NEW APPROACHES AND ESTIMATES OF FOREIGN BORN EMIGRATION

Jennifer Van Hook

Bowling Green State University

Jeffrey Passel

The Urban Institute

Weiwei Zhang

Bowling Green State University

Frank D. Bean

University of California—Irvine

NEW APPROACHES AND ESTIMATES OF FOREIGN BORN EMIGRATION

Abstract

Post-censal population estimates depend on the accuracy of estimates of demographic components of change. Of these components, emigration of the foreign born is probably the least well known. Indirect methods (i.e., the residual method) are particularly poor at estimating emigration among recent arrivals. We introduce a new method for estimating emigration that takes advantage of the unique sample design of the CPS. The CPS follows housing units—but not necessarily individuals—over a period of 16 months. Individuals in the March CPS not successfully followed up include those who died, internal migrants, and emigrants. We use statistical methods to estimate the proportion of emigrants among those not followed up. Our method produces emigration estimates that are comparable to residual-based methods in the case of earlier arrivals (immigrants who arrived more than ten years ago), but yields much higher estimates for earlier arrivals.

NEW APPROACHES AND ESTIMATES OF FOREIGN BORN EMIGRATION

Post-censal population estimates depend on the accuracy of estimates of demographic components of change: births, deaths, immigration, and emigration. Of these components, emigration of the foreign born is probably the least well known. Yet population estimates built up through the cohort component method will be too low if emigration is over-estimated and too high if emigration is under-estimated. In the case of residual estimates of unauthorized migrants, the accuracy of emigration rates among the legal foreign born is critical. If the emigration estimates of the legal foreign-born are too high, the legal population will be too low and the unauthorized migrant population will be overestimated.

Official statistics on emigration from the United States are virtually non-existent. The INS collected data from departing foreign born emigrants from 1908 to 1957 (Woodrow-Lafield 1998), but discontinued this practice due to concerns about the quality of emigration data (Kraly 1998). In contrast, the number of births, deaths, and arrivals of legal immigrants are known with relative accuracy because the United States vital registration system and the former Immigration and Naturalization Service (now "U.S. Citizenship and Immigration Services") collects data on these events. Out of necessity, emigration has been estimated with a variety of indirect demographic methods, most prominent of which is the residual method. We develop a new method for estimating emigration that takes advantage of the longitudinal nature of the Current Population Survey. By producing new foreign-born emigration estimates with which residual estimates can be compared, we offer a new way to evaluate and update the emigration estimates currently used by the Census Bureau and others for the production of population estimates.

Previous Research

Warren and Peck (1980) first developed indirect methods for estimating emigration. Their technique—referred to here as the residual method—has since served as the major method used by the U.S. Census Bureau for developing and updating foreign-born emigration estimates. Warren and Peck's (1980) originally estimated an annual number of emigrants of 114,000 for the 1960-70 decade. This figure was later increased to 133,000 on the basis of Warren and Passel's (1987) analysis of INS Alien Address Registration data for 1965-1980. The 133,000 figure was used as an official "point estimate" by the U.S. Census Bureau until the mid-1990s, when the number was increased to 195,000 based on the residual estimates of Ahmed and Robinson (1994). The number may be updated again on the basis of Mulder's (2003) more recent work, which estimated an annual number of 225,000 foreign-born emigrants during the 1990s.

The residual method for estimating emigration in the decade between two censuses involves the comparison of two population estimates: the "expected" population if no emigration had occurred during the decade, and the enumerated population at the end of the decade. Residual estimates of emigration during the 1990s (Mulder 2003), for example, were constructed by surviving immigrants who arrived prior to 1990 forward to 2000 (i.e., by aging all cohorts by ten years and subtracting the estimated numbers of deaths) and then comparing the survived population with the enumerated population in 2000. The difference is attributed to emigration. The number of emigrants among immigrants who arrived between 1990 and 2000 is estimated by applying emigration rates from the earlier arrivals.

Prior emigration estimates based on the residual method as well as other methods have been reviewed in detail elsewhere (Kraly 1998; Woodrow-Lafield 1998; Mulder 2003). To synthesize this literature, we compare the results of various studies in a table showing estimates of the number of emigrants together with the emigration rates and the rate at which immigration is offset by emigration as would be implied by the various estimates (Table 1). Even though the residual estimates of the annual number of emigrants has increased over time from 114, to 195, and then to 225 thousand), the associated rates of emigration (relative to the mid-year foreign born population) and ratio of emigrants to newly-admitted legal immigrants appears to have declined (from 1.18% during the 1960s, 1.15% in the 1980s, and 0.88% in the 1990s). Among recently-arrived immigrants, the emigration rates are inconsistent across studies. Mulder's (2003) estimate of 21,000 emigrants per year among 1990s arrivals implies a much lower emigration rate than prior studies. Emigration rates for Mexican immigrants are even less consistent, with the residual estimates showing much lower levels of emigration than the estimates made by Massey and his colleagues. Massey and his colleagues estimate emigration through the analysis of detailed accounts of trips by Mexican migrants to the United States. The detail in the data permit the identification of each separate trip as contributing in- and out-migration, and thus capture many more emigrants (and return migrants) over time.

A major drawback of the residual method is that the estimates are sensitive to differences in census coverage. For example, coverage error was higher in the 1990 Census than the 2000 Census (Robinson et al., 1993; Hogan 1993; U.S. Census Bureau 2001), so in many cases the "expected" populations in 2000 turned out to be smaller than the enumerated populations in 2000 and thus implied a negative emigration rate (an impossibility). This problem is especially evident for country-of-origin groups that contain large proportions of unauthorized migrants, such as Mexico. For Mexicans, the "expected" population is significantly lower than the enumerated population for both the 1980-90 and 1990-2000 decade (Ahmed and Robinson 1994; Mulder 2003). Ahmed and Robinson's (1994) analysis suggests that census coverage of migrants from Mexico and Central America may have been much lower in the 1980 Census than in the 1990 Census by which time many unauthorized migrants had legalized under IRCA. Ahmed and Robinson (1994) handle the problem of differential undercount and negative emigration estimates by estimating emigration rates by race/ethnicity (not country-of-origin) while excluding those country groups with negative rates. They use the race-specific rates as proxies for rates of countries, for example, matching Hispanic rates to countries sending high proportions of Hispanic immigrants. Mulder (2003) handles the problem by adjusting the census estimates for undercount. This results in a set of final emigration estimates that are highly sensitive to the coverage estimates on which the adjustments are based.

A second drawback of the residual method is that it depends on the accurate comparison of age- and year-of-entry cohorts across two censuses. The consistency of reporting on year of entry is of particular concern. In the 1990 Census, one-third of immigrants who reported having come to the U.S. between 1985 and 1990 is likely to have been resident in the United States prior to 1985 (Ellis and Wright 1998). If a significant number of recently-arrived immigrants understate the length of time they have lived in the United States in one census, this would lead to upwardly biased estimates of recent arrivals in the first census, and upwardly biased estimates of the "expected" population ten years later. If reporting on year of entry were more accurate in the later census, it would appear that more recent arrivals emigrated than was in fact the case.

A third weakness of the residual method lies in its inability to estimate emigration for recently arrived immigrants. This is evident by the inconsistent and, in some cases, unrealistically low emigration residual-based estimates for recently arrived immigrants (Table 1). The residual method compares cohorts between two censuses and therefore produces emigration estimates for those who were present in the country at the time of the first census. But the emigration rates for immigrants who arrived during the intercensal period are not derived from the data, but instead, are assigned the rates that were calculated for earlier arrivals.

Finally, the residual method is that it does not permit the estimation of emigration by migration status because migration status is not identified in the Census[1]. Massey and Singer (1985) make a significant contribution by using Mexican Migration Project (MMP) data to estimate in-, out-, and net unauthorized migration flows from Mexico to the United States. Their work suggests that the unauthorized flow from Mexico consists of large proportions of temporary and circular migrants whose emigration rates are likely to be much higher than legal permanent residents. However, they do not prepare comparable statistics for legal immigrants or unauthorized migrants from countries other than Mexico. It remains important to compare the emigration patterns of unauthorized and legal immigrants. One reason is that the residual method for estimating the number of unauthorized migrants depends on the emigration patterns of *legal* migrants, not all foreign born.

We supplement residual estimates for the 1990s with estimates based on a new approach we refer to as the "CPS Matching Method." This method takes advantage of the unique sample design of the CPS, which follows housing units—but not necessarily individuals—over a period of one year and 4 months. Individuals in the March CPS not successfully followed up in the March CPS in the following year include those who died, internal migrants, and emigrants. We use statistical methods to estimate the proportion of emigrants among those not followed up.

An advantage of the CPS Matching method is that it does not depend on assumptions about differential census coverage or consistent reporting on year-of-entry since in follows *individuals*, not groups, over time. Another advantage is that it treats recently-arrived

[1] The Census Bureau and others produce estimates of the number and characteristics of unauthorized migrants enumerated in the Census. However, these estimates rest in part on assumptions about foreign-born emigration. Their subsequent usage in the production of emigration estimates involves a certain amount of circularity in logic.

immigrants in the same manner as earlier arrivals. It therefore is more likely than the residual method to produce comparable estimates across different period-of-entry groups. Finally, the CPS Matching Method permits us to estimate emigration by legal status. We use the migration status imputations developed earlier by Clark and Passel. The imputation procedure assigns foreign-born non-citizens legal status on a probabilistic basis given place of birth, occupation, sex, gender, welfare recipiency, state of residence, and other characteristics. Once we assign individuals as legal and unauthorized, we analyze one-year follow-up rates separately by legal status in order to produce emigration estimates by legal status.

As will become clear below, the CPS matching method depends on the accuracy of certain assumptions, most significantly that emigration among the third-or-higher generation is very low, and that immigrants and the third-or-higher generation have similar patterns of non-follow-up due to residual causes (while controlling for a number of socioeconomic factors). Although our assumptions could be subject to the same kind of scrutiny as Mulder's assumptions about differential undercount, both sets of emigration estimates would gain credibility if they were consistent with each other.

Our emigration estimates are likely to be larger than those based on the residual method because the residual method tends to miss emigrants who return to the United States. Massey and his colleagues estimate that the average duration of a Mexican labor migrant's first trip to the United States is only 21 months and that one-third of these migrants return to the United States in a second trip within ten years of the first trip (Massey, Durand, and Malone 2002). Eight-five percent of entries into the U.S. by unauthorized Mexican migrants between 1965 and 1989 were offset by exits (Massey and Singer 1995). This type of circular migration is not typically captured by the residual method, which counts the number of *net* rather than *total* emigrants.

The number of *net* emigrants from 1990 to 2000 is equal to the sum of the number of emigrations each year (E_y) minus the number of return trips to the U.S. in each year among those who emigrated between 1990 and 2000 (R_y):

$$E = \sum_{y=1990}^{y=2000} E_y - \sum_{y=1990}^{y=2000} R_y$$

Taking the annual average, $\frac{E}{10} = \overline{E}_y - \overline{R}_y$. So, $E = 10 \cdot (\overline{E}_y - \overline{R}_y)$, and $\overline{E}_y = \frac{E}{10} + \overline{R}_y$.

This shows that the difference between estimates using a ten-year interval and a one-year interval (as does the CPS matching method introduced here) is equivalent to the average annual number of return trips.

Data

To estimate foreign-born emigration rates for the late 1990s and early 2000s, we use data drawn from the 1998, 1999, 2000, 2001, 2002, 2003, and 2004 March Current Population Survey (CPS) files. The March CPS samples offer several advantages in that they follow housing units over time and contain information about internal migration and international immigration. In addition, we have already developed methodology for imputing migration legal status for the foreign-born in the CPS. Thus the groundwork has already been laid that makes it possible to produce foreign-born emigration estimates separately for legal and unauthorized migrants.

An underused feature of the Current Population Survey is that it follows housing units over a period of 16 months or more. Most CPS respondents are eligible to be interviewed eight times. The number of months a sampling unit has been in the sample is identified by the variable "Month-in-sample," which ranges from 1 to 8. When most respondents first enter the CPS sample, they are interviewed four months in a row (months-in-sample 1-4). Eight months later, they are interviewed again for four months (months-in-sample 5-8). For example, a respondent who is interviewed in each month from January through April in one year is eligible to be re-interviewed in the same months in the following year. This means that all respondents in the March sample with a month-in-sample code of 1-4 are eligible to be followed up in March of the following year.

The March supplement includes an oversample of Hispanics that follows a slightly different interview schedule than the regular sample. All Hispanics in the October CPS sample are eligible to be interviewed in March in addition to the eight months they would normally be interviewed. For example, Hispanics who are interviewed for the first time in October are eligible to be interviewed not only in November, December, and January, but also in March. All respondents who appear for the first time in the March Demographic supplement, including those in the oversample, are eligible to be followed up in the March sample in the following year.

A key feature of the CPS sample design—one that is critical for our purposes—is that it follows housing units rather than individuals. If a respondent moves to a new address, the new occupants of the original housing unit are interviewed and the original respondent is dropped from the sample. This feature of the CPS sample design permits us to use follow-up rates (i.e., the proportion of persons in the March sample who are successfully followed up in the March sample in the following year) as a basis for estimating emigration.

Components of non-follow-up.

Madrian and Lefgren (1999) estimate that 29 percent of those eligible for follow-up in the March CPS 1980-1998 surveys were not successfully followed up. There are many reasons a person may not be followed up. Based on known rates of internal migration and mortality (derived from the CPS and NCHS statistics), Madrian and Lefgren (1999) estimate that 16.3 percent moved to another address in the United States and 0.86 percent died, leaving 11.8 percent who were not followed up for other reasons. Of the residual 11.8 percent, some may have moved to another country, while others were not followed up due to non-response, refusals, coding error, or the inability to contact a person in the housing unit.

Thus, the proportion of eligible persons not followed up (u), consists of the proportion who migrated within the United States (m), the proportion who died (d), the proportion who emigrated (e), and the proportion who were not followed up for other reasons (r). For immigrants,

$$u^I = m^I + d^I + r^I + e^I \tag{1a}$$

and for natives:

$$u^N = m^N + d^N + r^N + e^N. \tag{1b}$$

Subtracting (1b) from (1a) and solving for e^I yields:

$$e^I = u^I - u^N + m^N - m^I + d^N - d^I + r^N - r^I + e^N. \tag{1c}$$

We are able to estimate non-follow-up rates (u^I and u^N) and internal migration rates (m^I and m^N) directly from the CPS data. The mortality component is likely to be small except in the older age groups (Madrian and Lefgren 1999). We are able to estimate mortality rates by nativity from the National Health Interview Survey (Palloni and Aries 2004). In addition, we make the simplifying assumption that emigration among natives is negligible (i.e., that the value of e^N is close to zero). Fernandez (1995) estimates that during the 1980s, roughly 48,000 U.S. born emigrated per year. This amounts to an annual rate of about .02 percent. However, estimating residual non-follow-up rates (r^I and r^N) is more challenging. Our strategy is to assume that immigrants and natives have identical non-follow-up rates after controlling statistically for factors associated with attrition. We describe below how we use predicted values generated from multivariate logistic regression models to approximate e^I.

Estimation Strategy

The probability of non-follow-up for each person can be modeled with logistic regression as a function of a vector of socioeconomic and demographic characteristics:

$$u_i^I = \frac{\exp(\mathbf{X_i}'\boldsymbol{\beta}^\mathbf{I})}{1+\exp(\mathbf{X_i}'\boldsymbol{\beta}^\mathbf{I})} = F(\mathbf{X_i}'\boldsymbol{\beta}^\mathbf{I}), \qquad \text{(immigrants)}$$

.

$$u_n^N = \frac{\exp(\mathbf{X}_n'\boldsymbol{\beta}^N)}{1+\exp(\mathbf{X}_n'\boldsymbol{\beta}^N)} = F(\mathbf{X}_n'\boldsymbol{\beta}^N), \qquad \text{(natives)}$$

.

and each component of non-follow-up can be similarly modeled:

$$\left.\begin{array}{l} m_i^I = F(\mathbf{X_i}'\boldsymbol{\mu}^\mathbf{I}) \\ d_i^I = F(\mathbf{X_i}'\boldsymbol{\delta}^\mathbf{I}) \\ r_i^I = F(\mathbf{X_i}'\boldsymbol{\rho}^\mathbf{I}) \\ e_i^I = F(\mathbf{X_i}'\boldsymbol{\lambda}^\mathbf{I}) \end{array}\right. \text{(immigrants)} \qquad \left.\begin{array}{l} m_n^N = F(\mathbf{X}_n'\boldsymbol{\mu}^N) \\ d_n^N = F(\mathbf{X}_n'\boldsymbol{\delta}^N) \\ r_n^N = F(\mathbf{X}_n'\boldsymbol{\rho}^N) \\ e_n^N = F(\mathbf{X}_n'\lambda^N) \end{array}\right. \text{(natives)}$$

The likelihood of non-follow-up can therefore be expressed as the sum of predicted values generated from each component model. For example, for immigrants:

$$u_i^I = F(\mathbf{X_i}'\boldsymbol{\mu}^\mathbf{I}) + F(\mathbf{X_i}'\boldsymbol{\delta}^\mathbf{I}) + F(\mathbf{X_i}'\boldsymbol{\rho}^\mathbf{I}) + F(\mathbf{X_i}'\boldsymbol{\lambda}^\mathbf{I}) \qquad (3)$$

Averaging (3a) across all immigrants yields the non-follow-up rate among immigrants:

$$
\mathrm{u}^{\mathrm{I}} = \frac{\sum_{i=1}^{i=I} F(\mathbf{X_i}'\beta^{\mathbf{I}})}{I} = \frac{\sum_{i=1}^{i=I} F(\mathbf{X_i}'\boldsymbol{\mu}^{\mathbf{I}}) + F(\mathbf{X_i}'\boldsymbol{\delta}^{\mathbf{I}}) + F(\mathbf{X_i}'\boldsymbol{\rho}^{\mathbf{I}}) + F(\mathbf{X_i}'\boldsymbol{\lambda}^{\mathbf{I}})}{I} \tag{4a}
$$

where I is the number of immigrants. Our strategy is to estimate generational differences in each component of non-follow-up (internal migration, mortality, and residual non-follow-up) while controlling statistically for compositional differences. We accomplish this by comparing predicted values for immigrants (equation 4a) with predicted values for natives assuming both groups had identical characteristics. The non-follow-up rate of natives *if they had the same characteristics as immigrants* ($\mathrm{u}^{\mathrm{N}*}$) is obtained by replacing the coefficients in equation (4a) with native coefficients:

$$
\mathrm{u}^{\mathrm{N}*} = \frac{\sum_{i=1}^{i=I} F(\mathbf{X_i}'\beta^{N})}{I} = \frac{\sum_{i=1}^{i=I} F(\mathbf{X_i}'\boldsymbol{\mu}^{N}) + F(\mathbf{X_i}'\boldsymbol{\delta}^{N}) + F(\mathbf{X_i}'\boldsymbol{\rho}^{N}) + F(\mathbf{X_i}'\boldsymbol{\lambda}^{N})}{I} \tag{4b}
$$

By subtracting equation 3b from 3a and solving for $\sum_{i=1}^{i=I} F(\mathbf{X_i}'\boldsymbol{\lambda}^{\mathbf{I}})/\mathrm{I}$, we estimate the foreign-born emigration rate as:

$$
\mathrm{e}^{\mathrm{I}} = \sum_{i=1}^{i=I} F(\mathbf{X_i}'\boldsymbol{\lambda}^{\mathbf{I}})/\mathrm{I}
$$

$$
= \sum_{i=1}^{i=I} \frac{1}{I}[F(\mathbf{X_i}'\beta^{I}) - F(\mathbf{X_i}'\beta^{N}) + F(\mathbf{X_i}'\mu^{N}) - F(\mathbf{X_i}'\mu^{I}) + F(\mathbf{X_i}'\delta^{N}) - F(\mathbf{X_i}'\delta^{I})
$$

$$
+ F(\mathbf{X_i}'\rho^{N}) - F(\mathbf{X_i}'\rho^{I}) + F(\mathbf{X_i}'\lambda^{N})] \tag{5a}
$$

At this point, we introduce a number of simplifying assumptions. First, we assume that—if given the same socioeconomic, health, and demographic characteristics—immigrants would exhibit the same residual non-response rates as natives. That is $\sum_{i=1}^{i=I} \frac{1}{I}\left[F(\mathbf{X_i}'\rho^N) - F(\mathbf{X_i}'\rho^I)\right] = 0$. Also, we assume that emigration among natives is negligible for the reasons noted above. In other words, $\sum_{i=1}^{i=I} \frac{1}{I}\left[F(\mathbf{X_i}'\lambda^N)\right] = 0$. Therefore,

$$\mathrm{e}^{\mathrm{I}} = \sum_{i=1}^{i=I} \frac{1}{I}\left[F(\mathbf{X_i}'\beta^I) - F(\mathbf{X_i}'\beta^N) + F(\mathbf{X_i}'\delta^N) - F(\mathbf{X_i}'\delta^I) + F(\mathbf{X_i}'\mu^N) - F(\mathbf{X_i}'\mu^I)\right] \quad (5b)$$

where the first two terms gives the predicted nativity difference in non-follow-up, the second two terms equals the predicted difference in internal migration, and the last two give the predicted difference in mortality. Equation 4b is particularly useful because all components can be estimated with CPS and NHIS data.

To obtain values for the components of equation (5b), we first estimate three sets of logistic regression models, the first predicting non-follow-up among those eligible to be followed up, the second predicting internal migration (i.e., living at a different address from the year before) among those who were living in the United States the year before, and the third predicting the one-year probability of dying. The models of non-follow-up and internal migration are estimated separately by sex and race/ethnicity for immigrants and the second generation and include as independent variables identical sets of socio-demographic variables: age, education, housing tenure (rent vs. own), type of residence (mobile home, barracks, dorm, etc.), school enrollment status, and general health status.

The second generation rather than all natives is used as the comparison group. The reason is to reduce the predicted nativity difference in residual non-follow-up. Because of cultural and social similarities, immigrants' residual non-follow-up rates are more likely to be similar to the second generation than the third-or-higher generation's rates. This is especially clear in the case of blacks. Third-or-higher generation blacks include very few in the third generation. Most are descendants of slaves whose families have been in the United States for centuries and differ from black immigrants (and their children) in many important ways, including perhaps, in their rates of residual non-follow-up. One problem with using the second generation as the comparison group is that both immigrant and second generation children are children of immigrants, often share the same households, and are therefore likely to have similar emigration rates. This means that, in the case of children, the estimated emigration rates of the second generation are unlikely to be anywhere close to zero. If we were to use the methodology outlined above for children, we would almost certainly underestimate immigrant children's emigration rates. For this reason, we treat children ages 0 to 14 differently by assigning them the predicted values of their parent on the assumption that most children emigrate with their families.

For adults, we generate from each model predicted values of the likelihood of non-follow-up and internal migration. Four predicted values are calculated for each immigrant:

(1) non-follow-up using immigrant coefficients,
(2) non-follow-up using second generation coefficients,
(3) internal migration using immigrant coefficients, and
(4) internal migration using second generation coefficients.

Predicted values of non-follow-up and internal migration for each sex and race/ethnic group are derived from each groups' corresponding models. For example, the predicted values for Mexican males come from the "Mexican male" models. In the case of Asians, there were not enough cases in the second (or third) generation to obtain stable estimates, so we obtained their predicted values from models estimated on all race-ethnic groups combined.

Predicted one-year probabilities of death for adults ages 15 and older are obtained from the National Health Interview Survey 1989-1993 (NHIS). The NHIS is linked to the National Death Index in order to ascertain whether and the age at which respondents die. Using a person-year file, we estimate logistic regression models of whether a person died during the year separately for immigrants and natives. We do not estimate models for the second generation apart from all natives due to data limitations of the NHIS. The independent variables include sex, age, race, ethnicity, and general health. We use the coefficients from the mortality models to generate both "immigrant" and "native" predicted probabilities for immigrants in the CPS data, as shown in the last two terms in equation 5b.

Finally, we subtract the immigrant predicted values from the second generation predicted values (native predicted values in the case of mortality) as shown in the brackets in equation (5b). We then average the results to obtain an estimate of e^I for all immigrants and for immigrant subgroups by age, sex, country-of-origin, year-of-entry, and migration status (i.e., we average the bracketed portion of (5b) for the different groups). As noted above, children are assigned the predicted values of their parent (identified by the PARENT identifier in the CPS).

Matching CPS Surveys. To determine whether a respondent in the March CPS in one year (t) is successfully followed up the following year (t+1), we match those eligible for follow-up[2] in the 1998-2003 March CPSs with respondents in the following years' CPS (1999-2004). We use the methodology and STATA code developed by Madrian and Lefstrom (1999) for matching cases across CPS files, matching on household identification number and person line number. Because matched cases may not represent the same individual due to coding errors on the person or household identification variables, we also require consistency in sex or age

[2] For most cases, those with month-in-sample codes 1-4 are eligible for follow-up. For those in the Hispanic oversample in most years, month-in-sample is erroneously reverse-coded (personal communication with Greg Weyland). For these cases, we select months-in-sample 5-8 as eligible for follow-up.

(person at year t can be no more than 2 years younger than the matched case in year t) before considering a case a "true" match. We do not take into account consistency in race because of changes between 2002 and 2003 in the race question (now multiple racial identities are permitted).

Internal migration. The CPS asks respondents whether he/she lived in a different residence one year before. We define the internal migration rate as the proportion of movers among those who reported having lived in the United States one year before. Since the internal migration question is retrospective, we estimate internal migration rates for year t (1998-2003) from respondents in the CPS in year t+1 (1999-2004). For example, the proportion of internal migrants in 1998 is estimated from the 1999 March CPS. This is different from how we calculate follow-up rates, which are the proportion of CPS respondents in year t who are followed up in year t+1.

Return Migration. Foreign born who lived abroad one year before but who reported having come to stay in the U.S. more than 2 years before are defined here as "return" migrants.

Migration Status. [coming soon].

Results

As shown in Table 2, we estimate an annual foreign born emigration rate of 2.39 percent. For a population of 30 million foreign born (as in the 2000 March CPS), this translates into roughly 716 thousand emigrants per year. At the same time, we estimate a return migration rate of 0.87 percent, or 260 thousand. Subtracting return migration from total emigration yields an annual net emigration rate of 1.52 percent, or 456 thousand per year.

Males are nearly twice as likely to emigrate as females (3.13 versus 1.65 percent) and are more likely to return to the United States (but not enough to offset their higher emigration rates).

Emigration and return migration rates tend to be relatively high for younger immigrants, decline with age, and in the case of emigration but not return migration, increase in the oldest age group (65+). Taking emigration and return migration together, net emigration appears highest among children and working aged (25-44) adults. Net emigration dips for teenagers and young adults (age 15-24) and middle-aged adults (45-64), but increases somewhat for the elderly (65+) perhaps a result of retirees returning to their countries of origin.

Among selected country or region of origin groupings, immigrants from Mexico, India, Africa, North America, and Europe appear to have the highest emigration rates and immigrants from China the lowest. Of the high-emigration groups, the relatively high return migration rates of Mexicans stand out, reflecting the circular migration patterns commonly observed for this group. On balance, net migration rates appear lowest for groups from Southeast Asia (China, Philippines, "other" Asia) and Central and South America, and highest for immigrants from south Asia (India), Africa, Europe, and North America (Canada). Because of their high return migration rates, Mexicans fall between these groups.

Emigration rates vary with migration/legal status in expected ways. Refugees (many of whom cannot return to their countries of origin), and immigrants who have naturalized (who have made the greatest legal/political commitments to living in the United States) exhibit the lowest emigration rates (both total and net). Legal non-citizens have the next highest emigration rates. Unauthorized migrants, many of whom engage in circular migration for temporary work, exhibit relatively high emigration and return migration rates. Finally, legal non-immigrants (primarily those on temporary work and student visas) exhibit by far the highest emigration rates.

In general, emigration rates are highest for recent arrivals and decline with time in the United States. Return migration rates, as expected, are relatively low for both recent and earlier arrivals (0 to 4 years and 10 or more years), but higher for those who have lived in the U.S. for a moderate length of time (5 to 9 years).

How do these estimates compare with residual-based estimates? We compare our net emigration rates with the residual-based estimates prepared by Mulder (2003) (Table 3). Mulder's estimates pertain to the 1990-2000 decade. To make a fair comparison with our estimates, we derive net emigration rates from Mulder's results and then apply these rates to the 2000 population to obtain estimates of the number of emigrants for 2000. Our estimates are very similar to Mulder's in the case of earlier arrivals (in the U.S. ten or more years). However, we estimate many more emigrants among recent arrivals (1.81 versus 0.32 percent, or a difference of 210 thousand emigrants). In light of the methodological problems associated with residual-based estimates of emigration among recent arrivals, this suggests that the CPS-matching method, while comparable with residual-based estimates for earlier arrivals, offers an improved method for estimating emigration among recent arrivals.

We also find that the CPS-matching method tends to produce higher emigration estimates for countries or regions sending many unauthorized, temporary, and recently arrived immigrants, such as Mexico (a source of unauthorized and circular migrants), India (a source of many temporary migrants on H1B visas), and Africa (a relatively new source of immigration).

Conclusions and Further Work

Refinements. We plan to explore the sensitivity of the foreign-born emigration estimates to various assumptions about the level of U.S.-born emigration. The method we outline above assumes no emigration for the second generation. If emigration for the second generation were not negligible, this would mean that the foreign-born emigration estimates are too low. We will re-estimate a range of foreign-born emigration estimates under varying assumptions about second generation emigration rates. A plausible range of assumptions about U.S.-born emigration will be drawn from the research literature (e.g., Fernandez 1995).

Also, we plan to produce more detailed emigration estimates by region/state of residence and marital status, and we plan to produce estimates for the U.S. born minor children of the foreign-born.

References

Ahmed, Bashir and J. Gregory Robinson. 1994. "Estimates of Emigration of the Foreign-Born Population: 1980-1990." Population Division Technical Working Paper No. 9. U.S. Census Bureau: Washington, DC.

Ellis, M. and R. Wright. 1998. "When Immigrants are not Immigrants: Counting Arrivals of the Foreign Born Using the U.S. Census." *International Migration Review* 32(1): 127-144.

Fernandez, Edward W. 1995. "Estimates of the Annual Emigration of U.S. Born Persons by using Foreign Censuses and Selected Administrative Data: Circa 1980." Population Division Technical Working Paper No. 10. U.S. Census Bureau: Washington, DC.

Hogan. 1993. "The 1990 Post-Enumeration Survey: Operations and Results," *JASA: Journal of the American Statistical Association* 88(423): 1,047-1,060.

Kraly, Ellen Percy. 1998. "Emigration: Implications for U.S. Immigration Policy Research." Pp. 587-618 in *Migration Between Mexico and the United States: Binational Study, Vol. 2 Research Reports and Background Materials.* Mexican Ministry of Foreign Affairs and the U.S. Commission on Immigration Reform: Mexico City and Washington, DC.

Madrian, Brigitte C. and Lars John Lefgren. 1999. "A Note on Longitudinally Matching Current Population Survey (CPS) Respondents." Technical Working Paper 227. National Bureau of Economic Research: Cambridge, MA.

Massey, Douglas S. and Audrey Singer. 1985. "New Estimates of Undocumented Mexican Migration and the Probability of Apprehension." *Demography* 32: 203-213.

Massey, Douglas S., Jorge Durand, and Nolan J. Malone. 2002. *Beyond Smoke and Mirrors: Mexican Migration in an Era of Economic Integration*. Russell Sage Foundation: New York.

Mulder, Tammany. 2003. "Foreign-born Emigration from the United States: 1990 to 2000." Paper presented at the 2003 annual meetings of the Population Association of America, Minneapolis, MN.

Palloni, Alberto and Elizabeth Arias. 2004. "Paradox Lost: Explaining the Hispanic Adult Mortality Advantage." *Demography* 41(3): 385-416.

Passel, Jeffrey and Rebecca Clark. 1998. Immigrants in New York: Their Legal Status, Incomes, and Taxes. The Urban Institute: Washington, DC.

Robinson, J. Gregory, Bashir Ahmed, Prithis Das Gupta, and Karen A. Woodrow. 1993. "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis," *JASA: Journal of the American Statistical Association* 88(423): 1,061-1,079.

U.S. Census Bureau. 2001. ESCAP II October 2001 Report on Adjustment decision. www.census.gov/dmd/www/EscapRep2.html.

Warren, Robert and Jeffrey Passel. 1987. "A Count of the Uncountable." *Demography* 24: 375-393.

Warren, Robert and Jennifer M. Peck. 1980. "Foreign-born Emigration from the United States: 1960 to 1970." *Demography* 17(1): 71-84.

Woodrow-Lafield, Karen. 1998. "Viewing Emigration at Century's End." Pp. 683-694 in *Migration Between Mexico and the United States: Binational Study, Vol. 2 Research Reports and Background Materials.* Mexican Ministry of Foreign Affairs and the U.S. Commission on Immigration Reform: Mexico City and Washington, DC.

Table 1
Emigration Estimates in Prior Research

	Time Period of Estimate	Annual Number of Emigrants (in thousands)	Emigration Rate	Emigrants per Arrivals
All Foreign Born				
Warren and Peck (1980)*	1960-1970	114	0.0118	0.3432
Ahmed and Robinson (1994)*	1980-1990	195	0.0115	0.2783
Mulder (2003)*	1990-2000	225	0.0088	0.2474
Recent Arrivals: In U.S. Less Than 10 Years				
Warren and Peck (1980)*	1960-1970	62	0.0439	0.1864
Borjas and Bratsberg	1970-1980	89	0.0320	0.1971
Mulder (2003)*	1990-2000	21	0.0032	0.0231
Recent Arrivals: In U.S. 10-20 Years				
Mulder (2003)*	1990-2000	142	0.0166	0.1393
Mexican Foreign Born				
Ahmed and Robinson (1994)*	1980-1990	20	0.0062	0.1212
Mulder (2003)*	1990-2000	27	0.0040	0.1609
Massey and Singer (1995)	1965-1989	1,306	0.5153	0.8588
Recent Arrivals: In U.S. Less Than 10 Years				
Mulder (2003)*	1990-2000	3	0.0012	0.0165
Massey, Durand, and Malone (2002)a	1965-1985	629	0.2875	0.5750
Massey, Durand, and Malone (2002)b	1965-1985	182	0.1250	0.2500
Recent Arrivals: In U.S. 10-20 Years				
Mulder (2003)*	1990-2000	15	0.0068	0.0509

* Residual estimate
[a] unauthorized Mexican Migrants
[b] legal Mexican Migrants

Table 2
Emigration Estimates of the Foreign Born, Circa 2000 (Based on CPS Matching Method)

	Emigration Rates (%)			*Population Estimates for 2000 (in thousands)*			
	Annual Emigration Rate	Annual Return Migration Rate	Annual Net Emigration Rate	2000 Population*	Annual No. Emigrants	Annual No. Return Migrants	Annual No. Net emigrants
All	2.39	0.87	1.52	29,988	716	260	456
Male	3.13	1.10	2.03	15,160	474	167	307
Female	1.65	0.71	0.94	14,828	244	105	139
Age							
0-14	3.66	1.25	2.41	2,034	74	25	49
15-24	2.42	1.24	1.18	4,353	105	54	51
25-34	3.82	1.08	2.73	6,810	260	74	186
35-44	3.70	1.07	2.63	6,455	239	69	170
45-64	0.64	0.53	0.11	7,171	46	38	8
65+	1.45	0.39	1.06	3,164	46	12	34
Country/region of Origin							
Mexico	3.02	1.24	1.78	8,535	258	106	152
Central & South Am.	1.76	0.88	0.88	4,411	78	39	39
Caribbean	1.87	0.68	1.20	4,849	91	33	58
China	1.13	0.39	0.74	1,548	18	6	11
Philippines	1.41	0.67	0.74	1,352	19	9	10
India	3.85	0.72	3.13	1,121	43	8	35
Other Asia	1.73	0.99	0.75	3,128	54	31	23
Africa	3.45	0.71	2.73	711	25	5	19
Europe	2.91	0.44	2.47	3,628	105	16	89
North America	3.36	0.69	2.67	705	24	5	19
Migration Status							
Naturalized	1.61	0.34	1.26	9,743	157	34	123
Legal Immigrant	2.58	1.11	1.47	9,535	246	106	140
Unauthorized Migrant	3.33	1.26	2.06	7,503	250	95	155
Refugee	1.34	0.76	0.58	2,259	30	17	13
Legal Non-Immigrant	4.46	0.90	3.56	948	42	8	34
Years in U.S.							
0-4	3.61	0.98	2.63	4,618	167	45	122
5-9	3.01	1.60	1.41	9,479	285	152	133
10+	1.79	0.67	1.12	15,891	285	107	178

* calculated from 2000 March CPS using Census 2000 weights.

Table 3
Comparison of Emigration Estimates Based on the Residual Method vs. the CPS-Matching Method

	Net Emigration Rate (%)			*Annual No. Net Emigrants, 2000*		
	Residual Method[a]	CPS-Matching Method	Difference	Residual Method[a]	CPS-Matching Method	Difference
All foreign born	0.88	1.52	0.64	265	456	191
Recent Arrivals (0-10 yrs)	0.32	1.81	1.49	45	255	210
Longer-term residents	1.39	1.27	-0.12	220	201	-19
Country/region of Origin						
Mexico	0.42	1.78	1.37	35	152	117
Central & South Am.	0.98	0.88	-0.10	43	39	-4
Caribbean	0.96	1.20	0.23	47	58	11
China	0.87	0.74	-0.14	14	11	-2
Philippines	0.64	0.74	0.11	9	10	1
India	0.40	3.13	2.73	4	35	31
Other Asia	1.83	0.75	-1.08	57	23	-34
Africa	1.13	2.73	1.60	8	19	11
Europe	1.05	2.47	1.42	38	89	51
North America	1.15	2.67	1.52	8	19	11

[a]Mulder (2003)