Distr.: General
10 October 2019

English

United Nations Economic Commission for Europe

Conference of European Statisticians

Work Session on Migration Statistics
Geneva, Switzerland
29-31 October 2019
Item 6 of the provisional agenda
Measuring emigrants and hard-to-reach groups of migrants

United States Census Bureau methods for estimating emigration of the foreign born

Note by the United States Census Bureau*

Summary

The United States has few data sources that measure residents who move out of the country. Emigration has been the most challenging measure of international migration and is a necessary demographic component for producing official population estimates. The Census Bureau historically used change-in-stock measures from the decennial censuses to estimate annual emigration levels. After 2010, the Census Bureau switched to a residual method using data on the foreign-born population collected by the American Community Survey, a large sample household survey that contains demographic, migration, and socioeconomic variables. The residual method measures cohort change in the foreign-born stock measured between two survey years, decomposed by deaths and immigration. The method assumes any remaining change not accounted for by deaths and immigration reflects emigration. In this paper, we provide an overview of the current emigration methodology used to produce the official population estimates and provide examples using publicly available data from the Census Bureau. We also discuss weaknesses in the method and ongoing research into using survey and administrative records to measure emigration. This paper is to be informative and provides a case study of a country that has few administrative data to measure emigration.

*Prepared by Mr. Anthony Knapp

I. Introduction

1. As a key component of population change, emigration is required for the postcensal population estimates, population projections, and coverage evaluation programs at the Census Bureau. However, this is the most difficult component to estimate since the United States (U.S.) lacks data sources to directly measure residents who move out of the country. The Census Bureau for decades used a residual method on information from decennial censuses to indirectly measure emigration of residents who are foreign born (persons who are not a U.S. citizen at birth). The residual method measures cohort change in the resident foreign-born population between two points in time, decomposed by deaths and international migration. The residual is any remaining change not accounted for by deaths and immigration, which subsequently is assumed to represent emigration flow. Although the residual yields an estimate of emigration occurring throughout the decade, it is not always suitable for applications that require yearly emigration estimates. A criticism has been the inability for the residual to track annual patterns in emigration, especially for recent migrants, as well as underestimating total emigration occurring between census years (Van Hook et al., 2006). It also only measures emigration based on cohorts that were living in the country at the time of the census, but not those who entered the country between censuses. The Population Estimates Program at the Census Bureau has devoted substantial effort into improving methods for estimating emigration. In this paper, we provide a brief overview of the emigration methodology in the United States, recent efforts by the Census Bureau to improve existing methods, an example of using survey data to produce emigration rates, and a discussion of lessons learned from using this methodology for the population estimates.

II. Background

A. Emigration Statistics in the United States

2. For the first half of the 20th century, the United States collected information on emigrants, but this was discontinued in 1957 over concerns about data quality (Warren and Peck, 1980; Kraly, 1998). Another source of data was the Alien Address Report Program from 1952 to 1981. The Alien Address Report Program came about from the Alien Registration Act of 1940, which required non-citizens living in the U.S. to register with the federal government. In addition, immigrants were required to report address changes while foreigners were required to verify their address every 90 days (U.S. Government Accountability Office, 2002). The program expanded under the Internal Security Act of 1950 and the Immigration and Nationality Act of 1952, which required resident immigrants to register with the Immigration and Naturalization Service (INS) and annually update their address information on 1 January every year. This resulted in a micro-level, longitudinal dataset, known as the Alien Address Form I-53 file, which produced statistics on the annual stock of resident non-citizens. This also provided an independent dataset to evaluate coverage of the foreign born in the census (Warren, 1979). Furthermore, adjusting for immigration and mortality, stock estimates from the Form I-53 file could be converted into gross international migration flows, from which emigration could be derived (Jasso and Rosenzweig, 1982; Warren and Passel, 1987). However, this dataset, similar to population registers, was prone to noncompliance in reporting, as well as lack of reporting when a person moved out of the country (U.S. Government Accountability Office, 2002). The program was discontinued under the Immigration and Nationality Act of 1981. Internal documents at the time cited

initiatives to cut costs and reduce respondent burden as justifications for discontinuation (U.S. Government Accountability Office, 2002).

3. Currently, the U.S. lacks comprehensive administrative data from which statistics on emigration can be produced. Data collected from Form I-94 non-immigrant arrival and departure records by the Department of Homeland Security provide a possible data source, given these forms collect visa information, thereby differentiating resident migrants from tourists and business visitors. However, the departure records are incomplete since the U.S. does not have strict departure control (Baker, 2017), with most departure records coming from commercial air and sea carrier departure manifests (Baker, 2018), rather than land border crossings. Researchers have investigated the utility of social security[1] records to estimate emigration, but more definitive statistics are limited to legal immigrants who are eligible for social security (Duleep, 1994), reported positive earnings (Schwabish, 2011), or persons who are registered at a foreign address (Turra and Elo, 2008). By far, decennial censuses for 2000 and prior and annual household surveys after 2000 provide the most complete information on the foreign born. Though censuses and surveys do not ask questions about persons who emigrated, indirect estimation may be done instead.

B. Indirect Estimates of Emigration, 1980-present

4. Given the lack of emigration statistics from direct sources during this period, Warren and Peck (1980), of the INS and Census Bureau respectively, developed a residual method to indirectly estimate emigration using census data on the foreign-born population. This method estimated cohort change in the foreign-born population between the 1960 and 1970 censuses by accounting for deaths and immigration over this period. Under some circumstances, immigration flows are difficult to measure. However, Warren and Peck (1980) were able to estimate period-specific immigration using the year of immigration question from the 1970 census long form (which sampled households for additional questions as part of the census enumeration). The method assumes the residual reflects emigration flow but is impacted by any coverage differences between the censuses and measurement error from the year of entry and nativity or citizenship questions. Their work established the foundation of the residual method used by the Population Estimates Program.

5. Estimates on emigration levels between 1970 and 1980 were based on analyses of the INS Form I-53 file from Warren and Passel (1987). However, the residual method was utilized by Ahmed and Robinson (1994), a highly-cited working paper from the Census Bureau, based on 1980 and 1990 census data to estimate emigration levels and rates by country of birth and demographic characteristics. The results were updated further by Oosse (1998) for the 1990 postcensal estimates and 2000 demographic analysis program evaluation at the Census Bureau. Mulder (2003) updated the residual with data from the 1990 and 2000 censuses. However, extensive coverage differences between the 1990 and 2000 censuses resulted in negative emigration numbers for certain countries of birth, a demographic impossibility. Census 2000 was the final year that the decennial census collected information on country of birth and migration. The American Community Survey (ACS) has since replaced the decennial census for collecting this type of information. The ACS is an annual large sample household survey which asks detailed questions on demographic

[1] Social Security is a national government-administered program that manages retirement, disability, and survivors benefits. Most workers pay Social Security taxes on their earnings. Legal immigrants, as well as nonimmigrants, are allowed to apply for a social security number.

characteristics, migration, household structure, housing units, and socioeconomic variables. The sample was fully expanded in 2006 and has been conducted every year. An advantage of this survey is that it provides more current estimates than the decennial censuses. However, sample sizes are smaller, with final interviews from about 2 million housing units in the 2010 ACS (U.S. Census Bureau, n.d.) compared to about 19 million in the 2000 long-form (U.S. Census Bureau, 2002). In the 2000s, the Population Estimates Program incorporated both census long form and survey data to produce emigration. For the post-2010 estimates, the program used data entirely from the ACS.

III. The Residual Method

A. The Residual and Emigration

6. The residual method is used for measuring net migration flow for geographic areas where population registers or other administrative data are limited or non-existent. At the national level, this method yields an estimate of net international migration. For sub-national areas, the estimate represents net migration, which reflects both international and internal migration. This method compares the population measured between two censuses and accounts for population change due to births, deaths, and net migration. This measure is based on the national-level cohort component model, updated to measure the foreign-born population:

$$\hat{P}_2 = P_1 - D_{1\text{-}2} + M_{1\text{-}2} \qquad [1]$$

7. Where $\hat{P}_2$ is a derived population at time 2, P_1 is the starting population measured at time 1, $D_{1\text{-}2}$ is the number of deaths occurring between time 1 and time 2, and $M_{1\text{-}2}$ is a measure of net international migration occurring between time 1 and time 2. Unlike a traditional cohort component model this equation omits births ($B_{1\text{-}2}$) since the population of interest is the foreign born.

8. If the value for P_2 is known a priori, then the components may be rearranged to derive an implied estimate of net international migration ($\hat{M}_{1\text{-}2}$):

$$\hat{M}_{1\text{-}2} = P_2 - (P_1 - D_{1\text{-}2}) \qquad [2]$$

9. Where P_2 is the "observed" population at time 2 and ($P_1 - D_{1\text{-}2}$) represents the "expected" population at time 2 if no migration has occurred. $\hat{M}_{1\text{-}2}$ is the "residual" estimate. The residual is an implied estimate since it represents migration along with any coverage, sampling, non-sampling, and measurement error in the input data used to estimate the population and death variables.

10. Warren and Peck (1980) expanded the above formula to convert net migration flow into gross migration flow by using known measures on immigration into the U.S. The following formula expresses emigration as a positive value:

$$\hat{E}_{1\text{-}2} = I_{1\text{-}2} - \hat{M}_{1\text{-}2} \qquad [3]$$

11. Where $\hat{E}_{1\text{-}2}$ represents implied foreign-born emigration occurring between time 1 and time 2. Under this modified method, $\hat{E}_{1\text{-}2}$ replaces $\hat{M}_{1\text{-}2}$ as the residual. $I_{1\text{-}2}$ represents immigration measured between time 1 and time 2. Substituting $\hat{M}_{1\text{-}2}$ from Equation [3] with Equation [2] and rearranging the terms produces the logical formula for estimating implied emigration:

$$\hat{E}_{1\text{-}2} = (P_1 - D_{1\text{-}2}) - (P_2 - I_{1\text{-}2}) = P_{exp,2} - P_{obs,2} \qquad [4]$$

12. Where $(P_1 - D_{1\text{-}2})$ represents the "expected" population measured at time 2 if no migration has occurred ($P_{exp,2}$), and $(P_2 - I_{1\text{-}2})$ represents the "observed" population measured at time 2 ($P_{obs,2}$). The residual, $\hat{E}_{1\text{-}2}$, is the difference between the observed and expected populations. Equation [4] decomposes emigration into components of population change that can be calculated using input data from surveys, censuses, and vital records.

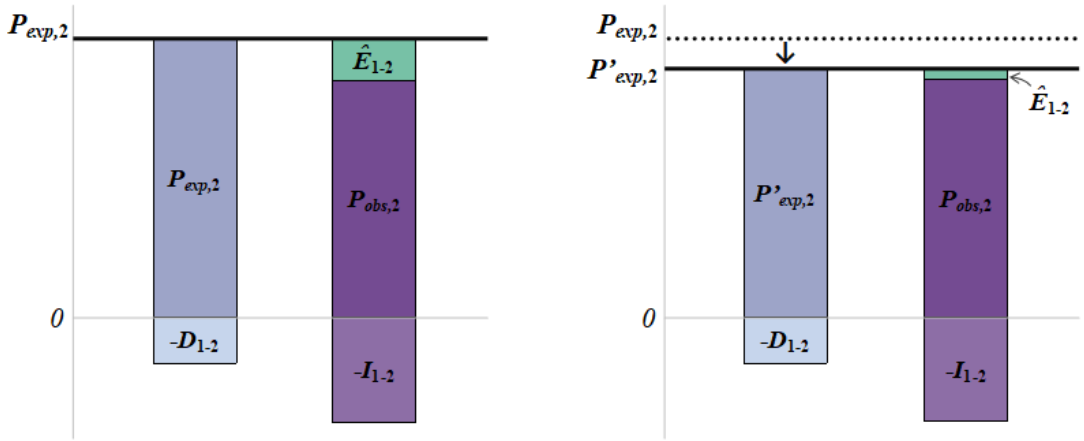
B. Data Considerations and Limitations

13. The residual method may be applicable to countries that lack longitudinal databases that track individuals over time. An advantage is that the population variables may be estimated from cross-sectional data from two censuses or survey years. Summary tables may be used instead of micro-data. Deaths may come from a national vital registration system or from mortality rates produced from the national statistical office or from international organizations.

14. However, the residual method is highly sensitive to coverage and measurement error in the input data. The following graphs are visual representations of the residual method from Equation [4] and show how data quality can influence the residual. Please note, the following graphs exaggerate the scale of deaths, immigration, and emigration for better visibility. Figure 1a shows an example of an unbiased residual estimate, which assumes any error from the input data is negligible. The residual ($\hat{E}_{1\text{-}2}$) is the difference between the expected and observed populations. Figure 1b shows a scenario where the expected population at time 2 is reduced from $P_{exp,2}$ to $P'_{exp,2}$. This may occur if the starting population at time 1 is underestimated or deaths are overstated. Assuming all other components remain the same, the model decreases the residual to account for the lower expected population.

Figure 1
Components of the Foreign Born and Residual Calculation

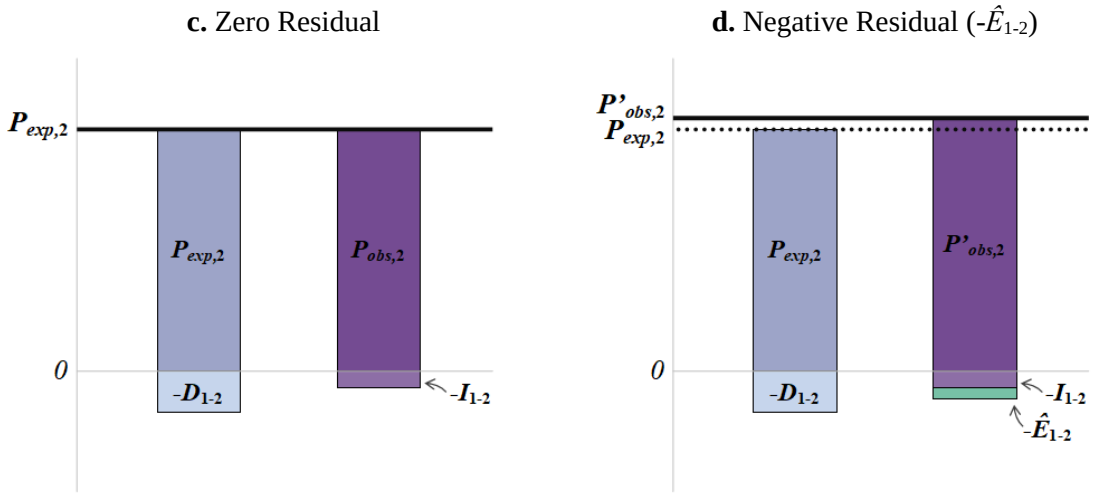
a. An Unbiased Residual ($\hat{E}_{1\text{-}2}$) **b.** Underestimated Expected Population ($P'_{exp,2}$)



15. Figure 2 shows components for another population, which exhibits low levels of migration. Figure 2c represents the components for an immigrant population that exhibits near-zero emigration. A low residual may be acceptable if this is known to be a relatively closed population and there is high confidence in the input data. However, under these

circumstances, the residual method is susceptible to estimating negative emigration if there is any slight deviation in the population, death, or immigration estimates. Figure 2d shows a scenario where the observed population at time 2 was overestimated. This outcome may be due to an overestimation of the population measured at time 2 or underestimation of immigration. The method scavenges from the residual in order to offset the higher observed population. However, this results in a negative residual ($-\hat{E}_{1\text{-}2}$), which is demographically impossible.

Figure 2
Components of the Foreign Born and Residual Calculation, continued



16. The residual method typically performs better for immigrant groups with higher propensities to emigrate. These may include persons born in countries adjacent to the U.S. (e.g. Canada and Mexico), recent arrivals (e.g. temporary migrants), immigrants who have better coverage in the censuses and surveys, as well as those within certain age groups where migration is more prevalent (e.g. young adults and sometimes post-labor migrants or retirees). On the other hand, the residual method may be weaker for certain immigrant groups with lower propensities to emigrate, such as non-recent immigrants.

17. The residual method is prone to diminished emigration rates as the interval between time 1 and time 2 increases because the method fails to measure recent migration patterns occurring between time 1 and time 2 (Leach and Jensen, 2014), such as those for temporary and circular migrants. For example, an immigrant may be present in the census at time 1, return to their home country where they reside for several years thereafter, and finally return to the U.S. before the next census at time 2. The residual method will fail to identify this person as an emigrant. As recent as 2013, the Population Estimates Program was still using Census 2000 long-form data to estimate the starting population at time 1 while using ACS data to estimate the population at time 2. This led to an interval of up to 13 years, which exceeded the conventional 10-year interval used in prior decades. Leach and Jensen (2014) updated the residual method to incorporate more recent ACS data to measure both the time 1 and time 2 populations. Since the ACS publishes data annually instead of once every 10 years like censuses, the interval can be reduced considerably. Theoretically, the interval may be reduced to one year. However, this is not done because the magnitude of survey errors at time 1 and time 2 may exceed the size of any measurable emigration occurring over a one-year period. Previous research has shown that 1-year intervals are prone to producing

negative residuals (Leach and Jensen, 2014). As the time interval increases, so might the numeric size of emigration, in which case it becomes more likely to surpass the size of potential survey error.

IV. Census Bureau Methods for Estimating Emigration

A. Current Estimation Procedure

18. Research from Leach and Jensen (2013, 2014), led to the development of what is called the *ACS-ACS residual method* currently used to estimate the foreign-born emigration component for the population estimates and projections programs. The method is based on Equation [4] from the previous section, in which the expected population is the ACS estimate of the foreign-born population at time 1, minus deaths to the foreign-born occurring between time 1 and time 2. The observed population is the ACS estimate of the foreign-born population at time 2, minus the year-of-entry estimate of recent immigration occurring between time 1 and time 2. Since we do not have complete information on individual deaths by nativity from the national vital records system, we use survival rates calculated from National Center for Health Statistics (NCHS) life tables to approximate period deaths to the foreign born. Official life tables on the foreign-born population living in the U.S. do not exist. We instead use life tables on the Hispanic[2] population, which we assume to be a reasonable substitute for approximating mortality patterns of the total foreign born. We apply survival rates by single year of age to the starting population measured at time 1 to derive the expected population at time 2:

$$\hat{E}_{1\text{-}2} = (P_1 \times s) - (P_2 - I_{1\text{-}2}) \qquad [5]$$

19. Where s represents the survival ratio of a population age x at time 1 that will survive to age $x+k$ at time 2. The term k is the number of years between time 1 and time 2. ($P_1 \times s$) represents the expected foreign-born population at time 2 if no migration had occurred. ($P_2 - I_{1\text{-}2}$) represents the observed foreign-born population at time 2 that would have been residing in the country at time 1.

B. Emigrant Groups and Different Propensities to Emigrate

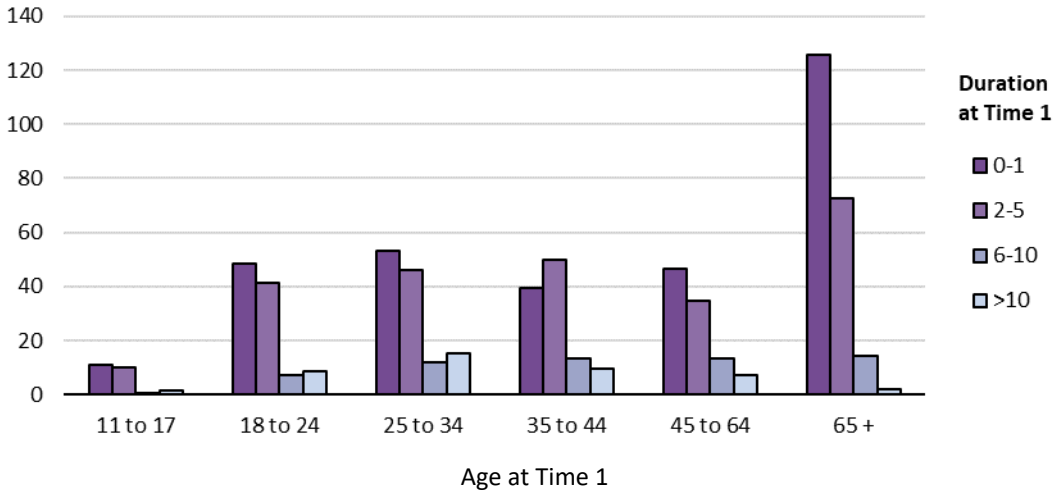
20. The previous sections show how we use the residual method to produce an estimate of foreign-born emigration for the total foreign-born population. However, producing a single emigration rate for the total foreign born is not done given immigrant subpopulations will have vastly different propensities to emigrate (Van Hook and Zhang, 2011). A simplistic approach looks solely at demographic factors, rather than various socioeconomic, household, political, historical, and international factors, to differentiate emigration patterns. Under this more sophisticated framework, age, place of birth, and duration of residence can be important indicators of emigration.

21. According to one estimate, about 80 percent of foreign-born emigration occurs within the first 10 years of living in the U.S. (Duleep, 1994), with most emigration occurring within the

[2] The U.S. definition of Hispanic includes persons of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race. Among the foreign born, this may include persons born in a Spanish-speaking country in the Americas or Spain.

first five years. We also expect emigration rates to vary by age. Furthermore, in the U.S. context, we expect most recent arrivals, as well as immigrants born in certain countries (e.g. in close proximity to the U.S.), to have the highest emigration rates. Figure 3 shows derived emigration rates from a previous evaluation of the ACS-ACS residual method by age and duration of residence at time 1 (Knapp et al., 2017). Immigrants who have been living in the U.S. for a shorter period (i.e. more recent immigrants) typically exhibit the highest emigration rates. This pattern is consistent across most age groups, and immigrants 65 and older exhibit the highest rates.

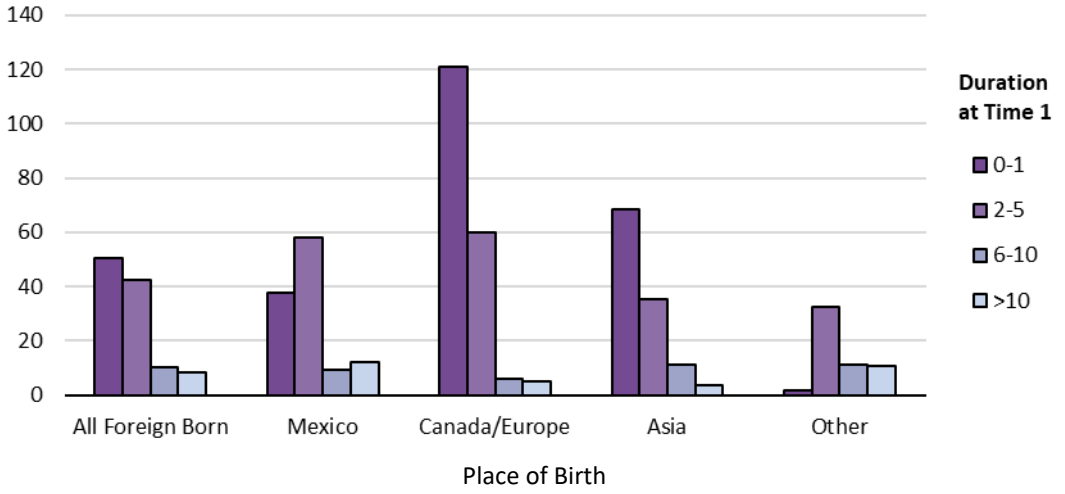
Figure 3
Annual Emigrants per 1,000 by Age and Duration in the U.S.



NOTES: Duration represents the number of years living in the U.S. as of time 1.
SOURCE: Reproduced from Knapp et al. (2017), from 2010-2014 5-year ACS.

22. Figure 4 shows the relationship of emigration to place of birth and duration of residence. Mexico, Canada and Europe, and Asia are major contributors of foreign-born immigration to the U.S. Similar to Figure 3, emigration rates differ by duration of residence. However, a duration of one year and less does not yield the highest rates for the Mexican born and the population born outside of Canada, Europe, and Asia.

Figure 4
Annual Emigrants per 1,000 by Place of Birth and Duration in the U.S.



NOTES: Duration represents the number of years living in the U.S. as of time 1.
SOURCE: Reproduced from Knapp et al. (2017), from 2010-2014 5-year ACS.

23. Therefore, we estimate emigration by subpopulations in order to better capture the heterogeneity of emigration patterns and characteristics of the foreign born. The Population Estimates Program uses emigrant groups defined by place of birth, duration of U.S. residence, and sex. Residuals are calculated for each emigrant group. Previously, the method used four groups based on Mexican/Non-Mexican born and duration of residence. This has since increased to seven groups shown in Table 1.

Table 1
Emigrant Groups for Residual and Emigration Rate Calculations

Place of Birth	**Sex**	**Duration of U.S. Residence (in Years)**	
		More Recent Immigrant	**Less Recent Immigrant**
Mexico	Male	≤ 10	
Mexico	Female	≤ 10	
Mexico	All		> 10
Canada and Europe	All	≤ 10	
Asia	All	≤ 5	
Other	All	≤ 10	> 5 (Asian) and > 10 (Non-Asian)

NOTES: The seven emigrant groups represent mutually-exclusive subpopulations of the foreign born, the sum of which reconstitutes the total foreign-born population.

24. The groups in Table 1 are defined to increase between-group differences in emigration while similar groups are combined. Sample sizes from the surveys limit the number of groups that can be used. Canadian and European-born populations exhibited similar patterns and

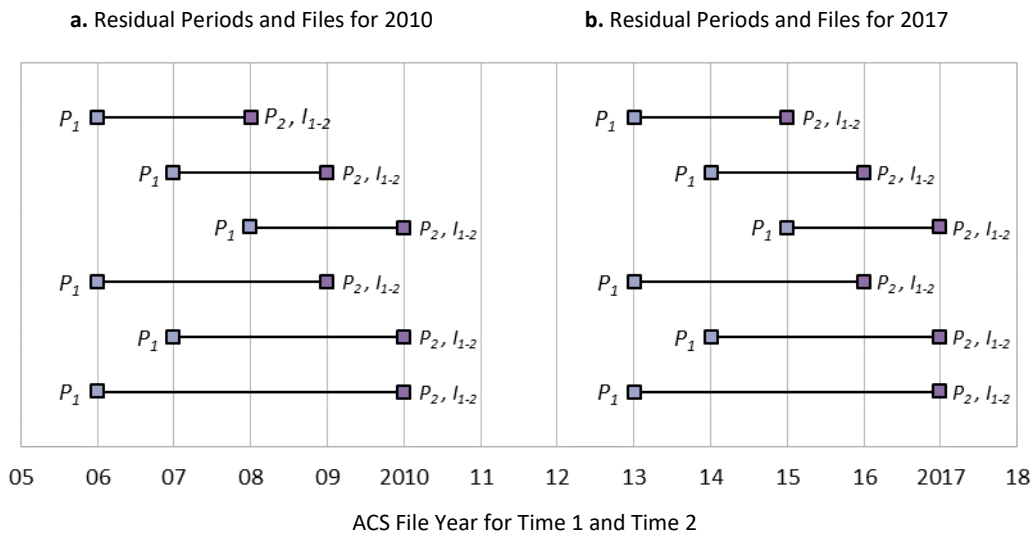
therefore are combined into a single group, while Mexico was split into separate groups by sex due to differences between Mexican males and females. In addition to estimating residuals, demographic characteristics and geographic distributions are estimated separately for each group.

C. Residual Periods and Averaging

25. Negative residuals do occur from the estimation process. In order to mitigate the effects of sampling and non-sampling error from the surveys, Leach and Jensen (2014) used averaged estimates from various combinations of survey years for time 1 and time 2. They tabulated five consecutive survey years of data on the foreign-born population to estimate emigration for one full year. From these five survey years, three 2-year residuals, two 3-year residuals, and one 5-year residual are produced. Notionally, four 1-year residuals could be derived; however, as described earlier, these tend to result in negative emigration (Leach and Jensen, 2014) and hence are not included in our methodology.

26. Figure 5 shows an example of ACS file selection for calculating six residuals by estimation year. P_1 comes from the ACS file for time 1, and P_2 and $I_{1\text{-}2}$ come from the ACS file for time 2. In Figure 5a, the top three line segments depict the 2-year residual periods. The next two line segments depict the 3-year residual periods, and the bottom line segment depicts the 4-year residual period. The residual periods lag behind the estimation year. The procedure uses 2006-2010 ACS files to estimate emigration for the 2010 estimation year. All six residual periods increment forward by one year for each subsequent estimation year. Figure 5b shows the final collection of ACS files for estimating annual emigration in 2017. Since the 2018 ACS was not released prior to our Vintage 2018 estimates production cycle, the procedure uses rates from 2017 to estimate annual emigration in 2018.

Figure 5
Residual Periods with ACS Variable and File Selection: 2010 and 2017



NOTES: P_1 and P_2 are foreign-born populations measured at time 1 and time 2 respectively. $I_{1\text{-}2}$ is foreign-born immigration measured between time 1 and time 2.
SOURCE: Based on the survey file selection from Leach and Jensen (2014).

27. A residual is estimated for each residual period. The residual only reflects emigration of those who were present in the country at time 1 and not present at time 2. We assume the emigration patterns are representative of the more recent foreign born who were not present at time 1. We convert the residual into an annualized emigration rate. The current procedure accomplishes this by dividing the residual by the number of person years for the period. The six emigration rates are averaged to produce the final rate, which is then applied to an at-risk population for each of the seven emigrant groups to calculate the annual emigration for the nation. The next section provides an example of the complete method explained thus far, with data on Canadian and European-born population residing in the U.S.

D. Example using Public Use Microdata Sample from the ACS

28. The previous section described the framework for estimating emigration for the population estimates and projections programs. These programs use data from internal ACS microdata files. For the following example, we use publicly available microdata downloaded from the Census Bureau website. The purpose of this exercise is to provide an example for countries that lack administrative records or other higher-quality data sources for estimating emigration. Table 2 lists the variables, descriptions, and data sources to be used for the residual method.

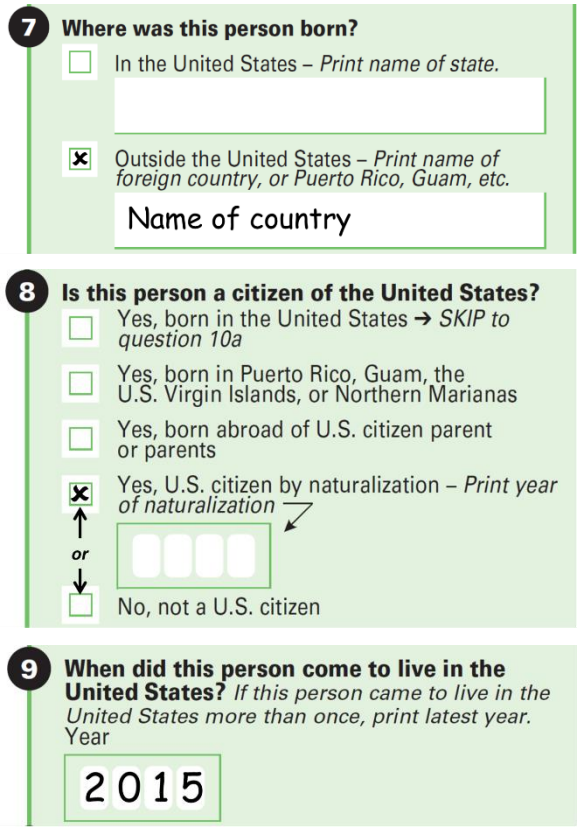
Table 2
Description of the Variables and Data Sources

Variable	Description	Data Source
P_1	Foreign-born population observed at time 1	Census Bureau, 1-year ACS, Citizenship and Country of Birth
P_2	Foreign-born population observed at time 2	Census Bureau, 1-year ACS, Citizenship and Countryof Birth
$I_{1\text{-}2}$	Recent immigration: foreign-born population observed at time 2 that entered the U.S. after time 1	Census Bureau, 1-year ACS, Year of Entry
s_x	Survival ratio by sex and age x	NCHS, Hispanic life tables
$D_{1\text{-}2}$	Estimated deaths	$= P_1 \times (1 - s_x)$
$P_{obs,2}$	Observed foreign-born population at time 2 that was present at time 1	$= (P_2 - I_{1\text{-}2})$
$P_{exp,2}$	Expected foreign-born population at time 2 that was present at time 1	$= (P_1 \times s_x)$ or $(P_1 - D_{1\text{-}2})$
$\hat{E}_{1\text{-}2}$	Residual: resident foreign-born population observed at time 1 that emigrated between time 1 and time 2	$= P_{exp,2} - P_{obs,2}$
$PY_{1\text{-}2}$	Person-years lived between time 1 and time 2	-
r	Annualized emigration rate	$= \hat{E}_{1\text{-}2} / PY_{1\text{-}2}$

29. ACS questions for place of birth, citizenship, and year of entry are provided in Figure 6. Country of birth is used for estimating foreign-born populations for time 1 and 2 and to define the seven emigrant groups by place of birth. The citizenship question is used to limit

the sample to the foreign born only (year of naturalization is not used). The year of entry question is used for estimating immigration and to define the seven emigrant groups by duration of residence.

Figure 6
ACS Place of Birth, Citizenship, and Year of Entry Questions



SOURCE: U.S. Census Bureau, 2015 American Community Survey. Available online at: http://www2.census.gov/programs-surveys/acs/methodology/questionnaires/2015/quest15.pdf

30. Table 3 shows the calculation of emigration rates for the recent Canadian and European group, using data from the 2006 to 2010 ACS. File selection following the logic from Figure 5a is used to estimate annual emigration flow in 2010. The $PY_{1\text{-}2}$ column in Table 3 represents person-years survived and is based on the mid-point population between time 1 and time 2 multiplied by the number of years in the residual period. The first step converts the residual into a rate while the second step annualizes the rate. Alternatively, dividing the residual by the starting population less one-half of deaths and one-half of the residual may produce similar results. The average of the six rates is 0.031 (31 emigrants per 1,000 U.S. residents born in Canada or Europe and residing in the U.S. for 10 years or less as of time 1). Negative rates are retained prior to averaging in order to offset potentially over-inflated rates. If the average emigration rate is negative, emigration is set to zero. The Canadian and European populations tend to exhibit higher propensities to emigrate compared to other groups. The residual method typically does not produce negative rates for this group.

Table 3
Residual and Emigration Rate Calculation for Recent Canadian and European Immigrants Living in the United States: 2006-2010

Survey Year time 1, time 2	P_1	$D_{1\text{-}2}$	$P_{exp,2}$	P_2	$I_{1\text{-}2}$	$P_{obs,2}$	$\hat{E}_{1\text{-}2}$	$PY_{1\text{-}2}$	r
	a	b	c = a - b	d	e	f = d - e	g = c - f	h	i = g/h
2006, 2008	1,597	9	1,588	1,890	403	1,488	100	3,134	0.032
2007, 2009	1,534	8	1,526	1,759	355	1,404	122	2,999	0.041
2008, 2010	1,479	8	1,471	1,726	348	1,378	93	2,904	0.032
2006, 2009	1,597	14	1,583	1,891	474	1,417	166	4,687	0.035
2007, 2010	1,534	12	1,522	1,873	459	1,414	108	4,530	0.024
2006, 2010	1,597	19	1,578	2,007	582	1,425	153	6,274	0.024
								Average =	**0.031**

NOTES: Totals are in thousands. Calculations may not be exact due to rounding.
SOURCE: U.S. Census Bureau, Population Division. Simulation of the net international migration estimates production code on 2006-2017 ACS PUMS file.

31. The process is repeated so that an average rate is produced for each year in the decade. An average rate was calculated from 2010 to 2018 as shown in Table 4. The at-risk population is the population that may experience emigration and is estimated from the previous year's survey file. In this example, the at-risk population is defined as the resident U.S. population of persons born in Canada or Europe who entered the country up to 10 years prior to the survey (i.e. have been residing in the U.S. for 10 years or less). The average rate is applied to the at-risk population to produce an estimate of annual emigration. Based on requirements for the population estimates, the 2010 estimate reflects only a quarter year.[3] In this example, 442 thousand recent Canadian and European arrivals had emigrated from the U.S. since 2010 (not an official estimate). Since the 2018 survey is not available, we simply use the rate from 2017 for 2018.

Table 4
Annual Emigration (in Thousands) of Recent Canadian and European Immigrants: 2010–2018

Estimates Year	Estimates Period	6-Rate Average	At-Risk Population	Emigration
a	b	c	d	e = c × d
Cumulative	**1 April 2010 – 30 June 2018**	**-**	**-**	**442**
2010	1 April 2010 – 30 June 2010	0.031	1,454	11[a]
2011	1 July 2010 – 30 June 2011	0.033	1,389	46
2012	1 July 2011 – 30 June 2012	0.036	1,263	45
2013	1 July 2012 – 30 June 2013	0.033	1,208	40

[3] 1 April was the date of the 2010 census. The population estimates represent the mid-year population on 1 July. The full estimates year is from 1 July to 30 June.

Estimates Year	Estimates Period	6-Rate Average	At-Risk Population	Emigration
a	b	c	d	e = c × d
2014	1 July 2013 – 30 June 2014	0.037	1,211	45
2015	1 July 2014 – 30 June 2015	0.040	1,175	47
2016	1 July 2015 – 30 June 2016	0.053	1,234	65
2017	1 July 2016 – 30 June 2017	0.060	1,184	71
2018	1 July 2017 – 30 June 2018	0.060[b]	1,201	72

NOTES: These estimates are for demonstration purposes only and are not official.
[a] The 2010 estimate represents a quarter year
[b] The 2017 rate is held constant for 2018.

SOURCE: U.S. Census Bureau, Population Division. Internal simulation of the population estimates on 2006-2017 1-year ACS PUMS data.

E. Subnational Characteristics

32. We distribute annual emigration totals into the demographic and geographic cells. Since sample size limitations prevent us from estimating the residual by age, sex, race/ethnicity categories, and subnational geographies, we obtain this information from a *proxy universe*, which is a population with demographic and geographic characteristics assumed to be similar to that of recent emigrants. The proxy universe for this example is the resident U.S. population born in Canada or Europe who were living in the U.S. 10 years or less. This universe is the same as the at-risk population; however, the estimate comes from pooled, instead of 1-year, ACS files in order to increase sample sizes. We use three-year pooled data to obtain national and state distributions and five-year ACS files for county distributions. The county is the smallest level of geography for which the Population Estimates Program estimates. Proportions are generated from the proxy universe, which allocate the national emigration total for each year into the single age, sex, race, Hispanic origin, state, and county data cell.

F. Lessons Learned this Decade

33. This section describes some of the experiences working with the residual method for calculating population estimates. While previous decades used constant emigration rates to estimate annual emigration, the ACS-ACS residual method was intended to be a data-driven method that produces emigration rates that can vary from year-to-year. However, this method can be volatile due to sampling and non-sampling variability from survey data. Due to smaller sample sizes relative to the Census 2000 long form, coverage, and occasional methodological changes to the surveys and population controls, the derived rates can deviate substantially each year. For example, the 2008 survey cancelled the operation known as Failed Edit Follow-Up for most of the year (U.S. Census Bureau, 2009). This survey operation contacts households that returned an incomplete survey questionnaire as well as large households that have more members than space available on the questionnaire (the questionnaire collects person characteristics for up to six members in the household). In another example, the ACS updated the respondent instructions for the year of entry question in 2015. This led to higher immigration in 2015 relative to earlier years. Assuming all other

components were unaffected, increased immigration reduces the size of the observed population and thus produces a higher residual.

34. The residual method works reasonably well for immigrants who exhibit high propensities to emigrate, but less so for those with lower propensities. For example, immigrants who are living in the country over 10 years (whom we classify as non-recent) are expected to have the lowest emigration rates. But non-recent groups are especially susceptible to negative rates. In addition, these groups have the largest at-risk populations. The Non-Recent Other emigrant group from Table 1 represents almost half of the foreign-born population living in the U.S. Any minute deviation to the rate to the one-thousandth can produce an additional 100 thousand emigrants per year. One may assume that the true rate is likely close to zero and should not vary considerably from year to year given this is a relatively well-established population. Due to implausible fluctuations in emigration rates, our method uses constant, rather than variable, emigration rates for the non-recent groups.

35. Rate averaging is intended to reduce the effect of outliers in the six residuals. Outliers may inexplicably arise from use of certain combinations of survey years used at time 1 and time 2. However, certain survey years will have disproportionately larger effects on the final result because of the current survey file selection procedure. For example, in Figure 5a, three out of six of the residuals used 2006 survey data to measure the population at time 1. Similarly, half of the residuals used the 2010 survey for the population at time 2. Furthermore, the six residual periods are not evenly distributed across the year of the estimate. The 2006-2010 ACS files are used to estimate emigration in 2010 although 2008 is the mid-point for these files. This is potentially a holdover from older residual methods which used the previous two censuses to estimate emigration for the current decade (e.g. use of 1980 and 1990 census long form data to estimate emigration for the 1990s).

36. Finally, like many surveys and censuses, the ACS single-year-of-age data contain heaping in which respondents exhibit a digit preference for ages ending in zero or five. We do not smooth age distributions since we summarize these data to produce total, instead of age-specific, emigration. However, a potentially more serious weakness is heaping of the year of entry responses on: (1) reported years of entry ending in zero and five, and (2) reported years of entry five and ten years prior to the survey year. The second scenario occurs when respondents formulate their year of entry response based on the number of years from the time of survey. Both cases show that the year of entry question is susceptible to recall bias. However, we do not apply a method to adjust heaping in year of entry as annual immigration trends are not smoothly distributed across years, unlike age. This may have a sizeable effect on emigration rates and estimates of the at-risk population, especially for immigrants who fall close to the 5-year or 10-year cut-offs that are used to define duration of U.S. residence for the emigrant groups in Table 1.

V. Conclusions

37. The Census Bureau uses residual methods for its official population estimates and projections due to the lack of emigration data collection in the U.S. This working paper provided a high-level overview of the development and implementation of the current foreign-born emigration methodology used for the population estimates and projections programs. Though the implementation of the annual ACS greatly reduces the length of the residual period and incorporates more recent input data compared to decennial censuses, additional improvements can be made to the overall emigration methodology. There are also unresolved questions about the coverage of the most recent migrants in the surveys.

Coverage in the data, and necessary adjustments, will have key effects on the quality of the emigration estimates using a residual methodology. We are exploring other national datasets that could provide more information on emigration trends, demographic composition, and subnational geographic distributions of emigration. Though the U.S. lacks a population register, there are possibilities for other administrative data collected by the federal government to track changes in the foreign-born stock and potentially derive emigration flow. An example is the Social Security Administration Numident file for persons who have obtained social security numbers by country of birth, age, and sex. Furthermore, the Internal Revenue Service immigrant tax identification number (ITIN) file includes non-citizens ineligible for social security numbers, which may help supplement the incomplete coverage of the foreign born in the Numident file. Finally, domestic and international address information from the federal income tax files, matched to country of birth information in the Numident file, may estimate annual movement between the U.S. and abroad.

References

Ahmed, B. and Robinson, J.G. 1994. "Estimates of Emigration of the Foreign-Born Population: 1980-1990." Population Division Working Paper 9. Washington, DC: U.S. Census Bureau.

Baker, B. 2017. "Nonimmigrants Residing in the United States: Fiscal Year 2015." Office of Immigration Statistics, Population Estimates. Washington, DC: U.S. Department of Homeland Security.

— 2018. "Nonimmigrants Residing in the United States: Fiscal Year 2016." Office of Immigration Statistics, Population Estimates. Washington, DC: U.S. Department of Homeland Security.

Duleep, H. 1994. "Social Security and the Emigration of Immigrants." Social Security Bulletin, 57(1): 37–52.

Jasso, G. and Rosenzweig, M.R. 1982. "Estimating the Emigration Rates of Legal Immigrants using Administrative and Survey Data: The 1971 Cohort of Immigrants to the United States." Demography, 19(3): 279–290.

Knapp, A.; Benetsky, M.J.; and Jensen, E.B. 2017. "Evaluating Foreign-Born Emigration Rates for the Postcensal Estimates: Tenure in the United States versus Age," Presented at the Annual Meeting of the Population Association of America, 27-29 April, Chicago, IL.

Kraly, E.P. 1998. "Emigration: Implications for U.S. Immigration Policy Research." in Migration between Mexico and the United States: Binational study, Vol. 2 Research Reports and Background Materials, pp. 587–618. Mexico City and Washington, DC: Mexican Ministry of Foreign Affairs and the U.S. Commission on Immigration Reform.

Leach, M.A. and Jensen, E.B. 2013. "Estimating Foreign-Born Emigration from the United States using Data from the American Community Survey." Presented at the Federal Committee on Statistical Methodology Research Conference, 4-6 November, Washington, DC.

— 2014. "Revisiting the Residual Method: Using the American Community Survey to Estimate Foreign-Born Emigration." Presented at the Annual Meeting of the Population Association of America, 1-3 May, Boston, MA.

Mulder, T.J. 2003. "Foreign-Born Emigration from the United States: 1990-2000." Presented at the Annual Meeting of the Population Association of America, 1-3 May, Minneapolis, MN.

Oosse, M. 1998. "Procedures Used to Estimate Annual Emigration Rates of the Foreign-Born by Age, Sex, and 14 Country of Birth Groups, 1980-1990." Unpublished Population Division Memorandum. Washington, DC: U.S. Census Bureau.

Schwabish, J.A. 2011. "Identifying Rates of Emigration in the United States Using Administrative Earnings Records." International Journal of Population Research, 2011: 1-17.

Turra, C.M. and Elo, I.T. 2008. "The Impact of Salmon Bias on the Hispanic Mortality Advantage: New Evidence from Social Security Data." Population Research and Policy Review, 27(5): 515–530.

U.S. Census Bureau. 2002. "Census 2000 Data for the United States, Summary File 3." Census 2000. September 2002.

— 2009. "2008 ACS Failed Edit Follow-up Operation." American Community Survey User Note. September 2009.

— n.d. "Sample Size: Initial Addresses and Sample Selected and Final Interviews." American Community Survey.

U.S. Government Accountability Office. 2002. "INS Cannot Locate Many Aliens Because It Lacks Reliable Address Information." GAO-03-188 Report to Congressional Requestors. Washington, DC: Government Printing Office.

Van Hook, J. and Zhang, W. 2011. "Who Stays? Who Goes? Selective Emigration among the Foreign-Born." Population Research and Policy Review, 30(1): 1-24.

Van Hook, J., Zhang, W., Bean, F.D., and Passel, J.S. 2006. "Foreign-Born Emigration: A New Approach and Estimates Based on Matched CPS Files." Demography, 43(2): 361-382.

Warren, R. 1979. "Alien Emigration from the United States, 1963-1974." Presented at the Annual Meeting of the Population Association of America, Philadelphia, PA.

Warren, R. and Peck, J.M. 1980. "Foreign-Born Emigration from the United States: 1960 to 1970." Demography, 17(1): 71-84.

Warren, R. and Passel, J.S. 1987. "A Count of the Uncountable: Estimates of Undocumented Aliens Counted in the 1980 United States Census." Demography, 24(3): 375-393.