A Sample Coverage Approach to Multiple-System Estimation With Application to Census Undercount

Anne CHAO and P. K. TSAY

The concept of "sample coverage" used in animal abundance estimation is modified to evaluate the undercount of a census. Lack of independence between the census and its postenumeration survey leads to correlation bias for the standard estimator of population size. An additional recapture sample (besides the census and postenumeration survey) can be used to estimate the correlation bias due to two types of dependences. This work expresses the correlation bias for the three-sample model as a function of expected sample coverage and measures of dependence between lists. A nonparametric population size estimator that incorporates the correlation bias is proposed. A simulation study investigates the performance of the proposed procedure. Data from a 1988 dress rehearsal study for the 1990 census conducted by the U.S. Bureau of the Census are used to illustrate the proposed three-system estimation procedure and to compare the resulting estimates with those given by Darroch, Fienberg, Glonek, and Junker and Zaslavsky and Wolfgang.

KEY WORDS: Capture–recapture sampling; Correlation bias; Dual-system estimator; Heterogeneity; List dependence; Population size.

1. INTRODUCTION

The U.S. census undercount problem has been widely discussed in the literature. An overview of this topic was provided in a collection of articles on the 1990 census in the September 1993 issue of the *Journal of the American Statistical Association*. A series of papers by Fienberg (1992 and references therein) in *Chance* described the controversial issue of the adjustment of the census results. More recently, three theme papers with discussions appeared in the November 1994 issue of *Statistical Science*. The debate on this topic has also attracted the interest of non-Americans, as well as people in other research fields. This article provides an outside view on a special type of census undercount problem from the standpoint of animal abundance estimation.

It is well known that there are two principal approaches to evaluating the U.S. census undercount: (1) demographic analysis using vital data of birth, death, immigration, and emigration records (Robinson, Ahmed, Das Gupta, and Woodrow 1993), and (2) dual-system estimation using a postenumeration survey (PES) (Hogan 1993). We focus on the second approach, for which a PES is conducted after the census. In this article a "system" refers to a census, PES, list, or a "capture" sample. A dual-system estimator (DSE) based on the census and its PES for each poststratum defined by geographical and demographic variables is commonly used to estimate the undercount by assuming that these two lists are independent.

The independence assumption in wildlife studies is usually expressed in terms of "equal catchability"—all animals have the same probability of capture in each trapping sample. Ecologists and biologists have long recognized that this assumption is rarely valid in most practical applications. Because individuals can be cross-classified according to their presence and absence in each sample, the data can be summarized by a 2×2 table. Hence the independence assumption for human populations is usually interpreted in terms of the concept of independence from a 2×2 categorical data analysis. Lack of independence between the census and its PES leads to a bias for the DSE, usually referred to as "correlation bias." Similar bias arises if there are more than two systems. This bias may be due to the following two types of dependences:

- Dependence within each individual. Conditional on any individual, inclusion in the census has a direct causal effect on his or her inclusion in the PES; for example, if the probability of being listed in the PES given his or her presence in the census is higher than that given his or her absence in the census, then the census and the PES become positively dependent. This type of dependence is usually referred to as "list dependence" in the literature; it means that the probability for inclusion in any sample for a given individual depends on his or her records of other samples. See Section 2.1.2 for details. In a stratified analysis, the foregoing arguments can also be applied to each substratum instead of each individual.
- Heterogeneity between individuals. Even if the two lists are independent within individuals, the lists may become dependent if the capture probabilities are heterogeneous among individuals. This phenomenon is similar to Simpson's paradox in categorical data analysis. That is, aggregating two independent 2×2 tables might result in a dependent table. Hook and Regal (1993) provided an example in epidemiological applications. Further discussion is given in Section 2.1.1.

These two types of dependences are confounded and cannot be separated in a data analysis unless more information is provided. Several authors have made significant

Anne Chao is Professor and P. K. Tsay is graduate student, Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043. The authors thank an associate editor and two referees for carefully reading the manuscript, improving the exposition, and providing thoughtful suggestions. This research was supported by the National Science Council of Taiwan under contract NSC85-2121-M007-004.

contributions to relaxing the independence assumption in the context of census undercount estimation. Isaki (1986) proposed several alternative DSEs to incorporate the correlation bias due to list dependence. Bell (1993), Isaki and Schultz (1986), and Wolter (1990) suggested the use of additional demographic information (e.g., population totals or sex ratio) to weaken the independence assumption. Zaslavsky and Wolfgang (1993) proposed using a third system, besides the census and PES, to provide additional information to assess the dependence. Their method is illustrated by using the data from a 1988 dress rehearsal census, its PES, and an administrative data source as the third sample. Assuming independence within each individual, Darroch, Fienberg, Glonek, and Junker (1993) also proposed the use of a third sample using a log-linear model approach with the Rasch model (Rasch 1961). Alho, Mulry, Wurdeman, and Kim (1993) modeled the heterogeneity using a logistic model containing several explanatory variables under the assumption of independence within individuals. Agresti (1994) considered the applications of the Rasch model to capture–recapture models. His results can be adapted easily to a census undercount estimation as well.

The idea of sample coverage (Good 1953) has been used in species estimation (see Bunge and Fitzpatrick 1993 and Chao and Lee 1992 for reviews). The basic idea is that the sample coverage can be well estimated in nonindependent cases. Thus an estimator of population size can be derived via the relation of sample coverage and population size. Chao, Lee, and Jeng (1992) extended this idea to estimate wildlife population size from capture–recapture data. This work cannot be directly applied to the census undercount problem, however, because (1) there are usually more trapping samples in wildlife studies whereas in the census problem only two or three lists are available, (2) the Rasch model and its generalizations (see Sec. 2.1) commonly used for human populations cannot be expressed in the form of any model adopted for animal populations, and (3) although a PES is conducted after the census, there is usually no temporal or sequential time order for the third list; such order usually exists in typical animal experiments. Thus the methods used in wildlife studies need to be modified to handle the census undercount estimation. This is our basic motivation for the research in this article.

In Section 2.1 we define a general capture–recapture model that encompasses the Rasch model and some models proposed by Otis, Burnham, White, and Anderson (1978) and Pollock (1976) as special cases. We discuss two-sample and three-sample cases separately in Sections 2.2 and 2.3. We show that at least three samples are needed to measure the possible dependence. We propose an estimator of population size that allows for dependence among the lists and heterogeneity across individuals for three samples and also briefly discuss extension to four samples. In Section 3 we use data from a 1988 dress rehearsal study for the 1990 census conducted by the U.S. Bureau of the Census to illustrate the proposed three-system estimation procedure and to compare the resulting estimates with those given by Darroch et al. (1993) and Zaslavsky and Wolfgang (1993). In Section 4 we report simulation results to study the performance of the proposed estimator and to compare it with estimates based on log-linear models. We give concluding remarks in Section 5.

2. MODELS AND ESTIMATORS

2.1 General Model

Assume that the true population size is N and that there are t systems (samples, lists) including the census. Let the individuals be indexed by $1, 2, \ldots, N$ and the samples indexed by $1, 2, \ldots, t$. The ordering here is just for convenience of mathematical treatment, and our resulting estimator is invariant to the ordering of the lists. The data can be conveniently expressed by a $N \times t$ matrix $\mathbf{X} = (X_{ij})$, where

$$X_{ij} = I[\text{the } i\text{th individual is listed in the } j\text{th sample}]$$

and $I[A]$ is the usual indicator function. Individuals are assumed to act independently. For a fixed sample j, each X_{ij} conditional on the records of other samples is a Bernoulli random variable. The probability $P\ (X_{ij} = 1)$ could be treated by either fixed-effects or random-effects arguments. In the following we focus mainly on the fixed-effects approach. All of the derivations for random-effects arguments are parallel, and the resulting estimators are exactly the same. The parameters that must be identified in a fixed-effects analysis are $\mu_j = N^{-1} \sum_{i=1}^{N} E(X_{ij})$ (the average inclusion probability for the jth sample), as well as the dependence measures between or among samples. Define the coefficient of covariation (CCV) of samples j and k as

$$\gamma_{jk} = \frac{1}{N} \sum_{i=1}^{N} E\{(X_{ij} - \mu_j)(X_{ik} - \mu_k)\}/(\mu_j \mu_k). \quad (1)$$

The interpretation of the CCV is elaborated in later discussions. The magnitude of γ_{jk} measures the degree of the dependence of samples j and k. In the independent case all CCVs are 0. Similarly, define the CCV for samples $k_1, k_2, \ldots, k_m$ as

$$\gamma_{k_1 k_2 \ldots k_m} = \frac{1}{N} \sum_{i=1}^{N} E\{(X_{ik_1} - \mu_{k_1})(X_{ik_2} - \mu_{k_2}) \cdots (X_{ik_m} - \mu_{k_m})\}/(\mu_{k_1} \mu_{k_2} \cdots \mu_{k_m}). \quad (2)$$

This general setup includes most ecological models as special cases. The most general model for animal populations discussed by Pollock (1976, 1991) is called model M_{tbh}, where time variation, behavioral response, and heterogeneity affect the capture probabilities. For this model, a commonly used form of the capture probability is

$$P\ (X_{ij} = 1 | \mathcal{H}_{ij}) = h_i e_j I\left(\sum_{m=1}^{j-1} X_{im} = 0\right) + b_i e_j I\left(\sum_{m=1}^{j-1} X_{im} > 0\right), \quad (3)$$

where $\mathcal{H}_{ij}$ denotes the capture records excluding sample j for the ith individual. The foregoing capture probability depends on individual's previous capture histories; the

probability is $h_i e_j$ for a first capture and $b_i e_j$ for a recapture. There exists a "behavioral" response to capture, which causes list dependence. The probabilities also vary among individuals; h_i in (3) denotes the "heterogeneity" effect, and e_j represents the "sample" or "time" effect. For this model, both types of dependences arise. We now discuss two special cases to clarify the difference between the two types of dependences.

2.1.1 Only Heterogeneity Between Individuals Exists. If there is no list dependence within an individual, then conditional on individual i, the capture histories $\{X_{i1}, X_{i2}, \ldots, X_{it}\}$ are independent. The probability $P(X_{ij} = 1|\mathcal{H}_{ij})$ on the left side of (3) does not depend on the history $\mathcal{H}_{ij}$. Therefore, we can let $P(X_{ij} = 1|\mathcal{H}_{ij}) = p_{ij}$. Thus $\mu_j = \sum_i p_{ij}/N$, and the CCV of samples j and k reduces to

$$\gamma_{jk} = \left\{ \frac{1}{N} \sum_{i=1}^{N} (p_{ij} - \mu_j)(p_{ik} - \mu_k) \right\} \Big/ (\mu_j \mu_k)$$

$$= \frac{1}{N} \sum_{i=1}^{N} p_{ij} p_{ik} / (\mu_j \mu_k) - 1. \tag{4}$$

Hence γ_{jj} reduces to the square of the usual coefficient of variation (CV) defined for a single "sample" $\{p_{1j}, p_{2j}, \ldots, p_{Nj}\}$. Here the magnitude of γ_{jk} measures the degree of the dependence due purely to heterogeneity in catchabilities. Note that $\gamma_{jk} = \rho_{jk}(\gamma_{jj}\gamma_{kk})^{1/2}$, where ρ_{jk} is the "sample" correlation coefficient of the two "samples" $\{p_{1i}, p_{2i}, \ldots, p_{Ni}\}$ and $\{p_{1j}, p_{2j}, \ldots, p_{Nj}\}$. The concept of γ_{jk} was also used by Kadane, Meyer, and Tukey (1992). The CCV for samples $k_1, k_2, \ldots, k_m$ now becomes

$$\gamma_{k_1 k_2 \ldots k_m}$$

$$= \left\{ \frac{1}{N} \sum_{i=1}^{N} (p_{ik_1} - \mu_{k_1})(p_{ik_2} - \mu_{k_2}) \cdots (p_{ik_m} - \mu_{k_m}) \right\}$$

$$\div (\mu_{k_1} \mu_{k_2} \cdots \mu_{k_m}).$$

Note that if there exists a sample k_j with homogeneous catchability (call it sample k) then $p_{ik} = \mu_k$ for all i, and any CCV of samples including sample k is 0.

This special class of models includes the following three widely used models:

- The Rasch (1961) model:

  $$P(X_{ij} = 1) = \exp(w_i' + a_j')/[1 + \exp(w_i' + a_j')],$$

  which can be reparameterized as

  $$P(X_{ij} = 1) = w_i a_j/(1 + w_i a_j), \tag{5}$$

  where w_i denotes the heterogeneity effect of the ith individual and a_j denotes the sample effect of the jth list. The values of w_i and a_j are defined only up to a multiplicative constant. In a fixed-effects model all w_i's are treated as parameters, whereas in a random-effects model $(w_1, w_2, \ldots, w_N)$ are a random sample from a distribution. It is understood that in the latter case, $P(X_{ij} = 1)$ is a conditional probability; that is, $P(X_{ij} = 1|w_i)$.
- The generalized Rasch model (Darroch et al. 1993; Stegelman 1983):

  $$P(X_{ij} = 1)$$

  $$= \begin{cases} w_i a_j/(1 + w_i a_j), & j = 1, 2, \ldots, k \\ w_i^* a_j/(1 + w_i^* a_j), & j = k+1, \ldots, t. \end{cases} \tag{6}$$

  That is, the patterns of heterogeneity effects are different in two separate groups of samples. This model can evidently be extended to more than two types of heterogeneity effects.
- Model M_{th} (Pollock 1976; also Otis et al. 1978):

  $$P(X_{ij} = 1) = h_i e_j, \tag{7}$$

  which is a special case of (3). As in the two foregoing models, there is no response to capture within each individual. The individual's heterogeneity effect h_i and the sample effect e_j are defined up to a multiplicative constant. Using (4), we can show for all j and k that $\gamma_{jk} = [\sum(h_i - \bar{h})^2/N]/\bar{h}^2$, where $\bar{h} = \sum h_i/N$. Thus the CCVs for all pairs of samples are identical and are equal to the square of the CV of $(h_1, h_2, \ldots, h_N)$. Similarly, we have $\gamma_{ijk} = [\sum(h_i - \bar{h})^3/N]/\bar{h}^3$ for all i, j, and k. The effects $(h_1, h_2, \ldots, h_N)$ can also be modeled as random effects so that they represent a random sample from a certain distribution.

2.1.2 Only List Dependence Exists. If there is no heterogeneity, then the conditional probability $P(X_{ij} = 1|\mathcal{H}_{ij})$ defined in (3) is independent of i. This class of model includes models for which the probability of presence in one sample depends on capture histories, but this probability does not vary among individuals. For example, in a two-sample capture–recapture model which allows for time variation and behavioural response to capture, we have for all i that $P(X_{i2} = 1|X_{i1}) = he_2 I(X_{i1} = 0) + be_2 I(X_{i1} = 1)$, where $h \neq b$. The marginal probability $P(X_{ij} = 1) \equiv \mathcal{P}_j$ is the probability of being listed in sample j. $P(X_{ij} = 1, X_{ik} = 1) \equiv \mathcal{P}_{jk}$ is the probability of being listed in both lists j and k, and $\mathcal{P}_{jkm}$ is defined similarly. Then we have

$$\gamma_{jk} = \frac{\mathcal{P}_{jk}}{\mathcal{P}_j \mathcal{P}_k} - 1 = \frac{P(X_{ik} = 1|X_{ij} = 1)}{P(X_{ik} = 1)} - 1 \tag{8}$$

and

$$\gamma_{jkm} = \frac{\mathcal{P}_{jkm}}{\mathcal{P}_j \mathcal{P}_k \mathcal{P}_m} - \gamma_{jk} - \gamma_{jm} - \gamma_{km} - 1.$$

Other analogous expressions can be derived for higher order CCVs.

For notational simplicity, let

$$Z_{s_1, s_2 \ldots s_t} = \sum_{i=1}^{N} I[X_{i1} = s_1, X_{i2} = s_2, \ldots, X_{it} = s_t]$$

be the number of individuals with "capture" history $s_1, s_2, \ldots, s_t$, where $s_j = 0$ denotes absence in sample j and $s_j = 1$ denotes presence in sample j. For example, when $t = 2$, there are three observed cells Z_{10}, Z_{01}, and Z_{11}, where Z_{10} is the number of individuals listed in sample 1 but not in sample 2, Z_{01} is the number of individuals listed in sample 2 only, and Z_{11} is the number of individuals listed in both samples. The unobserved cell Z_{00} denotes the uncounted. When $t = 3$, there are seven observed cells, $Z_{100}, Z_{010}, Z_{001}, Z_{110}, Z_{011}, Z_{101}$, and Z_{111}. A similar interpretation pertains to these histories. When we add over a sample, the subscript corresponding to that sample is replaced by a "+" sign; for example, $Z_{+11} = Z_{011} + Z_{111}$ and $Z_{++1} = Z_{001} + Z_{011} + Z_{101} + Z_{111}$. Let $n_j, j = 1, 2, \ldots, t$ be the number of individuals listed in sample j. Thus $n_1 = Z_{1+}$ and $n_2 = Z_{+1}$ for $t = 2$ and $n_1 = Z_{1++}, n_2 = Z_{+1+}$, and $n_3 = Z_{++1}$ for $t = 3$.

2.2 Two-Sample Case

As shown by Chao and Lee (1992), the concept of sample coverage is closely related to population size estimation. The basic idea relies on the fact that it is difficult to estimate the number of uncounted directly, but the sample coverage can be well estimated in any nonindependent situation. Thus we attempt to estimate the number of uncounted by way of estimating the sample coverage.

The common definition for the sample coverage of a given sample is the probability-weighted fraction of the population discovered in that sample. Now this concept must be slightly modified in a census problem, because there is more than one set of "population probabilities." For illustrative purposes, assume for the moment that there is no list dependence but there is heterogeneity, and $P(X_{ij} = 1) = p_{ij}$ is defined as in Section 2.1.1. For a two-list case, there are two populations: population I with capture probabilities $\{p_{11}, p_{21}, \ldots, p_{N1}\}$, which corresponds to list 1, and population II with probabilities $\{p_{12}, p_{22}, \ldots, p_{N2}\}$, which corresponds to list 2. For given records of list 1, define the sample coverage of list 1 with respect to population II, $C^*_{\mathrm{II}}(L_1)$, as

$$C^*_{\mathrm{II}}(L_1) = \frac{\sum_i p_{i2} I[X_{i1} > 0]}{\sum_i p_{i2}},$$

where L_1 denotes list 1. The quantity $C^*_{\mathrm{II}}(L_1)$ represents the probability-weighted fraction of population II that appears in list 1. Here the sample coverage of list 1 is defined with respect to the associated population of the other list, so we can obtain the overlap or jointly covered information, which is essential to the estimation of undercount. Intuitively, the greater the overlap, the fewer the uncounted.

If list dependence also exists, then the probabilities for population II given list 1 must be changed to $\{E(X_{i2}|X_{i1}); i = 1, 2, \ldots, N\}$. Replacing p_{i2} in the foregoing definition of $C^*_{\mathrm{II}}(L_1)$ by $E(X_{i2}|X_{i1})$, we have the following generalized definition:

$$C_{\mathrm{II}}(L_1) = \frac{\sum_i E[X_{i2}|X_{i1}] I[X_{i1} > 0]}{\sum_i E[X_{i2}|X_{i1}]}.$$

We can define $C_{\mathrm{I}}(L_2)$ similarly. The sample coverage of the two lists is taken as the average of these two quantities. We have $C = [C_{\mathrm{II}}(L_1) + C_{\mathrm{I}}(L_2)]/2$; that is,

$$C = \frac{1}{2}\left\{\frac{\sum_i E[X_{i2}|X_{i1}] I[X_{i1} > 0]}{\sum_i E[X_{i2}|X_{i1}]} + \frac{\sum_i E[X_{i1}|X_{i2}] I[X_{i2} > 0]}{\sum_i E[X_{i1}|X_{i2}]}\right\}.$$

We could consider a weighted average of $C_{\mathrm{II}}(L_1)$ and $C_{\mathrm{I}}(L_2)$ in the definition of C, but simulation results have shown the differences with respect to population estimation are quite limited. Because an expectation of a ratio is approximately equal to a ratio of expectations, we have $E(C) \approx (EZ_{11}/En_2 + EZ_{11}/En_1)/2$. Thus an estimator of $E(C)$ that is also valid in the nonindependent cases is

$$\hat{C} = \frac{1}{2}\left(\frac{Z_{11}}{n_1} + \frac{Z_{11}}{n_2}\right) = 1 - \frac{1}{2}\left(\frac{Z_{10}}{n_1} + \frac{Z_{01}}{n_2}\right). \quad (9)$$

In the special case in which neither type of dependence exists, the sample coverage reduces to $C = D/N$, where $D = \{\sum_i I[X_{i2} > 0] + \sum_i I[X_{i1} > 0]\}/2 = (n_1 + n_2)/2$. Therefore, an estimator of N in this case is $\tilde{N}_0 = D/\hat{C} = (n_1 + n_2)/(2\hat{C})$. Substituting $\hat{C} = Z_{11}(1/n_1 + 1/n_2)/2$ into it, $\tilde{N}_0$ is simply the well-known estimator for two samples $\tilde{N}_0 = n_1 n_2 / Z_{11}$. If any type of dependence occurs, then we can easily verify the following identity:

$$N = \frac{(En_1)(En_2)}{EZ_{11}}(1 + \gamma_{12}). \quad (10)$$

As shown in (8), $\gamma_{12} = \mathcal{P}_{12}/(\mathcal{P}_1\mathcal{P}_2) - 1$ in the case of list dependence only, and the identity (10) under this circumstance was noted by the International Society of Disease Monitoring and Forecasting (1995, p. 1050). If only heterogeneity exists, then $\gamma_{12} = N^{-1}\sum_i p_{ij}p_{ik}/(\mu_j\mu_k) - 1$, and (10) was previously derived by Seber (1982, p. 86), Sekar and Deming (1949, p. 107), and Wolter (1986, p. 341). In the latter case, if the probabilities of inclusion for one sample are constant, then $\gamma_{12} = 0$; the other sample could be highly heterogeneous without inducing any correlation bias, provided that the independence within individuals is valid. When both types of dependences occur, γ_{12} as defined in (1) is a measure of confounded dependence effect.

In practical applications, we need to estimate γ_{12} to remove the correlation bias. However, there is an unidentifiability problem in a dual-system approach, because there are four parameters, μ_1, μ_2, N, and γ_{12} but only three observed cells in the data. That is, the information is not sufficient for estimating γ_{12} for a dual-system approach unless γ_{12} is set to some value. The usual estimator for two samples is obtained by setting $\gamma_{12} = 0$, which is the independence assumption—yet in many practical situations the value of γ_{12} can be much different from 0, and any assumption about γ_{12} is untestable. Thus the independence assumption has become the main weak point in the use of the usual estimator for two samples.

It follows from the definition of γ_{12} or (10) that $\gamma_{12} = NE(Z_{11})/[E(n_1)E(n_2)] - 1$. Therefore, if an initial estimate $\tilde{N}$ of N is available based on some external information, then the estimate $\hat{\gamma}_{12} = \tilde{N}Z_{11}/(n_1n_2) - 1$ measures the overall dependence. Although the interpretations for γ_{12} are different for various situations, they are estimated by the same quantity. Thus a large value of the CCV estimate could be due to list dependence and/or to heterogeneity across individuals. However, any external information cannot help in obtaining a bias-corrected estimator of N based on (10), because substitution of $\tilde{N}Z_{11}/(n_1n_2) - 1$ for $\hat{\gamma}_{12}$ will result in $\tilde{N}$ itself, the bias-corrected estimate.

2.3 Three-Sample Case

We now extend the concept of the sample coverage to a three-list case. Given the combined records of the first two lists, $L_1 \cup L_2$, the sample coverage of the two lists with respect to population III having probabilities $\{E(X_{i3}|X_{i1}, X_{i2}); i = 1, 2, \ldots, N\}$ associated with list 3 can be analogously defined as

$$C_{\mathrm{III}}(L_1 \cup L_2) = \frac{\sum_i E[X_{i3}|X_{i1}, X_{i2}]I[X_{i1} + X_{i2} > 0]}{\sum_i E[X_{i3}|X_{i1}, X_{i2}]}. \tag{11}$$

We can also consider the other two possible combinations of two samples and obtain the sample coverage as $C = \{C_{\mathrm{III}}(L_1 \cup L_2) + C_{\mathrm{II}}(L_1 \cup L_3) + C_{\mathrm{I}}(L_2 \cup L_3)\}/3$. If neither type of dependence exists, then $C = D/N$, where

$$D = \frac{1}{3}\left\{\sum_i I[X_{i2} + X_{i3} > 0] + \sum_i I[X_{i1} + X_{i3} > 0] + \sum_i I[X_{i1} + X_{i2} > 0]\right\}, \tag{12}$$

which is the average of the number of individuals listed in the combination of two lists omitting the third. From the definition of C, we can derive as in a two-sample case that

$$E(C) \approx 1 - \frac{1}{3}\left\{\frac{\sum_i P\,[X_{i1} = 1, X_{i2} = X_{i3} = 0]}{\sum_i E(X_{i1})} + \frac{\sum_i P\,[X_{i2} = 1, X_{i1} = X_{i3} = 0]}{\sum_i E(X_{i2})} + \frac{\sum_i P\,[X_{i3} = 1, X_{i1} = X_{i2} = 0]}{\sum_i E(X_{i3})}\right\}. \tag{13}$$

An estimator of the sample coverage is

$$\hat{C} = 1 - \frac{1}{3}\left(\frac{Z_{100}}{n_1} + \frac{Z_{010}}{n_2} + \frac{Z_{001}}{n_3}\right), \tag{14}$$

which is an obvious extension of (9). Thus when all three samples are independent, a natural estimator is

$$\hat{N}_0 = D/\hat{C}. \tag{15}$$

If any type of dependence arises, then it follows from (13) and the expectation of D that

$$N = \frac{E(D)}{E(C)} + \frac{N}{3E(C)} \times [(\mu_1 + \mu_2)(1 - \mu_3)\gamma_{12} + (\mu_1 + \mu_3)(1 - \mu_2)\gamma_{13} + (\mu_2 + \mu_3)(1 - \mu_1)\gamma_{23} - (\mu_1\mu_2 + \mu_1\mu_3 + \mu_2\mu_3)\gamma_{123}]. \tag{16}$$

This identity expresses the population size as a function of expected sample coverage, average probabilities, and dependence measures. When there is no list dependence and heterogeneity exists only in one sample, it is interesting to note that the usual estimator assuming independence is still valid, because all CCVs are 0. Hence the correlation bias arises only when at least two samples are heterogeneous in capture probabilities.

Rewrite (16) as

$$N = \frac{E(D)}{E(C)} + \frac{1}{3E(C)} \times [E(Z_{1+0} + Z_{+10})\gamma_{12} + E(Z_{10+} + Z_{+01})\gamma_{13} + E(Z_{01+} + Z_{0+1})\gamma_{23}] + \frac{R}{3E(C)}, \tag{17}$$

where

$$R/N = \mu_1\mu_2[\gamma_{12}(\gamma_{13} + \gamma_{23}) - \gamma_{123}] + \mu_1\mu_3[\gamma_{13}(\gamma_{12} + \gamma_{23}) - \gamma_{123}] + \mu_2\mu_3[\gamma_{23}(\gamma_{12} + \gamma_{13}) - \gamma_{123}]. \tag{18}$$

There are several situations under which the remainder term R vanishes:

- If neither list dependence nor heterogeneity exists, then $\gamma_{12} = \gamma_{13} = \gamma_{23} = \gamma_{123} = 0$ and thus $R = 0$.
- If there is no list dependence but heterogeneity only exists in one sample, then all CCVs are 0, as mentioned earlier. Therefore, we have $R = 0$.
- There is no heterogeneity and list dependence only occurs in two samples, for example, in lists 1 and 2. Then $\gamma_{13} = \gamma_{23} = \gamma_{123} = 0$, which implies $R = 0$.
- In a heterogeneous random-effects model M_{th} [see (7)], where $P\,(X_{ij} = 1|h_i) = h_ie_j$ and $(h_1, h_2, \ldots, h_N)$ are a random sample from a gamma distribution with density $\beta^\alpha \exp(-\beta h)h^{\alpha-1}/\Gamma(\alpha)$, theoretically it can be proved that $\gamma_{12} = \gamma_{13} = \gamma_{23} = 1/\alpha$ and $\gamma_{123} = 2/\alpha^2$, which implies $R = 0$. Because the gamma distribution can cover a wide range of types of heterogeneity, as shown by Fisher, Corbet, and Williams (1943), this is our main motivation for ignoring R.

Some relevant discussion in Section 4 regarding the numerical values of R/N under the Rasch models and other situations further justifies using (17) without considering the remainder term R. Thus our basic assumption $R = 0$ is

equivalent to setting the right side of (18) to be 0, which implies that γ_{123} is a function of $\mu_1, \mu_2, \mu_3, \gamma_{12}, \gamma_{13}$, and γ_{23}. Hence it avoids the estimation of γ_{123}, and there is no problem in identification because there are seven cell counts corresponding to seven free parameters $\mu_1, \mu_2, \mu_3, \gamma_{12}, \gamma_{13}, \gamma_{23}$, and N. Replacing all of the expected values by the observed values in (17) and ignoring the remainder term, we can express N as a function of the sample coverage estimate and the CCVs between two samples:

$$N \approx \frac{D}{\hat{C}} + \frac{1}{3\hat{C}}\left[(Z_{1+0} + Z_{+10})\gamma_{12} + (Z_{10+} + Z_{+01})\gamma_{13} + (Z_{01+} + Z_{0+1})\gamma_{23}\right]. \quad (19)$$

Note that $\gamma_{12} = NE(Z_{11+})/[E(n_1)E(n_2)] - 1$, and similar expressions for γ_{13} and γ_{23}. Substituting $\gamma_{12} = NZ_{11+}/(n_1 n_2) - 1, \gamma_{13}$, and γ_{23} with the same forms into (19) results in an estimating equation of N. The solution of that estimating equation is our proposed estimator, which turns out to be

$$\hat{N} = \frac{Z_{+11} + Z_{1+1} + Z_{11+}}{3\hat{C}} \times \left\{1 - \frac{1}{3\hat{C}}\left[\frac{(Z_{1+0} + Z_{+10})Z_{11+}}{n_1 n_2} + \frac{(Z_{10+} + Z_{+01})Z_{1+1}}{n_1 n_3} + \frac{(Z_{0+1} + Z_{01+})Z_{+11}}{n_2 n_3}\right]\right\}^{-1}. \quad (20)$$

We remark that simulation experiences have suggested that if the sample coverage is high enough (say, over 55%), which is generally valid for census applications, then $\hat{N}$ performs well. However, when the sample coverage is low, $\hat{N}$ behaves unstably (see Chao, Tsay, Shau, and Chao 1996).

As N tends to infinity, we can show that

$$\frac{\hat{N}}{N} \xrightarrow{P} \alpha = 1 - \frac{R/N}{A + R/N}, \quad (21)$$

where $\xrightarrow{P}$ denotes convergence in probability, R/N is given in (18), and

$$A = \mu_1\mu_2(\gamma_{12} + 1) + \mu_1\mu_3(\gamma_{13} + 1) + \mu_2\mu_3(\gamma_{23} + 1).$$

Here R is exactly the remainder term given in (17). The smaller the magnitude of the remainder term, the smaller the absolute bias of the proposed estimator. In the cases of $R = 0$ as discussed earlier, $\hat{N}/N$ converges to 1 in probability as N tends to infinity.

A variance estimator of the proposed estimators $\hat{N}$ can be constructed using a nonparametric bootstrap procedure. The "capture" histories, as shown by Darroch et al. (1993), are approximately distributed as a multinomial distribution in many situations. A bootstrap replication $\{Z^*_{000}, Z^*_{100}, \ldots, Z^*_{111}\}$ for capture histories is generated from the multinomial distribution with parameter $\hat{N}$ and cell probabilities $\{\hat{Z}_{000}/\hat{N}, Z_{100}/\hat{N}, \ldots, Z_{111}/\hat{N}\}$, where $\hat{Z}_{000} = \hat{N} - \sum Z_{ijk} I[i + j + k > 0]$ is the estimated unidentified. For each bootstrap replication, the bootstrap estimate $\hat{N}^*$ and the bootstrap missing count $\hat{N}^* - \sum Z^*_{ijk} I[i + j + k > 0]$ are obtained based on the seven observable cells $\{Z^*_{100}, \ldots, Z^*_{111}\}$. After B replications, the bootstrap standard error of $\hat{N}$ is simply the sample standard error of those B bootstrap population estimates. The performance of the bootstrap variance estimates is investigated in Section 4.

The method developed for three samples conceptually can be extended to four samples using parallel arguments. Define the sample coverage as

$$C = \frac{1}{4}\sum_{l=1}^{4}\left\{\sum_{i=1}^{N} E[X_{il}|X_{ij}, j \neq l] I\left[\sum_{j\neq l}^{4} X_{ij} > 0\right] \div \sum_{i=1}^{N} E[X_{il}|X_{ij}, j \neq l]\right\},$$

and let

$$D = \frac{1}{4}\sum_{l=1}^{4}\left\{\sum_{i=1}^{N} I\left[\sum_{j\neq l}^{4} X_{ij} > 0\right]\right\}.$$

A more general approximation formula similar to (19) can be derived. Parallel arguments result in an analogous estimator for four samples. However, further efforts are needed to investigate the performance of this estimation procedure for four samples.

3. EXAMPLE

We use the data discussed by Darroch et al. (1993) and Zaslavsky and Wolfgang (1993) to illustrate our method. The data consists of a population subgroup from the 1988 dress rehearsal census in St. Louis and its PES. The additional sample was compiled from precensus administrative records of state and federal government agencies (see Zaslavsky and Wolfgang 1993 for details of the data). The data are reproduced in Table 1. There are two sampling strata (stratum 11 and strata 11, 12, and 13; stratum 11 is involved in both cases) and three lists: the E list, the dress rehearsal census; the P list, the PES; and A list, the administrative records. For each sampling stratum, there are four poststrata defined by housing tenure type and age: O2 (owners, 20–29 years), R2 (renters, 20–29 years), O3 (owners, 30–44 years), and R3 (renters, 30–44 years). The total counted, the sample coverage estimate $\hat{C}$ [see (14)], and D [given in (12)] are also tabulated for each poststratum. In these sets, the sample coverage estimates are between 71.7% and 81.1%, the estimated average probabilities (unreported) are in the range .2–.6, and all of the estimated two-sample CCVs are positive, with most less than .5.

Darroch (1981) and Fienberg (1981) discovered that the Rasch model was equivalent to (except for some moment restrictions) a quasi-symmetric log-linear model in a contingency table. Moreover, Darroch et al. (1993) built the equivalence for the generalized Rasch model and the par-

Table 1. Three-List Data From the 1988 Dress Rehearsal Census (Zaslavsky and Wolfgang 1993)

Lists			Stratum 11				Strata 11, 12, 13			
E	P	A	O2	R2	O3	R3	O2	R2	O3	R3
0	0	1	59	43	35	43	59	43	35	43
0	1	0	8	34	10	24	65	70	69	53
0	1	1	19	11	10	13	19	11	10	13
1	0	0	31	41	62	32	75	73	77	71
1	0	1	19	12	13	7	19	12	13	7
1	1	0	13	69	36	69	217	144	262	155
1	1	1	79	58	91	72	79	58	91	72
Column total			228	268	257	260	533	411	557	414
D			195	229	221	227	467	349	500	358
$\hat{C}$ (%)			79.3	74.4	79.7	79.0	76.7	71.7	81.1	75.6

tial quasi-symmetric model. Based on these equivalence relations, they analyzed these data using log-linear models. The data for each poststratum were regarded as a form of an incomplete 2^3 contingency table for which the cell corresponding to those individuals uncounted by all three lists was missing. Then various log-linear models were fitted to the observed cells, and the chosen model was projected onto the unobserved cell by assuming that there was no three-factor interaction. In other words, the exponential of the three-factor interaction, ρ, where

$$\rho = \frac{EZ_{111}EZ_{100}EZ_{010}EZ_{001}}{EZ_{000}EZ_{011}EZ_{101}EZ_{110}}, \tag{22}$$

is assumed to be 1. This yields an extrapolation formula for the missing cell; that is, $\hat{Z}_{000} = \hat{Z}_{111}\hat{Z}_{100}\hat{Z}_{010}\hat{Z}_{001}/(\hat{Z}_{011}\hat{Z}_{101}\hat{Z}_{110})$, where $\hat{Z}_{ijk}$'s are the fitted values of the chosen model. All of the estimates and their associated standard errors based on various log-linear models are presented in Table 2 (see Darroch et al. 1993 for details). The estimates obtained from fitting independent models are considerably lower than other estimates. Except for the O2 and O3 poststrata in stratum 11, the estimates obtained by fitting partial quasi-symmetric and saturated models are comparable, and both have very large variation. Darroch et al. (1993) generally suggested using a partial quasi-symmetric model based on the goodness-of-fit test for the observed cells.

Table 2 also lists three estimates given by Zaslavsky and Wolfgang (1993). The DSE denoted by $(E \cup P) \times P$ treats the combined census and PES as the first sample and the A list as the second sample; the DSE denoted by $(E\Delta P) \times P$ is obtained similarly, but persons in both the census and the P list are excluded in the first combined list; the estimate $\alpha_{EP} = \alpha_{EP|A}$ is obtained by assuming that the marginal odds ratio between the E and P lists is the same as the conditional odds ratio given in the A list. The standard error estimates were obtained using a jackknife method based on blocks (see Zaslavsky 1994).

Our proposed estimates (20) for the uncounted are also given in Table 2. The estimated standard errors were calculated using a bootstrap method based on 200 replications. Our estimates are much lower than results based on the partial quasi-symmetric model and are generally between the DSE $(E \cup P) \times P$ and the DSE $(E\Delta P) \times P$ given by Zaslavsky and Wolfgang (1993). The relative merits of the log-linear model approach and the sample coverage technique are discussed in the next section.

4. SIMULATION STUDY

We carried out a simulation for several populations with

Table 2. Various Estimates and Standard Errors for the Uncounted (Standard Errors in Parentheses)

Estimator	O2	R2	O3	R3
	Stratum 11			
Log-linear model				
Independent model	14(3)	28(5)	14(3)	18(3)
Saturated model	246(150)	382(203)	422(253)	379(222)
Quasi-symmetric model	553(294)	126(55)	508(254)	102(46)
Partial quasi-symmetric model	378(212)	384(204)	867(472)	352(200)
Zaslavsky and Wolfgang*				
DSE $(E \cup P) \times A$	26(10)	76(27)	33(10)	58(29)
DSE $(E\Delta P) \times A$	61(26)	140(53)	110(55)	120(83)
$\alpha_{EP} = \alpha_{EP\|A}$	130(64)	312(171)	254(202)	305(432)
Proposed (20)	51(12)	114(24)	59(14)	80(17)
	Strata 11, 12, 13			
Log-linear model				
Independent model	45(6)	48(7)	28(4)	35(5)
Saturated model	290(119)	670(335)	497(247)	826(450)
Quasi-symmetric model	47(14)	102(36)	42(13)	82(29)
Partial quasi-symmetric model	292(119)	670(334)	490(242)	768(403)
Zaslavsky and Wolfgang*				
DSE $(E \cup P) \times A$	180(60)	152(52)	125(31)	130(58)
DSE $(E\Delta P) \times A$	217(69)	267(95)	222(94)	267(170)
$\alpha_{EP} = \alpha_{EP\,A}$	285(122)	601(319)	458(368)	729(1039)
Proposed (20)	199(36)	223(43)	154(31)	183(36)

* Standard errors are obtained using a jackknife method based on blocks.

varying degrees of dependence. The focus was on comparing our proposed estimator (20) with the log-linear model approach. We fixed $N = 200, t = 3$ and considered the following three types of sample effects:

- type A: $(a_1, a_2, a_3) = (1, 1, 1)$.
- type B: $(a_1, a_2, a_3) = (1.5, 1, .5)$.
- type C: $(a_1, a_2, a_3) = (3, 2, 1)$.

Four categories (a total of 13 cases) of heterogeneity patterns given herein are reported. Cases 1–9 consider the Rasch model and its generalization; dependence is due purely to heterogeneity. In Cases 10–13, two types of dependences arise, but no sample effects are assumed. The parameters of the following heterogeneity patterns were chosen so that the sample coverages and average catchabilities would be comparable to those of the real datasets discussed in Section 3. Other trials with lower sample coverages have been discussed by Chao et al. (1996). We report the following cases herein:

Random-Effects Rasch model. $p_{ij} = w_i a_j/(1 + w_i a_j)$, where $(w_1, w_2, \ldots, w_{200})$ for each simulation trial are a random sample from a certain distribution:

- Case 1: $\log w \sim \mathrm{N}(0, 1)$
- Case 2: $w \sim$ exponential distribution with density $\exp(-x), x > 0$
- Case 3: $w \sim$ gamma distribution with density xe^{-x} for $x > 0$

Fixed-Effects Rasch model. $p_{ij} = w_i a_j/(1 + w_i a_j)$ where $w_1, \ldots, w_{200}$ are constants:

- Case 4: $w_i = 2$ for $i =$ 1–100, and $w_i = 1/2$ for $i =$ 101–200
- Case 5: $w_i = 7$ for $i =$ 1–50, and $w_i = 2/5$ for $i =$ 51–200
- Case 6: there are four groups of 50 individuals each: $w_i = 7$ for $i =$ 1–50, $w_i = 1/2$ for $i =$ 51–100, $w_i =$ $1/2.5$ for $i =$ 101–150, and $w_i = 1/3$ for $i =$ 151–200

Generalized (Fixed-Effects) Rasch model. $p_{ij} = w_i a_j/(1 + w_i a_j)$ for $j = 1, 2$ and $p_{ij} = w_i^* a_j/(1 + w_i^* a_j)$ for $j = 3$:

- Case 7: values of w_i's are the same as in Case 4; $w_i^* = 3$ for $i =$ 1–100 and $w_i^* = 1/3$ for $i =$ 101–200.
- Case 8: values of w_i's are the same as in Case 5; $w_i^* = 2$ for $i =$ 1–50 and $w_i^* = 1/2$ for $i =$ 51–200.
- Case 9: values w_i's are the same as in Case 6; there are four groups for w_i^*: $w_i^* = 1.5$ for $i =$ 1–50, $w_i^* = 1$ for $i =$ 51–100, $w_i^* = 1/1.5$ for $i =$ 101–150, and $w_i^* = 1/2$ for $i =$ 151–200.

Populations with two types of dependences.

- Case 10: $P(X_{i1} = 1) = .8$ for $i =$ 1–50 and $P(X_{i1} = 1) = .4$ for $i =$ 51–200: $P(X_{i2} = 1|X_{i1}) = P(X_{i1} = 1)I[X_{i1} = 0] + \phi P(X_{i1} = 1)I[X_{i1} = 1]$, $\phi = 1.2$; and $P(X_{i3} = 1) = P(X_{i1} = 1)$.
- Case 11: same as in Case 10 except that $\phi = .8$.
- Case 12: $P(X_{i1} = 1) = .7$ for $i =$ 1–100 and $P(X_{i1} = 1) = .3$ for $i =$ 101–200: $P(X_{i2} = 1|X_{i1}) = P(X_{i1} = 1)I[X_{i1} = 0] + \phi P(X_{i1} = 1)I[X_{i1} = 1]$. $\phi = 1.2$; and $P(X_{i3} = 1|X_{i1}) = P(X_{i2} = 1|X_{i1})$.
- Case 13: same as in Case 12 except that $\phi = .8$.

In Cases 10 and 11, besides the heterogeneity among individuals, list dependence also occurs for samples 1 and 2; in Cases 12 and 13, there is additional list dependence for samples 1 and 3. For each type of sample effect (A, B, C) and the first nine cases of the heterogeneity patterns, we tabulate in Table 3 the theoretical values $(\mu_1, \mu_2, \mu_3, \gamma_{12}, \gamma_{13}, \gamma_{23}, \gamma_{123})$, the remainder term R/N [see (18)], the value of α [limiting value of $\hat{N}/N$; see (21)], the value of ρ [see (22)], and the expected sample coverage estimate $E(\hat{C})$. The same quantities are also given for Cases 10–13, but there are no sample effects. The parameters of the simulation cases were chosen so that the CCV for any pair of samples is generally positive to mimic the dependence structure of dress rehearsal data. Except for the exponential case (Case 2), all the values of R/N are relatively small, which numerically justifies using (17) without considering the remainder term. Consequently, it follows from (21) that the asymptotic values of $\hat{N}/N$ (value of α) are close to 1 except for Case 2. An explanation for the different behavior of Case 2 is given near the end of this section.

For each fixed type of sample effect and heterogeneity pattern, 200 datasets were generated. Then for each generated dataset, the proposed sample coverage estimator $\hat{N}$ and the bootstrap standard error estimate (based on 200 replications) were calculated. Finally, these 200 estimates and standard errors were averaged, and the sample standard error and the sample root mean squared error (RMSE) were obtained. The computer program was written in C language and carried out on an IBM RISC 6000 work station. The log-linear model fitting and estimates were obtained using S-plus software. The standard errors associated with each estimate for various log-linear models were provided in S-plus using a delta method. Because the three partial quasi-symmetric models yield very close estimates, only one such estimate is given for each case. Table 4 shows the results for type B sample effects. The conclusions for the other two types of sampling effects are generally consistent and so are not reported. For each case, the average value of the number of individuals listed in at least one sample (M) is also tabulated.

The following summarization is based on broader numerical studies, including the cases described in this section and results of Chao et al. (1996) as well as in previous simulations. First, notice that the estimates under the independent models have severe negative bias when all two-sample CCVs are positive. This result is well known in animal abundance estimation (Carothers 1973; Gilbert 1973) and in the context of census undercount estimation (Kadane et al. 1992). Under the Rasch model (Cases 1–6) and Cases 10–13, estimates based on nonindependent models (i.e., quasi-symmetric, partial quasi-symmetric, and saturated) are generally very close, except for those cases with very large variation (Cases 5 and 6). Under the generalized Rasch model (Cases 7–9), the estimates under both partial quasi-symmetric and saturated models are comparable.

Table 3. Relevant Parameters for Various Cases

Case	μ_1	μ_2	μ_3	γ_{12}	γ_{13}	γ_{23}	γ_{123}	α	R/N	ρ	$E(\hat{C})$
					$(a_1, a_2, a_3) = (1, 1, 1)$						
1	.50	.50	.50	.174	.174	.174	.000	.951	.045	1.000	.793
2	.40	.40	.40	.295	.295	.295	−.003	.880	.086	.734	.739
3	.60	.60	.60	.084	.084	.084	−.017	.972	.033	.869	.854
4	.50	.50	.50	.111	.111	.111	.000	.978	.018	1.000	.778
5	.43	.43	.43	.347	.347	.347	.236	.996	.003	3.126	.740
6	.44	.44	.44	.343	.343	.343	.222	.991	.007	2.979	.743
7	.50	.50	.50	.111	.167	.167	.000	.964	.032	1.000	.787
8	.44	.44	.42	.347	.204	.204	.139	1.013	−.009	1.850	.725
9	.44	.44	.46	.343	.113	.113	.063	1.003	−.002	1.353	.733
					$(a_1, a_2, a_3) = (1.5, 1, .5)$						
1	.58	.50	.36	.145	.184	.224	.005	.951	.041	.997	.783
2	.48	.40	.28	.261	.305	.349	.005	.880	.078	.730	.727
3	.68	.60	.45	.070	.091	.108	−.016	.973	.031	.867	.842
4	.59	.50	.35	.091	.117	.143	.000	.977	.018	.980	.765
5	.51	.43	.32	.269	.379	.488	.258	.990	.008	3.057	.737
6	.51	.44	.32	.267	.373	.480	.243	.985	.011	2.918	.734
7	.59	.50	.37	.091	.168	.205	.000	.963	.031	.942	.779
8	.51	.43	.28	.269	.216	.278	.147	1.014	−.008	1.851	.713
9	.51	.44	.30	.267	.118	.150	.066	1.004	−.002	1.353	.715
					$(a_1, a_2, a_3) = (3, 2, 1)$						
1	.71	.64	.50	.076	.101	.126	−.013	.971	.037	1.019	.871
2	.62	.54	.40	.169	.203	.236	−.038	.910	.094	.721	.829
3	.80	.73	.60	.035	.048	.060	−.012	.984	.025	.849	.919
4	.73	.65	.50	.041	.059	.077	.000	.994	.008	1.123	.869
5	.65	.57	.43	.102	.161	.220	.070	1.019	−.019	2.750	.809
6	.65	.57	.44	.103	.161	.219	.064	1.014	−.015	2.648	.804
7	.73	.65	.50	.041	.088	.115	.000	.989	.014	1.170	.809
8	.65	.57	.42	.102	.095	.129	.041	1.016	−.015	1.774	.807
9	.65	.57	.46	.103	.055	.074	.018	1.005	−.005	1.339	.816
10	.50	.56	.50	.209	.120	.134	.057	1.009	−.009	1.845	.790
11	.50	.44	.50	.009	.120	.103	.037	1.026	−.019	1.312	.744
12	.50	.56	.56	.247	.247	.193	.024	.939	.070	1.599	.828
13	.50	.44	.44	.050	.050	.144	−.031	.963	.027	.804	.739

NOTE: α is the asymptotic value of $\hat{N}/N$, see (21); R/N is the remainder term, see (18); and ρ is the exponential of the three-factor interaction, see (22).

As expected, the standard error associated with an estimate based on a less-restrictive model is larger than the standard error associated with an estimate derived from a restrictive model. For example, the standard error generally increases when the data are fitted in the sequential order of independent, quasi-symmetric, partial quasi-symmetric, and saturated models. Another anticipated phenomenon is that if the sample effects are doubled, (i.e., the sample effect is changed from type B to type C), then the sample coverage increases and all estimators are improved in both standard error and RMSE.

It is clear that the behavior of the log-linear model estimates depends on the value of ρ; see (22). In Cases 1, 3, 4, 7, 9, 11, and 13, where the values of ρ are near 1, the estimates for nonindependent log-linear models generally work well with respect to bias. In other cases where the values of ρ are somewhat higher than 1, the estimates derived from dependent models severely overestimate and also have very large variation. For example, the ρ values in Cases 5 and 6 are about 3, and the sample standard errors become extremely large. This seems to be a main weakness in the log-linear model approach. This may also be due to the nonrobustness of the extrapolation formula based on (22) and to the departure from the assumption $\rho = 1$.

For any nonindependent log-linear estimator, our proposed estimator has better precision for any value of ρ and smaller bias when ρ is large. This implies that the proposed estimator (20) is generally preferable to any of log-linear estimates in terms of RMSE. It is important to note that we do need a sufficient amount of data to obtain stable estimates. As mentioned in Section 2, numerical results have suggested that the proposed estimator performs satisfactorily if the sample coverage is over 55%, which is generally valid in census applications. In other applications such as epidemiological studies, the sample coverage may be lower than 55%; an alternative estimator using this approach has been proposed by Chao et al. (1996).

The simulated average values of $\hat{N}/N$ in Table 4 are very close to the theoretical asymptotic α values given in Table 3. In addition to the sample coverage, the performance of the proposed estimator depends also on the remainder term R/N, which we have ignored in the derivation of (17). The numerical results have shown that if $R/N > (<)\ 0$, then the proposed estimator $\hat{N}$ has a negative (positive) bias, and the larger the magnitude of the remainder term, the larger the magnitude of the bias. These results can be verified theoretically from (21). For each type of sample effect, the largest magnitude of R/N appears in the exponen-

Table 4. Comparison of Various Estimators of Population Size, 200 Trials

Case	Estimator	Estimate	Sample SE	Estimated SE	Sample RMSE
1: M = 161	Independent	176	7.1	5.3	25.5
	Quasi-symmetric	204	28.3	30.2	28.5
	Partial quasi-symmetric	204	30.4	33.8	30.5
	Saturated	205	32.1	36.1	32.4
	Proposed (20)	189	13.2	12.5	17.0
2: M = 139	Independent	159	8.1	6.7	42.3
	Quasi-symmetric	186	29.4	35.2	32.3
	Partial quasi-symmetric	187	32.1	39.2	34.6
	Saturated	188	34.7	42.4	36.6
	Proposed (20)	176	20.8	24.3	31.2
3: M = 178	Independent	188	4.9	4.0	12.5
	Quasi-symmetric	202	16.2	16.6	16.3
	Partial quasi-symmetric	203	18.5	19.1	18.6
	Saturated	203	19.7	20.2	19.9
	Proposed (20)	194	9.0	8.1	10.4
4: M = 165	Independent	184	8.0	6.2	17.8
	Quasi-symmetric	203	22.4	25.1	22.5
	Partial quasi-symmetric	204	25.4	28.8	25.6
	Saturated	205	27.5	30.9	27.9
	Proposed (20)	195	14.4	14.7	15.0
5: M = 144	Independent	160	7.8	5.7	40.8
	Quasi-symmetric	372	167.0	186.0	239.3
	Partial quasi-symmetric	370	187.5	213.0	252.8
	Saturated	383	209.1	241.4	277.5
	Proposed (20)	197	18.9	21.8	19.0
6: M = 145	Independent	161	7.7	5.7	40.1
	Quasi-symmetric	342	140.9	158.3	200.1
	Partial quasi-symmetric	350	167.9	193.4	224.9
	Saturated	375	214.6	253.6	276.4
	Proposed (20)	196	19.7	21.1	20.1
7: M = 164	Independent	181	6.6	5.7	20.4
	Quasi-symmetric	211	25.6	30.8	27.7
	Partial quasi-symmetric	204	27.4	30.4	27.7
	Saturated	206	34.7	36.1	35.2
	Proposed (20)	194	12.9	13.5	13.8
8: M = 146	Independent	167	8.6	6.9	34.1
	Quasi-symmetric	228	53.8	60.5	60.9
	Partial quasi-symmetric	276	95.0	116.1	121.4
	Saturated	285	106.6	133.1	135.8
	Proposed (20)	204	24.2	26.6	24.5
9: M = 150	Independent	174	8.4	7.3	27.8
	Quasi-symmetric	204	35.3	38.4	35.5
	Partial quasi-symmetric	233	72.4	76.0	79.5
	Saturated	241	92.7	98.2	101.0
	Proposed (20)	202	22.3	24.2	22.5
10: M = 167	Independent	181	6.6	4.9	20.5
	Quasi-symmetric	232	39.4	44.4	50.8
	Partial quasi-symmetric	236	44.3	49.1	57.1
	Saturated	239	46.5	52.8	60.9
	Proposed (20)	203	11.3	11.7	11.6
11: M = 167	Independent	190	7.6	6.8	13.4
	Quasi-symmetric	212	25.1	29.0	27.6
	Partial quasi-symmetric	214	28.9	32.5	32.3
	Saturated	216	29.5	33.5	33.3
	Proposed (20)	205	15.8	16.3	16.4
12: M = 163	Independent	172	6.2	3.7	28.8
	Quasi-symmetric	233	47.1	51.9	57.7
	Partial quasi-symmetric	233	47.0	52.4	57.5
	Saturated	232	46.3	52.1	55.9
	Proposed (20)	188	9.7	9.9	15.0
13: M = 163	Independent	188	8.1	7.4	14.9
	Quasi-symmetric	198	21.2	23.0	21.3
	Partial quasi-symmetric	197	20.9	23.0	21.1
	Saturated	196	20.2	22.5	20.5
	Proposed (20)	192	15.4	15.9	17.3

NOTE: True population size = 200; $(a_1, a_2, a_3) = (1.5, 1, .5)$. M = average of the number of individuals counted in the sample.

tial random-effects model (Case 2). Notice that if values of CCVs for two-sample and three-sample CCV are consistently positive, then the remainder term R would tend to be small; see (18) for the expression of R as a function of CCVs. For exponential cases, any pair of samples is strongly positively correlated, whereas the dependence measure for three samples becomes less evident or even negative. This inconsistent dependence structure might help explain the different behavior of the exponential case.

The standard error estimates for the proposed estimator calculated by the bootstrap procedure (column 5 in Table 4 under the heading "estimated SE") are generally satisfactory when compared with the sample standard errors (column 4 under the heading "sample SE"). We remark that for log-linear model estimates, the standard error estimates using the delta method consistently underestimate for the independent model but overestimate for other models.

5. CONCLUDING REMARKS

In conclusion, the numerical comparisons have shown that for three samples with dependence, the log-linear model estimates perform well concerning bias when the three-factor interaction is not much different from 0. However, estimates based on quasi-symmetric, partial quasi-symmetric, and saturated models seem to have great variation. Our proposed estimator, which incorporates the bias due to dependence, is more precise for any value of the three-factor interaction and also less biased for relatively large values of the three-factor interaction. Thus our proposed estimator is generally preferable in terms of RMSE.

The proposed estimation technique can easily be adapted for use in epidemiological studies. (See Chao et al. 1996 for applications to estimate the number of infected for some diseases by merging several incomplete lists of names.) Theoretical and analytic comparisons of the proposed sample coverage and log-linear approaches are still under investigation.

[Received August 1995. Revised July 1997.]

REFERENCES

Agresti, A. (1994), "Simple Capture–Recapture Models Permitting Unequal Catchability and Variable Sampling Effort," *Biometrics*, 50, 494–500.

Alho, J. M., Mulry, M. H., Wurdeman, K., and Kim, J. (1993), "Estimating Heterogeneity in the Probabilities of Enumeration for Dual-System Estimation," *Journal of the American Statistical Association*, 88, 1130–1136.

Bell, W. R. (1993), "Using Information From Demographic Analysis in Postenumeration Survey Estimation," *Journal of the American Statistical Association*, 88, 1106–1118.

Bunge, J., and Fitzpatrick, M. (1993), "Estimating the Number of Species: Recent Developments," *Journal of the American Statistical Association*, 88, 364–373.

Carothers, A. D. (1973), "The Effects of Unequal Catchability on Jolly–Seber Estimates," *Biometrics*, 29, 79–100.

Chao, A., and Lee, S.-M. (1992), "Estimating the Number of Classes via Sample Coverage," *Journal of the American Statistical Association*, 87, 210–217.

Chao, A., Lee, S.-M., and Jeng, S.-L. (1992), "Estimating Population Size for Capture–Recapture Data When Capture Probabilities Vary by Time and Individual Animal," *Biometrics*, 48, 201–216.

Chao, A., Tsay, P. K., Shau W.-Y., and Chao, D.-Y. (1996), "Population Size Estimation for Capture–Recapture Models With Applications to Epidemiological Data," in *Proceedings of the Biometrics Section, American Statistical Association*, pp. 108–117.

Darroch, J. N. (1981), "The Mantel–Haenszel Test and Tests of Marginal Symmetry: Fixed-Effects and Mixed Models for a Categorical Response," *International Statistical Review*, 49, 285–307.

Darroch, J. N., Fienberg, S. E., Glonek, G. F. V., and Junker, B. W. (1993), "A Three-Sample Multiple-Recapture Approach to Census Population Estimation With Heterogeneous Catchability," *Journal of the American Statistical Association*, 88, 1137–1148.

Fienberg, S. E. (1981), "Recent Advances in Theory and Methods for the Analysis of Categorical Data: Making the Link to Statistical Practice," *Bulletin of the International Statistical Institute*, 49, 763–7911.

——— (1992), "An Adjusted Census in 1990? The Trial," *Chance*, 5, 28–38.

Fisher, R. A., Corbet, A. S., and Williams, C. B. (1943), "The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population," *Journal of Animal Ecology*, 12, 42–58.

Gilbert, R. O. (1973), "Approximations of the Bias in the Jolly–Seber Capture–Recapture Model," *Biometrics*, 29, 501–526.

Good, I. J. (1953), "On the Population Frequencies of Species and the Estimation of Population Parameters," *Biometrika*, 40, 237–264.

Hogan, H. (1993), "The 1990 Postenumeration Survey: Operations and Results," *Journal of the American Statistical Association*, 88, 1047–1060.

Hook, E. B., and Regal, R. R. (1993), "Effects of Variation in Probability of Ascertainment by Sources ('Variable Catchability') Upon 'Capture–Recapture' Estimates of Prevalence," *American Journal of Epidemiology*, 137, 1148–1166.

International Society for Disease Monitoring and Forecasting (1995), "Capture–Recapture and Multiple-Record Systems Estimation I: History and Theoretical Development," *American Journal of Epidemiology*, 142, 1047–1058.

Isaki, C. T. (1986), "Bias of the Dual System Estimator and Some Alternatives," *Communications in Statistics, Part A—Theory and Methods*, 15, 1435–1450.

Isaki, C. T., and Schultz, L. K. (1986), "Dual System Estimation Using Demographic Analysis Data," *Journal of Official Statistics*, 2, 169–179.

Kadane, J. B., Meyer, M. M., and Tukey, J. W. (1992), "Correlation Bias in the Presence of Stratum Heterogeneity," Technical Report 549, Carnegie-Mellon University, Dept. of Statistics.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978), "Statistical Inference From Capture Data on Closed Animal Populations," *Wildlife Monographs*, 62, 1–135.

Pollock, K. H. (1976), "Building Models of Capture–Recapture Experiments," *The Statisticians*, 25, 253–260.

——— (1991), "Modeling Capture, Recapture, and Removal Statistics for Estimation of Demographic parameters for Fish and Wildlife Populations: Past, Present, and Future," *Journal of the American Statistical Association*, 86, 225–238.

Rasch, G. (1961), "On General Laws and the Meaning of Measurement in Psychology," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, ed. J. Neyman, Berkeley, CA: University of California Press, pp. 321–333.

Robinson, J. G., Ahmed, B., Das Gupta, P., and Woodrow, K. A. (1993), "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis," *Journal of the American Statistical Association*, 88, 1061–1079.

Seber, G. A. F. (1982), *The Estimation of Animal Abundance* (2nd ed.), London: Griffin.

Sekar, C., and Deming, W. E. (1949), "On a Method of Estimating Birth and Death Rates and the Extent of Registration," *Journal of the American Statistical Association*, 44, 101–115.

Stegelman, W. (1983), "Expanding the Rasch Model to a General Model Having More Than One Dimension," *Psychometrika*, 48, 259–267.

Wolter, K. M. (1986), "Some Coverage Error Models for Census Data," *Journal of the American Statistical Association*, 81, 338–353.

——— (1990), "Capture–Recapture Estimation in the Presence of a Known Sex Ratio," *Biometrics*, 46, 157–162.

Zaslavsky, A. M., and Wolfgang, G. S. (1993), "Triple-System Modeling of Census, Postenumeration Survey, and Administrative-List Data," *Journal of Business and Economic Statistics*, 11, 279–288.