Using Administrative Records to Predict Census Day Residency

Elizabeth Stuart *
Department of Statistics
Harvard University
1 Oxford St.
Cambridge, MA 02138

Alan Zaslavsky
Department of Health Care Policy
Harvard Medical School
180 Longwood Avenue
Boston, MA 02115

August 13, 2001

Abstract

Administrative records are a promising data source for estimating census coverage or identifying people missed in the census. An important unsolved problem in using records is determining which of them correspond to people actually resident on Census day. We propose a hierarchical model in which one level describes the migration process, and the other describes the probabilities of observation in each of the available record systems. The observation model uses the full information in the records, including the dates associated with the records and available covariate information, and can accommodate a variety of record types, such as tax records, Medicare claims, and school enrollment lists. In addition, multiple record systems can be modeled concurrently simply by multiplying the likelihood of observation for each type. Posterior distributions of the in- and out-migration dates are obtained, leading to an estimate of the probability of residency in the area on Census day. This work could be useful in the context of an administrative records census, or as a way of expanding the role of administrative records in triple system estimation.

1 Introduction

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

This work has its basis in the methods of multiple system estimation. The field of multiple recapture estimation was originally developed as a way of estimating animal populations, but has found application in the Census undercount estimation (Fienberg 1992), as well as a variety of other fields. The idea is to capture a set of animals, mark them in some way, release them, and then make further captures at later points in time. A 2^k contingency table can then be written, where k is the number of captures. The contingency table indicates how many individuals were caught in each possible combination of captures.

*Supported by the US Bureau of the Census and a National Science Foundation Graduate Research Fellowship

One cell will be missing: the cell for individuals missed by all captures. The role of modeling is to estimate the size of this cell, thus estimating the total population size. The situation with two captures is known as dual system estimation, and similarly, that with three captures is called triple system estimation. See Darroch et al. (1993), El-Khorazaty et al. (1977), Pollock (1991), and Seber (1982) for more information on multiple system estimation.

The early work in this field rested on a number of assumptions. These are: the population is closed (no birth, death, or migration), the captures are independent, each individual has the same probability of capture, and individuals can be perfectly matched between captures.

Work done more recently has tried to relax some of the above assumptions. Log-linear models have been used to model the cell counts of the contingency table (Bishop et al. 1975, Fienberg 1972). This allows the model to include dependencies between the lists. Unequal capture probabilities can be accommodated by calculating estimates by strata, or by the use of a Rasch model (Fienberg et al. 1999). Bayesian methods have also been employed in this problem. In particular, George and Robert (1992) use a Gibbs sampling approach, while Smith (1991) compares Bayes, empirical Bayes, and Bayes empirical Bayes solutions.

Because of the interest in migration, this work also relates to the literature on the estimation of migration parameters for animal populations. There is a large literature on this topic, mostly as an outgrowth of the capture-recapture work. Most of these papers assume that several (usually 3-5) geographical areas have been defined and they attempt to estimate the population size and the transition probabilities among the areas. This is done by capturing animals in each location at a set of time periods, and keeping a record of where and when each animal is observed. Estimates of the total population size and the migration rates are then obtained. 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 context of the US Census, triple-system estimation has been suggested as a way to estimate the total population size. The three systems are usually taken to be the Census itself, the Post-Enumeration Survey (PES), and a series of administrative lists. One of the drawbacks of the use of administrative records is that they usually cover a different time period than Census day, sometimes quite a bit earlier than Census day itself. We have thus developed a model of migration that allows prediction of whether someone is still a resident on Census day, given that we have a set of administrative records for them. As will be discussed later, since the administrative records are available nationally, the model can also be used to assist in small area undercount estimation across the country. For more information on administrative records and the Census, see Larsen (1999), Scheuren (1999), and Zaslavsky and Wolfgang (1993).

2 Overview of Model

We propose a hierarchical model in which one level describes the migration process, and the other describes the probabilities of observation in each of the available record systems. The observation 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 are modeled concurrently simply by multiplying the likelihood of observation for each type. The posterior distributions of the in- and out-migration dates are obtained, leading to an estimate of the probability of residency in the area on Census day for each individual.

Suppose we have a list of administrative records (possibly also with the Census and/or the PES) from a geographic area, covering the time period T_0 to T_1. Define a population consisting of all people living in this area at any point during this time interval, thus including individuals captured by some, all, or none of the systems. 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 size of the population at a particular point in time, usually Census day.

The model is a hierarchical model with 3 levels:

$$\text{Level 1:} \quad P(\text{observation history}|\text{migration dates, covariates, parameters})$$

$$\text{Level 2:} \quad P(\text{migration dates}|\text{covariates, parameters})$$

$$\text{Level 3: Priors on the parameters}$$

Level 1 models each individual's observation in the administrative records. The likelihoods of observation in each of the systems are multiplied together to obtain the full observation likelihood. Level 2 describes the migration history for each individual: the in- and out-migration dates. These migration events are observed through the observation history in Level 1. Level 3 describes the prior assumptions on the parameters. We make prior assumptions about the parameters by setting them equal to pre-specified values, or by specifying a distributional form for them. These prior assumptions may also depend on available covariates.

3 Details of Model

In this section we present specifics for the models at each level, giving examples of what each model might look like. These are simplified examples for the purpose of exposition, and more complicated models can be specified without changing the overall structure.

3.1 Migration Model

Level 2 describes the migration of the individuals, i.e. the time when the individual is in the area. This models what is actually happening for each individual in the population. Each individual's migration history is summarized by two variables: t_{0i}, the time person i moved to the area, and t_{1i}, the time person i moved out of the area.

One simple model assumes a certain percentage of the population is present at the beginning of the time period, and a constant hazard of moving for each individual. The two parameters that describe this are the migration rate (λ, the inverse of the average length of duration), and the proportion of people in the population who were in the area at the beginning of the time period of interest (q). This model implies an exponential distribution for the length of residency and a mixture for t_{0i}, with a mass at T_0 and a uniform distribution over the remaining time, to T_1.

An alternative would be to calculate the proportion of people in the area at the beginning using the model implied by the constant hazard of moving. However, this implies an additional assumption that the constant hazard model applies throughout the indefinite past. Although the constant hazard assumption is appropriate over a short time period, it is less appropriate over a longer time frame. Making this additional assumption would imply that the people who have been in the area for a long time are homogeneous with those who just moved in. Our more flexible model allows for possibilities such as a mixture of long- and short-stayers in the population, with different migration rates for the two groups.

3.2 Observation Model and Examples

The observation level (level 1) describes the process of observing the individuals in the record systems. The migration history is observed through these record systems as each person's opportunity to be observed depends on their migration history. We lay out a general framework for the probability of observing a set of records given the migration history and covariates. This general model can accommodate many types of record systems. A model for observation in each of the record systems is then generated, and the full likelihood of observation is the product of the likelihoods of observation in each of the systems.

This example does assume conditional independence of each individual's observation in each of the record systems. Although this is a fairly common assumption in multiple system estimation, many studies have shown that it is not a very good approximation. More complicated versions of our model can accommodate dependence among the systems in two ways. One way is to allow for individual level heterogeneity in the probabilities of observation, which would provide a link between the systems for each individual. A second way is to model the joint distribution of observation in multiple systems directly.

We first give the general framework for a record system, and then go into a few examples of how the model can be used in specific cases. For notational convenience, let j index the types of records (could be the Census, the PES, or types of administrative records), and i index individuals.

The following variables are defined for each of the record systems ($j = 1, ..., J$):

w_{ji} = Bernoulli variable indicating person i being in record system j at any time
  $w_{ji}|\alpha_{ji} \sim$ Bernoulli($\alpha_{\rm ji}$)
  α_{ji} represents the probability of individual i having record type j
y_{ji} = Date associated with record type j for individual i
  Distribution will depend on record type
z_{ji} = Indicator for individual i being observed in file j
  Function of w_{ji}, y_{ji}, migration dates, and file coverage
  $z_{ji} = Z_j\left(w_{ji}, t_{0i}, t_{1i}, y_{ji}, T_{0j}, T_{1j}\right)$
T_{0j} = Beginning of time period covered by record type j
T_{1j} = End of time period covered by record type j

Define T_0 as $\min\{T_{0j}\}$, the beginning of the time period covered by any source, and T_1 as $\max\{T_{1j}\}$, the end of the time period covered by any source.

The pieces of the model are combined by multiplying the likelihoods of each record type. The full likelihood is thus of the following form:
$L(z|\theta) \propto \prod_i \left[\prod_j P(w_{ji}|\alpha_{ji}) P(y_{ji}|w_{ji}, \alpha_{ji}) P(z_{ji}|y_{ji}, w_{ji}, t_{0i}, t_{1i}, T_{0j}, T_{1j})\right] P(t_{0i}, t_{1i}|\lambda, q)$.

This notation for the observation model can accommodate a variety of record systems, including administrative records files, the Census, and the PES. The framework stays the same, but the specifics of the distributions change based on the type of record system. Below are a few examples of the types of records that can be modeled in this way.

3.2.1 Census

In the case of the Census, $w_{Ci} \sim \text{Bernoulli}(\alpha_{\text{Ci}})$ for all i. α_{Ci} depends on each individual's characteristics, as well as the undercount rate. Since the Census records cover just one day, $y_{Ci} = $ April 1 for everyone, and T_{0C} and T_{1C} are both April 1. The function for z_{Ci} is then $z_{Ci} = 1$ if $\{w_{Ci} = 1, t_{0i} \leq \text{April 1} \leq t_{1i}\}$, $z_{Ci} = 0$ otherwise.

3.2.2 Tax Returns

For tax returns, $w_{Ti} \sim \text{Bernoulli}(\alpha_{\text{Ti}})$, where α_{Ti} depends on personal characteristics, and represents the probability that someone with person i's characteristics files a tax return. This may depend on characteristics such as age or region of the country. Since tax returns are generally filed around April 15, the distribution of y_{Ti} is centered around April 15, with some spread for early or late filers. Two choices here would be to model this parametrically or non-parametrically. A parametric model might use a Gamma distribution. However, with a large data set, a non-parametric estimate of the distribution of filing dates is possible. T_{0T} is the beginning of the time period covered by the file, and T_{1T} is the end date of the period covered by the file. The function for z_{Ti} is then $z_{Ti} = 1$ if $\{w_{Ti} = 1, t_{0i} \leq y_{Ti} \leq t_{1i}, T_{0T} \leq y_{Ti} \leq T_{1T}\}$, and $z_{Ti} = 0$ otherwise.

3.2.3 Driver's Licenses

Although driver's licenses are unlikely to be used as a record system in the Census context because of complications of state level laws and data files, they are a good, intuitive example of how the model works. In this case, $w_{Di} \sim \text{Bernoulli}(\alpha_{\text{Di}})$, where α_{Di} depends on personal characteristics (in particular, age), and location in the country. α_{Di} represents the probability that someone with person i's characteristics has a driver's license. Since most driver's licenses are renewed every 3 or 5 years, on the individual's birthday, we assume that the distribution of renewal dates, y_{Di}, is Uniform. Since we are only concerned with the most recent renewal, the right endpoint of this distribution is T_{1D} (the endpoint of our observation interval), and the left endpoint is $T_{1D} - R$, where R is the length of time between renewals. We then assume that anyone with a driver's license who was in the area would have had to renew their license at some point in this interval. The function for z_{Di} is then $z_{Di} = 1$ if $\{w_{Di} = 1, t_{0i} \leq y_{Di} \leq t_{1i}, T_{0D} \leq y_{Di} \leq T_{1D}\}$, and $z_{Di} = 0$ otherwise.

3.2.4 Other Types of Records

Other types of records that could be modeled in this way include the Social Security Service's Master Beneficiary Record, which is a list of anyone entitled to Social Security Benefits, updated monthly. Each individual would have a probability of being a Beneficiary in each month, and their observation date would be modeled as uniform through the month. The monthly files could give us fairly precise information on when the individuals moved to or from the area.

Similarly, records such as Medicare can be modeled, but are somewhat more complicated. In this case we can still estimate the probability that an individual is a Medicare recipient. However, the distribution of claims will be more complicated since some people will have many claims in a short time period, while others may have claims very spread out. A mixture model may be of use here, to model the different types of people. The specifics still have to be worked out, but the general framework still applies.

4 Inference

The structure of the hierarchical model allows inference on each of the 3 levels: global parameters such as the migration rate, individual migration times, and individual observation and record histories. The level of inference will depend on the goal. For example, inference about the global migration parameters may be of interest to sociologists interested in studying migration patterns. This flexibility of levels of inference enables the model to be useful for a variety of purposes.

In the Census context, we are mostly interested in inference on the second level, regarding the migration dates for individuals. It is possible to obtain posterior estimates of individual's migration dates, which lead to estimates of the probability of residency, and in turn lead to an estimate of population size on Census day. An example of this is given below.

5 Simulation Example

5.1 Set-up

The first situation considered is a simple example where we have a file of driver's license records as well as the Census. As discussed above, this is not entirely realistic for Census population estimation, but it provides a simple example to illustrate the main ideas of the model, including inference about residency on Census Day.

Assume that the observation period starts at $T_0 = 0$ and ends at $T_1 = 365$ (measured in days). Census day is at the end of this time period, at day 365. The observation model for the Census is that described above, with a file covering Census day. We assume that $\alpha_{Ci} = \alpha_C$ for all i, implying that everyone has the same probability of being in the Census. This is an assumption that is easily relaxed.

The observation model for the driver's licenses is also described above. Again, we assume that $\alpha_{Di} = \alpha_D$ for all i. We assume that people have to renew their licenses every year ($R = 365$) and that the Driver's License file coverage is 1 year, ending at Census Day. The distribution of the most recent renewal date is thus approximated as Uniform$(0, 365)$. This set-up gives us more information on the migration dates and file coverage. If someone is not observed in the driver's license file, we know that it is either because they do not have a license ($w_{Di} = 0$) or because they were not in the area at the time of their renewal ($y_{Di} < t_{0i}$ or $y_{Di} > t_{1i}$).

The migration model is that described above, with a mixture model for t_{0i} and an exponential distribution for the time before moving out. For clarity and ease of exposition of the inference part of the model, q and λ are set to be constant during the initial simulations. The values are $q = .8$ and $\lambda = \frac{1}{1825}$, which correspond to an average duration of stay of 5 years. The prior distributions used for α_C and α_D are Beta$(1, 1)$

5.2 Details of Computations

The posterior distribution of the parameters is obtained by multipling the likelihood by the prior. 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. The Gibbs sampler iterates through the following steps:

Let $\Theta = \{\{t_{0i}\}, \{t_{1i}\}, \{w_{Ci}\}, \{w_{Di}\}, \{z_{Ci}\}, \{z_{Di}\}, \{y_{Ci}\}, \{y_{Di}\}, \lambda, q, \alpha_C, \alpha_D\}$

1. Global Parameters
   (a) $\alpha_C|\Theta_{\setminus\alpha_C} \sim \mathrm{Beta}(a_{\alpha_C} + \sum_i w_{Ci}, b_{\alpha_C} + N - \sum_i w_{Ci})$
   (b) $\alpha_D|\Theta_{\setminus\alpha_D} \sim \mathrm{Beta}(a_{\alpha_D} + \sum_i w_{Di}, b_{\alpha_D} + N - \sum_i w_{Di})$
   (c) $\lambda|\Theta_{\setminus\lambda} \sim \mathrm{Gamma}(a_\lambda + N, b_\lambda + \sum_i(t_{1i} - t_{0i}))$
   (d) $q|\Theta_{\setminus q} \sim \mathrm{Beta}(a_q + \sum_i \delta(t_{0i} = T_0), b_q + \sum_i \delta(T_0 < t_{0i} \leq T_1))$
   Note: In this simulation, λ and q were not drawn. The distributions provided here would be used if λ and q were not set to be constant.
2. Individual Migration Parameters
   (a) $t_{0i}|\Theta_{\setminus t_{0i}} \sim \lambda e^{\lambda t_{0i}} \left(q\left(\delta(t_{0i} = T_0)\right) + (1-q)(\delta(T_0 < t_{0i} \leq T_1))\right) I\{t_{0i}^L \leq t_{0i} \leq t_{0i}^U\}$
   t_{0i}^L and t_{0i}^U are bounds on t_{0i}, determined by the set of records observed
   (b) $t_{1i}|\Theta_{\setminus t_{1i}} \sim \mathrm{Exp}(\lambda)\mathrm{I}\{t_{1i}^L \leq t_{1i} \leq t_{1i}^U\}$
   t_{1i}^L and t_{1i}^U are bounds on t_{1i}, determined by the set of records observed
3. Individual Observation Parameters
   (a) $y_{Ci}|\Theta_{\setminus y_{Ci}}$ =Census Day, for those with $w_{Ci} = 1$
   y_{Ci} undefined if $w_{Ci} = 0$
   (b) $y_{Di}|\Theta_{\setminus y_{Di}} \sim \mathrm{Uniform}(y_{Di}^L, y_{Di}^U)$
   y_{Di}^L and y_{Di}^U are determined by the set of records observed
   y_{Di} undefined if $w_{Di} = 0$
   (c) $w_{Ci}|\Theta_{\setminus w_{Ci}} \sim \mathrm{Bernoulli}(\alpha_C)$, unless determined by observation history
   - e.g. $w_{Ci} = 1$ if $z_{Ci} = 1$
   (d) $w_{Di}|\Theta_{\setminus w_{Di}} \sim \mathrm{Bernoulli}(\alpha_D)$, unless determined by observation history

Note that some of these distributions involve N, the population size. Since not everyone in the area is observed in one of the systems, N is unknown. However, with good file coverage and enough administrative record files, the number of people observed in at least one of them should approach the true value of N. A later version of the model will allow for individuals who were in the area but missed by all files. This can be done by drawing a value of N from its posterior distribution, and then imputing individuals for whom $z_{Ci} = 0$ and $z_{Di} = 0$ but who were in the area at some point between T_0 and T_1.

The ranges of possible values for t_{0i}, t_{1i} and y_{Di} are determined by the records observed for each individual and the current values of the other parameters. The dates of observation also determine how much information there is in the data for each individual. For example, an individual with a Driver's License observed on day 25 and also observed on Census Day (day 365) has much more information than someone observed only on day 25. Likewise, an individual observed only in the Driver's License file, on day 350, has much more information than someone observed only in the Driver's Licenses but only on day 25.

This also leads to complications in the computations, as each individual has a different range of possible values. Many of the steps in the Gibbs sampler thus consist of a series of cases (observed in both, observed in just the census, or observed in just the driver's licenses). A few examples are given below, and more details on this are available upon request.

Case 1: Observed in Driver's License file on day y_{Di} and in the Census on day $y_{Ci} = 365$
For this individual, we know that $t_{0i} < y_{Di}$ and $t_{1i} > 365$. We do not consider the possibility of someone moving out and then back in in a short time period. We also know that for this individual, $w_{Ci} = 1$ and $w_{Di} = 1$.

Case 2: Not observed in Driver's License file, observed in Census on day $y_{Ci} = 365$
The possible range of t_{0i} and t_{1i} depends on the current value of w_{Di}.

Case a: $w_{Di} = 0$: Since the individual does not have a Driver's license, their absence from the file tells us nothing about their migration history. We thus only know that $t_{0i} < 365$ and $t_{1i} > 365$ since they were in the area on Census Day. Finally, since $w_{Di} = 0$, y_{Di} is undefined.

Case b: $w_{Di} = 1$: This implies that the individual did renew their driver's license, but not during the time that they were in the area. For a given value of y_{Di} (drawn from its posterior distribution, given above), we then know that the individual must have either moved in after y_{Di} or out before y_{Di}. This thus restricts the possible values of t_{0i} and t_{1i}.

Case 3: Observed in Driver's License file on y_{Di}, not in Census file
Again, there are two cases, depending on the current value of w_{Ci}.

Case a: $w_{Ci} = 0$: Since the person is not in any Census record, this gives us no information on the individual's migration history. We only know that $t_{0i} < y_{Di}$ and $t_{1i} > y_{Di}$. y_{Ci} is undefined.

Case b: $w_{Ci} = 1$: This implies that the person is in the Census, but was not in the area of interest on Census Day. Since our population is defined as everyone in the area at some point between T_0 and T_1, we thus know that the individual must have moved into the area before y_{Di}, and out of the area before Census Day (365).

Case 4: Observed in neither file
For now, we do not consider these individuals since they would not appear in our files. However, as discussed above, a later version of the model will impute individuals who are in this category.

5.3 Results

Since we do not have access to administrative or Census records, the data used to illustrate this example was simulated. The simulated data set consisted of only people observed in one or both of the systems, resulting in a sample size of 427. The "true" parameter values are shown in Table 1, as well as posterior estimates from the Gibbs sampler.

Table 1: Posterior Estimates of Parameter Values

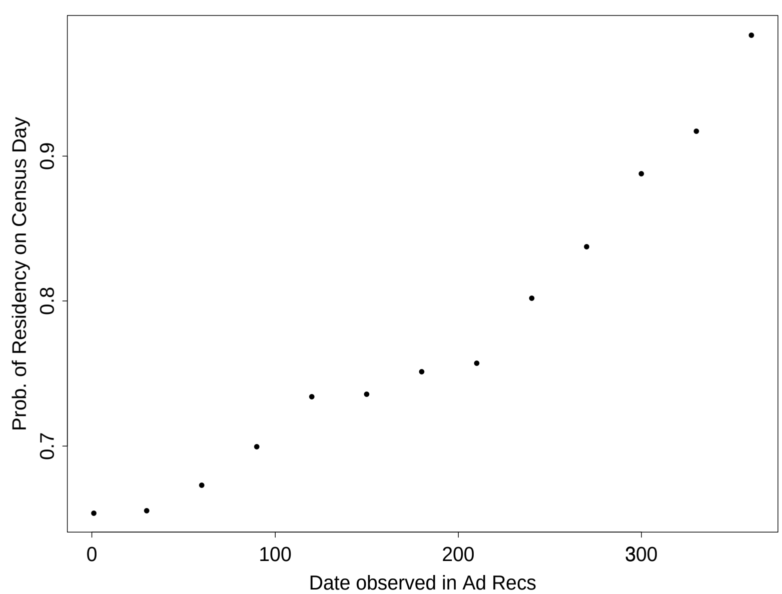
Parameter	True Value	Posterior Mean	95% Posterior Interval
q	0.8	NA	NA
λ	$\frac{1}{1825}$	NA	NA
α_C	0.9	0.89	(0.86, 0.92)
α_D	0.7	0.79	(0.74, 0.83)
N_C	408	412	(402, 420)

N_C = size of population on Census Day

The addition of just one record system, the Driver's License file, added 42 individuals to the Census Day population count (the Census file had 370 individuals observed on Census day). In addition, the posterior intervals for the three main parameters covered or were close to the true parameter values.

Since our goal is to determine the probability of residency on Census Day, we are primarily interested in the individual migration dates, t_{0i} and t_{1i}, and their implications regarding residency on Census Day. The main inference will be for individuals observed in the Driver's License file and not in the Census. Figure 1 illustrates the information gained by observing an individual in the Driver's License file at various points in time. As expected, individuals observed later are more likely to still be in the area on Census day.

Figure 1: Probability of Residency on Census Day for those observed only in Administrative Records



6 Discussion

The strength of this model lies in its flexibility. In addition to being able to model many types of records, multiple record systems can be modeled at the same time. This is done by multiplying the likelihoods for observation of each type of record. This could be of particular use with the new Census Bureau StARS data set that contains data from 5 different records systems.

In addition, as discussed above, the model does not assume that the record systems are independent. As described, the dependence of the systems may be accommodated by allowing dependence among the probabilities of an individual being observed in each of the systems. Alternatively, the joint distribution of observation in all of the record systems could be modeled directly.

Complexity can also be built up by adding covariates such as demographic or area characteristics. These could be added in by making the observation or migration models dependent on various covariates. On the migration side, seasonality in migration could be taken into account, as well as different rates of migration by demographic characteristics or location.

In a larger sense, administrative records have great potential to assist in the estimation of the undercount of the US Census. One possibility that has been discussed is to use administrative records as the third system in triple system estimation (the Census itself and the PES being the other two systems). However, even more potential may lie in using national administrative records as the "second" system in triple system estimation, with the PES serving as the third system. National administrative records could be matched to the Census, and then the PES would help estimate some of the parameters. Since both the Census and the administrative records are available nationally, the whole system could then be used to provide small area estimates across the country.

There are three main advantages to using administrative records as a second major source of individuals. The records could be used to add (or subtract) people for whom we have direct evidence that they were (or were not) in the area on Census day. The PES can do this as well, but not across the entire nation. There would also be less reliance on synthetic estimates of the undercount, as no assumptions of homogeneity across areas would be necessary and local undercount estimates could be obtained more reliably. Finally, this would be a major step forward in the use of administrative records in the Census. The StARS database currently under development and the corresponding AREX experiment in the 2000 Census should give some indication of the potential for this method.

The migration model described here could reduce some of the problems associated with the use of administrative records. In particular, it could help reduce the amount of field follow-up needed, as it could identify the people that were more or less likely to still be in the area on Census day. Finally, the model may be useful to deal with movers in the PES. In that case, we would observe an individual on a date after Census day, and use the model "backwards" to predict residency on Census day.

In addition, there is potential for the use of this model in fields such as demography and sociology, where human migration is a major research area. The model can be extended to describe a list of events for individuals, jointly with their movement patterns. It could also possibly be used to help identify the determinants of migration.

References

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

[2] Brownie, C., Hines, J.E., Nichols, J.D., Pollock, K.H., and Hestbeck, J.B. Capture-Recapture Studies for Multiple Strata Including non-Markovian Transitions. *Biometrics* 49: 1173–1187 (1993).

[3] Darroch, J.N. The Multiple-Recapture Census: I. Estimation of a Closed Population. *Biometrika* 45: 343–359 (1958).

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

[5] Dupuis, J.A. Bayesian Estimation of movement and survival probabilities from capture-recapture data. *Biometrika* 82,4: 761–772 (1995).

[6] El-Khorazaty, M.N., Imrey, P.B., Koch, G.G., and Wells, H.B. Estimating the Total Number of Events with Data from Multiple-record systems: A Review of Methodological Strategies. *International Statistical Review* 45: 129–157 (1977).

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

[8] Fienberg, S.E. Bibliography on Capture-Recapture Modeling With Application to Census Undercount Adjustment. *Survey Methodology* 18: 143–154 (1992).

[9] Fienberg, S.E., Johnson, M.S., and Junker, B.W. Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society A* 162, Part 3: 383–405 (1999).

[10] George, E.I. and Robert, C.P. Capture-Recapture Estimation Via Gibbs Sampling. *Biometrika* 79: 677–683 (1992).

[11] Hansen, K.A. Seasonality of Moves and Duration of Residence. US Census Bureau Current Population Reports, Household Economic Studies. P70-66 (1998).

[12] Hestbeck, J.B., Nichols, J.D., and Malecki, R.A. Estimates of Movement and Site Fidelity Using Mark-Resight Data of Wintering Canada Geese. *Ecology* 72:523–533 (1991).

[13] Larsen, M. Predicting the Residency Status for Administrative Records that Do Not Match Census Records. Administrative Records Research Memorandum Series #20 (1999).

[14] Pollock, K.H. 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 (1991).

[15] Scheuren, F. Administrative Records and Census Taking. *Survey Methodology* 25, No. 2: 151–160 (1999).

[16] Seber, G.A.F. *The Estimation of Animal Abundance and related parameters, 2nd edition.* MacMillan Publishing Co., Inc., New York. (1982).

[17] Smith, P.J. Bayesian Analyses for a Multiple Capture-Recapture Model. *Biometrika* 78: 399-407 (1991).

[18] Wittes, J.T. Applications of a Multinomial Capture-Recapture Model to Epidemiological Data. *Journal of the American Statistical Association* 69: 93–97 (1974).

[19] Zaslavsky, A.M, and Wolfgang, G.S. Triple-System Modeling of Census, Post-Enumeration Survey, and Administrative-List Data. *Journal of Business & Economic Statistics* 11: 279–288 (1993).