

This research was supported in part by National Institutes of Health, Institute of General Medical Sciences through Grant GM-70004-05, by the U.S. Bureau of the Census through Joint Statistical Agreement JSA 75-2, and by Burroughs-Wellcome Co. through Joint Research Agreement 75-1.

THE ASYMPTOTIC COVARIANCE STRUCTURE OF LOG-LINEAR MODEL  
ESTIMATED PARAMETERS FOR THE MULTIPLE RECAPTURE CENSUS

by

Gary G. Koch and Alcinda L. Lewis  
Department of Biostatistics  
University of North Carolina at Chapel Hill

and

M. Nabil El-Korazaty  
Carolina Population Center  
University of North Carolina at Chapel Hill  
Institute of Statistics Mimeo Series No. 1054  
February 1976

THE ASYMPTOTIC COVARIANCE STRUCTURE  
OF LOG-LINEAR MODEL ESTIMATED PARAMETERS  
FOR THE MULTIPLE RECAPTURE CENSUS

by

Gary G. Koch, M. Nabil El-Khorazaty,\* and Alcinda L. Lewis  
Department of Biostatistics  
University of North Carolina at Chapel Hill  
Chapel Hill, North Carolina, U.S.A.

SUMMARY

This paper presents a matrix formulation for log-linear model analysis of the incomplete contingency table which arises from multiple recapture census data. Explicit matrix product expressions are given for the asymptotic covariance structure of the maximum likelihood estimators of both the log-linear model parameter vector and the predicted value vector for the observed and missing cells. These results are illustrated for data pertaining to a population of children possessing a common congenital anomaly.

Some key words: log-linear model; asymptotic covariance matrix; incomplete contingency tables; multiple recapture census; functional asymptotic regression methodology (FARM); Deming-Stephan Iterative Proportional Fitting (IPF) algorithm.

\*now at Carolina Population Center, University of North Carolina, Chapel Hill, North Carolina 27514

## 1. INTRODUCTION

As indicated by Fienberg (1972), data from the multiple recapture census for a single closed population can be analyzed from the point of view of an incomplete  $2^d$ -th contingency table, with one missing cell, that conceptually reflects the overall experience of all individuals in the population. Given this framework, log-linear models can be fitted to the observed cells of this table and then used to predict the missing cell as well as the total population size. The basic assumptions which are required for the validity of this estimation procedure are

1. The cell frequencies in the  $2^d$ -th conceptual contingency table for all possible outcomes can be described by a multinomial distribution
2. The  $2^d$ -th contingency table is characterized by a log-linear model which involves no  $d$ -th order interaction effect.

Further discussion of the nature and implications of these assumptions are given in Fienberg (1972).

This paper is concerned with the asymptotic covariance structure of the estimated parameters associated with the log-linear model analysis of multiple recapture data. Fienberg (1972) indicates that one approach which can be generally applied to this problem is a method due to Haberman (1974) which expresses "the asymptotic variance in terms of orthogonal projections with respect to various inner products." An iterative procedure involving cyclic descent is then described for obtaining an estimator of this formulation of asymptotic variance. Alternatively, this paper presents explicit matrix product expressions for estimating the asymptotic covariance structures for both the log-linear model estimated parameter

vector and the predicted value vector for the observed and missing cells by using results based on the matrix formulation of log-linear models in Koch et al. (1975). This approach is illustrated in Section 3 for the same data with the same model as used by Fienberg (1972), and essentially the same numerical estimate for the asymptotic variance of the total population is obtained.

## 2. THEORY

Let  $\ell = 1, 2, \dots, N$  index the elements of a particular population with  $N$  being the total number of elements in the population. Let  $g = 1, 2, \dots, d$  index a set of record systems according to which each individual may or may not be registered (tagged). Let  $j_g = 1, 2$ , index the response categories corresponding to the presence ( $j_g = 1$ ) or absence ( $j_g = 2$ ) of the attribute corresponding to registration by the  $g$ -th record system. Let the vector subscript  $\underline{j} = (j_1, j_2, \dots, j_d)$  index the multivariate response profiles for simultaneous registration status with respect to the  $d$  record systems.

The registration experience of each element can be expressed in terms of indicator random variables

$$N_{j_1 j_2 \dots j_d, \ell} = \begin{cases} 1 & \text{if element } \ell \text{ from population has} \\ & \text{registration status } \underline{j} = (j_1, j_2, \dots, j_d) \\ 0 & \text{if otherwise} \end{cases} \quad (2.1)$$

which can be arranged in overall response vectors

$$\underline{N}'_{\ell} = (N_{11\dots 1, \ell}, \dots, N_{j_1 j_2 \dots j_d, \ell}, \dots, N_{22\dots 2, \ell}). \quad (2.2)$$

Similarly, the quantities

$$\tilde{n}_j = n_{j_1 j_2 \dots j_d} = \sum_{\ell=1}^d N_{j_1 j_2 \dots j_d, \ell} \quad (2.3)$$

which represent the number of individuals in the population with registration status  $\tilde{j}$  can be arranged in an overall response vector

$$\tilde{n}' = (n_{11\dots 1}, \dots, n_{j_1 j_2 \dots j_d}, \dots, n_{22\dots 2}) \quad (2.4)$$

which corresponds conceptually to a  $2^d$ -th contingency table. In this regard, however, it should be noted that the cell frequency  $n_{22\dots 2}$  is not observed because it pertains to those elements who are not registered by any of the  $d$  record systems and that  $N$  is an unknown parameter for which an estimator is sought.

The vector  $\tilde{n}$  is assumed to have the multinomial distribution with parameters  $N$  and  $\tilde{\pi}$  where

$$\tilde{\pi}' = (\pi_{11\dots 1}, \dots, \pi_{j_1 j_2 \dots j_d}, \dots, \pi_{22\dots 2}) \quad (2.5)$$

with  $\pi_{j_1 j_2 \dots j_d}$  denoting the probability that a population element has registration status  $\tilde{j} = (j_1, j_2, \dots, j_d)$ . Thus, the relevant multinomial probability model is

$$\phi = N! \prod_{j_1=1}^2 \prod_{j_2=1}^2 \dots \prod_{j_d=1}^2 \{ \pi_{j_1 j_2 \dots j_d}^{n_{j_1 j_2 \dots j_d}} / n_{j_1 j_2 \dots j_d} ! \} \quad (2.6)$$

with the constraint

$$\sum_{j_1=1}^2 \sum_{j_2=1}^2 \dots \sum_{j_d=1}^2 \pi_{j_1 j_2 \dots j_d} = 1. \quad (2.7)$$

Although the model (2.6) - (2.7) is not necessarily valid for all multiple recapture census situations, it is reasonable for those where the  $N_{\tilde{j}}$  are

mutually independent vector random variables, each of which has the same multinomial distribution with parameters  $N_{\lambda} = 1$  and  $\pi$ .

The population is also assumed to have known structure in the sense that the variation among the elements of  $\pi$  is assumed to be characterized by the log-linear model

$$\pi = \pi(\beta) = \{ \exp(X \beta) / 1'_r [\exp(X \beta)] \} \quad (2.8)$$

where  $r = 2^d$  denotes the total number of multivariate response profiles,  $X$  denotes an appropriate  $(r \times t)$  design (or independent variable) matrix of known coefficients whose  $t < r$  columns are linearly independent and represent a basis for the main effects and interactions which constitute the model,  $\beta$  denotes the corresponding  $(t \times 1)$  vector of unknown parameters,  $1_r$  denotes an  $(r \times 1)$  vector of 1's, and  $\exp$  transforms a vector to the corresponding vector of exponential functions. Finally, the columns of the matrix  $X$  are assumed without loss of generality to be jointly linearly independent of the vector  $1_r$  in the sense that

$$\text{Rank}\{1_r, X\} = 1 + \text{Rank}\{X\}. \quad (2.9)$$

Since the cell frequency  $n_{22\dots 2}$  is not observed, the model (2.8) cannot be fitted directly to the overall response vector  $n$  defined by (2.4).

However, if

$$n = N - n_{22\dots 2} \quad (2.10)$$

denotes the number of elements which are registered by at least one of the record systems, then the multinomial probability model (2.6) - (2.7) and the log-linear model (2.8) imply that the observable response vector  $n_0$  defined by

$$n' = (n'_0, n_{22\dots 2}) \quad (2.11)$$

has the conditional multinomial distribution

$$\phi_0 = \frac{n!}{\{1'_{(r-1)} \{n_0!\}\}} \frac{\exp(n'_{(r-1)} X_0 \beta)}{\{1'_{(r-1)} [\exp(X_0 \beta)]\}^n} \quad (2.12)$$

where  $\{n_0!\}$  denotes the vector of factorial functions of the elements of the vector  $n_0$  and  $X_0$  is the  $(r-1) \times t$  sub-matrix of  $X$  which corresponds to the elements of the observable response vector  $n_0$ ; i.e.,  $X_0$  consists of the first  $(r-1)$  rows of  $X$ . If it can be assumed that the matrix  $X_0$  has the same full rank  $t < (r-1)$  as the matrix  $X$  (even though it has one less row), then the model (2.12) for the observable response vector  $n_0$  has essentially the same structure and involves the same parameters  $\beta$  as the model (2.6) - (2.8) for the conceptual response vector  $n$ . On the other hand, if this assumption does not hold, then  $\phi_0$  becomes a  $(t-1)$ -dimensional model whereas  $\phi$  is a  $t$ -dimensional model, so that estimates based on  $\phi_0$  are no longer necessarily valid for functions like the predicted value of the missing cell of the parameters which pertain to  $\phi$ .

A more precise definition of the assumption that  $X$  and  $X_0$  have the same rank can be formulated when the columns of the matrix  $X$  are expressed in terms of the following types of indicator functions of the response profiles  $j$

$$x_k(j) = x_{k_1 k_2 \dots k_d}(j) = \prod_{g=1}^d \{x_{gk_g}(j)\} \quad (2.13)$$

where

$$x_{g0}(j) \equiv 1$$

$$x_{g1}(j) = \begin{cases} -1 & \text{if } j_g = 1 \\ 1 & \text{if } j_g = 2 \end{cases} \quad (2.14)$$

with  $k_g = 0, 1$  for  $g = 1, 2, \dots, d$ . In this framework, the variable  $x_{00\dots 0}(j) \equiv 1$  is excluded from the model by (2.9) and the variable

$$x_{11\dots 1}(j) = \prod_{g=1}^d \{x_{g1}(j)\} \quad (2.15)$$

is excluded in order to allow the assumption

$$\text{Rank}(X) = \text{Rank}(X_0) = t. \quad (2.16)$$

The variable  $x_{11\dots 1}(j)$  corresponds to the  $d$ -th order (log-linear) interaction effect  $\beta_{11\dots 1}$  in the sense analogous to definitions given by Bartlett (1935), Roy and Kastenbaum (1956), and others. Thus, the exclusion of  $x_{11\dots 1}(j)$  from  $X$  is equivalent to the assumption of no  $d$ -th order interaction in the sense that  $\beta_{11\dots 1} = 0$ . In summary, the assumption of no  $d$ -th order interaction of the attributes corresponding to registration by the  $d$  record systems is a sufficient condition for the rank identity assumption (2.16). However, it should be noted that this type of assumption may not be necessarily valid nor can it be tested from the observable response vector  $n_0$ . Thus, its adoption should be governed by considerations pertaining to substantive knowledge of the particular nature of the record systems and the relationships among them and previous experience with their use. Finally, this same basic issue applies to other possible sufficient conditions for the rank identity assumption (2.16) which are based on alternative formulations of the matrix  $X$  different from (2.13) - (2.14).

Given the previous discussion, attention will be henceforth directed at the model (2.12) with  $X$  being formulated in terms of the indicator functions (2.13) - (2.14) with the  $d$ -th order interaction effect excluded so that (2.16) holds. As such, this framework is the same as that used by



Fienberg (1972). Proceeding similarly, the method of maximum likelihood will be used to obtain estimators  $\hat{\beta}$  for  $\beta$ . These estimators are characterized implicitly by the following set of equations

$$\left[ \frac{d \log_e \phi_0}{d\beta} \right]_{\beta=\hat{\beta}} = 0_t \quad (2.17)$$

where  $0_t$  is a  $(t \times 1)$  vector of 0's. By matrix differentiation methods similar to those used by Forthofer and Koch (1973), it follows that

$$\begin{aligned} \frac{d}{d\beta'} [\log_e \phi_0] &= \frac{d}{d\beta'} [n'X_0\beta - n \log_e \{1'_{(r-1)}[\exp(X_0\beta)]\}] \\ &= n'X_0 - \frac{n[\exp(X_0\beta)]'X_0}{1'_{(r-1)}[\exp(X_0\beta)]} \\ &= n'X_0 - n[\pi_0(\beta)]'X_0 \end{aligned} \quad (2.18)$$

where

$$\pi_0 = \pi_0(\beta) = \frac{\exp(X_0\beta)}{1'_{(r-1)}[\exp(X_0\beta)]} \quad (2.19)$$

denotes the vector of conditional probabilities for the observable response profiles based on the model (2.8). Thus, the equations (2.17) may be compactly written as

$$X_0'\hat{\pi}_0 = X_0'p_0 \quad (2.20)$$

where  $\hat{\pi}_0 = \pi_0(\hat{\beta})$  represents the maximum likelihood estimator for  $\pi_0(\beta)$ , and  $p_0 = (n_0/n)$  represents the conditional observed vector of sample proportions. Although the equations (2.20) appear straightforward, their non-linear structure often does not permit explicit solution. For these situations, successive search algorithms are required to determine  $\hat{\beta}$  and/or  $\hat{\pi}_0$ . In this regard, if  $X$  has an hierarchical structure which includes in

the model with any given interaction variable all corresponding lower order interaction variables, then the Deming-Stephan Iterative Proportional Fitting (IPF) algorithm indicated in Fienberg (1972) can be used to calculate  $\hat{\pi}_0$  which can be subsequently used to obtain the estimated parameter vector  $\hat{\beta}$  and predicted values for the observed and missing cells.

As a consequence of the full rank matrix formulation of the log-linear model (2.12), the asymptotic covariance matrix of the estimator  $\hat{\beta}$  is determined by forming the negative inverse of the Fisher Information Matrix. By matrix differentiation methods similar to those used with (2.18), it follows that the Fisher Information Matrix is given by

$$\begin{aligned} \frac{d}{d\beta d\beta'} [\log_e \phi_0] &= -n X_0' \left[ \frac{d}{d\beta} \{\pi_0(\beta)\} \right] \\ &= -n X_0' [D_{\pi_0} - \pi_0 \pi_0'] X_0 . \end{aligned} \quad (2.21)$$

Thus, the asymptotic covariance matrix  $V_{\hat{\beta}}(\pi_0)$  for  $\hat{\beta}$  is

$$V_{\hat{\beta}}(\pi_0) = \frac{1}{n} \{X_0' [D_{\pi_0} - \pi_0 \pi_0'] X_0\}^{-1} . \quad (2.22)$$

Since  $\hat{\pi}_0 = \pi_0(\hat{\beta})$  is a consistent estimator for  $\pi_0$ , a consistent estimator for the covariance matrix  $V_{\hat{\beta}}(\pi_0)$  in (2.22) is

$$V_{\hat{\beta}} = V_{\hat{\beta}}(\hat{\pi}_0) = \frac{1}{n} \{X_0' [D_{\hat{\pi}_0} - \hat{\pi}_0 \hat{\pi}_0'] X_0\}^{-1} . \quad (2.23)$$

The asymptotic covariance matrix for the estimator  $\hat{\pi}_0 = \pi_0(\hat{\beta})$  of the vector of conditional probabilities of the response profiles and the estimator  $\hat{\gamma} = \gamma(\hat{\beta})$  of the ratio

$$\gamma = \frac{\pi_{22\dots 2}}{[1'_{(r-1)}, 0] \pi} = \frac{\pi_{22\dots 2}}{(1 - \pi_{22\dots 2})} \quad (2.24)$$

of the probabilities of the missing cell vs. the overall set of observable

cells is obtained by the well known  $\delta$ -method as based on the first order Taylor series approximations for these estimators. In this regard, the compound function notation used in Forthofer and Koch (1973) is used to express  $\hat{\pi}_{\sim 0}$  and  $\hat{\gamma}$  in the form

$$\begin{bmatrix} \hat{\pi}_{\sim 0} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \pi_{\sim 0}(\hat{\beta}) \\ \gamma(\hat{\beta}) \end{bmatrix} = \exp[A_{\sim 3}(\log\{A_{\sim 2}[\exp(A_{\sim 1}\hat{\beta})]\})] \quad (2.25)$$

where

$$A_{\sim 1} = X, \quad A_{\sim 2} = \begin{bmatrix} I_{\sim r} \\ 1'_{\sim(r-1)}, 0 \end{bmatrix}, \quad A_{\sim 3} = [I_{\sim r}, -1_{\sim r}] \quad (2.26)$$

with  $I_{\sim r}$  being an  $(r \times r)$  identity matrix,  $1'_{\sim(r-1)}$  being a  $[1 \times (r-1)]$  row vector of 1's, and  $1_{\sim r}$  being an  $(r \times 1)$  column vector of 1's. As a result, a consistent estimator for the corresponding asymptotic covariance matrix can be determined as the matrix product

$$\begin{bmatrix} V_{\sim 0}^{\hat{\pi}} & V_{\sim 0, \hat{\gamma}}^{\hat{\pi}} \\ V_{\sim 0, \hat{\gamma}}^{\hat{\pi}} & V_{\sim \hat{\gamma}}^{\hat{\gamma}} \end{bmatrix} = D_{\sim y_3} A_{\sim 3} D_{\sim a_2}^{-1} A_{\sim 2} D_{\sim y_1} A_{\sim 1} [V_{\sim \hat{\beta}}(\hat{\pi}_{\sim 0})] A_{\sim 1}' D_{\sim y_1}^{-1} A_{\sim 2}' D_{\sim a_2}^{-1} A_{\sim 3}' D_{\sim y_3} \quad (2.27)$$

where

$$y_{\sim 1} = \exp(A_{\sim 1}\hat{\beta}), \quad a_{\sim 2} = A_{\sim 2}y_{\sim 1}, \quad y_{\sim 3} = \exp\{A_{\sim 3}[\log_e(a_{\sim 2})]\}. \quad (2.28)$$

Finally, this approach yields the following estimators

$$\hat{n}_{22\dots 2} = n \hat{\gamma} \quad (2.29)$$

$$\hat{N} = n(1 + \hat{\gamma}) \quad (2.30)$$

for the missing cell and total population size respectively. Since  $n$  itself is a random variable which is assumed to have a binomial distribu-

tion with parameters  $N$  and  $(1 - \pi_{22\dots 2})$ , the methods indicated in Darroch (1958) and Fienberg (1972) can be used to produce the following estimators for the asymptotic variances of  $\hat{n}_{22\dots 2}$  and  $\hat{N}$

$$V_{\hat{n}_{22\dots 2}} = n^2 V_{\hat{\gamma}} + \hat{N} \hat{\gamma}^2 \hat{\pi}_{22\dots 2} (1 - \hat{\pi}_{22\dots 2}) = n^2 V_{\hat{\gamma}} + \{\hat{n}_{22\dots 2}^3 / n \hat{N}\} \quad (2.31)$$

$$V_{\hat{N}} = n^2 V_{\hat{\gamma}} + \hat{N} (1 + \hat{\gamma})^2 \hat{\pi}_{22\dots 2} (1 - \hat{\pi}_{22\dots 2}) = n^2 V_{\hat{\gamma}} + \{\hat{n}_{22\dots 2} \hat{N} / n\}. \quad (2.32)$$

These estimators for the asymptotic variance of  $\hat{n}_{22\dots 2}$  and  $\hat{N}$  are essentially the same as those given by Fienberg (1972). However, here the quantity  $V_{\hat{\gamma}}$  is obtained by the matrix multiplication operations given in (2.23) and (2.27), while Fienberg, in general, uses the orthogonal projection methods of Haberman (1974) which require cyclic descent iterative computations.

### 3. EXAMPLE

The observed data in column 6 of Table 1 were used by Fienberg (1972) to illustrate the application of log-linear model methods to multiple recapture census data. They have also been previously analyzed by Wittes (1970), Wittes, Colton, and Sidel (1974), El-Khorazaty (1975), and El-Khorazaty and Sen (1976). The data pertain to the number of children in Massachusetts (born between 1 January 1955 and 31 December 1959 and still alive on 31 December 1966) for whom a positive diagnosis of a specific congenital anomaly (Down's syndrome) was made. In this regard, registration information for the following five record systems is used:

1. Other hospital records (OHR)
2. Obstetric records (OBR)
3. Schools (S)
4. Massachusetts Departments of Mental Health (MDMH)
5. Massachusetts Department of Health (MDH).

Table 1. Observed and log-linear model predicted frequencies for positive diagnoses of a specific congenital anomaly based on five record systems; OHR, OBR, S, MDMH, and MDH

Record System					Observed frequency $n_0$	Fienberg Model X MLE $\tilde{\pi}_0(\hat{\beta})$	Fienberg Model X WLS $\tilde{\pi}_0(\hat{b})$	Reduced Model X <sub>R</sub> FARM $\tilde{\pi}_0(\hat{\beta}_R)$
OHR	OBR	S	MDMH	MDH		Predicted frequency $n \tilde{\pi}_0(\hat{\beta})$	Predicted frequency $n \tilde{\pi}_0(\hat{b})$	Predicted frequency $n \tilde{\pi}_0(\hat{\beta}_R)$
Yes	Yes	Yes	Yes	Yes	2	1.30	1.49	1.51
Yes	Yes	Yes	Yes	No	8	9.05	9.12	9.88
Yes	Yes	Yes	No	Yes	2	2.55	2.86	3.64
Yes	Yes	Yes	No	No	18	17.75	17.53	23.79
Yes	Yes	No	Yes	Yes	5	3.52	3.88	3.64
Yes	Yes	No	Yes	No	25	24.46	23.82	23.79
Yes	Yes	No	No	Yes	1	2.69	2.92	3.64
Yes	Yes	No	No	No	19	18.69	17.88	23.79
Yes	No	Yes	Yes	Yes	0	0.50	0.63	0.41
Yes	No	Yes	Yes	No	23	16.96	17.43	15.22
Yes	No	Yes	No	Yes	0	0.99	1.21	0.98
Yes	No	Yes	No	No	34	33.28	33.51	36.65
Yes	No	No	Yes	Yes	3	1.36	1.65	0.98
Yes	No	No	Yes	No	37	45.86	45.53	36.65
Yes	No	No	No	Yes	1	1.04	1.24	0.98
Yr	No	No	No	No	37	35.03	34.18	36.65
N	Yes	Yes	Yes	Yes	3	1.33	1.62	1.51
No	Yes	Yes	Yes	No	5	9.27	9.91	9.88
No	Yes	Yes	No	Yes	5	3.90	4.46	3.64
No	Yes	Yes	No	No	36	27.14	27.38	23.79
No	Yes	No	Yes	Yes	1	3.60	4.22	3.64
No	Yes	No	Yes	No	22	25.07	25.90	23.79
No	Yes	No	No	Yes	4	4.11	4.55	3.64
No	Yes	No	No	No	27	28.57	27.93	23.79
No	No	Yes	Yes	Yes	0	0.94	1.17	0.98
No	No	Yes	Yes	No	30	31.65	32.23	36.65
No	No	Yes	No	Yes	3	2.75	3.22	2.36
No	No	Yes	No	No	83	92.65	89.00	88.23
No	No	No	Yes	Yes	2	2.54	3.05	2.36
No	No	No	Yes	No	97	85.59	84.19	88.23
No	No	No	No	Yes	4	2.89	3.29	2.36

## Model goodness of fit statistics

1. Asymptotic chi-square criterion	25.81 <sup>a</sup>	19.42 <sup>b</sup>	34.94 <sup>c</sup>
2. Degrees of freedom	21	21	27

Notes: a. Log-likelihood ratio chi-square statistic

b. Weighted least squares log-linear chi-square statistic

c. Sum of log-likelihood ratio chi-square statistic for model  $\tilde{X}$  and weighted least squares FARM chi-square statistic for reducing model  $\tilde{X}$  to model  $\tilde{X}_R$

The log-linear model fitted by Fienberg (1972) may be expressed in the matrix form (2.8) with

$$\tilde{X} = \begin{bmatrix} -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\tilde{\beta} = \begin{bmatrix} \beta_{10000} \\ \beta_{01000} \\ \beta_{11000} \\ \beta_{00100} \\ \beta_{00010} \\ \beta_{00001} \\ \beta_{10010} \\ \beta_{00110} \\ \beta_{00001} \\ \beta_{01001} \end{bmatrix} \quad (3.1)$$

which involves the first order main effects for each of the record systems and the pairwise interaction effects for 1 vs 2, 1 vs 4, 3 vs 4, and 2 vs 5. The maximum likelihood estimators for the predicted frequencies  $\hat{n}_{\sim 0} = n \hat{\pi}_{\sim 0}$  corresponding to the observed cells were obtained by the Deming-Stephan IPF algorithm described in Fienberg (1972) and are shown in column 7 of Table 1. More specifically, these quantities are computed by starting with a  $2^5$  table in which each of the observable cells is initially assigned the value 1 and the missing cell is assigned the value 0 and then adjusting

this table to conform successively with the observed two-dimensional tables which correspond to the set of highest order interaction effects defining the model; namely, the marginal configurations corresponding to sources 1 vs 2, 1 vs 4, 3 vs 4, and 2 vs 5.

Although more direct methods may be available, the estimators  $\hat{\beta}$  and  $V_{\hat{\beta}}$  as well as  $V_{\hat{\pi}_0}$ ,  $\hat{\gamma}$ , and  $V_{\hat{\gamma}}$  were determined by using the Functional Asymptotic Regression Methodology (FARM) computing procedures described in Koch et al. (1975). This approach is of supplementary interest because it permits non-hierarchical models which are defined in terms of linear combinations of the columns of the matrix  $X$  in (3.1) to be fitted to the data. For this purpose, let

$$F(\hat{\pi}_0) = K[\log_e(A \hat{\pi}_0)] \quad (3.2)$$

where

$$A = I_{31}, \quad K = [-I_{30}, I_{30}] \quad (3.3)$$

denote the vector of log ratios for  $\hat{\pi}_{0,11111}$  with respect to each of the other elements of  $\hat{\pi}_0$  and let

$$V_F(\hat{\pi}_0) = K D_{\hat{\pi}_0}^{-1} K' \quad (3.4)$$

denote the consistent estimator (based on the  $\delta$ -method) for the asymptotic covariance matrix for the functions  $F(p_0)$  of the conditional observed vector  $p_0 = (n_0/n)$  evaluated at  $\hat{\pi}_0$ . Then fit the linear regression model

$$E_A\{F(\hat{\pi}_0)\} = F(\pi_0) = K X_0 \beta \quad (3.5)$$

where " $E_A$ " means asymptotic "expectation" to the functions  $F(\hat{\pi}_0)$  by weighted least squares with  $V_F(\hat{\pi}_0)$  as the weight matrix. Thus, in this framework,  $\hat{\beta}$  and  $V_{\hat{\beta}}$  are obtained as

$$\hat{\beta} = \{X_0' K' [K D_{\hat{\pi}_0}^{-1} K']^{-1} K X_0\}^{-1} X_0' K' [K D_{\hat{\pi}_0}^{-1} K']^{-1} [F(\hat{\pi}_0)] \quad (3.6)$$

$$\hat{V}_{\tilde{\beta}} = \{X_0' K' [K D^{-1} K']^{-1} K X_0\}^{-1}. \quad (3.7)$$

Strictly speaking, the FARM procedures described in Koch et al. (1975) require that  $X_0$  be orthogonal to  $1_{(r-1)}$  whereas the application here only requires that  $X_0$  and  $1_{(r-1)}$  are jointly linearly independent in the sense of (2.9). However, this difficulty is bypassed by an argument given in the Appendix which demonstrates that the computational strategy (3.2) - (3.7) yields the same results as would be obtained by applying the FARM procedures to fit a model which is equivalent to  $X_0$  via (2.8) but also is orthogonal to  $1_{(r-1)}$ . For the data in Table 1, the computational procedures (3.6) and (3.7) yielded

$$\hat{\beta} = \begin{bmatrix} 0.262 \\ 0.069 \\ 0.150 \\ 0.262 \\ 0.201 \\ 0.100 \\ -0.236 \\ 1.365 \\ 0.395 \end{bmatrix} \quad (3.8)$$

$$\hat{V}_{\beta} = \begin{bmatrix} 0.2191 & & & & & & & & & \\ -0.0044 & 0.8208 & & & & & & & & \\ -0.0580 & -0.0374 & 0.2256 & & & & & & & \\ 0.0245 & 0.0110 & 0.0345 & 0.2453 & & & & & & \\ 0.0073 & 0.0078 & 0.0245 & 0.0944 & 0.2340 & & & & & \\ -0.0018 & 0.0081 & 0.0253 & 0.0421 & -0.0235 & 0.2105 & & & & \\ 0.0245 & 0.0110 & 0.0345 & 0.0041 & -0.0166 & 0.0421 & 0.2453 & & & \\ 0.0107 & -0.2291 & 0.0150 & 0.0249 & 0.0177 & 0.0183 & 0.0249 & 0.8167 & & \\ 0.0107 & -0.6944 & 0.0150 & 0.0249 & 0.0177 & 0.0183 & 0.0249 & 0.1951 & 0.8167 & \end{bmatrix} \times 10^{-2}$$

Symmetric

These results were then used to estimate  $\hat{V}_{\tilde{\pi}_0}$ ,  $\hat{\gamma}$ , and  $\hat{V}_{\hat{N}}$  via (2.25) - (2.28). In this regard, the square roots of the diagonal elements of  $\hat{V}_{\tilde{\pi}_0}$  which represent estimated standard errors for the elements of  $\hat{\pi}_0$  are shown in Table 2 together with  $\hat{\pi}_0$ . Also included in Table 2 are  $\hat{\gamma}$ ,  $\hat{n}_{22\dots 2}$ , and  $\hat{N}$  as well as their estimated standard errors based on  $\sqrt{\hat{V}_{\hat{\gamma}}}$ ,  $\sqrt{\hat{V}_{\hat{n}_{22\dots 2}}}$ , and  $\sqrt{\hat{V}_{\hat{N}}}$ . Thus, it can be noted that the value shown for  $\sqrt{\hat{V}_{\hat{N}}} = 18.3$  is the same as that obtained by Fienberg (1972).



Table 2. Observed and log-linear model estimators for the conditional probability vector  $\pi_0$ ,  
the missing vs observed ratio  $\gamma$ , the missing cell  $n_{22...2}$ ,  
and the total population  $N$  together with their corresponding standard errors

Record System					Observed proportions $P_U$	Estimated s.e.,	Fienberg Model X		Fienberg Model X		Reduced Model $X_R$	
OHR	OBR	S	MDMH	MDH			MLE Predicted proportions $\pi_0(\hat{\beta})$	Estimated s.e.	WLS Predicted proportions $\pi_0(\hat{b})$	Estimated s.e.	FARM Predicted proportions $\pi_0(\hat{\beta}_R)$	Estimated s.e.
Yes	Yes	Yes	Yes	Yes	0.0037	0.0026	0.0024	0.0006	0.0028	0.0007	0.0028	0.0005
Yes	Yes	Yes	Yes	No	0.0149	0.0052	0.0168	0.0028	0.0170	0.0028	0.0184	0.0018
Yes	Yes	Yes	No	Yes	0.0037	0.0026	0.0048	0.0011	0.0053	0.0013	0.0068	0.0011
Yes	Yes	Yes	No	No	0.0334	0.0071	0.0331	0.0048	0.0326	0.0047	0.0443	0.0029
Yes	Yes	No	Yes	Yes	0.0093	0.0041	0.0065	0.0015	0.0072	0.0017	0.0068	0.0011
Yes	Yes	No	Yes	No	0.0464	0.0091	0.0456	0.0061	0.0444	0.0059	0.0443	0.0029
Yes	Yes	No	No	Yes	0.0019	0.0019	0.0050	0.0012	0.0054	0.0013	0.0068	0.0011
Yes	Yes	No	No	No	0.0353	0.0080	0.0348	0.0050	0.0333	0.0048	0.0443	0.0029
Yes	No	Yes	Yes	Yes	0.0009	0.0013	0.0009	0.0003	0.0012	0.0003	0.0008	0.0002
Yes	No	Yes	Yes	No	0.0427	0.0087	0.0316	0.0046	0.0325	0.0045	0.0283	0.0036
Yes	No	Yes	No	Yes	0.0009	0.0013	0.0018	0.0006	0.0023	0.0007	0.0018	0.0003
Yes	No	Yes	No	No	0.0631	0.0105	0.0620	0.0076	0.0624	0.0076	0.0682	0.0034
Yes	No	No	Yes	Yes	0.0056	0.0032	0.0025	0.0008	0.0031	0.0009	0.0018	0.0003
Yes	No	No	Yes	No	0.0687	0.0109	0.0854	0.0094	0.0848	0.0098	0.0682	0.0034
Yes	No	No	No	Yes	0.0019	0.0019	0.0019	0.0006	0.0023	0.0006	0.0018	0.0003
Yes	No	No	No	No	0.0687	0.0109	0.0652	0.0079	0.0637	0.0076	0.0682	0.0034
No	Yes	Yes	Yes	Yes	0.0056	0.0032	0.0025	0.0006	0.0030	0.0007	0.0028	0.0005
No	Yes	Yes	Yes	No	0.0093	0.0041	0.0173	0.0027	0.0185	0.0030	0.0184	0.0018
No	Yes	Yes	No	Yes	0.0093	0.0041	0.0073	0.0016	0.0083	0.0018	0.0068	0.0011
No	Yes	Yes	No	No	0.0669	0.0108	0.0505	0.0059	0.0510	0.0056	0.0443	0.0029
No	Yes	No	Yes	Yes	0.0019	0.0019	0.0067	0.0015	0.0079	0.0018	0.0068	0.0011
No	Yes	No	Yes	No	0.0409	0.0085	0.0467	0.0059	0.0482	0.0063	0.0443	0.0029
No	Yes	No	No	Yes	0.0074	0.0037	0.0076	0.0018	0.0085	0.0020	0.0068	0.0011
No	Yes	No	No	No	0.0501	0.0094	0.0532	0.0071	0.0520	0.0070	0.0443	0.0029
No	No	Yes	Yes	Yes	0.0009	0.0013	0.0017	0.0005	0.0022	0.0006	0.0018	0.0003
No	No	Yes	Yes	No	0.0557	0.0099	0.0589	0.0074	0.0600	0.0076	0.0682	0.0034
No	No	Yes	No	Yes	0.0056	0.0032	0.0051	0.0014	0.0060	0.0016	0.0044	0.0008
No	No	Yes	No	No	0.1541	0.0156	0.1725	0.0144	0.1658	0.0148	0.1643	0.0090
No	No	No	Yes	Yes	0.0037	0.0026	0.0047	0.0014	0.0057	0.0016	0.0044	0.0008
No	No	No	Yes	No	0.1801	0.0166	0.1594	0.0132	0.1568	0.0123	0.1643	0.0090
No	No	No	No	Yes	0.0074	0.0037	0.0054	0.0016	0.0061	0.0017	0.0044	0.0008
Missing vs Observed ratio $\gamma$					0.0142	0.0514	0.1816	0.0275	0.1691	0.0262	0.1643	0.0090
Missing cell $n_{22...2}$					7.61	27.60	97.52	14.85	90.79	14.16	88.23	5.06
Total population $N$					544.61	27.74	634.52	18.25	627.79	17.45	625.23	11.24

Alternatively, if the computational procedures (3.6) and (3.7) are applied with  $\hat{\pi}_0$  replaced by  $p_0$ , the resulting estimators  $\underline{b}$  and  $\underline{V}_b$  correspond to the linearized modified chi-square criterion of Neyman (1949) as described by Grizzle et al. (1969). In large samples, these estimators are asymptotically equivalent to the maximum likelihood estimators  $\hat{\beta}$  and  $\underline{V}_{\hat{\beta}}$ ; they also can be obtained directly from the observed data by weighted least squares (WLS) without iteration. However, they encounter computational difficulties as well as dubious statistical properties if the observed cell frequencies are small. In this context, 0-values cause the most problems because they must be replaced by a small number like (1/2) as described in Berkson (1955) in order to avoid " $\log_e(0)$ " calculations. Given these considerations, the estimators  $\underline{b}$  and their respective standard errors based on  $\underline{V}_b$  are shown in Table 3 for purposes of comparison with the estimators  $\hat{\beta}$  and their respective standard errors. Moreover, analogous estimators for  $\pi_0$ ,  $\gamma$ ,  $n_{22\dots 2}$ , and  $N$  based on  $\underline{b}$  and their respective standard errors are shown in Table 2; and for the observed cell frequencies, in Table 1. Thus, it can be seen that the results based on the WLS estimators  $\underline{b}$  and  $\underline{V}_b$  are reasonably similar to those based on the MLE estimators  $\hat{\beta}$  and  $\underline{V}_{\hat{\beta}}$ .

A by-product of the FARM approach is that it permits the testing of hypotheses pertaining to model parameters. In particular, an appropriate test statistic for the hypothesis

$$H_0: \underline{C} \underline{\beta} = \underline{0}_q \quad (3.8)$$

where  $\underline{C}$  is a  $(q \times t)$  matrix of full rank  $q \leq t$  is the generalized Wald (1943) statistic

$$Q_C = \hat{\beta}' \underline{C}' [\underline{C} \underline{V}_{\hat{\beta}} \underline{C}']^{-1} \underline{C} \hat{\beta} \quad (3.9)$$

which has approximately a chi-square distribution with D.F. =  $q$  in large

Table 3. Estimated parameters, estimated standard errors,  
and tests of significance for Fienberg model  $\tilde{X}$

Parameter	MLE Estimator	Estimated s.e.	Test Statistic $Q_C$ (D.F.=1)	WLS Estimator	Estimated s.e.	Test Statistic $Q_C$ (D.F.=1)
$\beta_{10000}$	0.262	0.047	31.37	0.265	0.047	31.91
$\beta_{01000}$	0.069	0.091	0.59	0.080	0.088	0.84
$\beta_{11000}$	0.150	0.048	9.95	0.133	0.048	7.70
$\beta_{00100}$	0.262	0.050	27.88	0.245	0.049	24.56
$\beta_{00010}$	0.201	0.048	17.29	0.182	0.049	14.00
$\beta_{10010}$	0.100	0.046	4.75	0.091	0.047	3.73
$\beta_{00110}$	-0.236	0.050	22.69	-0.235	0.049	22.68
$\beta_{00001}$	1.365	0.090	228.00	1.283	0.088	212.31
$\beta_{01001}$	0.395	0.090	19.08	0.376	0.088	18.23

Since the model (3.11) is not hierarchical, the maximum likelihood methods given by Fienberg (1972) cannot be used to determine  $\hat{\beta}_{\sim R}$  although other numerical algorithms perhaps could. Alternatively, having obtained  $\hat{\pi}_0$  for the model  $X$  in (3.1), the FARM computational procedures can be applied to determine an estimator  $\bar{\beta}_{\sim R}$  for  $\beta_{\sim R}$  which is asymptotically equivalent to  $\hat{\beta}_{\sim R}$ . For this purpose, expressions (3.6) and (3.7) are used with  $X_0$  replaced by  $X_{\sim R,0}$ . The resulting estimators  $\bar{\beta}_{\sim R}$  and  $V_{\bar{\beta}_{\sim R}}$  were

$$\bar{\beta}_{\sim R} = \begin{bmatrix} 0.220 \\ 1.375 \\ 0.436 \end{bmatrix}, \quad V_{\bar{\beta}_{\sim R}} = \begin{bmatrix} 0.0439 & & \\ 0.0015 & 0.7448 & \\ -0.0059 & -0.0091 & 0.2141 \end{bmatrix} \times 10^{-2}. \quad (3.12)$$

Estimators for  $\pi_0$ ,  $\gamma$ ,  $n_{22\dots 2}$ , and  $N$  based on  $\bar{\beta}_{\sim R}$  are shown in Table 2; and for the observed cell frequencies, in Table 1. Finally, it should be noted that the motivation underlying the formulation of  $X_{\sim R}$  was an attempt to identify a model with as few parameters as possible which fitted the data in order to minimize the variance for the estimator of the population total  $N$ . As such, this model is perhaps too aggressive in the sense of being overly data-dependent, and thus the results which are based on it should be interpreted carefully. In this context, however, its application is still of interest from the point of view of illustrating the use of the FARM computational procedures for fitting a non-hierarchical model.

In summary, the matrix formulation (2.8) for log-linear models and the FARM computational procedures (3.2) - (3.7) permit the application of a broad range of comprehensive analyses to multiple recapture census data. However, as stated previously, these methods do require certain fundamental assumptions which may not always apply and thus should be used with caution.

Since the model (3.11) is not hierarchical, the maximum likelihood methods given by Fienberg (1972) cannot be used to determine  $\hat{\beta}_{\sim R}$  although other numerical algorithms perhaps could. Alternatively, having obtained  $\hat{\pi}_0$  for the model  $\tilde{X}$  in (3.1), the FARM computational procedures can be applied to determine an estimator  $\bar{\beta}_{\sim R}$  for  $\beta_{\sim R}$  which is asymptotically equivalent to  $\hat{\beta}_{\sim R}$ . For this purpose, expressions (3.6) and (3.7) are used with  $X_0$  replaced by  $X_{\sim R,0}$ . The resulting estimators  $\bar{\beta}_{\sim R}$  and  $V_{\bar{\beta}_{\sim R}}$  were

$$\bar{\beta}_{\sim R} = \begin{bmatrix} 0.220 \\ 1.375 \\ 0.436 \end{bmatrix}, \quad V_{\bar{\beta}_{\sim R}} = \begin{bmatrix} 0.0439 & & \\ 0.0015 & 0.7448 & \\ -0.0059 & -0.0091 & 0.2141 \end{bmatrix} \times 10^{-2}. \quad (3.12)$$

Estimators for  $\pi_0$ ,  $\gamma$ ,  $n_{22\dots 2}$ , and  $N$  based on  $\bar{\beta}_{\sim R}$  are shown in Table 2; and for the observed cell frequencies, in Table 1. Finally, it should be noted that the motivation underlying the formulation of  $X_{\sim R}$  was an attempt to identify a model with as few parameters as possible which fitted the data in order to minimize the variance for the estimator of the population total  $N$ . As such, this model is perhaps too aggressive in the sense of being overly data-dependent, and thus the results which are based on it should be interpreted carefully. In this context, however, its application is still of interest from the point of view of illustrating the use of the FARM computational procedures for fitting a non-hierarchical model.

In summary, the matrix formulation (2.8) for log-linear models and the FARM computational procedures (3.2) - (3.7) permit the application of a broad range of comprehensive analyses to multiple recapture census data. However, as stated previously, these methods do require certain fundamental assumptions which may not always apply and thus should be used with caution.

## REFERENCES

- Bartlett, M. S. (1935). Contingency table interactions. Journal of Royal Statistical Society. Supplement 2, 248-52.
- Berkson, J. (1955). Maximum likelihood and minimum  $\chi^2$  estimates of the logistic function. Journal of the American Statistical Association 50, 130-62.
- Darroch, J. N. (1958). The multiple-recapture census, I. Estimation of a closed population. Biometrika 46, 343-59.
- El-Khorazaty, M. N. (1975). Methodological strategies for the analysis of categorical data from multiple-record systems. University of North Carolina Institute of Statistics, Mimeo Series No. 1019.
- El-Khorazaty, M. N. and Sen, P. K. (1976). The capture-mark-recapture strategy as a method for estimating the number of events in a human population with data from dependent sources. Submitted to the Journal of the American Statistical Association.
- Fienberg, S. E. (1972). The multiple-recapture census for closed populations and incomplete  $2^k$  contingency tables. Biometrika 59, 591-603.
- Forthofer, R. N. and Koch, G. G. (1973). An analysis for compounded functions of categorical data. Biometrics 29, 143-57.
- Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. Biometrics 25, 489-504.
- Haberman, S. J. (1974). The analysis of frequency data. Chicago, University of Chicago Press.
- Koch, G. G., Freeman, D. H., Jr., and Tolley, H. D. (1975). The asymptotic covariance structure of estimated parameters from contingency table log-linear models. University of North Carolina Institute of Statistics, Mimeo Series No. 1046.
- Neyman, J. (1949). Contribution to the theory of the  $\chi^2$  test. Proceedings of Berkeley Symposium in Mathematical Statistics and Probability. University of California Press, Berkeley and Los Angeles, 239-73.
- Roy, S. N. and Kastenbaum, M. A. (1956). On the hypothesis of no "interaction" in a multi-way contingency table. Annals of Mathematical Statistics 27, 749-57.
- Wald, A. (1943). Tests of statistical hypothesis concerning general parameters when the number of observations is large. Transactions of the American Mathematical Society 54, 426-82.
- Wittes, J. T. (1970). Estimation of population size: the Bernoulli census. Unpublished Ph.D. thesis, Harvard University, Cambridge, Massachusetts.
- Wittes, J. T., Colton, T., and Sidel, V. W. (1974). Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. Journal of Chronic Disease 27, 25-36.

APPENDIX

The proof that the computational strategy (3.2) - (3.7) yields the same results as the FARM procedures in Koch et al. (1975) is based on the following general argument which can be applied specifically to the multiple recapture census situation by replacing  $\underline{X}$  with  $\underline{X}_0$ ,  $r$  with  $(r-1)$ , and  $(r-1)$  with  $(r-2)$ .

1. Let  $\underline{K} = [-\underline{1}_{(r-1)}, \underline{I}_{(r-1)}]$ . Thus,  $\underline{K}\underline{1}_{(r-1)} = \underline{0}_{(r-1)}$ .

In addition,  $\underline{K}\underline{K}' = \{\underline{1}_{(r-1)}, (r-1) + \underline{I}_{(r-1)}\}$ , so that  $(\underline{K}\underline{K}')^{-1} = \{\underline{I}_{(r-1)} - \frac{1}{r}\underline{1}_{(r-1)}, (r-1)\}$ .

2. Consider the model  $\underline{K}[\log \underline{\pi}] = \underline{K}\underline{X}\underline{\beta}$ .
3. Let  $\underline{X}_C$  be an ortho-complement matrix to the matrix  $[\underline{1}_r, \underline{X}]$  so that  $\underline{X}_C'[\underline{1}_r, \underline{X}] = \underline{0}$ .

4. Then the model (2) is equivalent to the model

$$\begin{bmatrix} \underline{X}' \\ \underline{X}'_C \end{bmatrix} [\underline{K}' (\underline{K}\underline{K}')^{-1} \underline{K}] [\log \underline{\pi}] = \begin{bmatrix} \underline{X}' \\ \underline{X}'_C \end{bmatrix} [\underline{K}' (\underline{K}\underline{K}')^{-1} \underline{K}]\underline{X}\underline{\beta}$$

where  $\underline{K}' (\underline{K}\underline{K}')^{-1} \underline{K} = (\underline{I}_r - \frac{1}{r}\underline{1}_{rr})$  because the columns of  $\underline{X}$  are jointly linearly independent of  $\underline{1}_r$  and  $\underline{X}_C$  is orthogonal to  $\underline{1}_r$ .

5. Let  $\underline{Z}'_1 = \underline{X}'\underline{K}' (\underline{K}\underline{K}')^{-1} \underline{K} = \underline{X}'(\underline{I}_r - \frac{1}{r}\underline{1}_{rr})$  and let  $\underline{Z}'_2 = \underline{X}'_C\underline{K}' (\underline{K}\underline{K}')^{-1} \underline{K} = \underline{X}'_C(\underline{I}_r - \frac{1}{r}\underline{1}_{rr}) = \underline{X}'_C$ . It then follows that  $\underline{Z}'_1\underline{1}_r = \underline{Z}'_2\underline{1}_r = \underline{0}$  and  $\underline{Z}'_1\underline{Z}'_2 = \underline{X}'\underline{X}_C = \underline{0}$  so that  $\underline{1}_r$ ,  $\underline{Z}'_1$ , and  $\underline{Z}'_2$  are mutually orthogonal. Moreover,  $\underline{X}'\underline{K}' (\underline{K}\underline{K}')^{-1} \underline{K}\underline{X} = \underline{Z}'_1\underline{Z}'_1 = \{\underline{X}'\underline{X} - \frac{1}{r}\underline{X}'\underline{1}_{rr}\underline{X}\}$  and  $\underline{X}'_C\underline{K}' (\underline{K}\underline{K}')^{-1} \underline{K}\underline{X} = \underline{X}'_C\underline{X} = \underline{0}$ .

On the basis of all of these results, the model given in (4) may be written in the form

$$\begin{bmatrix} \underline{Z}'_1 \\ \underline{Z}'_2 \end{bmatrix} [\log \underline{\pi}] = \begin{bmatrix} (\underline{Z}'_1\underline{Z}'_1) \\ \underline{0} \end{bmatrix} \underline{\beta}$$

where  $Z_1$ ,  $Z_2$ , and  $1_r$  are mutually orthogonal to each other. However, this framework is identical to that given in Koch et al. (1975) for fitting the log-linear model  $Z_1$  to the data. Thus, the results given there apply to the corresponding estimators  $\hat{\beta}$  and  $V_{\hat{\beta}}$  so obtained. But the weighted least squares fit of the model  $Z_1$  with respect to  $(\log \pi)$  in this framework is identical to the weighted least squares fit of the model  $K X$  in (2) with respect to  $K(\log \pi)$ .

Finally, the fact that

$$\begin{aligned} \frac{\exp(Z_1 \beta)}{1_r' [\exp(Z_1 \beta)]} &= \frac{\exp\{(I_r - \frac{1}{r} 1_r 1_r') X \beta\}}{1_r' [\exp\{(I_r - \frac{1}{r} 1_r 1_r') X \beta\}]} \\ &= \frac{\{\exp(-\frac{1}{r} 1_r' X \beta)\} \{\exp(X \beta)\}}{\{\exp(-\frac{1}{r} 1_r' X \beta)\} \{1_r' [\exp(X \beta)]\}} \\ &= \frac{\exp(X \beta)}{1_r' [\exp(X \beta)]} \end{aligned}$$

demonstrates the equivalence of the models based on  $X$  and  $Z_1$ .