Estimates of International Migration for United States Natives

Christopher Dick, Eric B. Jensen, and David M. Armstrong

United States Census Bureau
christopher.dick@census.gov, eric.b.jensen@census.gov, david.m.armstrong@census.gov

Abstract

Net international migration (NIM) is a component of change in the Census Bureau's annual population estimates. This paper focuses on one aspect of NIM, native migration. It is difficult to estimate native migration because of a lack of data on both immigration and emigration of U.S. natives. In the Population Estimates Program, we use a residual method and data on U.S. natives collected by censuses in other countries. In this paper we present two areas of research into improving this estimate. The first was part of 2010 Demographic Analysis (DA) and focuses on the assumption of who is enumerated in the foreign country residual method. The second area of research focuses on an alternative approach that involves calculating a residual using data from the American Community Survey (ACS), Census 2000, and vital statistics from the National Center for Health Statistics (NCHS).

There are both benefits and limitations to using the alternative approach instead of the method currently used in the Population Estimates Program. For example, estimates of native migration are small relative to the size of the native population residing in the United States. Since the alternative approach calculates the residual using data on U.S. natives, even small errors in the estimate of natives may have a large impact on the residual-based estimate of native migration. However, the alternative approach provides two major benefits. First, it uses data from the United States – instead of a large number of countries – allowing for greater understanding of the quality and limitations of the data. Second, we can more frequently and easily update our estimates when using U.S.-based data.

In sum, in this paper we will describe the data and methods that can be used to estimate the international migration of U.S. natives since 2000, provide results from our ongoing research, and discuss the benefits and limitations of each approach.

Introduction

The Population Division of the Census Bureau annually produces postcensal population estimates. Each year a new "vintage" of estimates is released. A vintage includes annual population estimates dating from the prior Census forward. Two vintages of estimates for the same period (e.g., Vintage 2008 estimates for estimate year 2008 and Vintage 2009 estimates for estimate year 2008) may vary because of changes in the methodology used in the estimates program and/or incorporation of new or revised data. These population estimates include a component of net international migration (NIM), which is composed of foreign-born immigration, foreign-born emigration, net native migration, migration between Puerto Rico and the United States, and net military movement overseas. This presentation focuses on one of these sub-components, net native migration.

Native migration is currently estimated using a residual method and data from other countries (foreign country residual method) (Schachter 2008).[1] However, there are several limitations of this method including uncertainty of data quality, different measures of nativity across countries, and the lack of recent data for many countries. In this paper we focus on two ways to improve the estimate of net native migration. The first is a modification of the foreign country residual method to deal with the fact that it is possible that not all those who are born abroad of U.S. citizen parents (who would thus be U.S. citizens themselves) are included in the foreign country residual method. In the 2010 Demographic Analysis (DA) we included three differing assumptions about the extent to which those who are born abroad of U.S. citizen parents are included in the foreign country residual method: (1) they are fully

---

[1] The native-born population includes anyone who is a U.S. citizen at birth, including those who were born in the United States, Puerto Rico, a U.S. Island Area (American Samoa, Guam, the Commonwealth of the Northern Mariana Islands, or the U.S. Virgin Islands), or abroad of a U.S. citizen parent or parents.

included, (2) they are not included at all, and (3) they are included in some countries (that collect information on citizenship) and not in others (that have information on place of birth but not citizenship). It is likely that those who are born abroad of U.S. citizen parents in countries that collect data on place of birth only would be enumerated as being born in that particular foreign country, not the United States.

A second approach for improving the estimate of net native migration (NNM) is by using an alternative measure. Instead of collecting data from foreign censuses and population registers, we instead use data from the United States to calculate a residual based measure of NNM. We use data from Census 2000, the National Center for Health Statistics (NCHS), and the American Community Survey (ACS). The U.S. data based alternative measure has several benefits including: (1) certainty in the quality of the data, (2) the estimate is easier to update and thus we are able to keep the estimate more up to date, (3) we publish data in the United States at a more regular interval than in some other countries, and (4) the entire population of interest is included in the measure which may or may not be the case in the foreign country residual method (e.g., those born abroad of U.S. citizen parents may not be included).

The paper will proceed as follows. First, I will discuss the foreign country residual method, the data used to calculate net native migration using the foreign country residual method, and the benefits and limitations of the foreign country residual method. Second, I will discuss work completed during the 2010 DA to underline the uncertainty around the coverage of those born abroad of U.S. citizen parents in the foreign country residual method. Third, I will focus on the alternative U.S.-based method and its strengths and weaknesses. I conclude with a discussion of the foreign country residual and alternative methods and directions for further research.

Foreign Country Residual Method and Data

The foreign country residual method employed by the U.S. Census Bureau's Population Estimates Program to estimate net native migration was developed by Schachter (2008) as an update to Gibbs and colleagues (2003) work. Both of these estimates use the foriegn country residual method to determine the net number of U.S. native migrants. To better understand how this method works it is useful to describe the process that is currently used in the Population Estimates Program (PEP).
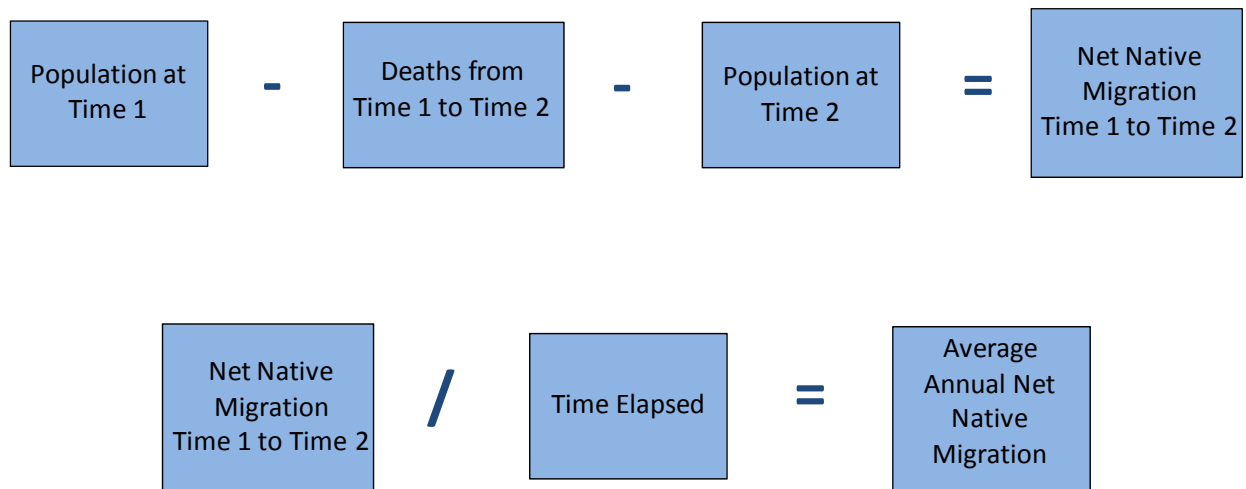
In the PEP, we use data from 84 countries to obtain an estimate of net native migration (see Appendix 1 for a listing of countries). Data are used from at least two time points for each country and come from either census (70 countries) or population register (14 countries) sources. Information is taken from the two most recent time points available. For instance, data from Germany are from a register of foreigners and come from the 2006 to 2007 time period, whereas data for Guam come from the 1980 and 1990 censuses. Depending upon the country where data were collected, we measure U.S. nativity in one of two ways: (1) born in the United States (59 countries) and (2) U.S. citizenship (25 countries).

Schachter (2008) used a residual cohort-component method to arrive at an average annual net international migration estimate for U.S. natives (see Figure 1). The population who were either U.S. citizens or were born in the United States were estimated by age at two time points. The measurements are usually two consecutive census dates (e.g., Time 1 is Census 1990 and Time 2 is Census 2000). The population at Time 1 is then survived forward using NCHS life tables for the resident population of the United States to Time 2 by age.[2] In the example using Census data from 1990 and 2000, the population would be survived forward 10 years. The difference between the survived Time 1 population and the actual Time 2 population is net native migration (NNM) over the entire time period. The NNM estimate is then annualized by dividing by the number of years between Time 1 and Time 2 to give average annual NNM. This average annual NNM is calculated for each of the 84 countries for which data was available. After the average annual NNM estimate is calculated for the 84 countries, the estimates are summed to create an overall average annual NNM estimate for the United States. Schachter (2008) estimated that approximately 45,000 U.S. natives leave the country per year and this estimate was used in the Vintage 2010 Population Estimates.

---

[2] Schachter (2008) determined that surviving the population by age and sex did not make a significant difference, and thus survived the population only by age.

Figure 1. The Foreign Country Residual Method

Population at Time 1	−	Deaths from Time 1 to Time 2	−	Population at Time 2	=	Net Native Migration Time 1 to Time 2

Net Native Migration Time 1 to Time 2	/	Time Elapsed	=	Average Annual Net Native Migration

For each new vintage of population estimates that comes from the Census Bureau's Population Estimates Program, the net number of native migrants must be distributed by demographic characteristics (age, sex, race, and Hispanic origin) at three geographic levels (nation, state, and county). In Vintage 2010 (V2010) this process was achieved through the use of the Census 2000 sample edited file, the American Community Survey 3-year file (2005-2007), and the use of proxy universes. A proxy universe is a population from a dataset that both is thought to represent the estimate that it is being used to distribute (in this case NNM) and is large enough to distribute demographic characteristics at the county level. For NNM, the Population Estimates Program uses the proxy universe of all natives.[3]

There are several limitations with calculating NNM using the foreign country method. First, the process of obtaining the data and calculating residuals for so many countries is prohibitive, and thus the estimate cannot be updated regularly. For instance, when updating the estimate by Gibbs et al. (2003), Schachter (2008) collected data on 128 countries – often having to contact officials from the countries or agencies outside the United States. Of these 128 countries, only 84 had two time-points with usable data. Second, with the number of foreign countries included in the estimate it is hard to ascertain the quality of the data. The method is sensitive to errors in the calculation of the residual. Differences in data quality or collection methods between the Time 1 and Time 2 data points could have a substantial impact on the native migration estimate. The difference would then look as if a large number of U.S. natives either moved in or out of that particular country when really they were just covered in the enumerations differentially. Third, since our determination of who is a U.S. native is based on different population universes (U.S. citizens or those born in the United States) in different countries, error is introduced by differential enumeration. Fourth, countries undertake census enumeration at more or less regular time intervals. Some countries are extremely regular about census enumeration, whereas others will have a census once every twenty years. Calculating residuals with highly differing lengths of time between Time 1 and Time 2 will invariably lead to instability in the estimate. Further, if countries have a long period of time between the two measurement points error will likely increase in the calculation of deaths.

Demographic Analysis

Demographic analysis (DA) is a population estimation technique that the U.S. Census Bureau Population Estimates Program uses each census year as a comparison to the decennial census. DA uses historical data on births and deaths as well as information on net international migration to obtain national-level population estimates that are independent of the census. In 2010, DA was a little different than 2000 DA. That is, instead of providing one estimate of the national-level population by age, sex, and race (Black/non-Black), a sensitivity analysis of several of

---

[3] We are currently researching alternative proxy universes for potential incorporation into our Vintage 2012 population estimates.

the component measures was undertaken and a series of five plausible estimates was produced.

The net international migration component for DA 2010 included a NNM subcomponent, which like many of the other subcomponents had assumptions varied. As in the yearly production of population estimates, data calculated by Schachter (2008) were used as the base of the NNM estimate in DA 2010. However, it is possible that the foreign country residual method, as carried out in this estimation procedure, may underestimate those who are born abroad of U.S. citizen parents. To try and assess the possible variance in the estimate due to assumptions about those born abroad of U.S. citizen parents, three estimates of NNM were calculated. The first estimate, assumed that all those born abroad of U.S. citizen parents were enumerated in foreign country census or population register data and included only the estimate put forth by Schachter (2008) of approximately 45,000 net out migration of natives from the United States each year. On the other end of the continuum, the second estimate assumed that none of those born abroad of U.S. citizen parents were enumerated in foreign country census or population register data. This series included the estimate of -45,000 net native migration plus a net migration estimate of the population born abroad of U.S. citizen parents. The third calculation made the assumption that some of those born abroad of U.S. citizen parents were enumerated in foreign country census or population register data while others were not (or they were not identified as natives of the United States). Specifically, we made the assumption that those countries that collect data by country (or place) of birth did not enumerate or mis-identified those born abroad of U.S. citizen parents whereas those who collected data by citizenship would count this population correctly. Thus, we looked at net migration of those born abroad of U.S. citizen parents from only those countries that collected data by country of birth in Schacter's (2008) estimate and then added these back into the total estimate of NNM.

Figure 2. Method for Estimating Migration of the Population Born Abroad of United States Citizen Parents
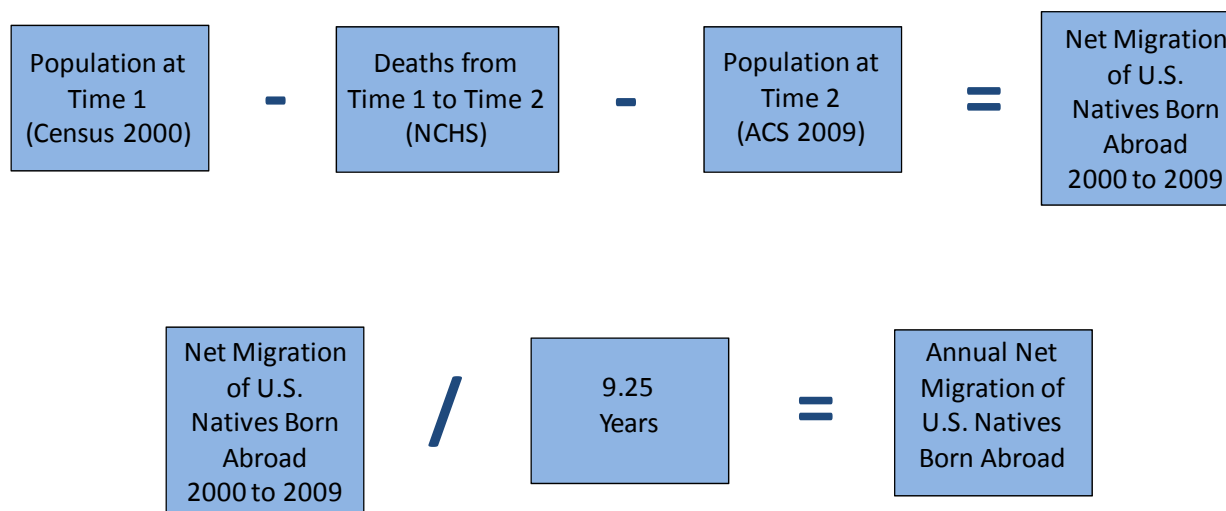
A residual method was also used in the calculation of net migration of those born abroad of U.S. citizen parents (see Figure 2). However, instead of using foreign country census or population register data we used Census 2000 sample edited data, single year ACS data (2009)[4] and NCHS resident population lifetables. The population of those born abroad of U.S. citizen parents from Time 1 (Census 2000) was survived forward using the NCHS resident population lifetables by age and sex from April 1, 2000 to July 1, 2009. We then compared this expected population to the observed population at Time 2 (ACS 2009). The difference is the net migration of those born abroad of U.S. citizen parents over the 9.25 year time period. We then divided by 9.25 to get an annual estimate. An estimate was produced both for all those born abroad of U.S. citizen parents and also those born abroad of U.S. citizen parents who were born in countries that collect data by place of birth (see Appendix 1). Table 1 reports the residual NNM, born abroad of U.S. citizen parents, and total NNM used in each of the five DA 2010 series of estimates and in the Population Estimates Program.

---

[4] See http://www.census.gov/acs/www/data_documentation/documentation_main/ for information on the accuracy of ACS estimates.

Table 1. Net Native Migration in 2010 Demographic Analysis and the Vintage 2010 Population Estimates

Estimate Series	Residual Net Native Migration	Born Abroad of U.S. Citizen Parents	Total Net Native Migration
DA Low	-45,228	0	-45,228
DA Low-Middle	-45,228	0	-45,228
DA Middle	-45,228	39,705	-5,523
DA High-Middle	-45,228	39,705	-5,523
DA High	-45,228	62,424	17,196
Vintage 2010	-45,228	0	-45,228

Sources: U.S. Census Bureau, Population Division tabulation of data from the Census 2000 sample edited data file and the annual 2009 American Community Survey.

Table 1 shows that the inclusion of those born abroad of U.S. citizen parents changes the number of net native migrants leaving or entering the country annually from approximately 45,000 leaving the country per year in the Low and Low-Middle Series (assumes migration of the population born abroad of U.S. citizen parents is included in foriegn country residual method) to approximately 5,500 leaving the country in the Middle and High-Middle Series (includes an additional estimate for those born in countries where data is collected by country of birth), and approximately 17,000 entering in the High Series (includes an additional estimate for those born abroad of U.S. citizen parents in all countries). In the Population Estimates Program, it is assumed that all those born abroad of U.S. citizen parents are enumerated in the foreign country residual net native migration calculation. Thus, like the Low and Low-Middle DA Series, the Vintage 2010 annual net native migration estimate was approximately 45,000 natives leaving the country.

An Alternative Method

Estimating native migration using data collected in the United States eliminates the impact of variation in the frequency, quality, and consistency in data collection across countries on the native migration estimate. Here we present an alternative method using information from Census 2000, the ACS, and NCHS (see Figure 3).

The calculation begins with an estimate of the native population at Time 1 using data from Census 2000. The Census 2000 native population is survived forward to a Time 2 using the NCHS resident population life tables by single year of age and sex. In this calculation we use four different Time 2 points: July 1, 2006, 2007, 2008, and 2009. These are the years of ACS data that were available for the resident population when this research was completed. Since we are working with the native population, we also add in births obtained from the NCHS for the time period to obtain an expected population at Time 2. The expected Time 2 population is then compared to an observed Time 2 population from the single year ACS (2006, 2007, 2008, and 2009). The difference between the expected and observed populations is the residual net native migration over the entire time period. This population is then annualized by dividing the estimate by the number of years between Time 1 and Time 2 (e.g., 6.25 years between Census 2000 and ACS 2006). The next step of the calculation is to create an annual NNM rate. To do this we take the annualized residual and divide by the mid-period population and multiply the result by 1,000. Since the rate increases or decreases from year to year partly due to survey error in the ACS, we take an average of NNM rates from the calculations ending in the years 2006, 2007, 2008, and 2009 and apply the result to the population at risk of being a migrant in each single-year ACS file from 2000 to 2009.[5] We define the population at risk of being a migrant as all natives in the household sample of the ACS.[6] Table 2 reports the yearly and cumulative totals for the

---

[5] If the rate calculation came out as a positive, our definition of the population at risk of migrating would be problematic. We are currently researching a way to deal with this issue should it arise.
[6] We use the household population because the alternative approach is being researched for potential incorporation in the Population Estimates Program. The Population Estimates Program includes a separate component that measures changes in the group quarters population.

population at risk of being a migrant, the NNM rate, and the NNM estimate.

Figure 3. Net Native Migration Calculated Using the Alternative Method
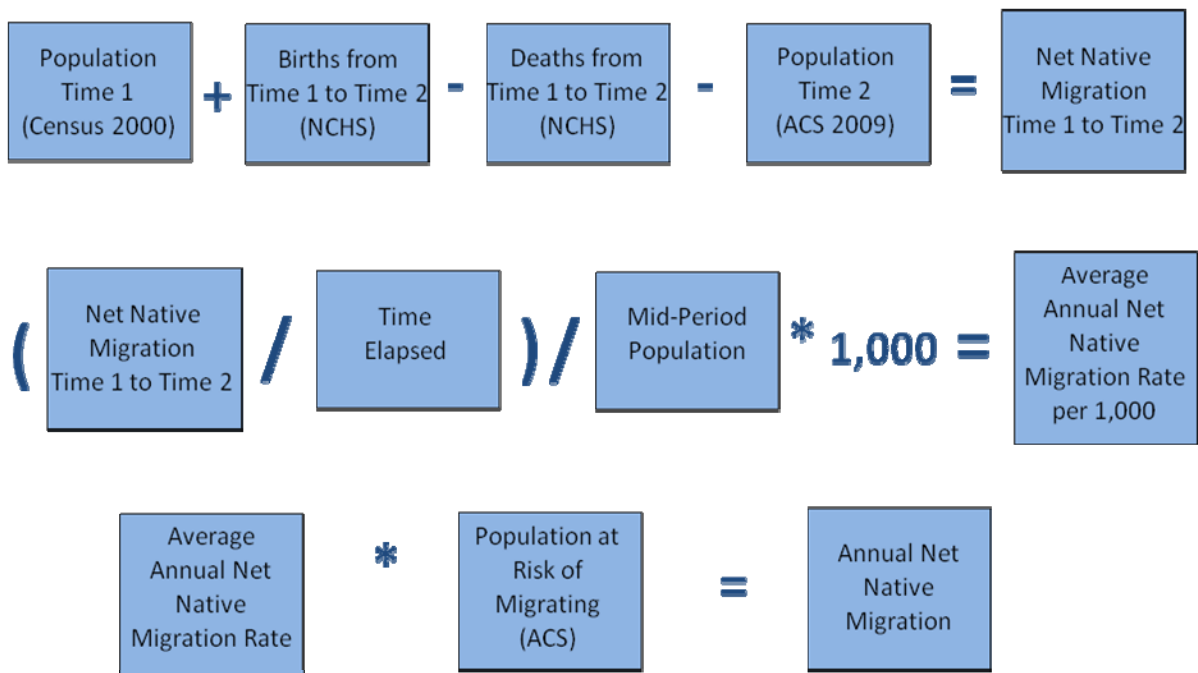
One thing to note from Table 2 is that the alternative method for calculating NNM allows the estimate to vary due to changes in the size of the population in the United States at risk of emigrating. However, as more years of ACS data become available, we may be able to also vary the NNM rate across time.

Comparing the annual estimates in Table 2 to those in Table 1, we see that the alternative estimate falls between the native migration estimate using the foreign country residual method (approximately -45,000 per year) and the NNM estimate using the foreign country residual method plus an additional estimate for those who were born abroad of U.S. citizen parents in countries that collected data by country of birth (approximately -5,500 per year). Let us make the assumption that both of these methods work without any error and the only difference is the inclusion of those born abroad of U.S. citizen parents (which is clearly not the case). If this were true we could surmise that the foreign country residual method includes some but not all of those born abroad of U.S. citizen parents.

There are several limitations associated with calculating net native migration using the alternative method. First, since we use the Census 2000 as the beginning point and the ACS as the end point, we potentially violate the assumption of constant error between Time 1 and Time 2 (this is also potentially an issue with the foreign country residual method). Second, we apply resident population survival rates to the native resident population to get deaths. The foreign-born population is included in the calculation of the survival rates, and differential mortality between the foreign-born and natives would produce error in our calculation of deaths. Third, applying rates to such a large population increases the impact of even small errors. When using the foreign country residual method, the populations used to calculate the residual are small for each country compared to the alternative method. Therefore error in one country will impact the total much less than an error in the U.S.-data-based calculation. Finally, sampling and non-sampling error in the ACS may have a substantial impact on the alternative estimate of native migration[7].

---

[7] See http://www.census.gov/acs/www/data_documentation/documentation_main/ for information on the accuracy of ACS estimates.

Table 2. Annual Net Native Migration Estimates From the Alternative Method: 2000-2009

Year	ACS Estimate of the Population at Risk of Migrating (in thousands)	Net Native Migration Rate	Net Native Migration Estimate
2000	243,370		-17,425
2001	245,535		-17,580
2002	247,491		-17,720
2003	249,376		-17,855
2004	251,412	-0.0716	-18,001
2005	252,688		-18,092
2006	254,398		-18,215
2007	256,081		-18,335
2008	258,445		-18,504
2009	260,836		-18,676
Total	**N/A**	**N/A**	**-180,403**

Notes: Net native migration rates are reported per 1,000 population. The population at risk of emigrating is the native household population. Year refers to the time period from July 1 of the year indicated to June 30 of the following year.
Sources: U.S. Census Bureau, Population Division tabulation of data from the Census 2000 sample edited data file and the annual 2000-2009 American Community Survey.

Discussion

The estimation of net native migration is problematic due to the lack of data on native immigration and emigration. The current method used by the U.S. Census Bureau's Population Estimates Program uses data from other countries to estimate NNM. As noted, there are several limitations with this technique, both methodologically (uncertainty about error) and pragmatically (difficulty in updating the estimate). In the 2010 Demographic Analysis, some steps forward were made in challenging the assumptions of this estimate as well as questioning the efficiency of the method. Specifically, in DA 2010 we varied assumptions about the coverage of those born abroad of U.S. citizen parents in the foreign country residual method. In this paper we have taken a further step and calculated an alternative estimate of net native migration using data from the United States.

While there are limitations of both the foreign country residual based and alternative (U.S. data source based) methods, there are several reasons that when taken together favors the U.S. based approach. First, after taking into account those born abroad of U.S. citizen parents, the alternative method falls in the middle of the range of the foreign country residual method and the estimates calculated in DA 2010. This may mean that the alternative method is picking up on something that the foreign country residual method is missing. Second, it is much easier and more cost effective to update the alternative method. The alternative method has many of the same qualities as the Population Estimates Program's method for estimating foreign-born emigration, and thus the method has been tested. Further, the alternative method includes the entire population of interest whereas the foriegn country residual method may or may not include those born abroad of U.S. citizen parents (and may include part of this population, but we have no way of knowing how much). Third, since the alternative method is easier to update, we do not have

to hold NNM constant for long periods of time. Fourth, on balance, data for more recent time points are used in the calculation of the alternative method. Indeed, some of the data used in the calculation of the foreign country residual method begins with data in the 1970's. The alternative method uses a starting point of 2000 with an ending point in the decade between 2000 and 2010. Further, as will be discussed in the next section, it may also be possible to calculate NNM using the ACS as both the Time 1 population and the Time 2 population, which would produce migration rates that are even more up to date.

Future Research

The work in this paper points us towards several avenues of future research. First, if we continue to use the foreign country residual method we should incorporate new data as many countries have completed censuses in the past couple of years. Second, we need to continue research on the alternative method. As noted above, one of the most promising areas of future research is to explore if it is possible to use the ACS as both the observed Time 1 and observed Time 2 populations. We need to evaluate the impact of sampling variability in the ACS and its impact on the net native migration rates.

A final area of future research will focus on the proxy universe used to distribute our total estimate of net native migration by demographic characteristics (age, sex, race, and Hispanic origin) and geography (nation, state, and county). In our forthcoming Vintage 2011 series of population estimates, we plan to apply the characteristics and geographic distribution of natives in the ACS (multi-year files) to our national-level estimate of native migration. We know that the likelihood of migrating varies within the native population. We plan to evaluate alternative proxy universes for incorporation in future series of population estimates. Overall, while we do not have a comprehensive data source to estimate the migration of U.S. natives, information from the ACS will be valuable in our continuing efforts to improve our estimates of international migration.

References

Gibbs, J., G. Harper, M. Rubin, and H. Shin. 2003. "Evaluating Components of International
        Migration: Native-Born Emigrants," Population Division Working Paper No. 63. U.S. Census Bureau.

Schachter, J. 2008. "Estimating Native Emigration from the United States," Memorandum dated
        December 24, developed during contract work for the U.S. Census Bureau.

U.S. Census Bureau. Accuracy of the Data of the American Community Survey.
        http://www.census.gov/acs/www/data_documentation/documentation_main/.

Appendix 1. Countries Included in the Foreign Country Residual Method

Citizenship
Andorra
Anguilla
Austria
Belgium
Czech Republic
Egypt
Faeroe Islands
Germany
Greece
Guinea
Hong Kong
Japan
Jordan
Kiribati
Liechtenstein
Luxembourg
Netherlands
Norway
Papua New Guinea
Philippines
Portugal
Republic of Korea (South)
Seychelles
Spain
Zambia

Country of Birth
American Samoa
Argentina
Aruba
Australia
Bahamas
Barbados
Belize
Bolivia
Brazil
Canada
Cayman Islands
Chile
Costa Rica
Croatia
Denmark

Dominica
Ecuador
El Salvador
Falkland Islands
Finland
France
French Polynesia
Guam
Guatemala
Guyana
Honduras
Hungary
Iceland
India
Ireland
Israel
Italy
Jamaica
Malta
Mexico
Netherlands Antilles
New Zealand
Nicaragua
Niue
Northern Mariana Islands
Palau
Panama
Paraguay
Peru
Samoa
Slovenia
South Africa
St. Kitts and Nevis
St. Lucia
St. Vincent and the Grenadines
Sweden
Switzerland
Trinidad and Tobago
Turkey
Turks and Caicos Islands
Uganda
United Kingdom
Uruguay
Venezuela