AN EMPIRICAL EVALUATION OF THE USE OF ADMINISTRATIVE RECORDS TO PREDICT CENSUS DAY RESIDENCY

Elizabeth A. Stuart, Harvard University and D.H. Judson, U.S. Census Bureau
Elizabeth Stuart, Department of Statistics, 1 Oxford St., Cambridge, MA 02138
(stuart@stat.harvard.edu)

Key Words: Bayesian hierarchical model, administrative records census, nonresponse followup, migration, multiple systems estimation.

Abstract:

Administrative records (AR) databases offer the prospect of reduced respondent burden and substantial cost savings in census taking and are promising for identifying people missed in the census. However, available AR databases have not yet been shown to be sufficient for strict enumeration purposes, mainly because of the lack of synchronicity between census day and the dates that data enter into the AR files. An important unsolved problem in using AR data is determining which correspond to people actually resident on census Day. This paper uses a hierarchical model developed by Stuart and Zaslavsky (2001, 2002) to predict census day residency by modeling migration and observation in the record systems. There are two unique features to the model: it is at an individual level, and uses all data available, including record dates and covariate information. In this application, we use data from the Statistical Administrative Records System, a simulated "administrative records census," to make block-level total population estimates. We illustrate by comparing estimates to Census 2000.

1. Introduction

This work utilizes administrative records to help predict census day residency by using a Bayesian hierarchical model both of migration and of observation in each of the available record systems. This is useful in the context of an administrative records census, or to expand the use of administrative records in multiple system estimation.

The motivation for this work arises from recent research in the use of administrative records in census enumeration. Administrative records have been suggested as a way to enhance conventional census operations, ranging from nonresponse follow-up (e.g. Zanutto and Zaslavsky, 2001; Judson and Bauder, 2003), to a full administrative records census (Scheuren, 1999). An administrative records census could reduce census cost, provide more frequent population counts, and improve the coverage rates of populations traditionally undercounted. However, there are many research questions still to be worked out regarding the use of administrative records (see, e.g., Scheuren, 1999). The Census Bureau's administrative records AREX 2000 experiment and other ongoing evaluations are examining the use of administrative records as a primary source of information.

One of the complications with administrative records is that they do not usually correspond to census day itself, and individuals may move in between the administrative record date and census day. As discussed in Heimovitz (2002), "There is no explicit means of recording migration in administrative records. Migration is captured by address changes that are dependent upon the type of participant and their active involvement in that federal program." This paper uses a model of migration and observation in administrative records to address this problem.

The theoretical work grows out of the multiple system methods literature. Multiple system estimation was originally developed to estimate animal populations, but has found application in census undercount estimation (e.g., Wolter, 1986; Fienberg, 1992), and a variety of other fields. The main idea is to predict the missing cell in the 2^k contingency table that results from k captures, thus estimating the number of unobserved individuals (see, e.g., Pollock, 1991; Seber, 1982). Loglinear or Rasch models

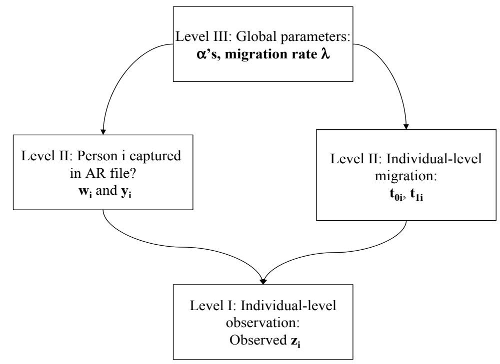
This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. The authors would like to express great appreciation for the contributions of Dean Resnick and Kevin M. Shaw of the Planning, Research and Evaluation Division of the U.S. Census Bureau for their assistance.

are often used to model the cell counts of the contingency table (Fienberg 1972; Fienberg et al. 1999). Bayesian methods have been employed in this problem by George and Robert (1992) and Smith (1991).

Our approach is also related to methods for estimating migration parameters for animal populations, which obtain estimates of the total population size and the migration rates. Much of this work involves modeling migration using Markov Chains (Brownie et al., 1993; Hestbeck et al., 1991). Dupuis (1995) provides a Bayesian approach.

In the US Census context, triple-system estimation using the census, a Post-Enumeration Survey (PES), and a series of administrative records has been suggested as a way to estimate the total population size. Zaslavsky and Wolfgang (1993) discuss the details of using triple system estimation for the census, using 1988 Census dress rehearsal data.

Figure 1: Diagram of Hierarchical Model



2. Overview of Model

Suppose we have a series of record systems (types of administrative records, possibly a census and/or a PES) from a geographic area. Each record is dated, providing evidence of a person being a resident in the area on that date. The total time period covered is T_0 to T_1. Define a population consisting of all people living in this area at some point during this time interval who were captured by at least one of the systems. The model uses the full information in the records, including the dates associated with the records and available covariate information, and accommodates a variety of record types, such as tax records, Medicare claims, and school enrollment lists. In addition, multiple record systems can be modeled concurrently.

We are interested in modeling the in- and out-migration times from the area: t_{0i} (the time person i moved in) and t_{1i} (the time person i moved out). The goal of the inference is to estimate the population size at a particular point in time, usually census day. This is done by obtaining the posterior distributions of the in- and out-migration dates for each individual, leading to an estimate of the probability of residency in the area on census day for each individual.

The model used was developed in Stuart and Zaslavsky (2001, 2002). The hierarchical model is a system of specific models, organized on three levels. Level I describes the observed data (whether someone is observed or not), Level II describes latent migration and record system observation variables, and Level III specifies the distributions governing the global parameters. Figure 1 diagrams how the levels of the hierarchical model relate.

We summarize each level here and give further details in Section 4. Level III consists of the global parameters: α_j represents the probability of capture in record system j, and λ represents a global migration rate. These parameters feed into two models at level II: (1) a model of individual i's presence in the file (represented by variables w_i and y_i, with w_i representing presence in that record system anywhere and y_i representing the date of observation in that system), and (2) a model of individual migration into (t_{0i}) and out of (t_{1i}) the area of interest. Level I represents whether an individual is observed in the file; if she was in that record system anywhere ($w_i = 1$) and was in the area of interest on the date associated with that record system ($t_{0i} < y_i < t_{1i}$), then she will be observed ($z_i = 1$). Otherwise she will not be observed ($z_i = 0$).

3. Data background and preparation

The main administrative records files used at the Census Bureau are the Statistical Administrative Records System (StARS). We utilize StARS 2000 data, and summarize the main record systems in Columns 1-5 of Table 1. In addition, the Census Bureau numident file, an edited version of the Social Security Administration's Numerical Identification File, provided much of the demographic information, death data, and social security numbers used in StARS.

The StARS 2000 file was linked with elements of the 100% Census Unedited File (HCUF). This allows for the use of observation in the census either for evaluation, or as another observation system in the

Table 1: Record Systems Utilized

File Name	Description	Target Population	Date	Usefulness	w	y	α
Center for Medicare and Medicaid Services Enrollment Database (MEDB)	All Medicare beneficiaries alive at cut date or with death date after 1989.	Primarily over age 65 and those with disabilities.	Start date of entitlement, date of 65th birthday, or most recent address change.	Good. Good coverage of over 65 population. Good date information.	Bern(α_{iM})	Uniform(0, 1096) 0=Jan 1, 1998 1096=Dec 31, 2000	0.95 if age > 65, 0.01 otherwise
Internal Revenue Service Individual Master File (IRS 1040)	Tax return data for all individual returns filed in 2000.	All individuals required to submit returns, and dependents. Does not include low-income individuals.	IRS processing date (to nearest week).	Excellent. Good population coverage, and good date information.	Bern(α_I)	Normal(850, 70) 850=Apr 1, 2000	0.9
Internal Revenue Service Information Returns [a]	All information returns filed by employers.	All individuals with income in 1999.	January 1, 2000	Limited. High potential due to population, coverage, but limited dates.	NA	NA	NA
HUD Tenant Rental Assistance Certification System File (HUD-TRACS)[a]	All persons who receive HUD subsidy payments.	Low income individuals.	July 1, 2000	Limited. Low population coverage and limited dates.	NA	NA	NA
HUD Multifamily Tenant Characteristics System File (HUD-MTCS)[a]	All HUD tenants in file.	Low income individuals.	Coverage effective date.	Limited. Very low coverage. Unclear date meaning.	NA	NA	NA
Indian Health Service Patient Registration File (IHS)[a]	All patients in IHS alive on cut date.	American Indians and Alaska natives living in IHS covered areas.	No valid date.	Limited. Low coverage and no date.	NA	NA	NA
Selective Service System Registration File (SSS)	All males who registered with the SSA.	Males age 18-25.	Date of last update.	Fair. Good coverage of target population. Dates may be old.	Bern(α_{iV})	Uniform(0, 821) 0=Jan 1, 1998 821=Apr 1, 2000	0.8 if young male, 0.01 otherwise
US Census [b]	Census file	All individuals	Apr 1, 2000	Very good.	Bern(α_C)	Apr 1, 2000	0.95

[a]Not used in modeling due to coverage or date problems.
[b]Not from StARS file. Included in table to give modeling information.

model. Future work may also include records from the Accuracy and Coverage Evaluation. For this exploratory work, individuals in the final data file included anyone who had any record (either census or any type of administrative record) in the AREX test sites. Each observation contains all records for the individual, including census (HCUF) and all administrative records in StARS 2000.

For this exploration the model was run at the block level, to make the calculations a reasonable size. Future use of the model may be at a larger geographic area, such as block groups or tracts, to obtain population estimates of larger geography units.

For this implementation we used the dates available on the files, making the implicit assumption that they are correct and relevant for the address given. In addition, the lack of unique, relevant dates on some files precluded their use in the model, as indicated in Table 1. Because dates are critical both for this model and for general use of administrative records, this should be an issue for future work; both to assess the validity of the dates, and to explore ways of obtaining better date and address information. We excluded proxy and commercial addresses (as determined by either the census or administrative records) from the data on which the model was developed.

4. Details of the Model

In this section we present specific examples for the model levels. More complex models can also be specified within this overall structure. For more details, see Stuart and Zaslavsky (2001, 2002).

4.1 Migration Model

The migration model describes the migration of the individuals, i.e. the time when each individual resided in the area. Each individual's migration history is summarized by two variables: t_{0i}, the time person i moved into the area, and t_{1i}, the time person i moved out of the area.

The population is modeled as a mixture of two types of people: stayers, who never move in or out of the area, and movers, who migrate to or from the area (although not necessarily during the time period of observation). The in- and out-migration dates are modeled using mixture distributions to account for the two types of individuals. The parameter s represents the proportion of people in the area at a given point in time who are stayers (considered to be constant across time). For the movers, we assume a stationary process with a constant hazard of moving (λ) that is the same for each individual. This implies a censored exponential distribution for the length of residency and a mixture for t_{0i}, with a mass q at T_0 and a uniform distribution over the remaining time, to T_1.

By including stayers in the model, we allow for two migration patterns with only one extra parameter (as opposed to constraining all individuals to have the same migration rate, or adding multiple extra parameters). A more general approach could post-stratify the individuals on the basis of covariates, thus forming cells with homogeneous migration rates.

4.2 Observation Model

The observation model describes the process of observing the individuals in the record systems. A generic approach has one indicator variable for whether an individual was in that record system type (if she filed a tax return, was in Medicare, etc.). Another variable indicates the date when that individual would appear. The interaction of these and the migration dates determines whether the individual would be observed in the record system file available. The exact interpretation of these variables is specific to each record system.

Each record system has its own observation model, described in Table 1. For reference purposes, the beginning of the observation period is defined to be January 1, 1998 (day 0). The end of the observation period is defined to be December 31, 2000. There is a general framework for all of the observation models. Letting j index record types, and i index individuals, the following variables are defined for each of the record systems ($j = 1, ..., J$):

T_{0j}: Beginning of time period covered by record type j

T_{1j}: End of time period covered by record type j

w_{ij}: Indicator for whether individual i has a record of type j ($w_{ij} = 1$ does not necessarily imply they are observed in our file, but they are observed if their observation date and migration dates are appropriate.)

y_{ij}: Date associated with record system j for individual i

z_{ij}: Indicator for whether individual i observed in record system j ($z_{ij} = 1$ if $w_{ij} = 1$, y_{ij} in the record system's observation period and the individual was in the area of interest at y_{ij}; $z_{ij} = 0$ otherwise)

For both the MEDB and SSS files, the endpoints of the uniform distributions are determined by the file coverage dates. The IRS 1040 distribution was estimated by examining the empirical distribution of dates in the files. The dates were centered around April 30 (2 weeks after the tax filing deadline of April 15), possibly due to processing and mail lags. In later work, the full distribution could be estimated empirically rather than using the normal approximation. This could also be possible with other large files such as the MEDB.

These models are combined under the assumption that, conditional on migration dates, age, and gender, observation in the systems is independent. Future versions of the model may relax this independence assumption, as discussed by Stuart and Zaslavsky (2001, 2002).

4.3 Parameter Values

For this illustration of the model, the α parameters and the migration parameters are set constant at values estimated from external sources. In future work, these parameters should be modeled with a prior distribution and drawn from their posterior distribution to allow for some uncertainty in the estimates. In the results shown in Table 2, the values shown in Column 8 of Table 1 were used.

The proportion of stayers in the population was set at $s = 0.25$ and the migration rate for the movers was set at $\lambda = 1/1825$. This reflects an average length of residency of five years, which is based on a Census Bureau report (Hansen, 1998). Sensitivity to these parameters is discussed in Section 6.1.

5. Inference

5.1 Levels of Inference

Due to the hierarchical structure of the model, as described in Section 2, the model allows inference on three levels: global parameters of coverage probabilities and migration, individual migration times, and individual observation and record histories.

In the census context, we are mostly interested in inference on the second level, regarding the migration dates for individuals. Posterior estimates of each individual's migration dates lead to estimates of the probability of residency, which in turn lead to an estimate of total population size on census day.

5.2 Computational Methods

Draws from the joint posterior are obtained by running a Gibbs sampler, which iteratively draws from each of the full conditional posterior distributions and converges to the joint posterior (Geman and Geman, 1984). Convergence to the posterior distribution is assessed using plots of the draws as well as multiple chains. For more detail on the use of Gibbs sampling in this model, see Stuart and Zaslavsky (2001,2002).

6. Exploratory Results

Exploratory runs on a small group of blocks give interesting results and exhibit the potential for this method. When the census is used only for diagnostic purposes, to assess the coverage of the StARS systems, the estimate of the population size on census day is just slightly larger than the number caught by the census. Results for a few blocks are shown in Table 2 below.

The two columns of estimates result from two different ways of running the model. The first estimates column models the census and StARS 2000 observations together. The second estimates column uses only StARS 2000. The comparison of these estimates with the number caught by the census is a diagnostic for how well the administrative records estimate the census day population size, under the assumption that the census enumeration is correct.

In general the results are in the correct range, although are somewhat higher than the census counts. The overestimation in comparison with direct census counts is particularly true for Baltimore City. This is not surprising, as Baltimore City is considered a "hard to enumerate" area, and possibly had census undercount. This parallels results found by Heimovitz (2002). Similarly, but to a lesser extent, the estimates for all areas appear to overcount, compared with the census enumeration. This discrepancy could be caused by either census undercount or by the StARS or AREX processing, and is an issue that should be explored further. The issue of using census counts as a benchmark is discussed further in Section 7.1.

Because we assume that all census enumerations are correct, when the census is included, the population estimates are slightly higher than when the census is not used. This extra information the census provides also results in tighter intervals when the census is utilized.

Interestingly, the AREX evaluation found that the StARS population estimates tended to undercount the population. It is possible that our method, which seems to overcount, could provide a counterbalance. We of course get a higher estimate than StARS since this model is a form of multiple system estimation, rather than simply the tabulation done in StARS.

The model also provides estimates of the probability of residency on census day for each individual. We thus can use the model to assess which observation histories provide the most information for census residency. These results are not shown here due to confidentiality reasons, however the results are as we would expect. An individual observed close to census day (for example, with a tax return on March 17, 2000), has a very high probability of still being in the area on census day. However, an individual that is observed only on February 25, 1998 is less likely to still be in the area on census day. For individuals not observed by the census, their probabilities of being there on census day are much lower when the census is included in the model run. This makes sense, as when we use the census file and they are not in it, that is much stronger evidence that they were in fact not there on census day. In comparison, when the census file is not used, whether they are observed or not by the census does not affect this probability.

When the census is utilized, the probabilities of census day residency are split between two modes; either very low ($< .2$) or very high ($> .8$). There are not many individuals in the middle range, which makes sense if we assume that the census has few, if any, errors (a reasonable assumption). Being captured (or not) by the census is a very good proxy for actual census day residency!

6.1 Sensitivity to Parameter Values

In the current model runs, the file coverage parameters (the α's) are estimated in a very crude way, from external data sources and documentation. Due to the currently ad hoc nature of their values, a sensitivity analysis was conducted to assess model sensitivity to these values. A 10% increase in these parameter values (which corresponds to substantial differences in the prior beliefs about coverage rates) resulted in a decrease in the census day population size estimate of approximately 2.5%-3.5%. The probabilities of census day residency for individuals were even less affected. A similar analysis assessed sensitivity to the migration parameters, and the results are even less sensitive, with a 10% change in the migration parameters resulting in less than a 1% change in the census day population estimate.

In parallel work, a theoretical examination of parameter sensitivity and frequentist coverage rates is underway. It appears that the model is not overly sensitive to these values and that the estimated census day population size is at least in the right range for a wide variety of parameter values. However, these parameters should be better estimated and this topic further explored.

Table 2: Block Estimates

Location of Block	Number Caught by Census	Census plus StARS 2000 Population Estimate and 95% Posterior Interval	StARS 2000 Population Estimate and 95% Posterior Interval
Baltimore City, Maryland	224	301 (295, 308)	282 (272, 291)
Baltimore County, Maryland	298	333 (327, 339)	303 (295, 313)
Douglas County, Colorado	284	313 (310, 317)	298 (292, 305)
El Paso County, Colorado	172	184 (181, 187)	178 (173, 184)
Jefferson County, Colorado	270	321 (315, 325)	297 (288, 305)

7. Discussion and Future Work

7.1 Conclusions and Ongoing Work

This work indicates the potential for this model in estimating census day population size. However, there is also much evidence of the need for future work. As discussed above, sensitivity to parameter values should be further examined. In addition, methods to estimate the necessary parameters should be explored. This may involve external studies, or the use of pre-existing data.

Since the blocks discussed in Section 6 are not representative of the country as a whole, work is underway to construct population estimates using this model in a larger sample of Census 2000 collection blocks. This work will give an indication for how well the model works in a more general setting, and perhaps provide guidance on when the results are more trustworthy, by identifying the characteristics of blocks in which the model works well.

The trend of possible overestimation of population size should be examined, to determine if it is unique to the blocks examined so far, or more systematic. If it appears to be more systematic, the model should be examined to determine the cause (and possible solution) of this. One suggested cause is mis-estimation of the file coverage and/or migration parameters.

This possibility of overcount also depends on the true size of the population on census day. If the census is deemed to be the truth, then the model does overestimate the population size. However, if in fact the census is an undercount, it may be that this model is correctly estimating the population size. Since the motivation for this work is to improve census enumeration, using the census as the benchmark has serious limitations. In particular, it is difficult to determine the true population size without intensive field work (as in Zaslavsky and Wolfgang, 1993). How to properly set standards or evaluate methods such as this is thus an open topic.

Finally, this work with actual administrative records data has suggested new aspects of the theoretical model. One area is to differentiate between household and individual level moves. The current model is at an individual level, however most moves are actually at a household level; incorporating this fact into the model could lead to better estimation. Similarly, mobility is often dependent on geographic area, and incorporating local information and migration patterns could be of great benefit.

7.2 Limitations and Extensions

There are a number of limitations to the model. One is that the implementation is currently very computationally intensive. One of the goals of the ongoing work discussed above is determining in which blocks the model can be of the most use, and if there are ways of easily generalizing the results.

In addition, the individual record observation models are combined assuming independence across systems, which may be unrealistic. This work was done under the assumption of high data quality and exact matching between the data sources. In reality, matches are often imperfect and the dates and addresses available may also be incorrect.

Due to the large number of parameters in the full model, many simplifying assumptions are made. For example, individuals are assumed to have the same probabilities of capture, conditional on a few eligibility criteria (such as age for the Medicare file and gender for the Selective Service file), and we assume only the two types of migration patterns. Future work should relax some of these assumptions. The model is very general and can allow for more complicated models, including models of inexact matching, erroneous enumerations, or heterogeneity of capture or migration probabilities. These ideas are discussed further in Stuart and Zaslavsky (2001, 2002).

7.3 Applications

One possible use of the model is in generating yearly population estimates. Since the IRS 1040, Medicare, and Selective Service files (the 3 StARS 2000 files utilized here) are available on a yearly basis, this model could be used to assist in intercensal population estimates.

However, due to the mixed results found in replicating census population counts, this model may be particularly useful for targeting individuals or blocks for more intense follow-up. Individuals with very high or very low probabilities of census day residency could be given low probability of field follow-up, and resources instead could be concentrated on individuals that have more ambiguous results. Similarly on a block level, the model could be used to estimate the census day population size for blocks throughout the country and field follow-up resources could be concentrated on blocks that have initial census returns very different from those predicted by the model.

The US Census Bureau currently uses hot-deck methods to impute the occupancy status, household size, and demographics for $1-2\%$ of housing units that have unknown status after follow-up. This method could be useful for this problem, as a model based alternative to the hot-deck procedure.

We recommend that further work be undertaken to improve the quality of dates and addresses available on the administrative records, to make them more contemporaneous with census day. We echo Heimovitz' (2002) finding in the AREX 2000 evaluation that "Demographic events [such as births, deaths, migration] and/or reporting lag impacted the accuracy of AREX counts. StARS processing needs to synchronize dates in administrative data to replicate census place-time reporting requirements, perhaps obtaining quarterly updates from data providers." Although an "actual enumeration" may be required and administrative records may or may not be able to be used for direct census population estimation, they nonetheless have great promise in assisting with census procedures.

References

[1] Brownie, C., Hines, J.E., Nichols, J.D., Pollock, K.H., and Hestbeck, J.B. (1993). Capture-recapture studies for multiple strata including non-Markovian transitions. *Biometrics* 49: 1173–1187.

[2] Darroch, J.N., Fienberg, S.E., Glonek, G.F.V., and Junker, B.W. (1993). A three-sample multiple-recapture approach to census population estimation With heterogeneous catchability. *Journal of the American Statistical Association* 88: 1137–1148.

[3] Dupuis, J.A. (1995). Bayesian estimation of movement and survival probabilities from capture-recapture data. *Biometrika* 82: 761–772.

[4] Fienberg, S.E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika* 59: 591–603.

[5] Fienberg, S.E. (1992). Bibliography on capture-recapture modeling with application to census undercount adjustment. *Survey Methodology* 18: 143–154.

[6] Fienberg, S.E., Johnson, M.S., and Junker, B.W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society A* 162: 383-405.

[7] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721–741.

[8] George, E.I. and Robert, C.P. (1992). Capture-recapture estimation via Gibbs sampling. *Biometrika* 79: 677–683.

[9] Hansen, K.A. (1998). Seasonality of moves and duration of residence. US Census Bureau Current Population Reports, Household Economic Studies. P70-66.

[10] Heimovitz, H.K. (2002) *Administrative Records Experiment in 2000 (AREX 2000): Outcomes*. Administrative Records Research Staff. Washington, DC: Bureau of the Census.

[11] Hestbeck, J.B., Nichols, J.D., and Malecki, R.A. (1991). Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. *Ecology* 72:523–533.

[12] Judson, D.H. and Bauder, Mark (2003). *Administrative Records Experiment in 2000 (AREX 2000): Household Level Analysis*. Administrative Records Research Staff. Washington, DC: Bureau of the Census.

[13] Pollock, K.H. (1991). Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *Journal of the American Statistical Association* 86: 225-238.

[14] Scheuren, F. (1999). Administrative records and census taking. *Survey Methodology* 25: 151-160.

[15] Seber, G.A.F. (1982). *The estimation of animal abundance and related parameters, 2nd edition*. New York: MacMillan Publishing Co., Inc.

[16] Smith, P.J. (1991). Bayesian analyses for a multiple capture-recapture model. *Biometrika* 78: 399-407.

[17] Stuart, E., Zaslavsky, A.M. (2001). Using administrative records to predict census day residency. *ASA Proceedings of the Joint Statistical Meetings*. Alexandria, VA: American Statistical Association.

[18] Stuart, E., Zaslavsky, A.M. (2002). Using administrative records to predict census day residency. In *Case Studies in Bayesian Statistics, Volume VI*, Edited by Gatsonis, C., Kass, R.E., Carriquiry, A., Gelman, A., Higdon, D., Pauler, D.K., Verdinelli, I. New York: Springer, 335-349.

[19] Wolter, K.M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association* 81: 338–346.

[20] Zanutto, E. and Zaslavsky, A.M. (2002). Using administrative records to impute for nonresponse. In Groves, R.M, Dillman, D.A., Eltinge, J.L, Little, R.J.A., Eds., *Survey Nonresponse*. New York: John Wiley and Sons, 403-415.

[21] Zaslavsky, A.M, and Wolfgang, G.S. (1993). Triple-system modeling of census, Post-Enumeration Survey, and administrative-list data. *Journal of Business & Economic Statistics* 11: 279–288.