{ij} = \alpha_i + \beta_j + e_{ij},$$

where  $E_\xi[e_{ij}] = 0$ ,  $V_\xi(e_{ij}) = \sigma_{ij}^2$  and  $C_\xi(e_{ij}, e_{i'j'}) = 0$ . Then

$$E_\xi[V_{SY}(\hat{t}_\pi)] \geq E_\xi[V_{MC}(\hat{t}_\pi)]$$

for all MC designs.

**Proof** From Proposition 1, the only term of interest is  $V_{MC}[\sum_{i=1}^n \mu_{iR_i}]$ , which under SY is

$$V_{SY} \left[ \sum_{i=1}^n \mu_{iR_i} \right] = V_{SY} \left[ \sum_{i=1}^n \alpha_i + n\beta_{R_1} \right] = n^2 V(\beta_{R_1}),$$

while under a general MC design,

$$V_{MC} \left[ \sum_{i=1}^n \mu_{iR_i} \right] = \sum_{i=1}^n \sum_{i'=1}^n C_{MC}(\beta_{R_i}, \beta_{R_{i'}}).$$

Since  $C_{MC}(\beta_{R_i}, \beta_{R_{i'}}) \leq V(\beta_{R_1})$ , the proposition follows.  $\square$

Some specific models are considered in the next five subsections.

### 3.1 Random Permutation Model

A model for a population in random order is a permutation model, in which a realization of the measurements  $y_1, \dots, y_N$  is given by one of the  $N!$  equally likely permutations of  $N$  fixed values. This model can be written as

$$\xi_1 : y_{ij} = \bar{y}_U + e_{ij},$$

where  $\bar{y}_U = \sum_U y_k / N$ . See Rao (1975) for more details. The following result is then a consequence of Theorem 2.1 of Rao and Bellhouse (1978).

**Result 2** Under the random permutation model,

$$E_{\xi_1} [V_{MC}(\hat{t}_\pi)] = (N^2/n)(1 - n/N) \sum_U (y_k - \bar{y}_U)^2 / (N - 1)$$

for any MC design.

Thus, the average variance over all permutations is exactly  $V_{SI}(\hat{t}_\pi)$ , where SI denotes (unstratified) simple random sampling without replacement. For SY, this result is originally due to Madow and Madow (1944). See also Sedransk (1969).

### 3.2 Stratification Effects Model

A model for a population with stratification effects is

$$\xi_2: y_{ij} = \alpha_i + e_{ij},$$

where the  $\alpha_i$  are fixed constants and  $e_{ij}$  are uncorrelated random variables with mean zero and variance  $\sigma^2$ . Note that if  $\alpha_i \equiv \mu$ , then  $\xi_2$  is an alternative to  $\xi_1$  as a model for a population in random order.

**Result 3** Under the stratification effects model,

$$E_{\xi_2} [V_{MC}(\hat{t}_\pi)] = na(a - 1)\sigma^2$$

for any MC design.

### 3.3 Linear Trend Model

A model for a population with a linear trend is

$$\xi_3: y_{ij} = \beta_0 + \beta_1[(i - 1)a + j] + e_{ij},$$

where  $\beta_0$  and  $\beta_1$  are fixed constants and  $e_{ij}$  are uncorrelated  $(0, \sigma^2)$  random variables.

**Result 4** Under the linear trend model  $\xi_3$ ,

$$E_{\xi_3} [V_{MC}(\hat{t}_\pi)] = \beta_1^2 a^2 V_{MC} \left[ \sum_{i=1}^n R_i \right] + na(a - 1)\sigma^2 \quad (2)$$

for any MC design. Since  $\xi_3$  is additive, no MC design has a larger expected variance under a linear trend model than SY.

The only design-dependent term in (2) is  $V_{MC}[\sum_{i=1}^n R_i]$ . Under SY,  $\sum_{i=1}^n R_i = nR_1$ , so that

$$V_{SY} \left[ \sum_{i=1}^n R_i \right] = n^2 V(R_1),$$

while under ST,

$$V_{ST} \left[ \sum_{i=1}^n R_i \right] = nV(R_1).$$

Under BA, for  $n$  even,

$$V_{BA} \left[ \sum_{i=1}^n R_i \right] = V_{BA} \left[ \frac{n}{2} R_1 + \frac{n}{2} (a + 1 - R_1) \right] = 0.$$

This implies that if the population is perfectly linear ( $\sigma^2 = 0$ ), then

$$E_{\xi_3} [V_{BA}(\hat{t}_\pi)] = 0,$$

so that  $\hat{t}_\pi = t$  for all samples, as noted by Murthy (1967, p. 165).

**Result 5** Under the linear trend model  $\xi_3$ ,

$$\begin{aligned} E_{\xi_3} [V_{BA}(\hat{t}_\pi)] &\leq E_{\xi_3} [V_{ST}(\hat{t}_\pi)] \\ &\leq E_{\xi_3} [V_{G_\rho}(\hat{t}_\pi)] \\ &\leq E_{\xi_3} [V_{SY}(\hat{t}_\pi)] = \max_{MC} E_{\xi_3} [V_{MC}(\hat{t}_\pi)], \end{aligned} \quad (3)$$

where the middle term is monotone increasing with decreasing  $\rho \in [0, 1]$ . If  $n$  is even, the left-hand side of (3) equals  $\min_{MC} E_{\xi_3} [V_{MC}(\hat{t}_\pi)]$ .

### 3.4 Periodic Population Model

A simple model for a population showing a deterministic periodicity with period  $p$  is the sine wave model

$$\xi_4: y_{ij} = \alpha \sin \left\{ \frac{2\pi}{p} [(i - 1)a + j] \right\} + e_{ij},$$

where  $e_{ij}$  are uncorrelated random variables with mean zero and variance  $\sigma^2$ .

**Result 6** Under the periodic population model  $\xi_4$ ,

$$\begin{aligned} E_{\xi_4} [V_{MC}(\hat{t}_\pi)] &= a^2 \alpha^2 V_{MC} \left[ \sum_{i=1}^n \sin \frac{2\pi}{p} [(i - 1)a + R_i] \right] \\ &\quad + na(a - 1)\sigma^2 \end{aligned}$$

for any MC design.

Denote the sine wave model  $\xi_4$  with  $p = a$  by  $\xi_{4a}$ . Under  $\xi_{4a}$ ,

$$\sin \left\{ \frac{2\pi}{p} [(i - 1)a + j] \right\} = \sin \frac{2\pi j}{a},$$

so that the model is additive and no MC design has larger expected design variance under  $\xi_{4a}$  than SY, highlighting the fact that SY is inappropriate for a population containing a periodicity with period equal to the sampling interval (Madow and Madow 1944). This result generalizes as follows.

**Result 7** If  $\mu_{ij} \equiv \beta_j$  in  $\xi$ , then  $\xi$  is a model for a population showing a deterministic periodicity with period equal to the sampling interval,  $a$ . The model  $\xi$  is additive and so no MC design has larger expected design variance under  $\xi$  than SY.

### 3.5 Autocorrelated Model

Beginning with Cochran (1946), many authors have compared ST, SY and simple random sampling under an autocorrelated superpopulation model. See Bellhouse (1988, §4) for a review.

Consider the following autocorrelation model due to Cochran (1946):

$$\xi_5: y_{ij} = \mu + e_{ij}$$

where  $\sigma_{ij,i'j'} = \gamma[(i' - i)a + j' - j]$  for  $i' \geq i$ .

**Result 8** Under the autocorrelated model  $\xi_5$ ,

$$E_{\xi_5}[V_{MC}(\hat{t}_\pi)] = na(a-1)\gamma(0) - 2n \sum_{h=1}^{a-1} \gamma(h)(a-h) + 2a \sum_{h=1}^{n-1} \sum_{j=1}^a \sum_{j'=1}^a \gamma(ha + j' - j)(n-h) \left( P_{jj'}^{(h)} - \frac{1}{a} \right)$$

for any MC design.

**Result 9** If, for  $h \geq 0$ ,  $\gamma(h)$  is non-negative, non-increasing and convex, *i.e.*,

$$\begin{aligned} \gamma(h) &\geq 0, \gamma(h) \geq \gamma(h+1) \quad \text{and} \\ \gamma(h+2) - 2\gamma(h+1) + \gamma(h) &\geq 0, \end{aligned}$$

then  $E_{\xi_5}[V_{SY}(\hat{t}_\pi)] = \min_{MC} E_{\xi_5}[V_{MC}(\hat{t}_\pi)]$ .

This result is a corollary of a theorem due to Hájek (1959), given as Theorem 4.1 of Bellhouse (1988); Bellhouse clarified the conditions under which the theorem holds. Hájek's theorem generalized an earlier result due to Cochran (1946), who compared SY, ST and simple random sampling.

## 4. EFFICIENCY: SOME NUMERICAL EXAMPLES

An important class of models for time series and spatial processes consists of a low-order polynomial trend plus an autocorrelated error sequence. A simple example is

$$\xi_{(\beta,\phi)}: y_{ij} = \beta_0 + \beta_1[(i-1)a + j] + e_{ij},$$

where the autocorrelation structure is that of a first-order autoregressive (AR) model,

$$\sigma_{ij,i'j'} = \gamma[(i' - i)a + j' - j] = \sigma^2 \phi^{(i' - i)a + j' - j}$$

for  $i' \geq i$  and  $|\phi| < 1$ . The average design variance under this model is obtained from Results 4 and 8. For different choices of  $\beta_1$  and  $\phi$ , the ratio of expected design variances,

$$E_{\xi}[V_{MC}(\hat{t}_\pi)]/E_{\xi}[V_{SY}(\hat{t}_\pi)], \quad (4)$$

is given in Table 2 for various MC designs. Also tabled is the optimal  $G_\rho$  design, obtained by minimizing (4) with respect to  $\rho$ . Use of this design is only feasible if superpopulation parameters are known, so it is tabled merely as a benchmark and not as a competitor.

When  $\beta_1 \neq 0$  and  $\phi = 0$ , the model is  $\xi_3$  and the tabled values agree with Result 5: SY is the worst MC design and BA is the best, with  $G_{1/3}$ ,  $G_{2/3}$  and ST falling between them. Though BA does extremely well for this model, any non-SY MC design would be a good choice.

When  $\beta_1 = 0$  and  $\phi \neq 0$ ,  $\xi_{(\beta,\phi)}$  is a special case of model  $\xi_5$ . For  $\phi > 0$ , Result 9 and the table agree that SY is most efficient since it makes the sample as "spread out" as possible, but for weak autocorrelation, the other MC designs are competitive. BA is very poor for this model, because the design ensures that every other pair  $R_i, R_{i+1}$  will be no more than  $a$  units apart. (For the same reason, BA is good for a negatively autocorrelated population.) AK,  $G_{1/3}$  and  $G_{2/3}$  outperform ST, because each of these designs encourages state transitions of around length  $a$ .

Similar results are obtained for the superpopulation model

$$\xi_{(\alpha,\phi)}: y_{ij} = \alpha \sin \frac{2\pi j}{a} + e_{ij},$$

where  $\sigma_{ij,i'j'}$  is as above. Table 2 gives the ratio of expected design variances (4) under this model, obtained from Results 6 and 8.

When  $\alpha \neq 0$  and  $\phi = 0$ , the model is  $\xi_{4a}$  and SY performs badly, as indicated by Result 7. Even for  $\phi \neq 0$ , SY performs well only when the periodicity is swamped by highly-correlated noise.

Note that no design dominates Table 2: each of SY,  $G_{1/3}$ ,  $G_{2/3}$ , ST, BA and AK is the best at least once among those considered. For a moderate trend and high autocorrelation, AK,  $G_{1/3}$  and  $G_{2/3}$  can beat standard MC designs. Overall, Table 2 suggests that some non-standard MC designs, such as  $G_{2/3}$  and AK, do reasonably well for a variety of populations: retaining much of the efficiency of SY against an autocorrelated population, while still guarding against systematic effects in other kinds of populations.

**Table 2**

Ratio of expected design variance under MC to expected design variance under SY for superpopulation consisting of trend (line with slope  $\beta_1$  or sine wave with period  $a$  and amplitude  $\alpha$ ) plus autoregressive (AR) errors ( $N = 1,000, \sigma^2 = 100, a = 10$ ). Here  $G_{\rho^*}$  is the optimal compromise design, where  $\rho^*$  is a function of superpopulation parameters. Ratio for the best realizable design in each row (if not SY) is italicized.

Model	$\phi$	Markov Chain Design						
		$G_{1/3}$	$G_{2/3}$	ST	BA	AK	$G_{\rho^*}$	( $\rho^*$ )
Line + AR $\beta_1 = 0.7$	-0.5	0.2322	0.2085	0.2001	<i>0.1666</i>	0.2056	0.2001	(1.0000)
	0.0	0.2220	0.1983	0.1903	<i>0.1821</i>	0.1957	0.1903	(1.0000)
	0.1	0.2187	0.1950	0.1871	<i>0.1825</i>	0.1921	0.1871	(1.0000)
	0.5	0.1922	0.1702	<i>0.1645</i>	0.1754	0.1659	0.1645	(1.0000)
	0.9	0.0980	0.0778	<i>0.0742</i>	0.0768	0.0762	0.0742	(1.0000)
Line + AR $\beta_1 = 0.4$	-0.5	0.4504	0.4328	0.4262	<i>0.3647</i>	0.4304	0.4262	(1.0000)
	0.0	0.4344	0.4172	0.4114	<i>0.4054</i>	0.4153	0.4114	(1.0000)
	0.1	0.4291	0.4121	<i>0.4065</i>	0.4085	0.4094	0.4065	(1.0000)
	0.5	0.3853	0.3727	0.3724	0.4116	<i>0.3667</i>	0.3719	(0.8320)
	0.9	0.1876	<i>0.1835</i>	0.1914	0.2170	0.1848	0.1821	(0.5223)
Line + AR $\beta_1 = 0.1$	-0.5	0.9233	0.9190	0.9163	<i>0.7941</i>	0.9175	0.9163	(1.0000)
	0.0	0.9201	0.9177	0.9169	<i>0.9160</i>	0.9174	0.9169	(1.0000)
	0.1	0.9191	0.9175	0.9175	0.9349	<i>0.9156</i>	0.9174	(0.8156)
	0.5	<i>0.9160</i>	0.9289	0.9439	1.0606	0.9185	0.9135	(0.1997)
	0.9	<i>0.8621</i>	0.9787	1.0725	1.2710	1.0017	0.7888	(0.0981)
Pure AR	-0.5	0.9978	0.9956	0.9935	<i>0.8617</i>	0.9942	0.9935	(1.0000)
	0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	(---)
	0.1	1.0009	1.0019	1.0028	1.0228	1.0001	1.0000	(0.0000)
	0.5	1.0179	1.0357	1.0536	1.1852	1.0245	1.0000	(0.0000)
	0.9	1.2517	1.4380	1.5814	1.8798	1.4734	1.0000	(0.0000)
Sine + AR $\alpha = 0.1$	-0.5	0.9929	0.9906	0.9884	<i>0.8578</i>	0.9892	0.9884	(1.0000)
	0.0	0.9947	0.9946	<i>0.9945</i>	0.9950	0.9946	0.9945	(1.0000)
	0.1	0.9955	0.9963	0.9972	1.0175	<i>0.9945</i>	0.9954	(0.1925)
	0.5	1.0110	1.0285	1.0462	1.1775	1.0173	0.9977	(0.0364)
	0.9	1.2178	1.3980	1.5371	1.8294	1.4322	0.9999	(0.0018)
Sine + AR $\alpha = 1.0$	-0.5	0.6747	0.6634	0.6586	<i>0.6008</i>	0.6604	0.6586	(1.0000)
	0.0	0.6603	0.6499	<i>0.6464</i>	0.6770	0.6477	0.6464	(1.0000)
	0.1	0.6554	0.6455	0.6425	0.6863	<i>0.6421</i>	0.6425	(1.0000)
	0.5	0.6149	0.6133	0.6196	0.7320	<i>0.6041</i>	0.6121	(0.5079)
	0.9	<i>0.3570</i>	0.3832	0.4126	0.5527	0.3877	0.3560	(0.2852)
Sine + AR $\alpha = 10.0$	-0.5	0.0668	0.0384	<i>0.0287</i>	0.1101	0.0323	0.0287	(1.0000)
	0.0	0.0656	0.0372	<i>0.0275</i>	0.1115	0.0311	0.0275	(1.0000)
	0.1	0.0652	0.0368	<i>0.0271</i>	0.1115	0.0307	0.0271	(1.0000)
	0.5	0.0622	0.0339	<i>0.0245</i>	0.1106	0.0277	0.0245	(1.0000)
	0.9	0.0529	0.0247	<i>0.0154</i>	0.1016	0.0187	0.0154	(1.0000)

**5. DISCUSSION**

The class of Markov chain designs has been defined and shown to include systematic sampling, stratified simple random sampling and balanced systematic sampling as special cases. Some new designs have been introduced ( $G_{\rho}$ , AK) and shown to be competitive with standard one-per-stratum designs under a variety of superpopulation models. In particular, the new designs work well in numerical examples for trending superpopulations with autocorrelated errors. This is the kind of population of concern in many area sampling problems, such as the 1992 National Resources Inventory in Alaska. A two-dimensional MC design implemented for that survey shows that one-dimensional MC designs might be usefully extended to a spatial sampling context, though further work on this extension is necessary.

Further work on variance estimation for MC designs is also needed. Because these are one-per-stratum designs, design-unbiased estimation of the variance of the Horvitz-Thompson estimator is not possible. The problem of variance estimation for one-per-stratum designs, particularly for SY, has received much attention. For example, Wolter (1985) discusses in detail eight different biased variance estimators for SY and evaluates their biases under superpopulation models. Work in this direction for the collapsed strata variance estimator (e.g., Cochran 1977, p. 139) under general MC designs is in progress.

**ACKNOWLEDGEMENTS**

This work was supported by the U.S. Department of Agriculture's Soil Conservation Service under SCS Cooperative Agreement No. 68-3A75-2-64. Jeff Goebel (Soil Conservation Service) and Wayne Fuller (Iowa State University) developed the MC design for the state of Alaska. The author thanks the anonymous referees for their constructive comments on an earlier version of this paper.

**REFERENCES**

BELLHOUSE, D.R. (1988). Systematic sampling. In *Handbook of Statistics*. (Eds. P.R. Krishnaiah and C.R. Rao), Vol. 6. Amsterdam: North-Holland, 125-145.

BELLHOUSE, D.R., and RAO, J.N.K. (1975). Systematic sampling in the presence of a trend. *Biometrika*, 62, 694-697.

CHANDRA, K.S., SAMPATH, S., and BALASUBRAMANI, G.K. (1992). Markov sampling for finite populations. *Biometrika*, 79, 210-213.

COCHRAN, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Annals of Mathematical Statistics*, 17, 164-177.

COCHRAN, W.G. (1977). *Sampling Techniques*, 3rd Ed. New York: Wiley.

HÁJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Časopis Pro Pěstování Matematiky*, 84, 387-423.

HORVITZ, D.G., and THOMPSON, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

MADOW, W.G., and MADOW, L.H. (1944). On the theory of systematic sampling, I. *Annals of Mathematical Statistics*, 15, 1-24.

MURTHY, M.N. (1967). *Sampling Theory and Methods*. Calcutta: Statistical Publishing Society.

MURTHY, M.N., and RAO, T.J. (1988). Systematic sampling with illustrative examples. In *Handbook of Statistics*. (Eds. P.R. Krishnaiah and C.R. Rao), (Vol. 6). Amsterdam: North-Holland, 147-185.

- RAO, J.N.K. (1975). On the foundations of survey sampling. In *A Survey of Statistical Design and Linear Models*. (Ed. J.N. Srivastava). Amsterdam: North-Holland, 489-505.
- RAO, J.N.K., and BELLHOUSE, D.R. (1978). Optimal estimation of a finite population mean under generalized random permutation models. *Journal of Statistical Planning and Inference*, 2, 125-141.
- SEDRANSK, J. (1969). Some elementary properties of systematic sampling. *Skandinavisk Aktuarietidskrift*, 1-2, 39-47.
- TAYLOR, H.M., and KARLIN, S. (1984). *An Introduction to Stochastic Modeling*. Orlando: Academic Press.
- WOLTER, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.