CAPTURE-RECAPTURE MODELS WHEN BOTH SOURCES HAVE CLUSTERED OBSERVATIONS

Charles D. Cowan, U.S. Bureau of the Census and Donald J. Malec, Ohio State University

Introduction

The model most commonly used in capture-recapture estimation of population sizes assumes that individuals are missed or captured in each source independently of all other individuals in the population. However, this assumption may be inadequate for some studies: certain research designs would lead naturally to clustering of misses in the enumeration of a population. This paper will develop a model that will describe the clustering of misses, it will describe biases inherent in the traditional dual system estimator when the model involving clustering holds, and finally, the paper will outline the use of the EM algorithm to estimate the total population size.

Modeling Misses in Each Enumeration

The traditional capture-recapture model assumes that each person in the population has a probability, p_1, of being captured in the first source (the census), and a probability, p_2, of being captured in the second source (the PES), that the captures are independent between sources, and that captures are independent between individuals within sources (i.e. no clustering of individuals). There has been extensive work on relaxing the assumption that captures are independent between sources, but the only successful treatments of this problem have been for capture-recapture studies involving three or more data sources. A review of the methodology for multiple systems that are correlated can be found in Bishop, Fienberg, and Holland (1975). For the assumption that captures are independent between individuals within sources, there has been relatively little work done on relaxing this assumption. The reason is that if captures are correlated (implying in this setting clustered misses within enumerated households), the distributions for the number of misses are no longer binomial. The binomial distribution would only be appropriate for the sums of independent Bernoulli events, and we've lost the independence of the events if we allow captures to be clustered within enumerated housing units.

In the model involving clustering, there are two capture events for each source. There is the event of a household being captured in the first source, with probability h_1, and conditional on the household being captured, each person in the household being captured with probability p_1. For example, in the census a listing is made of all housing units on a block in an address register, and then within each enumerated housing unit a roster is made of all persons in the housing unit on a census questionnaire. Either the housing unit can be left off the address register (with probability $1-h_1$), or a person can be left off the census questionnaire (with probaility $1-p_1$). The same events can be described for source two. The data available for analysis in this model are much more extensive and much more complex than the simple set of data used in the traditional capture--recapture model. The observed data can be expressed in a three - dimensional table with entries m_{ijk} as in Table 2.

The first column and row of the table are counts of housing units observed in only the first or only the second sources respectively, distributed by observed household size (0, 1, 2, . . ., with zero signifying a vacant household). The cell in the upper left corner of this table is empty, the out - out cell. This cell represents housing units missed in both enumerations, and all persons in the housing units who were missed.

The remainder of the table is three-dimensional, with entries being counts of housing units enumerated (captured) in both sources. The counts, denoted by m_{ijk} are counts of households, not persons, but households displaying the the characteristics denoted by the subscripts. The subscripts for m_{ijk} can be interpreted as:

i = the number of persons who match in both households.

j = the number of persons observed in the first source.

k = the number of persons observed in the second source.

As an example, m_{022} would be the count of households for which we observed two persons in the census and two persons in the PES, but none of the people matched between the two households. By implication we can say (in the absence of matching errors) that there are at least four persons in each of the households in the m_{022} count, since there are at least two persons in each of the data sources, but none of the persons match. Note that there may be additional people missed by both sources, so we can only say that there are at least four persons in each household, but there may be five or more persons with some unobserved.

This implies that there are a series of unobserved variables that together comprise the m_{ijk}. These would be counts of households of specific sizes, in which only some or all of the persons are enumerated at each visit. We can denote these variables as X_{ijks}, where the subscripts i, j, and k have the same meaning as in the m_{ijk}, and the subscript s denotes the true but unknown household size. The m_{ijk} are then seen to be the sum of the X_{ijks} over s, or

$$m_{ijk} = \sum_{s=j+k-i}^{\infty} X_{ijks} \qquad (2)$$

Based on the description given previously, the probability for any household of true size s being enumerated twice and having a specific (i,j,k) distribution is:

$$h_1 h_2 \binom{s}{i \;\; j-i \;\; k-i} p_{11s}^{i} \, p_{12s}^{j-i} \, p_{21s}^{k-i} \, p_{22s}^{s-j-k+i} \qquad (3)$$

where h_1 and h_2 are the capture probabilities for households in source one and source two respectively,

the portion in parentheses is a combinatoric describing the ways s persons can be captured twice or only once in either source, and p_{jks} being the probability for each person in a household of size s of being observed:

p_{11s} is the probability of a person being observed both on source one and source two

p_{12s} is the probability of a person being observed in source one but but not in source two

p_{21s} is the probability of a person being observed in source two but not in source one

p_{22s} is the probability of a person being missed in both sources.

There is some evidence that the p_{jks} vary over s from editing operations in the census. Large families in highly urban, low SES areas have problems in some cases defining exactly who is a household member at a particular time. When several related families live near each other, it may also be hard to determine where a particular person should be counted. As a consequence, people in these settings are more likely to be missed in the census. However, it is also true that other variables are likely to be important to stratification of the estimates and are likely to be as important as household size.

We can define

$$p_{1+s} = p_{11s} + p_{12s} \qquad (4)$$

$$p_{+1s} = p_{11s} + p_{21s} \qquad (5)$$

as the marginal probabilities of capture for source one and source two respectively. If

$$p_{11s} = p_{1+s} \cdot p_{+1s} \qquad (6)$$

the captures in the two sources are independent conditional on the households being captured in both data sources.

Finally, we need to consider the distribution of households of size s. To describe the distribution of households, we have a parameter set $\underline{R}$ consisting of parameters $(R_0, R_1, R_2, \ldots)$ which are the proportions of households in the total population of size zero, one, two, etc.. The distribution of captured household sizes would then be multiple hypergeometric. To complete the model specification, we have a parameter N_H, which is the true number of households in the population, and we note that

$$\sum_{s=0}^{\infty} R_s = 1.0 \qquad (7)$$

$$N_H \sum_{s=0}^{\infty} sR_s = N_P \qquad (8)$$

where N_P in the true number of people in the population. As an approximation, we assume the distribution of captured household sizes is multinomial with the aforementioned parameters. The remainder of this paper evaluates the bias in the traditional dual system estimator when the above model holds, and methods of estimation for this model.

<u>Biases in the Traditional Dual System Estimates</u>

To evaluate the bias in the traditional dual system estimates, we will take a Taylor Series expansion of the dual system estimator presented in (1) above, and, taking expected values, retain the first order term in the expansion as an approximation to the true expected value of the estimator.

All the data that is needed for estimation is found in table 2 presented above in the m_{ijk}. To complete the notation started in the previous section, let $m_{.j.}$ be the number of households observed in the first source observed to be of size j that were missed in the second source, and $m_{..k}$ be the number of households in the second source observed to be of size k that were missed in the first source. Then using the totals as presented in table 1 for dual system estimation, the totals can be found as:

$$M = \sum_{ijk} i m_{ijk} \qquad (9)$$

$$N_1 = \sum_{ijk} j m_{ijk} + \sum_{j} j m_{.j.} \qquad (10)$$

and
$$N_2 = \sum_{ijk} k m_{ijk} + \sum_{k} k m_{..k} \qquad (11)$$

The estimator for the population size, N_p, is given by equation (1). From the foregoing, using the first term in the Taylor Series Expansion, we get:

$$E\left[\frac{N_1 \times N_2}{M}\right] \doteq \frac{E(N_1)E(N_2)}{E(M)} , \qquad (12)$$

and so

$$E(N_P) \doteq \frac{E(\sum_{ijk} j m_{ijk} + \sum_{j} j m_{.j.}) E(\sum_{ijk} k m_{ijk} + \sum_{k} k m_{..k})}{E(\sum_{ijk} i m_{ijk})} \qquad (13)$$

Wittes (1970) shows that second and higher order terms in the Taylor Series Expansion are of order $O(N^{-1})$ and can be discarded for large populations. We can evaluate each of these terms by using the fact that the m_{ijk} are simply sums of the unobserved X_{ijks} for which we know the distributions. Starting with the denominator:

$$E(\sum_{ijk} i m_{ijk}) = \sum_{ijk} iE(m_{ijk}) = \sum_{s} \sum_{ijk} iE(X_{ijks})$$

$$= N_H h_1 h_2 \sum_{s} R_s \sum_{ijk} i \binom{s}{i\ \ j-i\ \ k-i} p_{11s}^{i}\, p_{12s}^{j-i}\, p_{21s}^{k-i}\, p_{22s}^{s-j-k+i}$$

$$= N_H\ h_1h_2 \sum_s R_s s p_{11s} \qquad (14)$$

$$E(\sum_{ijk} jm_{ijk}) = \sum_{ijk} jE(m_{ijk}) = \sum_{sj}\sum jE(X_{ijks})$$

$$= N_H h_1 h_2 \sum_s R_s \sum_s j\binom{s}{i\ \ j-i\ \ k-i} p_{11s}^{\,i}\, p_{12s}^{\,j-i}\, p_{21s}^{\,k-i}\, p_{22s}^{\,s-j-k+i}$$

$$= N_H h_1 h_2 \sum_s R_s \left[\sum_{ijk} (j-i)\binom{s}{i\ \ j-i\ \ k-i} p_{11s}^{\,i} p_{12s}^{\,j-i} p_{21s}^{\,k-i} p_{22s}^{\,s-j-k+i} \right.$$

$$\left. + \sum_{ijk} i\binom{s}{i\ \ j-i\ \ k-i} p_{11s}^{\,i}\ p_{12s}^{\,j-i}\ p_{21s}^{\,k-i}\ p_{22s}^{\,s-j-k+i} \right]$$

$$= N_H\ h_1h_2\ \sum_s R_s (sp_{121} + sp_{11s})$$

$$= N_H\ h_1h_2\ \sum_s R_s sp_{1+s} \qquad (15)$$

For the observed variables $m_{.j.}$ and $m_{..k}$ there are corresponding underlying unobservable variables X_{js} and X_{ks} such that

$$m_{.j.} = \sum_j X_{js} \qquad (16)$$

and

$$m_{..k} = \sum_k X_{ks}, \qquad (17)$$

$$\text{with} \quad X_{js} \sim \binom{s}{j} p_{1+s}^{\,j} (1-p_{1+s})^{s-j} \qquad (18)$$

$$\text{and} \quad X_{ks} \sim \binom{s}{k} p_{+1s}^{\,k} (1-p_{+1s})^{s-k} \qquad (19)$$

Using these relationships, we get

$$E(Sjm_{.j.}) = \sum jE(m_{.j.}) = \sum\sum jE(X_{js})$$

$$= N_H h_1 (1-h_2) \sum_s R_s \sum_j j\binom{s}{j} p_{1+s}^{\,j} (1-p_{1+s})^{s-j}$$

$$= N_H h_1 (1-h_2) \sum_s R_s s p_{1+s} \qquad (20)$$

Combining (15) and (20) we get

$$E(\sum_{ijk} jm_{ijk} + \sum_j jm_{.j.}) = N_H h_1 \sum_s R_s s p_{1+s} \qquad (21)$$

Similarly, for the second source we get

$$E(\sum_{ijk} km_{ijk} + \sum_k km_{..k}) = N_H h_2 \sum_s R_s s p_{+1s} \qquad (22)$$

Substituting (14),(21), and (22) into (13) we get the approximation to the expected value of N_p to be:

$$E(N_p) \doteq \frac{(N_H h_1 \ \Sigma R_s s p_{1+s})(N_H h_2\ \Sigma R_s s p_{+1s})}{(N_H h_1 h_2\ \Sigma R_s s p_{11s})}$$

$$= N_H \left[\frac{\Sigma R_s s p_{1+s}\ \Sigma R_s s p_{+1s}}{\Sigma R_s s p_{11s}}\right] \qquad (23)$$

Recalling from (8) that the true value of N_p can be expressed as

$$N_p = N_H\ \Sigma R_s s,$$

it can be seen that the expected value in (23) will not equal the true value except in providential circumstances. Bias can be defined as:

$$E(N_p) - N_p = N_H \left[\frac{\Sigma R_s s p_{1+s}\ \Sigma R_s s p_{+1s}}{\Sigma R_s s p_{11s}} - \Sigma R_s s\right] \qquad (24)$$

The difference in brackets is the average perceived household size minus the true average household size (persons per household). A bias results if the average household size differs from what is observed as a result of the capture process.

Several special cases of interest can be examined to consider where and how biases can enter the estimates. In the case where $p_{11s} = p_{1+s}p_{+1s}$, independence of captures within households of size s for all s, there is no obvious expansion or simplification to determine the direction of the bias. In fact, the bias could still be positive or negative depending on the relationship (relative sizes) between the R_s, p_{1+s}, and p_{+1s}. Another special case is that in which all of the capture probabilities are equal across different sized households. Equation (23) reduces to

$$E(N_p) = N_H\, \Sigma R_s s \left[\frac{p_{1+.}p_{+1.}}{p_{11.}}\right] \qquad (25)$$

From (25), the bias in N_p is now seen to be only a function of correlation bias, whereas in (23) biases could arise from correlation between the sources, heterogeneity among the capture probabilities, or both. The size and direction of the bias can be determined from the size and direction of the correlation. Note that this model explicitly assumes that household captures are independent. If the correlation bias is positive, i.e.

$$p_{11s} = p_{1+.}p_{+1.} + a$$

then (25) demonstrates that the population size will be underestimated.

Finally, the simplest model of all sets $p_{11s} = p_{11.}$, implying total independence between sources within households. From (25) it can be seen that $\hat{N}_p$ is an unbiased estimator of N_p, at least to a first order term. This result says that even though there may be severe clustering of person misses, as long as the captures are independent between and within households between the two sources, then the dual system estimator is still unbiased. This is the same type of assumption currently used in the traditional dual system estimator.

Three examples are presented to give an appreciation for the extent of the bias possible in the estimates when the capture probabilities are heterogeneous, but there is no correlation bias within households (i.e. $p_{11s} = p_{1+s}p_{+1s}$). Table 2 presents examples which retain the same distribution of household sizes for all three examples, and permutations of the capture probabilities.

The examples serve to show that the heterogeneity can lead to an underestimate of the true population size or an overestimate of the true population size. Though the differences look small for average household size, differences between estimates of total population can be quite large. For example, for a state with about two million households, the traditional dual system estimator would underestimate the population by about 58,000 persons. This could have effects on allocations of revenue sharing monies, block grant funding, and other federal allocation programs if different states exhibit different patterns of within household captures.

Estimation

The way the model is formulated, it lends itself quite naturally to estimation of the parameter values using the EM algorithm. The EM algorithm is a two step iterative algorithm which generates maximum likelihood estimates of population parameters. The first step of each iteration is used to generate expected values of unobserved variables, conditional on the data actually observed. In this problem, this would mean generating expected values $E(X_{ijks}|m_{ijk})$ to substitute in place of the values of X_{ijks} which are never observed. The second step of each iteration is to reestimate the population parameters using maximum likelihood techniques and the estimates of the unobserved variables from the first step (known as the complete data).

For this problem, the conditional expected values of the variables in the first step of the $(t+1)^{st}$ iteration are calculated as:

$$E(\overset{t+1}{X}_{ijks}|m_{ijk}) = \frac{m_{ijk}\, \overset{t}{R}_s \binom{s}{i\;\; j-i\;\; k-i} \overset{t}{p}{}_{11s}^{\,i}\; \overset{t}{p}{}_{12s}^{\,j-i}\; \overset{t}{p}{}_{21s}^{\,k-i}\; \overset{t}{p}{}_{22s}^{\,s-j-k+i}}{\sum_s \overset{t}{R}_s \binom{s}{i\;\; j-i\;\; k-i} \overset{t}{p}{}_{11s}^{\,i}\; \overset{t}{p}{}_{12s}^{\,j-i}\; \overset{t}{p}{}_{21s}^{\,k-i}\; \overset{t}{p}{}_{22s}^{\,s-j-k+i}} \qquad (26)$$

$$E(\overset{t+1}{X}_{js}|m_{.j.}) = \frac{m_{.j.}\, \overset{t}{R}_s \binom{s}{j} \overset{t}{p}{}_{1+s}^{\,j} (1-\overset{t}{p}_{1+s})^{s-j}}{\sum_s \overset{t}{R}_s \binom{s}{j} \overset{t}{p}{}_{1+s}^{\,j} (1-\overset{t}{p}_{1+s})^{s-j}} \qquad (27)$$

$$E(\overset{t+1}{X}_{ks}|m_{..k}) = \frac{m_{..k}\, \overset{t}{R}_s \binom{s}{k} \overset{t}{p}{}_{+1s}^{\,k} (1-\overset{t}{p}_{+1s})^{s-k}}{\sum_s \overset{t}{R}_s \binom{s}{k} \overset{t}{p}{}_{+1s}^{\,k} (1-\overset{t}{p}_{+1s})^{s-k}} \qquad (28)$$

For the second step of the iteration, the M step, estimates of the parameter values are calculated as maximum likelihood estimates of the parameters as if the X_{ijks} were known. The calculation of the MLE's draws upon the complete data likelihood and theory already established for the product-multinomial distribution (Bishop, Fienberg and Holland (1975)). To obtain the complete data likelihood, we start with equations (3), (18), and (19), which are the probabilities of persons being captured or not captured within captured households. We then recognize that, since household events are independent between households, the products iX_{ijks}, jX_{ijks}, kX_{ijks}, jX_{js}, and kX_{ks} are total counts of persons who have particular characteristics (e.g. iX_{ijks} is the total number of persons matched in a household observed to have j persons in the first capture, k persons in the second capture, and s persons actually in the household).

An intuitive approach to the derivation can be seen by simply examining the quantities being estimated. Equation (30) attempts to estimate p_{1+s}, the proportion of cases in households of size s who are captured in the first capture effort, regardless of their status in the second capture. This can be seen by examination, where the numerator is the (estimated) number of persons captured in the first sampling for households of true size s, while the denominator is the total of all persons in households of size s.

$$\overset{t}{p}_{1+s} = \frac{\sum_{ijk} j\overset{t}{X}_{ijks} + \sum_{j} j\overset{t}{X}_{js}}{\sum_{ijk} s\overset{t}{X}_{ijks} + \sum_{j} s\overset{t}{X}_{js}} \qquad (30)$$

$$\overset{t}{p}_{+1s} = \frac{\sum_{ijk} k\overset{t}{X}_{ijks} + \sum_{k} k\overset{t}{X}_{ks}}{\sum_{ijk} s\overset{t}{X}_{ijkl} + \sum_{k} s\overset{t}{X}_{js}} \qquad (31)$$

$$\overset{t}{R}_s = \frac{\sum_{ijk} \overset{t}{X}_{ijks} + \sum_{j} \overset{t}{X}_{js} + \sum_{k} \overset{t}{X}_{ks}}{\sum_{ijks} \overset{t}{X}_{ijks} + \sum_{is} \overset{t}{X}_{js} + \sum_{ks} \overset{t}{X}_{ks}} \qquad (32)$$

To estimate p_{11s} we only have data available from the X_{ijks}, and not the row (X_{js}) or column (X_{ks}) variables involving only one capture. Yet to make full use of data in the table, the denominator of the ratio must use the data from the X_{js} and X_{ks}. One could establish a three dimensional table which is only partially complete and iterate to a solution for the p_{11s}, p_{1+s}, and p_{+1s} at each step of the EM algorithm, as recommended in Bishop, Fienberg, and Holland (1975), but this is unnecessary. An equivalent procedure is to allocate the $m_{.j.}$ and $m_{..k}$ to variables X_{ijks} using the proportions calculated in the previous iterate of the algorithm. This is done in the same way as $E(X_{ijks}|m_{ijk})$ for variables Y_{ijks} and Z_{ijks}, where Y_{ijks} is the set of variables which are the proportional allocation of the calculated X_{js}, as if the households had been observed twice and Z_{ijks} bears the same relationship to X_{ks}.

Using these "raked" values, we should be able to estimate the values as:

$$p_{11s}^{t} = \frac{\sum_{ijk} i(X_{ijks}^{t} + Y_{ijks}^{t} + Z_{ijks}^{t})}{\sum_{ijk} s(X_{ijks}^{t} + Y_{ijks}^{t} + Z_{ijks}^{t})} \quad (35)$$

Obvious modifications can be made for the MLE's of p_{11s}, p_{1+s}, and p_{+1s} when restrictions are put on the model, e.g. independence within households or homogeneity across household sizes.

The algorithm is completed by estimating the total number of households using the traditional dual system estimate:

$$N_H = (\sum_{ijk} m_{ijk} + \sum_{j} m_{.j.})(\sum_{ijk} m_{ijk} + \sum_{k} m_{..k})/(\sum_{ijk} m_{ijk}) \quad (36)$$

and

$$N_P = N_H \sum_{s} sR_s \quad (37)$$

For models in which

$$p_{11s} = p_{1+s}p_{+1s} \quad (38)$$

or

$$p_{11.} = p_{1+.}p_{+1.} \quad (39)$$

both independence models, the EM algorithm rapidly converges to a single correct solution no matter what starting points are used for the algorithm. The model described by the restriction in (36) is independence within household size groups, demonstrated earlier to lead to biases because of the heterogenerity of the capture probabilities. The interdependence of the values R_s, p_{1+s}, p_{+1s} makes it impossible to directly estimate the population size except through use of iterative procedures. In the case of the restriction described in (37) where all the capture probabilities are equal for all household sizes, the EM algorithm and the traditional dual system estimator both give unbiased estimates of the population size.

Unfortunately, estimates are not so easily obtained for the full model involving no restrictions on the p_{11s}, or the reduced model where

$$p_{11s} = p_{11.} \text{ for all } s. \quad (40)$$

When the restriction of independence is removed from the model, there is too much indeterminancy in the model for there to be a single solution. The EM algorithm always converges, but to different sets of final parameter values for different starting points. Another way of looking at the convergence of the EM algorithm is that there is a ridge in the likelihood of equally likely points that satisfy the algorithm. The restriction of independence serves to identify a single point on the ridge.

Conclusions and Future Research

The traditional dual system estimator does well under a clustering model at estimating the total population size if all of the assumptions about independence hold and capture probabilites are homogeneous across household size categories. If person misses are clustered within households and within household capture rates differ by households size, the traditional dual system estimates can be biased and a better (unbiased) estimate can be produced using the EM algorithm. Models which allow correlations between within household misses show that the traditional dual system estimator is biased, but the parameters in these models are not estimable.

The model developed above does not allow for variation in the capture probabilities between demographic or geographic subgroups. This can be an especially difficult problem because persons with differing characteristics can be in the same household, and one if forced to model the distribution of these person characteristics for each housing unit of a particular size. This should be the next step attempted in developing a comprehensive model.

Acknowledgements

The authors wish to thank Gary Chase, Bob Fay, Dan Horvitz, and Ken Pollock for suggestions and examples presented in this paper.

Bibliography

Wittes, Janet T. (1970) Estimation of Population Size: The Bernoulli Census. Unpublished Ph.D. dissertation, Harvard University.

Bishop, Yvonne M., Stephen E. Fienberg, and Paul W. Holland (1975) Discrete Multivariate Analysis. The MIT Press, Cambridge, Mass.

Seber, G.A.F (1982) The Estimation of Animal Abundance, MacMillen Publishing Co., Inc. New York N.Y.

Hook, Ernest B., Susan G. Albright, and Philip K. Cross (1980) "Use of Bernoulli Census and Log-Linear Methods for Estimating the Prevalence of Spina Bifida in Livebirths and the Completeness of Vital Record Reports in New York State," American Journal of Epidemiology, 112, pp. 750-758.

Hook, Ernest B. and Regal, Ronald R. (1982) "Validity of Bernoulli Census, Log-Linear, and Truncated Binomial Models for Correcting for Underestimates in Prevalence Studies," American Journal of Epidemiology, 116, pp. 168-176.

Statistical Policy Working Paper 1: Report on Statistics for Allocations of Funds. U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, Washington D.C. 1978.

Table 1: Observed Data For Household/Person Population Estimates

			First Source						
	Housing		Out	In					
		Persons		0	1	2	3	4	5 ...
	Out			$\{m_{.j.}\}$					
Second Source	In	0	$\{m_{..k}\}$	$\{m_{ijk}\}$					
		1							
		2							
		3							
		4							
		5							
		. . .							

Table 2: Examples of Bias in Traditional Dual System Estimation When Person Misses are Clustered by Household and Within Household Captures are Independent

Household Size (s)	Θ_s	Set 1		Set 2		Set 3	
		p_{1+s}	p_{+1s}	p_{1+s}	p_{+1s}	p_{1+s}	p_{+1s}
0	.05	-	-	-	-	-	-
1	.15	.98	.98	.98	.68	.68	.68
2	.20	.95	.95	.95	.71	.71	.71
3	.15	.92	.92	.92	.74	.74	.74
4	.13	.89	.89	.89	.77	.77	.77
5	.10	.86	.86	.86	.80	.80	.80
6	.07	.83	.83	.83	.83	.83	.83
7	.06	.80	.80	.80	.86	.86	.86
8	.04	.77	.77	.77	.89	.89	.89
9	.03	.74	.74	.74	.92	.92	.92
10	.01	.71	.71	.71	.95	.95	.95
11+	.01	.68	.68	.68	.98	.98	.98

	Set 1	Set 2	Set 3
$\left[\dfrac{\Sigma\Theta_s s p_{1+s}\ \Sigma\Theta_s s p_{+1s}}{\Sigma\Theta_s s p_{1+s} p_{+1s}}\right]$	= 3.631	3.691	3.628
$\Sigma\Theta_s s$	= 3.660	3.660	3.660
Relative Bias (line 1/line 2)	99.2%	100.9%	99.1%