A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability

John N. Darroch; Stephen E. Fienberg; Gary F. V. Glonek; Brian W. Junker

Journal of the American Statistical Association, Vol. 88, No. 423 (Sep., 1993), 1137-1148.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199309%2988%3A423%3C1137%3AATMATC%3E2.0.CO%3B2-S

Journal of the American Statistical Association is currently published by American Statistical Association.

A Three-Sample Multiple-Recapture Approach to Census Population Estimation With Heterogeneous Catchability

JOHN N. DARROCH, STEPHEN E. FIENBERG, GARY F. V. GLONEK, and BRIAN W. JUNKER*

A central assumption in the standard capture–recapture approach to the estimation of the size of a closed population is the homogeneity of the "capture" probabilities. In this article we develop an approach that allows for varying susceptibility to capture through individual parameters using a variant of the Rasch model from psychological measurement situations. Our approach requires an additional recapture. In the context of census undercount estimation, this requirement amounts to the use of a second independent sample or alternative data source to be matched with census and Post-Enumeration Survey (PES) data. The models we develop provide a mechanism for separating out the dependence between census and PES induced by individual heterogeneity. The resulting data take the form of an incomplete 2^3 contingency table, and we describe how to estimate the expected values of the observable cells of this table using log-linear quasi-symmetry models. The projection of these estimates onto the unobserved cell corresponding to those individuals missed by all three sources involves the log-linear model of no second-order interaction, which is quite plausible under the Rasch model. We illustrate the models and their estimation using data from a 1988 dress-rehearsal study for the 1990 census conducted by the U.S. Bureau of the Census, which explored the use of administrative data as a supplement to the PES. The article includes a discussion of extensions and related models.

KEY WORDS: Census undercount; Conditional multinomial model; Log-linear model; Multiple recapture; Quasi-symmetry; Rasch model.

1. INTRODUCTION

Concerns about the accuracy of census counts in the United States have existed almost as long as the census itself. The U.S. Census Bureau has documented the extent of the undercount by age, race, and sex in all of the censuses since 1940. Two basic quantitative techniques have been used to estimate the census undercount: demographic analysis and the dual-system, or capture–recapture, modeling technique.

Demographic analysis puts the overall undercount at 2.9% in 1970, at 1.4% in 1980, and at 1.9% in 1990. The national undercount rate for Blacks has remained roughly 5% higher than for Non-Blacks in every census since 1940 (see Fay, Passel, Robinson, and Cowan 1988; Fienberg 1991). Demographic analysis can be used only at the national level, because we have no method for tracking the movement of individuals among states or other political units.

Dual-system estimation is more useful for estimating the undercount at local levels. It was also the method of choice for the statisticians at the Census Bureau in 1990, although Secretary of Commerce Robert A. Mosbacher decided that there would be no adjustment of official 1990 Census data (Fienberg 1992b; Mosbacher 1991). Following the enumeration component of the 1990 decennial census, the Bureau conducted a sample survey of about 5,000 blocks across the nation and matched the information gathered on the inhabitants of these sample blocks with the results of the census preliminary enumeration. This second count is known as the Post-Enumeration Survey (PES). In all, the PES involved enumeration of the occupants of 165,000 households nationwide. Data from the enumeration process and the PES can be combined in a 2×2 table of counts cross-classifying presence or absence in the original census enumeration with presence or absence in the PES (see Table 1). One cell is missing—the count of those missed by both the census and the PES. The dual-system technique assumes a fixed odds-ratio for this table and estimates the missing cell from this odds-ratio. The estimated undercount (relative to the original enumeration) is then the sum of this estimate and the count of those present in the PES and absent in the original enumeration.

Hogan (1992) described the 1990 PES in detail, including the sampling plan, treatment of nonresponse and erroneous enumeration, and adaptations required to apply dual-systems methods to the census enumeration and PES. Fienberg (1992a) provided an annotated bibliography of work on dual-system and capture–recapture methodology.

After the 1980 U.S. decennial census, considerable attention was given to using extensions of the dual-system methodology for correcting the census for the differential undercount of Blacks and other minorities (see, for example, Ericksen and Kadane 1985; Mulry and Spencer 1991; Wolter 1986). Various perspectives on the problem were provided by Fienberg (1991), Freedman (1991), and Wolter (1991). Leaving aside the special features required to adapt it to the

* John N. Darroch is Professor and Gary F. V. Glonek is Lecturer, School of Information Science and Technology, Flinders University, Adelaide, S.A. 5042, Australia. Stephen E. Fienberg is Professor of Statistics and Law and Vice President for Academic Affairs, York University, North York, Ontario M3J 1P3, Canada. Brian W. Junker is Assistant Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213. Fienberg's work was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Junker's work was partially supported by National Institutes of Mental Health National Research Service Award Grant MH15758 and Office of Naval Research Cognitive Sciences Division Grant N00014-91-J-1208. The authors thank Alan Zaslavsky for providing the up-to-date data used in this article (and for tolerating our comparisons with the methods of Zaslavsky and Wolfgang, 1990) and also the associate editor and two referees for providing insightful comments and suggestions.

Journal of the American Statistical Association
September 1993, Vol. 88, No. 423, Undercount in 1990 Census

Table 1. Dual-System (Two-Sample) Census

	Second sample	
	1	0
First sample		
1	x_{11}	x_{10}
0	x_{01}	$x_{00} = ?$

census context, criticisms of the dual-system methodology focus largely on three basic assumptions:

1. Perfect matching: Individuals in the second list (the PES) can be matched with those in the first list (the census) without error. The matching algorithms used to link census and sample files are not perfect (Jaro 1989), and they can be reconsidered using probabilistic matching integrated with capture–recapture structures (Ding 1990; Ding and Fienberg 1992) or using the probabilistic imputation approach developed for the 1990 PES (Rubin, Schafer, and Schenker 1988; Schenker 1988). In this article we assume perfect matching for simplicity of presentation.
2. Independence of lists: The probability of an individual being included in the first list does not depend on whether he or she was included in the second list. The census and the follow-up sample have been widely viewed as positively related. This can be checked through the use of a third list using the log-linear methods described by Fienberg (1972) and by Bishop, Fienberg, and Holland (1975, chap. 6).
3. Homogeneity: The probability of inclusion on a list does not vary from individual to individual. Following the original advice of Chandrasekar and Deming (1949) and others, this assumption has been addressed by the Bureau of the Census through the use of an extensive poststratification scheme in 1990; but this approach is still subject to challenge. Isaki, Schultz, Diffendal, and Huang (1988) described some of the efforts that went into the development of the census stratification scheme for 1990.

In fact, Assumptions 2 and 3 are related. Suppose that the lists are independent within strata but the probability of capture or inclusion varies across strata. When the strata are combined, the resulting data will in general no longer exhibit independence. Kadane, Meyer, and Tukey (1992) described the impact of collapsing over strata on the resulting population estimates (see also Holland and Rosenbaum 1986).

Our approach is to build models accounting for varying catchability of individuals as well as varying levels of penetration into the target population of each sample or list. We begin with the smallest possible sampling stratum, the individual, and assume that within individuals the lists are indeed independent. These single-individual strata are then combined into more realistic strata, such as the sampling post-strata described by Zaslavsky and Wolfgang (1990, 1993) and summarized in Section 2; the resulting combined counts exhibit the positive dependence due to heterogeneity described by Kadane et al. (1992). The resulting log-linear model for the combined strata contains parameters that represent both list effects and artifacts of the averaging distribution of the individual effects.

With only two samples, all such models essentially reduce to the estimator based on the 2×2 odds-ratio described earlier. With three or more samples, more refined models can be entertained. As a check on the PES methodology in the 1988 census dress rehearsal, the Census Bureau compiled a third list from administrative records; some interesting uses of this third list to improve the undercount estimate from the 2×2 table have been explored by Zaslavsky and Wolfgang (1990, 1993). These data were gathered as part of the dress rehearsal for the 1990 census. In the remainder of this article we develop some varying-catchability models and apply them to this example.

2. DATA FROM THE 1988 DRESS-REHEARSAL CENSUS

Zaslavsky and Wolfgang (1990, 1993) considered a population subgroup from the 1988 Dress-Rehearsal Census in St. Louis, Missouri and its PES. A third source of information was the Administrative List Supplement (ALS), compiled from pre-census administrative records of state and federal government agencies, including Employment Security, driver's license, Internal Revenue Service, Selective Service, and Veteran's Administration records. Zaslavsky and Wolfgang's analyses focused on the 70 blocks of the PES sample design stratum in which most residents were expected to be Black renters. One aim of the Census in this project was to better identify Black male renters, a group believed to be seriously undercounted in past censuses.

The resulting data were compiled into three lists: the E list, the dress-rehearsal enumeration; the P list, compiled from the dress-rehearsal PES; and the A list, compiled from the ALS. The P list and A list were not identical to the raw lists obtained from the PES and ALS, because of matching and classification issues.

Zaslavsky and Wolfgang considered four post-strata of Black males in investigating their population estimators, depending on whether residents owned or rented homes and whether they were age 20–29 or 30–44. The post-strata O2, R2, O3, and R3 represent the cross-classification of these two variables. Finally, two sets of lists were created. One compilation includes all of the E- and P-source persons in the three sampling strata 11, 12, and 13, which the given post-strata cross; the other restricts E- and P-source persons to those living in the one sampling stratum, 11, in which

Table 2. Three-Source Data From the 1988 Dress Rehearsal Census in St. Louis, Missouri

Lists			Stratum 11				Strata 11, 12, 13			
E	P	A	O2	R2	O3	R3	O2	R2	O3	R3
0	0	0	—	—	—	—	—	—	—	—
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

Source: From Zaslavsky and Wolfgang (1990, 1993).

most of the targeted population would be expected to be found. These data were initially considered by Zaslavsky and Wolfgang (1990); a more detailed description and data analysis was presented in Zaslavsky and Wolfgang (1993). Table 2 gives the raw counts for the eight separate tables they considered.

Zaslavsky and Wolfgang presented various triple-system estimates of the number of uncounted people, including the unrestricted no-second-order-interaction model, based on log-linear or log-linear-like models applied to a full three-system table or to various marginal subtables of a three-system table. It is not clear whether formal goodness-of-fit tests are available for the models implied by the six estimates these authors considered, though they provided informal checks of some of the assumptions underlying their estimators. Their main estimation results are presented in Appendix B.

As we noted earlier, the stratum 11 data represent counts for a subset of the individuals counted in the strata 11, 12, and 13 data. Applying the same model at these different levels of aggregation raises questions about whether the (in-)dependence among lists, sampling scheme, and other components stays the same. The models we present in Section 4 contain parameters for the dependence structure, to be estimated from the data, so they should be adaptable to different dependence structures at different levels of aggregation.

3. DUAL- AND TRIPLE-SYSTEM ESTIMATION

The dual-system estimator (DSE) estimates the missing cell count x_{00} in the 2×2 table of counts (Table 1) as $\hat{x}_{00} = r \cdot (x_{10}x_{01}/x_{11})$. The 2×2 cross-product ratio r cannot be estimated and must be set to some ad hoc value; usually one assumes that $r = 1$ (i.e., that the samples are independent of one another).

The incomplete 2^3 table of counts for a triple-system census (Table 3) can be divided into one complete 2×2 subtable and one incomplete 2×2 subtable. If we assume that the cross-product ratio r is the same in both subtables, then the cross-product ratio for the incomplete subtable can be estimated from the complete one, as $\hat{r} = x_{111}x_{001}/x_{101}x_{011}$. Applying the DSE to the incomplete subtable, we obtain

$$\hat{x}_{000} = \hat{r} \cdot (x_{100}x_{010}/x_{110}) = \frac{x_{111}x_{100}x_{010}x_{001}}{x_{011}x_{101}x_{110}}. \quad (1)$$

Table 3. Triple-System (Three-Sample) Census

	Third sample			
	1		0	
	Second sample		Second sample	
	1	0	1	0
First sample				
1	x_{111}	x_{101}	x_{110}	x_{100}
0	x_{011}	x_{001}	x_{010}	$x_{000} = ?$

This is equivalent to assuming that $\rho = 1$, where

$$\rho \equiv \frac{p_{111}p_{100}p_{010}p_{001}}{p_{000}p_{011}p_{101}p_{110}} \quad (2)$$

(the p's are cell probabilities); that is, it is equivalent to assuming that there is no second-order interaction in Table 3. The no-second-order-interaction assumption for the 2^3 table is in some sense analogous to the assumption of independence for the 2×2 table but is one layer deeper. All pairs of sources can exhibit dependence, but the amount of dependence in each pair is assumed to be unaffected by conditioning on the third source. Also, because the cell x_{000} is missing, the assumption $\rho = 1$ is untestable in isolation—just as $r = 1$ is untestable for the incomplete 2×2 table.

The use of three or more lists with a multinomial sampling model was first explored by Darroch (1958) for independent lists and was extended with log-linear models to allow dependence among the lists by Fienberg (1972). El-Khorazaty, Imrey, Koch, and Wells (1977) reviewed the literature on methods for multiple-system estimation and noted the links with the literature on multiple-recapture approaches using log-linear models, and also identified parallel assumptions and issues in the applications to human and wildlife populations.

In Section 4 we build models that allow for varying catchability of individuals as well as varying levels of penetration into the target population of each sample or list. Within individuals—the smallest possible sampling stratum—we assume that the lists are independent; however, different individuals may in general have different probabilities of capture, and the models we propose reflect this possibility with different parameters for each individual's catchability effect. In addition, there are parameters reflecting the different "catch efforts" of the samples producing the lists. We then combine these single-individual strata into more realistic strata, such as the sampling post-strata described by Zaslavsky and Wolfgang (1990, 1993) and summarized in Section 2. Because the actual number of parameters grows with the number of individuals, however, the usual asymptotic statistical estimation theory fails. One might approach the population estimation problem directly in that context, but such an approach is technically complex and requires the number of lists, as well as the population size N, to tend to infinity (Haberman 1977). If we instead consider the observed 2^3 table of counts arising from random sampling of individuals (see, for example, Cressie and Holland 1983; Darroch 1981), or from collapsing the $N \times 2^3$ table cross-classifying individuals with capture patterns (see, for example, Fienberg and Meyer 1983), the likelihood for the observed 2^3 table of counts is a marginal likelihood, averaged over the individual effects.

There are both fixed-effects and random-effects arguments formalizing this approach to the catchability effects, which are really nuisance parameters for the purposes of estimating the count in the missing cell x_{000} or the total population size N. For simplicity of exposition, we restrict ourselves in this article to a fixed-effects argument and leave for a subsequent article the technical details and interpretations of the approach that allows for randomly varying components to represent individuals' contributions to the capture probabilities. Though the two approaches are rooted in different conceptual frameworks with different interpretations and involve very different mathematical arguments, they ultimately lead to identical formulations of models for the observable data of Table 3.

To simplify the models for the 2^3 table, we adapt some variants of the Rasch model, well known in educational statistics and psychological measurement (see, for example, Rasch 1980), to the multiple-recapture setting. This leads to a set of log-linear models for three (or more) lists—census, PES, and ALS—to obtain an estimate of the undercount (or, equivalently, of the size of the population) that accommodates heterogeneity among individuals. Psychometric models such as the Rasch model provide an intuitively appealing way to generate appropriate dependence models for a three-sample census, but the census problem is really different from the psychometric problem: There are only three observations on each individual, and the focus of inference changes from the individual effects (in psychometrics) to the missing-cell count x_{000} (in the multiple-recapture census).

Our modeling approach also has clearly related antecedants in the statistical literature. For example, Burnham and Overton (1978) illustrated a random-effects binomial model for varying catchability that does not accommodate varying list quality. Sanathanan (1972b, 1973) developed estimation methods for the Rasch model in the multiple-recapture setting that are more complicated than the marginal log-linear approach we take. Lee and Chao (1991) presented a number of approaches to heterogeneity problems in animal population problems. Viewed from a contingency table perspective, our methods are also closely related to ideas of Darroch and McCloud (1990).

The heterogeneity models we develop do not constrain the missing cell, x_{000}, so we must embed them in the no-second-order-interaction model. Derivation of the marginal model for the 2^3 table by summing over individual effects does impose certain inequality restrictions on the log-linear parameters; as the number of lists increases, Holland (1990a) conjectured that the inequality restrictions approximately rule out to interactions of second or higher order. Thus there is in principle no reason to stop at three lists, and there may actually be advantages in considering a framework that allows for four or more lists as well.

4. MODELS FOR HETEROGENEOUS CATCHABILITY

Let $\mathbf{j} = (j_1, j_2, j_3)$ represent the capture pattern of an arbitrary individual, so $j_i = 1$ if the individual is on list i and 0 otherwise. If there exist observable covariates that explain the heterogeneity, then a logistic regression model can be built, predicting $x_\mathbf{j}$ from the covariates for each $\mathbf{j} \neq (0, 0, 0)$. The logistic regression model can then be used to estimate the count in the missing cell, x_{000}. This was the approach taken by Alho (1990, 1991) for example. On the other hand, if there are no observable covariates explaining the heterogeneity, or if we are in a stratum defined by the available covariates and some residual heterogeneity exists in the stratum, then there will be "extra dependence" induced in the table due to heterogeneous catchability in addition to whatever dependence exists among the lists.

To develop models for this within-stratum heterogeneity, we assume a fixed, closed population of size N, where each individual h, for $h = 1, 2, \ldots, N$, has his or her own fixed catchability parameters, and consider the hypothetical repetition of the entire triple-system estimation experiment under independent identical conditions. Here repetition of the experiment involves applying the same capture techniques to the same population, where the individuals always retain the same fixed catchabilities. Another way to develop our models is to consider random individual effects wherein each individual samples his or her catchability effects from some distribution for each repetition of the experiment; this leads to precisely the same dependence models as we develop in the remainder of Section 4.

4.1 The Rasch Model

Let us suppose that each individual h has probability $p_h(\mathbf{j})$ of capture pattern $\mathbf{j}$, and actually experiences capture pattern $\mathbf{j}_h = (j_{h1}, j_{h2}, j_{h3})$. We assume independence across individuals, and we also assume that the captures or lists are independent, given h:

$$p_h(\mathbf{j}) = \pi_{1j_1}(h)\pi_{2j_2}(h)\pi_{3j_3}(h)$$
$$= \prod_{i=1}^{3} \pi_{i1}(h)^{j_i}\pi_{i0}(h)^{1-j_i}, \tag{3}$$

where $\pi_{i1}(h) = 1 - \pi_{i0}(h)$ is the probability that individual h is on list i.

The assumption of homogeneous catchability means that the probability of being on each list is independent of h: $\pi_{i1}(h) \equiv \pi_{i1}$. Letting $\beta_i \equiv \log \pi_{i1}/\pi_{i0}$, the probability $p(\mathbf{j})$ of observing the response pattern $\mathbf{j} = (j_1, j_2, j_3)$ is

$$\log p(\mathbf{j}) = \alpha + j_1\beta_1 + j_2\beta_2 + j_3\beta_3, \tag{4}$$

which is the model of independence for the table $x_\mathbf{j}$.

On the other hand, suppose that the individuals have heterogeneous catchability, so that the $\pi_{i1}(h)$ are allowed to depend on h. Continuing to allow for heterogeneity in the catchability of individuals, we assume that the pattern of heterogeneity is the same for all three samples. More precisely, we assume that for any two individuals h and h', the odds ratio

$$\frac{\pi_{i1}(h)\pi_{i0}(h')}{\pi_{i0}(h)\pi_{i1}(h')} \tag{5}$$

is constant with respect to i. This assumption is equivalent to the additive-logit model

$$\log \frac{\pi_{i1}(h)}{\pi_{i0}(h)} = t_h + \beta_i, \tag{6}$$

so that capture probabilities are characterized by the logistic function

$$\pi_{i1}(t) = \frac{e^{t+\beta_i}}{1 + e^{t+\beta_i}}.$$

In educational statistics and psychological measurement, (6) is called the Rasch model, after the pioneering work done with it by Georg Rasch in the 1950s (see, for example, Rasch 1980). In such a setting, presence or absence on census lists is replaced by positive and negative responses to examination questions, survey items, and so forth. Duncan (1984) gave several applications and extensions of the Rasch model in survey research problems.

4.2 Quasi-symmetry

The quantity $p_h(\mathbf{j}) = \pi_{1j_1}(h)\pi_{2j_2}(h)\pi_{3j_3}(h)$ is the cell probability for the larger $N \times 2^3$ table $w_{h,\mathbf{j}}$ with cell counts 1 if person h has capture pattern $\mathbf{j}$ and 0 otherwise (see, for example, Fienberg and Meyer 1983). It is easy to see that the cell probabilities for the marginal 2^3 table $x_\mathbf{j}$ must be

$$p(\mathbf{j}) \equiv p_{j_1j_2j_3} = \frac{1}{N}\sum_{h=1}^{N} \pi_{1j_1}(h)\pi_{2j_2}(h)\pi_{3j_3}(h). \quad (7)$$

In many situations the counts $x_\mathbf{j}$ will approximately follow the multinomial distribution

$$N! \prod_{(j_1,j_2,j_3)} \frac{p_{j_1j_2j_3}^{x_{j_1j_2j_3}}}{x_{j_1j_2j_3}!}. \quad (8)$$

Some details of this approximation in the case of mild heterogeneity are considered in Appendix A. The variance–covariance matrix for $x_\mathbf{j}$ under the approximating multinomial (8) is actually larger than it would be under a product-multinomial model for $w_{h,\mathbf{j}}$, so that standard errors of parameter estimates are if anything too large under (8).

Using (3) and (6), we may rewrite the cell probabilities in (7) as

$$\begin{aligned} p(\mathbf{j}) &= \frac{1}{N}\sum_{h=1}^{N}\prod_{i=1}^{3}\left(\frac{\pi_{i1}(h)}{\pi_{i0}(h)}\right)^{j_i}\pi_{i0}(h) \\ &= \frac{1}{N}\sum_{h=1}^{N}\prod_{i=1}^{3} e^{j_i(t_h+\beta_i)}\pi_{i0}(h) \\ &= \exp[j_1\beta_1 + j_2\beta_2 + j_3\beta_3]\frac{1}{N}\sum_{h=1}^{N}[e^{t_h}]^{j_+}p_h(\mathbf{0}); \quad (9) \end{aligned}$$

taking logarithms we obtain a log-linear model for the 2^3 table $x_\mathbf{j}$ of the form

$$\log p(\mathbf{j}) = \alpha + j_1\beta_1 + j_2\beta_2 + j_3\beta_3 + \gamma(j_+), \quad (10)$$

where $j_+ = j_1 + j_2 + j_3$. It follows from Holland (1990a) that

$$\gamma(k) = \log E[e^{kT} \mid \mathbf{j} = \mathbf{0}] \quad (11)$$

where T follows the posterior distribution of the catchability effects t conditional on not being caught in any sample ($\mathbf{j} = \mathbf{0}$). Cressie and Holland (1983) gave a simple proof that Equation (10) holds, with $\exp\gamma(k)$, $k = 0, 1, 2, 3$ being the moment sequence of a positive random variable, if and only if a random-effects model of the form (6) holds. In particular, the $\gamma(k)$ are restricted by the inequalities that any set of log-moments must satisfy, such as $\gamma(2) \geq 2\gamma(1)$, $\gamma(3) + \gamma(1) \geq 2\gamma(2)$, and so forth. The log-linear parameters $\gamma(k)$ identify exactly the dependence due to heterogeneous catchability modeled by (6).

Equation (10) is the model of quasi-symmetry of order one (preserving one-dimensional marginal totals) first proposed by Bishop et al. (1975, chap. 8), together with the moment restrictions on $\gamma(k)$. The representation (10) for the Rasch model has been discovered independently by many authors; among them, Darroch (1981) obtained (10) by randomly sampling strata (h) with constant odds ratio (5), Fienberg (1981) and Fienberg and Meyer (1983) linked the additive representation (6) to quasi-symmetry by collapsing the larger $N \times 2^3$ table, Tjur (1982) showed that maximum likelihood (ML) estimates of the β_i's under a Poisson sampling scheme are identical with conditional ML estimates, given the sums j_+ for each individual, and Cressie and Holland (1983) focused on the statistical modeling implications of (10) for educational measurement. A description of current estimation theory for the Rasch model was given by Lindsay, Clogg, and Grego (1991).

4.3 Conditional Estimation and the Assumption of No Second Order Interaction

Following Sanathanan (1972a) and Fienberg (1972), we analyze the incomplete 2^3 table conditionally. Thus instead of estimating parameters directly from the likelihood (8), we work with the likelihood based on the conditional probability of the observable frequencies, given $n \equiv x_{001} + x_{010} + x_{011} + x_{100} + x_{101} + x_{110} + x_{111}$; that is,

$$n! \prod_{(j_1,j_2,j_3)\neq(0,0,0)} \frac{[p_{j_1j_2j_3}/(1 - p_{000})]^{x_{j_1j_2j_3}}}{x_{j_1j_2j_3}!}. \quad (12)$$

Once the model parameters have been estimated (conditionally, given n) using (12), we must be able to write the cell probability p_{000} in terms of these parameters in order to generate an estimate $\hat{x}_{000}$ for the unobserved cell count.

For $J = 3$ lists, the quasi-symmetry model (10) is equivalent (except for moment restrictions) to the two constraints

$$p(011)p(100) = p(101)p(010) = p(110)p(001), \quad (13)$$

and does not relate the probability $p(000)$ to the other seven probabilities. Thus an additional assumption, such as no second-order interaction, is needed. Under the Rasch/quasi-symmetry model (10), the second-order interaction (2) becomes

$$\rho = \frac{e^{\gamma(3)}e^{3\gamma(1)}}{e^{\gamma(0)}e^{3\gamma(2)}}, \quad (14)$$

where $\gamma(k)$ is defined as in (11). This is of considerable help in seeing how no second-order interaction, $\rho = 1$, might plausibly occur.

If $U = e^T$ is exactly lognormal in (11), then

$$\gamma(k) = ak + bk^2$$

and $\rho = 1$ exactly. In this case the model (10) reduces to

$$\log p(\mathbf{j}) = \alpha + j_1\beta_1 + j_2\beta_2 + j_3\beta_3 + \gamma\cdot(j_+)^2, \quad (15)$$

where now the α, β_j's, and γ are all linear coefficients, with $\gamma > 0$. This is a submodel of the no-second-order-interaction / quasi-symmetry model; for $J = 3$ lists, the models are equivalent except for the restriction that $\gamma > 0$, which is a consequence of the normality assumption.

It turns out that $\rho \approx 1$ for other distributions for U in (11) as well. If U is distributed as a gamma random variable with parameters α and β, then $\exp\gamma(k) = [\Gamma(k + \alpha)]/[\beta^k\Gamma(\alpha)]$ and hence

$$\rho = \frac{\alpha(\alpha + 2)}{(\alpha + 1)^2},$$

so that if α is large, then ρ is close to 1. A similar conclusion obtains if U is inverse gamma or Weibull. More generally, Holland (1990a) conjectured that as the number of lists J grows, U will likely be asymptotically lognormal in (11). When U is bimodal, however—for example, a mixture of two lognormals—the value of ρ can be very different from 1.

The no-second-order-interaction assumption allows us to express $\hat{x}_{000}$ in terms of the conditional ML estimates. In particular, we estimate x_{000} as in (1), but now replace the $x_{j_1j_2j_3}$ with conditional (given n) ML estimates of the corresponding cell means. Finally,

$$\hat{N} = n + \hat{x}_{000} \tag{16}$$

estimates the total population, and $\hat{N} - x_{1++}$ estimates the undercount. As Sanathanan (1972a) showed, $\hat{N}$ and the "unconditional" estimate $\hat{N}_U$, formed by maximizing (8) directly with respect to the unknown parameters and x_{000}, are asymptotically equivalent—and $\hat{N}_U \le \hat{N}$—under suitable regularity conditions. In Section 5 we show how the calculations may be arranged to make the estimation of $\hat{x}_{000}$ and its (conditional) standard error particularly easy.

4.4 Partial Quasi-symmetry

The assumptions of quasi-symmetry and no second-order interaction make the cell probabilities $p_{001}, \ldots, p_{111}$ for the seven observed cell functions of the five parameters α, β_1, β_2, β_3, and γ in (15), leaving 2 degrees of freedom for assessing model fit. The deviances for testing quasi-symmetry in the eight contingency tables displayed in Table 2 are shown in Table 4. (Model fitting and moment restrictions are discussed in Section 5.) The manner in which each table conflicts with quasi-symmetry can be seen from the three frequency products corresponding to the probability products in (13). These are also shown in Table 4.

Table 4. Fit of Models (13) and (17)

Sampling- and Post-Strata	*Deviance from (13) (2 df)*	$x_{011}x_{100}$	$x_{110}x_{001}$	$x_{101}x_{010}$	*Deviance from (17) (1 df)*
Stratum 11					
O2	11.70	589	767	152	7.51
R2	41.09	451	2,967	408	.04
O3	25.99	620	1,260	130	8.27
R3	59.31	416	2,967	168	2.92
Strata 11, 12, 13					
O2	94.04	1,425	12,803	1,235	.15
R2	53.00	803	6,192	840	.01
O3	67.89	770	9,170	897	.11
R3	73.01	923	6,665	371	3.45

There is a consistently large difference between $x_{110}x_{001}$ and the other two products. In fact, the deviations from quasi-symmetry here are related to marginal tests for constant latent odds ratios proposed by Rosenbaum (1987) and by Tjur (1982). For example, their results imply that when the Rasch model holds, if an individual is caught in either the PES or the ALS, but not both, then which list the individual is on should be independent of whether that individual was caught in the original enumeration. This is a 1-degree-of-freedom test based on the cross-product ratio $x_{110}x_{001}/x_{101}x_{010}$; in all of the tables, this independence hypothesis would be rejected.

For several of the tables, the products $x_{011}x_{100}$ and $x_{101}x_{010}$ are fairly close together. Thus it seems reasonable to assume that

$$p(011)p(100) = p(101)p(010). \tag{17}$$

The deviances for testing (17) are also given in Table 4. Property (17) may be interpreted in terms of the individual capture logits (6): Equation (17) arises by assuming that

$$\log \frac{\pi_{i1}(h)}{\pi_{i0}(h)} = t_h + \beta_i, \qquad i = 1, 2$$
$$= s_h + \beta_3, \qquad i = 3 \tag{18}$$

(this specification is related to the generalization of the Rasch model discussed in Stegelman 1983). That is, (17) arises from the assumption (18) that the pattern of heterogeneity is the same for the E samples and P samples only (and different in the A sample).

Indeed, the derivation leading to (10) immediately generalizes to vector-valued catchability effects, as in (18). Following the arguments in Section 4.2, we get a "partial quasi-symmetry" model for the table $x_{\mathbf{j}}$, replacing (10):

$$\log p(\mathbf{j}) = \alpha + j_1\beta_1 + j_2\beta_2 + j_3\beta_3 + \gamma(j_1 + j_2, j_3), \tag{19}$$

where $\gamma(k_1, k_2) = \log E[U^{k_1}V^{k_2} \mid \mathbf{j} = \mathbf{0}]$ for positive random variables $U = e^T$ and $V = e^S$; compare (11). Except for moment restrictions such as $\gamma(2, 0) \ge 2\gamma(1, 0)$ and $\gamma(2, 1) + \gamma(0, 1) \ge 2\gamma(1, 1)$, the model (19) is equivalent to the model (17).

Assuming that the model (19) holds, we examine the difference between $p_{110}p_{001}$ and $p_{011}p_{100}$ ($=p_{101}p_{010}$) to understand how this difference can turn out to be large and positive. We find that

$$\frac{p_{110}p_{001}}{p_{011}p_{100}} = \frac{e^{\gamma(2,0)}e^{\gamma(0,1)}}{e^{\gamma(1,1)}e^{\gamma(1,0)}} = \frac{E[U^2]E[V]}{E[UV]E[U]} > 1$$

if and only if

$$E[V]\text{var}(U) > E[U]\text{cov}(U, V).$$

Keeping in mind that $U = e^T$ and $V = e^S$ for the catchability effects t and s in (18), conditional on not being caught in any sample, it is clear that if s and t are weakly or negatively associated among the uncaught individuals, then $p_{110}p_{001} > p_{011}p_{100}$ will result. The assumption that catchability in the E or P lists is weakly associated with catchability in the

A list is made plausible by the radically different way in which the A list was constructed—essentially, exhaustive searches of administrative records for a particular geographical area covered by the P list. Similarly, Zaslavsky and Wolfgang (1990, 1993) showed that, conditional on membership in the A list, the E and P lists are highly dependent, whereas conditional on membership in the E or P list, the A list is nearly independent of the remaining list.

5. FITTING THE MODELS

The models we have developed for $\mathbf{p} = (p_{000}, p_{001}, \ldots, p_{111})$ may be fit as conventional log-linear models $\log \mathbf{p} = \mathbf{M}\boldsymbol{\theta}$ for an appropriately chosen design matrix $\mathbf{M}$ and vector of parameters $\boldsymbol{\theta}$, provided that the moment restrictions on the γ's are ignored. This implies a potential loss of efficiency in estimating the model parameters (Cressie and Holland 1983). This is less of a concern for three-list data than it would be for more than three lists, where higher-order moment inequalities could seriously restrict acceptable values of the γ's.

Because most computer packages provide a list of the parameter estimates together with their standard errors, it is convenient to arrange for $\log p(000)$ to be one of these parameters. To make $\log p_{000} = \theta_1$, it is necessary and sufficient that we take the first row of M to be $(1, 0, 0, \ldots, 0)$. Then the (conditional, given n) standard error for $\log \hat{p}_{000}$ can simply be read off the computer output. For the quasi-symmetry/no-second-order-interaction model,

$$\log p(j_1, j_2, j_3) = \alpha + j_1\beta_1 + j_2\beta_2 + j_3\beta_3 + \gamma(j_1, j_2, j_3), \quad (20)$$

where, recalling (14),

$\gamma(j_1, j_2, j_3) \equiv \gamma(j_+)$ is symmetric and

$$\gamma_{111} + \gamma_{100} + \gamma_{010} + \gamma_{001} - \gamma_{110} - \gamma_{101} - \gamma_{011} - \gamma_{000} = 0. \quad (21)$$

The model of Equations (20) and (21) turns out to be equivalent to

$$\log p(j_1, j_2, j_3) = \alpha + j_1\beta_1 + j_2\beta_2 + j_3\beta_3 + \gamma \cdot (j_+)^2 \quad (22)$$

(see Darroch, Fienberg, Glonek, and Junker 1991), where now γ is a multiplicative constant. The design matrix can be read off as

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 4 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 4 \\ 1 & 1 & 1 & 0 & 4 \\ 1 & 1 & 1 & 1 & 9 \end{pmatrix}.$$

As remarked in Section 4.3, the model (22) may also be obtained by assuming a lognormal distribution for $U = e^T$ in (11), with the additional restriction that $\gamma > 0$. Our fits of (22) to the data in Table 2 were consistent with the log normal assumption, in that $\hat{\gamma} > 0$ in each of the eight post-strata. For $J > 3$ lists, the model defined by (20) and (21) is distinct from the model (22), and both of these are in general distinct from the model (10) with moment restrictions on the function $\gamma(k)$.

Because the model (19), in which $\gamma(j_+)$ is replaced by $\gamma(j_1 + j_2, j_3)$, provides a plausibly better fit, we also wish to estimate the population size based on it. We construct the design matrix in two stages: the main effects are the same as in the design matrix for (22), and the higher-order effects are represented as the intersection of the column spaces for no second-order interaction and for the additional structure of $\gamma(j_1 + j_2, j_3)$. The resulting design matrix is

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 2 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 2 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 2 \end{pmatrix}.$$

Fitting this model to the post-strata in Table 2 produced parameter estimates that were again consistent with the moment restrictions on $\gamma(k, l)$; for example, if it is assumed that $\exp\gamma(k, l) \equiv E[U^k V^l | \mathbf{j} = \mathbf{0}] = 1$ whenever $k + l \leq 1$, then $\hat{\gamma}(2, 0) > 2\hat{\gamma}(1, 0)$ and $\hat{\gamma}(2, 1) + \hat{\gamma}(0, 1) > 2\hat{\gamma}(1, 1)$. The models were fit using the `glim` and `glim.print` functions for S (Becker, Chambers, and Wilks 1988), which are available by electronic mail request from `statlib@stat.cmu.edu`; see also McCullagh and Nelder (1983). (It should be noted that for the incomplete table the first row of each design matrix is omitted, and for `glim` the first column is also omitted.) Further details on obtaining the design matrices and fitting the models may be found in Darroch et al. (1991).

6. RESULTS

6.1 Some Interesting Poststrata

Tables 5 and 6 give the conditional estimates for x_{000}, given n, and goodness-of-fit statistics for the four models we consider here:

- the no-second-order-interaction model, $\rho = 1$ in (2), without additional restrictions
- the submodel assuming partial quasi-symmetry (19) or, equivalently, (17)
- the submodel assuming full quasi-symmetry (10) or, equivalently, (13)
- the model of complete independence among the three lists (4).

We are mostly interested in the partial and full quasi-symmetry models.

The full quasi-symmetry model provides a much better fit to the post-strata in sampling stratum 11 than does the complete independence model; and the partial quasi-symmetry model provides comparable improvements in fit over the full quasi-symmetry model. Independence, however, does not fit very well (see Table 5). Full quasi-symmetry fits

Table 5. Log-Linear Model Estimates for $\hat{x}_{000}$ Stratum 11 Data

Sampling Stratum 11	*O2*	*R2*	*O3*	*R3*
$\hat{x}_{000}$: No second-order interaction	246.31	381.71	421.99	378.68
SE (delta method)	149.63	203.07	253.3	222.21
Fit/df	—/0	—/0	—/0	—/0
$\hat{x}_{000}$: Partial quasi-symmetry	377.66	384.34	866.99	351.78
SE (delta method)	211.62	204.14	472.33	199.85
Fit/df	7.51/1	.04/1	8.27/1	2.92/1
$\hat{x}_{000}$: Full quasi-symmetry	552.83	126.34	508.44	101.83
SE (delta method)	293.87	55.32	253.56	46.23
Fit/df	11.70/2	41.09/2	25.99/2	59.31/2
$\hat{x}_{000}$: Independent lists	13.79	28.43	14.32	18.21
SE (delta method)	2.70	4.78	2.67	3.22
Fit/df	72.59/3	54.83/3	90.19/3	76.20/3

post-stratum O2 (Black males, age 20–29, living in owned homes) the best. The partial quasi-symmetry model appears to overfit post-stratum R2; the next best fit is to post-stratum R3 with a model deviance of 2.92 on 1 degree of freedom. Turning to sampling strata 11, 12, and 13 combined (see Table 6), we see that full quasi-symmetry fits almost as poorly as the independence model in all of the post-strata, but partial quasi-symmetry does very well, with a marginal to poor fit only in post-stratum R3 and evidence of overfit elsewhere.

Measures of fit and standard errors were obtained from the `glim` analyses. Cormack and Jupp (1991) showed that log-linear parameter estimates and standard errors (up to order $N^{-1/2}$) obtained under `glim`'s Poisson sampling assumptions for the observed counts in the incomplete table $x_{\mathbf{j}}$ are the same as those obtained under the conditional multinomial sampling model (12). To obtain the standard errors of $\hat{x}_{000}$ given in Tables 5 and 6, we apply the delta method to get $\widehat{\mathrm{SE}}^2(\hat{x}_{000}|n) = \hat{x}_{000}^2\hat{\sigma}_{\hat{\alpha}}^2$, where $\alpha = \log Np_{000}$, $\hat{\alpha}$ is the MLE of α under each given log-linear model, and $\hat{\sigma}_{\hat{\alpha}}$ is the estimated standard error of $\hat{\alpha}$, conditional on n. Following the development in Bishop et al. (1975, chap. 6) of Darroch's (1958) methods for obtaining asymptotic unconditional standard errors for $\hat{N} = n + \hat{x}_{000}$, it is easy to see that $\mathrm{SE}(\hat{N}) = \{\mathrm{SE}^2(\hat{x}_{000}|n) + x_{000} + x_{000}^2/n + O(1)\}^{1/2}$ as n (and x_{000}) tends to infinity. Standard errors for $\hat{N}$ calculated in this manner for the four log-linear models in Tables 5 and 6 do not differ appreciably from the tabulated conditional standard errors for $\hat{x}_{000}$. To the extent that a multinomial sampling model is acceptable for the table $x_{\mathbf{j}}$, these measures provide a useful way of assessing the appropriateness of the dependence models we have constructed, as well as the uncertainty in our population estimators.

6.2 Comparison with Zaslavsky and Wolfgang

Comparing these results with the estimates that Zaslavsky and Wolfgang (1990, 1993) obtained (see App. B), we note first that their "$k_3 = 1$" estimator is exactly the unconstrained no-second-order-interaction model. Their "P + ALS" estimator is a DSE applied to the 2×2 marginal table with the P and A lists combined and produces estimates about like the independence model estimates here. This could be expected, because this marginal 2×2 table and estimate would result by assuming independence for all three lists and then combining the P and A lists as Zaslavsky and Wolfgang did. Their "k_2" estimator is constructed in a manner similar to the unconstrained no-second-order-interaction estimator, except that the cross-product ratio from the $A = 1$ subtable is applied to the *marginal* $E \times P$ table instead of to the *conditional* $E \times P$ table, given $A = 0$; this estimator behaves roughly like the unconstrained no-second-order-interaction estimator, as also might be expected. Finally, their "ratio r_1" and "ratio r_2" estimators, which are DSE's for particular marginal 2×2 tables, appear to produce estimates intermediate between the full and partial quasi-symmetry models examined here.

Thus, with the exception of the "ratio r_1" and "ratio r_2" estimators, all of the Zaslavsky and Wolfgang estimators correspond roughly to log-linear model-based estimators

Table 6. Log-Linear Model Estimates for $\hat{x}_{000}$ Strata 11, 12, 13 Data

Sampling strata 11, 12, 13	*O2*	*R2*	*O3*	*R3*
$\hat{x}_{000}$: No second-order interaction	290.06	670.47	496.83	825.97
SE (delta method)	118.96	335.01	247.31	449.62
Fit/df	—/0	—/0	—/0	—/0
$\hat{x}_{000}$: Partial quasi-symmetry	291.55	669.50	489.84	767.80
SE (delta method)	119.46	334.28	242.12	402.86
Fit/df	.15/1	.01/1	.11/1	3.45/1
$\hat{x}_{000}$: Full quasi-symmetry	46.69	102.06	42.33	81.91
SE (delta method)	14.40	35.76	13.41	28.72
Fit/df	94.04/2	52.99/2	67.89/2	73.01/2
$\hat{x}_{000}$: Independent lists	44.64	48.26	27.51	34.98
SE (delta method)	5.69	6.78	3.80	5.06
Fit/df	94.06/3	58.34/3	70.15/3	80.06/3

being considered here. The log-linear model-based estimators have two advantages: first, they are built from explicit consideration of heterogeneity among individuals and varying list quality; and second, they are based on models whose fit can be formally tested.

6.3 Heterogeneity Effects

Kadane et al. (1992) explored the correlation bias effects in dual-system estimation that result from combining strata when the lists are independent within strata and there exists an ordering of the strata for which the probability of inclusion on each list increases from stratum to stratum. The same effects are of course present in multiple-system estimation; see, for example, Holland and Rosenbaum (1986) for related results in educational statistics and psychological measurement. The quasi-symmetry and partial quasi-symmetry models we have used incorporate parameters for exactly the kind of dependence explored by Kadane et al. (1992), thus we have some confidence that our models are an appropriate way to model unaccounted-for heterogeneity within available post-strata.

Indeed, the principal effect they predict—that ignoring the positive association induced by combining strata under these assumptions negatively biases the estimated population size—can be seen in Tables 5 and 6. Instead of post-strata, consider for a moment the smallest possible stratum: the individual. Our models make exactly the independence and ordering assumptions of Kadane et al. (1992) at this level, and then aggregate across these smallest-possible strata. Because the heterogeneity and aggregation assumptions are built into the model, so also are parameters representing the correlation bias. The population estimates (equivalently, estimates of x_{000}) based on the heterogeneity models are much higher than the estimates from the independent-lists model, which ignores this correlation bias. Moreover, because the process that aggregates data across strata is treated nonparametrically in the heterogeneity models considered here, the Rasch model and its variants should lead to a better population estimate, even for the data aggregated across sampling strata 11, 12, and 13.

It seems clear that when $J \geq 3$ lists are available, one should stratify with whatever observable covariates of heterogeneous catchability are available and then use a model within each stratum, such as the Rasch model, that accommodates possible further heterogeneity within strata. Indeed, the models we have proposed can be combined with observable covariates, in a manner analogous to Fischer's (1983) linear-logistic latent trait model. This approach also requires more than two lists, however. With only two lists, there is simply not enough information to estimate parameters reflecting the unaccounted-for heterogeneity.

7. CONCLUSIONS

In this article we have reanalyzed data for undercount estimation obtained by the U.S. Census Bureau (Zaslavsky and Wolfgang 1990, 1993) in connection with the dress rehearsal for the 1990 Census. Individuals are cross-classified according to their presence or absence in each of three lists: a list generated from the original Census (E sample), a list generated from the PES (P sample) and a list generated from the ALS (A sample). Although the A-sample data were originally intended to improve the coverage of the P sample and cross-check dual-system analyses of the undercount, it can be combined with the P sample and E sample to allow a triple-system analysis. Zaslavsky and Wolfgang (1990, 1993) considered several interesting but ad hoc triple-system estimators for the missing cell x_{000}.

The models we have built are triple-system models that allow for heterogeneous catchability among individuals being counted as well as for unequal coverages in the lists. They may be viewed as approximate marginalizations to the 2^3 table $x_{j_1 j_2 j_3}$ from a sparse $N \times 2^3$ multinomial table recording each individual's "capture record" uniquely, or they may be viewed directly as log-linear models for the cell probabilities $p_{j_1 j_2 j_3}$ with fixed effects for lists and random "catchability" effects for individuals. If we assume that the pattern of heterogeneity of individuals (catchability) is the same across all three lists, then a model for the logits of the individual capture probabilities for each list results, in which fixed list effects and random catchability effects are additive. This is well known as the Rasch model in educational statistics and psychological measurement settings. The additive random effects structure imposes a dependence structure, quasi-symmetry, on the log-linear model for $p_{j_1 j_2 j_3}$.

To obtain an estimate for the unobserved cell x_{000} from cell estimates for the seven observed cells, we use the no-second-order-interaction assumption for the 2^3 table. Standard errors of the estimates, given n, may be obtained using standard computing packages for generalized linear models. The quasi-symmetry models have some advantages over the unconstrained no-second-order-interaction model and the model assuming complete independence for the three lists. They more accurately reflect the belief that individuals have varying catchability than does the independence model, and they provide a framework in which no second-order interaction might plausibly occur. Finally, whereas the unconstrained no-second-order-interaction model is saturated for the incomplete 2^3 table, conditional model-fit statistics, given n, are straightforward to obtain for the quasi-symmetry models.

Based on the observed cross-product ratios in the data from Zaslavsky and Wolfgang (1990, 1993), we also built a "partial quasi-symmetry" model, in which the pattern of heterogeneity (catchability) is the same in the E and P samples but different in the A sample. Indeed, the radically different way in which the A list was constructed—essentially, exhaustive searches of administrative records for a particular geographical area covered by the P list—makes it plausible that there is a low or negative association between catchability in the A list and catchability in the E and P lists. Replacing the A list with another field sample, such as a pre-enumeration survey, would likely produce data more in line with the basic quasi-symmetry model. This simpler model is conceptually appealing, because its derivation from the Rasch model allows the critical assumption of no second-order interaction to be interpreted as a property of the distribution of catchability in the population. Alternatively, it may be believed that having the A list independent of the other two lists is useful from the standpoint of capturing individuals not otherwise caught by the E and P lists. Special models such as partial quasi-symmetry must then be built to accommodate this structure. But because partial quasi-symmetry is not derived from as simple a model for catchability in the population, the assumption of no second-order interaction does not have as simple an interpretation in terms of the catchability distribution.

We can easily extend the modeling ideas we present in this article to $J > 3$ lists. We write the additive-logit (Rasch) model as before, and it once again leads to a log-linear model exhibiting quasi-symmetry preserving first-order marginal totals. Averaging over the catchability effects also produces inequality restrictions on some of the log-linear parameters: They must behave like the log-moment sequence of a positive random variable. If the catchability effects, conditional on not being caught in any sample, are normally distributed, then the resulting model also exhibits no second-order interaction, with the first-order effect restricted to be nonnegative; in general, the assumption of second (or higher)-order interaction must be made separately. But as the number of lists J grows, Holland (1990a) conjectured that these conditional catchability effects will be asymptotically normal, so that for large J the no-second-order-interaction assumption may not be severe.

APPENDIX A: THE MULTINOMIAL APPROXIMATION

Consider again the larger $N \times 2^3$ table $w_{h,\mathbf{j}}$ with cell counts 1 if person h has capture pattern $\mathbf{j}$ and 0 otherwise. We assume a product-multinomial likelihood for the table $w_{h,\mathbf{j}}$:

$$P(w_{h,\mathbf{j}} : \forall\, h,\, \mathbf{j}) = \prod_{h=1}^{N} \prod_{\mathbf{j}} p_h(\mathbf{j})^{w_{h,\mathbf{j}}}. \tag{A.1}$$

A likelihood for the table $x_\mathbf{j}$ may be obtained by summing across all tables $\mathbf{w}$ "compatible" with the fixed table of counts $x_\mathbf{j}$:

$$\begin{aligned} P(x_\mathbf{j} : \forall\, \mathbf{j}) &= \sum_{\{\mathbf{w}:\sum_h w_{h,\mathbf{j}} = x_\mathbf{j} \forall \mathbf{j}\}} \prod_{h=1}^{N} \prod_{\mathbf{j}} p_h(\mathbf{j})^{w_{h,\mathbf{j}}} \\ &= \sum_{\{\mathbf{w}:\sum_h w_{h,\mathbf{j}} = x_\mathbf{j} \forall \mathbf{j}\}} \prod_{h=1}^{N} \left[\prod_{\mathbf{j}} p(\mathbf{j})^{w_{h,\mathbf{j}}} \prod_{\mathbf{j}} \left(\frac{p_h(\mathbf{j})}{p(\mathbf{j})} \right)^{w_{h,\mathbf{j}}} \right] \\ &= \sum_{\{\mathbf{w}:\sum_h w_{h,\mathbf{j}} = x_\mathbf{j} \forall \mathbf{j}\}} \left\{ \prod_{\mathbf{j}} p(\mathbf{j})^{x_\mathbf{j}} \right\} \left\{ \prod_{h=1}^{N} \prod_{\mathbf{j}} \left(\frac{p_h(\mathbf{j})}{p(\mathbf{j})} \right)^{w_{h,\mathbf{j}}} \right\} \qquad \text{(A.2)} \\ &\approx \frac{N!}{\prod_\mathbf{j} x_\mathbf{j}!} \prod_{\mathbf{j}} [c \cdot p(\mathbf{j})]^{x_\mathbf{j}}, \qquad \text{(A.3)} \end{aligned}$$

where $p(\mathbf{j}) = (1/N) \sum_{h=1}^{N} p_h(\mathbf{j})$. The approximate equality in (A.3) holds as long as the last factor in braces in (A.2) is near enough to some constant c^N over all frequently occurring tables $w_{i,\mathbf{j}}$. This is plausible, for example, if the heterogeneity of capture probabilities is not very great (so that the factors $p_h(\mathbf{j})/p(\mathbf{j})$ are all close to 1).

In any case the approximating multinomial (A.3), with $c = 1$, gives exactly the same expected cell counts $\mathbf{x} = (x_{000}, x_{001}, \ldots, x_{111})$ as the exact model (A.1), as observed in Section 4.2. Moreover, it is easy to see that under the exact model (A.1), the covariance matrix of the cell counts is

$$\sum_{h=1}^{N} [\mathrm{diag}(\mathbf{p}_h) - \mathbf{p}_h \mathbf{p}_h'] = N\, \mathrm{diag}(\mathbf{p}) - \sum_{h=1}^{N} \mathbf{p}_h \mathbf{p}_h',$$

where $\mathbf{p}_h = (p_h(000), p_h(001), \ldots, p_h(111))$. On the other hand, under (A.3) with $c = 1$ we obtain

$$N[\mathrm{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}'],$$

where $\mathbf{p} = (1/N) \sum_{h=1}^{N} \mathbf{p}_h$. The difference between these two covariance matrices is nonnegative definite. Thus the approximating multinomial agrees with the exact distribution on first moments and produces standard errors that are, if anything, too large.

Alternatively, models like (10) and (19) may be obtained directly by building a model for cell probabilities in the multinomial model (8) for the table $x_\mathbf{j}$ with random effects for individuals. This approach avoids the approximation step in (A.3), but changes the sampling model. For example, one might imagine the total population of N individuals as being sampled from a larger superpopulation with varying catchabilities, so that the random-effects distribution of the capture effects is the sampling distribution in the superpopulation; see Holland (1990b) for an analogous development in educational testing models. In this development the last summation in (9) may be identified as proportional to a posterior moment of the random effects, a fact first recognized by Cressie and Holland (1983).

Table A.1. DSE Table for r_1 Estimator

	A list	
	1	0
$E \cup P$	$x_{111} + x_{101} + x_{011}$	$x_{100} + x_{100} + x_{010}$
$(E \cup P)^c$	x_{001}	$x_{000} = ?$

Table A.2. DSE Table for r_2 Estimator

	A list	
	1	0
$E \Delta P$	$x_{101} + x_{011}$	$x_{100} + x_{010}$
$(E \cup P)^c$	x_{001}	$x_{000} = ?$

APPENDIX B: ZASLAVSKY AND WOLFGANG ESTIMATORS

Zaslavsky and Wolfgang (1990) presented several triple-system estimators of the undercount:

DSE without A source. This is the ordinary DSE, using only the E and P lists and ignoring the A list completely:

$$\hat{x}_{00+} = x_{10+} x_{01+} / x_{11+}.$$

DSE with P + ALS. This is a DSE with the P and A sources combined to make a single second list (this is the use originally conceived for the A source):

$$\hat{x}_{000} = x_{100}(x_{10+} + x_{001})/(x_{11+} + x_{101}).$$

DSE with k_2. An $E \times P$ cross-product ratio k_2 is calculated from the subtable with $A = 1$, and this is applied to the *marginal* $E \times P$ table:

$$\hat{x}_{00+} = k_2 x_{10+} x_{01+} / x_{11+};$$

$$k_2 = x_{101} x_{011} / x_{111} x_{001}.$$

Ratio r_1. The odds ratio for coverage by the A source is estimated from all the cells enumerated in the E or P source as $r_1 = (x_{111} + x_{101} + x_{011})/(x_{110} + x_{100} + x_{010})$; then $\hat{x}_{000} = x_{001}/r_1$. This is a DSE estimator, assuming independence, applied to Table A.1.

Ratio r_2. The odds ratio for coverage by the A source is estimated from all the cells enumerated in the E or P source, but not both, as $r_2 = (x_{101} + x_{011})/(x_{100} + x_{010})$; then $\hat{x}_{000} = x_{001}/r_1$. This is also a DSE estimator, assuming independence, applied to Table A.2 (where Δ is the symmetric difference operator).

Table A.3. *Estimates for the Unobserved Cell Counts in Various Post-Strata*

	Stratum 11				Strata 11, 12, and 13			
	Post-stratum				Post-stratum			
Estimator	O2	R2	O3	R3	O2	R2	O3	R3
$\hat{x}_{00+}$: DSE without ALS*	−44	−24	−23	−33	−32	−9	−15	−20
SE (jackknife)	17	15	10	15	17	15	11	15
$\hat{x}_{000}$: DSE with P + ALS	24	26	24	17	34	42	24	33
SE (jackknife)	10	9	10	8	6	9	7	9
$\hat{x}_{00+}$: DSE with k_2	130	312	254	305	285	601	458	729
SE (jackknife)	64	171	202	432	122	319	368	1039
$\hat{x}_{000}$: ratio r_1	26	76	33	58	180	152	125	130
SE (jackknife)	10	27	10	29	60	52	31	58
$\hat{x}_{000}$: ratio r_2	61	140	110	120	217	267	222	267
SE (jackknife)	26	53	55	83	69	95	94	170
$\hat{x}_{000}$: estimate, k_3 = 1	246	382	422	379	290	670	497	826
SE (jackknife)	182	240	489	565	129	384	430	1220

* $\hat{x}_{00+} < 0$ indicates the number of people discovered in the *A* list beyond the $\hat{x}_{00+}$ estimate.
Source: From Zaslavsky and Wolfgang (1990, 1993).

Estimate, $k_3 = 1$. This is the unconstrained, no-second-order-interaction estimate. It is equivalent to applying the odds ratio estimated from the $A = 1$ subtable to the $A = 0$ subtable to estimate the missing cell:

$$\hat{x}_{000} = (x_{101}x_{011}/x_{111}x_{001})(x_{100}x_{010}/x_{110}).$$

Table A.3 gives Zaslavsky and Wolfgang's (1990) estimates for the unobserved cell counts in both Stratum 11 and the combined Strata 11, 12, and 13. The same estimates, with somewhat different names, were presented in Zaslavsky and Wolfgang (1993).

[*Received January 1992. Revised November 1992.*]

REFERENCES

Alho, J. M. (1990), "Logistic Regression in Capture–Recapture Models," *Biometrics,* 46, 623–635.

——— (1991), "Variance Estimation in Dual Registration Under Population Heterogeneity," *Survey Methodology,* 17, 123–130.

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988), *The New S Language: A Programming Environment for Data Analysis and Graphics,* Pacific Grove, CA: Wadsworth and Brooks/Cole.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, MA: The MIT Press.

Burnham, K. P., and Overton, W. S. (1978), "Estimation of the Size of a Closed Population When the Capture Probabilities Vary Among Animals," *Biometrika,* 65, 625–633.

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

Cormack, R. M., and Jupp, P. E. (1991), "Inference for Poisson and Multinomial Models for Capture–Recapture Experiments," *Biometrika,* 78, 911–916.

Cressie, N., and Holland, P. W. (1983), "Characterizing the Manifest Probabilities of Latent Trait Models," *Psychometrika,* 48, 129–141.

Darroch, J. N. (1958), "The Multiple-Recapture Census, I. Estimation of a Closed Population," *Biometrika,* 45, 343–349.

——— (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. (1991), "A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability," Technical Report 543, Carnegie Mellon University, Dept. of Statistics.

Darroch, J. N., and McCloud, P. I. (1990), "Separating Two Sources of Dependence in Repeated Influenza Outbreaks," *Biometrika,* 77, 237–243.

Ding, Y. (1990), "Capture–Recapture Census With Uncertain Matching," unpublished Ph.D. dissertation, Carnegie Mellon University, Dept. of Statistics.

Ding, Y., and Fienberg, S. E. (1992), "Estimating Population and Census Undercount in the Presence of Matching Error," unpublished manuscript.

Duncan, O. D. (1984), "Rasch Measurement in Survey Research: Further Examples and Discussion," in *Surveying Subjective Phenomena,* eds. C. F. Turner and E. Martin, New York: Russell Sage Foundation, pp. 367–403.

El-Khorazaty, M. N., Imry, P. B., Koch, G. G., and Wells, H. B. (1977), "Estimating the Total Number of Events With Data From Multiple Record Systems: A Review of Methodological Strategies," *International Statistical Review,* 45, 129–157.

Ericksen, E. P., and Kadane, J. B. (1985), "Estimating the Population in a Census Year: 1980 and Beyond" (with discussion), *Journal of the American Statistical Association,* 80, 98–131.

Fay, R., Passel, J. S., Robinson, J. G., and Cowan, C. D. (1988), "The Coverage of the Population in the 1980 Census," 1980 Census of Population and Housing, Evaluation and Research Reports PHC80-E4, U.S. Department of Commerce, Bureau of the Census.

Fienberg, S. E. (1972), "The Multiple Recapture Census for Closed Populations and Incomplete 2^k Contingency Tables," *Biometrika,* 59, 591–603.

——— (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, 763–791.

——— (1991), "An Adjusted Census in 1990? Commerce Says "No," *Chance,* 4, 44–52.

——— (1992a), "Bibliography on Capture–Recapture Modeling With Applications to Census Undercount Adjustments," *Survey Methodology,* 18, 143–154.

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

Fienberg, S. E., and Meyer, M. M. (1983), "Log-linear Models and Categorical Data Analysis With Psychometric and Econometric Applications," *Journal of Econometrics,* 22, 191–214.

Fischer, G. (1983), "Logistic Latent Trait Models With Linear Constraints," *Psychometrika,* 48, 3–26.

Freedman, D. A. (1991), "Policy Forum: Adjusting the 1990 Census," *Science,* 252, 1,233–1,236.

Haberman, S. J. (1977), "Maximum Likelihood Estimates in Exponential Response Models," *The Annals of Statistics,* 5, 815–841.

Hogan, H. (1992), "The 1990 Post-Enumeration Survey: An Overview," *The American Statistician,* 45, 261–269.

Holland, P. W. (1990a), "The Dutch Identity: A New Tool for the Study of Item Response Models," *Psychometrika,* 55, 5–18.

——— (1990b), "On the Sampling Theory Foundations of Item Response Theory Models," *Psychometrika,* 55, 577–601.

Holland, P. W., and Rosenbaum, P. R. (1986), "Conditional Association and Unidimensionality in Monotone Latent Trait Models," *The Annals of Statistics,* 14, 1523–1543.

Isaki, C. T., Schultz, L. K., Diffendal, G. J., and Huang, E. T. (1988), "On Estimating Census Undercount in Small Areas," *Journal of Official Statistics,* 4, 95–112.

Jaro, M. (1989), "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Test Census of Tampa, Florida," *Journal of the American Statistical Association,* 84, 414–420.

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.

Lee, S. M., and Chao, A. (1991), "Estimating Population Size via Sample Coverage," in *Proceedings of the 1990 Taipei Symposium in Statistics*, eds. M. T. Chao and P. E. Cheng, Taipei, Taiwan: Institute of Statistical Science, Academia Sinica, pp. 407–426.

Lindsay, B., Clogg, C. C., and Grego, J. (1991), "Semiparametric Estimation in the Rasch Model and Related Exponential Response Models, Including a Simple Latent Class Model for Item Analysis," *Journal of the American Statistical Association*, 86, 96–107.

McCullagh, P., and Nelder, J. A. (1983), *Generalized Linear Models*, New York: Chapman and Hall.

Mosbacher, R. A. (1991), "Decision on Whether or Not a Statistical Adjustment of the 1990 Decentennial Census of the Population Should be Made for Coverage Deficiencies Resulting in an Overcount or Undercount of the Population," Internal Report, Office of the Secretary, U.S. Department of Commerce.

Mulry, M. H., and Spencer, B. D. (1991), "Total Error in PES Estimates of Population" (with discussion), *Journal of the American Statistical Association*, 86, 839–863.

Sanathanan, L. (1972a), "Estimating the Size of a Multinomial Population," *Annals of Mathematical Statistics*, 43, 142–152.

——— (1972b), "Models and Estimation Methods in Visual Scanning Experiments," *Technometrics*, 14, 813–839.

——— (1973), "A Comparison of Some Models in Visual Scanning Experiments," *Technometrics*, 15, 67–78.

Rasch, G. (1980), *Probabilistic Models for Some Intelligence and Attainment Tests* (expanded edition of the 1960 work, with forward and afterward by Benjamin D. Wright), Chicago: University of Chicago Press.

Rosenbaum, P. R. (1987), "Comparing Item Characteristic Curves," *Psychometrika*, 52, 217–233.

Rubin, D. B., Schafer, J. L., and Schenker, N. (1988), "Imputation Strategies for Missing Values in Post-Enumeration Surveys," *Survey Methodology*, 14, 209–221.

Schenker, N. (1988), "Handling Missing Data in Coverage Estimation With Application to the 1986 Test of Adjustment-Related Operations," *Survey Methodology*, 14, 87–98.

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

Tjur, T. (1982), "A Connection Between Rasch's Item Analysis Model and a Multiplicative Poisson Model," *Scandinavian Journal of Statistics*, 9, 23–30.

Wolter, K. M. (1986), "Some Coverage Error Models for Census Data," *Journal of the American Statistical Association*, 81, 336–346.

——— (1991), "Policy Forum: Accounting for America's Uncounted and Miscounted," *Science*, 253, 12–15.

Zaslavsky, A. M., and Wolfgang, G. S. (1990), "Triple-System Modeling of Census, Post-Enumeration Survey, and Administrative List Data," in *Proceedings of the Survey Research Section, American Statistical Association*, pp. 668–673.

——— (1993), "Triple System Modeling of Census, Post-Enumeration Survey, and Administrative List Data," *Journal of Business and Economic Statistics*, 11, 279–288.